UNIVERSIDADE FEDERAL DE SANTA CATARINA
PÓS-GRADUAÇÃO EM ESTUDOS DA TRADUÇÃO

Robert James Coulthard

RETHINKING BACK-TRANSLATION FOR THE CROSS-CULTURAL ADAPTATION OF HEALTH-RELATED QUESTIONNAIRES:

EXPERT TRANSLATORS MAKE BACK-TRANSLATION UNNECESSARY

Tese submetido(a) ao Programa de Pós-Graduação *stricto sensu* em Estudos da Tradução da Universidade Federal de Santa Catarina para a obtenção do Grau de Doutor em Estudos da Tradução
Orientador: Prof. Dr. Markus Weininger

Florianópolis

2009 – 2013

Robert James Coulthard

RETHINKING BACK-TRANSLATION FOR THE CROSS-CULTURAL ADAPTATION OF HEALTH-RELATED QUESTIONNAIRES: EXPERT TRANSLATORS MAKE BACK-TRANSLATION UNNECESSARY

Esta tese foi julgada adequada para a obtenção do título de DOUTOR EM ESTUDOS DA TRADUÇÃO e aprovada em sua forma final pelo curso de Pós-graduação em Estudos da Tradução da Universidade Federal de Santa Catarina.

Florianópolis, 25 de setembro de 2013.

_____
Profa. Dra. Andréia Guerini
Coordenadora do curso

Banca examinadora:

_____
Prof. Dr. Markus Johannes Weininger
Orientador e Presidente
Universidade Federal de Santa Catarina

_____
Profa. Dra. Maria José Bocorny Finatto
Universidade Federal de Rio Grande do Sul

_____
Prof. Dr. Fabio Alves
Universidade Federal de Minas Gerais

_____
Prof. Dr. Lincoln Fernandes
Universidade Federal de Santa Catarina

_____
Profa. Dra. Maria Lúcia Vasconcellos
Universidade Federal de Santa Catarina

_____
Prof. Dr. Werner Heidermann
(Suplente)
Universidade Federal de Santa Catarina

UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO DE COMUNICAÇÃO E EXPRESSÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTUDOS DA TRADUÇÃO

DEFESA DE TESE DE DOUTORADO

ATA 026/2013

Aos (25) vinte e cinco dias do mês de setembro de dois mil e treze, às 8 horas, na sala 307, prédio B, do Centro de Comunicação e Expressão, da Universidade Federal de Santa Catarina, reunida a Comissão Examinadora, designada pela Portaria n° 083/PGET/2013, de dez de setembro de dois mil e treze e constituída pelos Professores Doutores Markus Weininger – orientador (UFSC/PGET), Fabio Alves (UFMG), Maria José Bocorny Finatto (UFRGS), Lincoln Fernandes (UFSC/PGET), Maria Lúcia Vasconcellos (UFSC/PGET) e Werner Heidermann – suplente (UFSC/PGET) realizou-se em sessão pública a defesa da Tese de Doutorado de Robert James Coulthard, intitulada: "RETHINKING BACK-TRANSLATION FOR THE CROSSCULTURAL ADAPTATION OF HEALTH-RELATED QUESTIONNAIRES: EXPERT TRANSLATORS MAKE BACK-TRANSLATION UNNECESSARY", a qual foi orientada pelo Professor Doutor Markus Weininger. Após o candidato apresentar seu trabalho, procedeu-se à argüição e à avaliação, feitas nos termos regimentais. A Comissão Examinadora ___aprovada___ a tese do doutorando. O mesmo deverá apresentar, cumpridas as formalidades, a versão final, segundo o padrão gráfico da UFSC, no prazo máximo de 90 (noventa) dias, à Coordenadoria do Programa de Pós-Graduação em Estudos da Tradução. Nada mais havendo a tratar a sessão foi encerrada, dela sendo lavrada a presente ATA que é assinada pelos membros da Banca Examinadora, pelo Presidente e pelo doutorando.
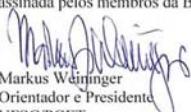
Markus Weininger
Orientador e Presidente
UFSC/PGET
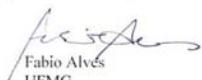
Fabio Alves
UFMG

Maria José Bocorny Finatto
UFRGS

Lincoln Fernandes
UFSC/PGET

Maria Lúcia Vasconcellos
UFSC/PGET

Werner Heidermann
Suplente
UFSC/PGET

Robert James Coulthard
Candidato

ACKNOWLEDGMENTS

CONTENTS

... since words are only names for things, it would be more convenient for all men to carry about them such things as were necessary to express a particular business they are to discourse on.

... it would serve as a universal language, to be understood in all civilised nations, whose goods and utensils are generally of the same kind, or nearly resembling, so that their uses might easily be comprehended. And thus ambassadors would be qualified to treat with foreign princes, or ministers of state, to whose tongues they were utter strangers.

Gulliver's Travels Into Several Remote Nations Of The World
Jonathan Swift, D.D., Dean of Saint Patrick's, Dublin, 1726

ABSTRACT

The primary objective of this thesis is to bring the existence of the practice of back-translation as used in healthcare to the attention of scholars studying translation.

Back-translation is a process that is primarily used for the cross-cultural adaptation of research instruments, particularly questionnaires. In back-translation an initial translation is translated back into its source language and then this back-translation is compared with the original questionnaire. The assumption is that discrepancies between the back-translation and the source questionnaire indicate problems with the forward translation, which are considered failures of equivalence.

The second objective of this thesis is to present a series of analyses of the back-translation process and some of its results and the conclusions that those analyses lead to.

Since the conclusions are uniformly unfavourable to back-translation, the third objective of this thesis is to demonstrate that competent specialist translators with experience in the area are capable of producing high quality translations of health-related questionnaires without back-translation.

In order to achieve these three objectives, the thesis is divided into an expository component an analytical component and an applied component.

The expository section presents the back-translation literature in detail, starting from the classic 1970 paper by Richard Brislin, covering contemporary cross-cultural adaptation processes that employ back-translation and ending with the scant literature containing criticism of back-translation, both positive and negative. The expository element also includes a brief introduction to translation theory.

The analytical section presents arguments that attempt to substantiate the following claims:
- back-translation was developed with amateur translators with suboptimal English competence;
- back-translation does not evaluate translation quality, rather it detects language deficiency;
- the value of back-translation as a step in cross-cultural adaptation was never proven;
- back-translation can lead to dysfunctional translations if followed blindly;

- back-translation is founded on a simplistic concept of language and of translation and on an *a priori* expectation of translation failure;
- back-translation methods perpetuate the illusion of symmetry in translation;
- back-translation anchors the translation to the source text and the source culture;
- back-translation enables monolingual control of a multilingual process and reinforces cultural insecurity in "peripheral" cultures;
- back-translation is itself perpetuated by past success as defined within the publication-dominated academic patronage system;
- back-translation is dismissive of translators' expertise, denying them the necessary conditions in which to exercise their expertise and enforcing anonymity;
- expert translators do not need back-translation, since they have adequate problem identification and solution skills;
- in cultures in which translators and translations are not low status, the imported back-translation paradigm is corrosive, since it both instils and reinforces mistrust of translations and translators;
- back-translation is not an acceptable means for giving translators a voice, since communication between translator and client should be proactive and ongoing, not reactive and defensive.

The applied component of this thesis describes two alternative methods for questionnaire translation. The first is a parallel method in which two expert translators each produced an initial draft and then worked in cooperation through a process of revision to produce a consensus version. Although the consensus-building process was not a complete success, both translators clearly demonstrated the validity of the claim that back-translation is unnecessary. The claim is further justified by the results of a back-translation of a preliminary draft which did not detect any problems that the expert translators had not discussed and failed to detect many problems that they did discuss, showing that back-translation is no substitute for competent expert translators. The second demonstration was unsuccessful in terms of producing a translation of acceptable quality because one translator was either not expert or not professional. Paradoxically, this further justifies the stricture that competent expert translators are a prerequisite for successful translation of health-related questionnaires.

In summary, this thesis presents a complete reappraisal of the back-translation process, its justifications and conceptual bases, from the perspective of a translator and translation scholar and in the light of

the great changes that have occurred since back-translation was introduced in 1970. Without exception, viewed from several different theoretical perspectives, the conclusion is invariably that back-translation is a tool that does not serve its purpose and, as Andrew Chesterman has pointed out, a tool with no useful function can be discarded.

RESUMO

O objetivo principal desta tese é chamar a atenção de estudiosos da tradução para a existência de uma prática de tradução conhecida como retrotradução, na forma como esta é utilizada na medicina e outras ciências da saúde.

Retrotradução é um processo usado primordialmente para a adaptação transcultural de instrumentos de pesquisa, especialmente questionários. Na retrotradução uma primeira versão é traduzida de volta para sua língua fonte e essa retrotradução é então comparada com o texto original. Supõe-se que discrepâncias entre a retrotradução e o texto fonte indiquem problemas com a primeira tradução, problemas esses considerados falhas de equivalência.

O segundo objetivo desta tese é apresentar uma série de análises do processo de retrotradução e alguns de seus resultado, e as conclusões às quais essas análises levaram.

Uma vez que as conclusões das análises são uniformemente desfavoráveis à retrotradução, o terceiro objetivo desta tese é demonstrar que tradutores especializados, competentes e com experiência na área (*expert*) são capazes de produzir traduções de alta qualidade de questionários relacionados à saúde, sem fazer uso da retrotradução.

A fim de alcançar esses três objetivos, a tese é dividida em três seções, a primeira expositiva, a segunda analítica e a ultima aplicada.

A seção expositiva apresenta, em detalhes, a literatura mais importante sobre retrotradução, iniciando com o artigo clássico de Richard Brislin, publicado em 1970, três processos contemporâneos de adaptação transcultural que empregam a retrotradução e a escassa literatura que contém críticas à mesma, tanto positivas quanto negativas. Essa seção inclui ainda uma breve introdução à alguns conceitos da teoria da tradução.

A seção analítica apresenta uma seria de argumentos cujo objetivo é embasar e justificar as seguintes afirmações:

- a retrotradução foi desenvolvida com tradutores amadores sem a devida competência na língua inglesa, a qual era sempre ou língua fonte ou língua alvo;
- o valor da retrotradução como um passo na adaptação transcultural nunca foi provado;
- a retrotradução não analisa a qualidade da tradução, mas sim detecta deficiências de competência lingüística;

- a retrotradução pode levar à traduções disfuncionais, se seguida cegamente;
- a retrotradução está alicerçada em um conceito simplista de linguagem e de tradução, e em uma expectativa, *a priori*, de falha tradutória;
- os métodos de retrotradução perpetuam a ilusão de simetria na tradução;
- a retrotradução ancora a tradução ao texto fonte e à cultura fonte;
- a retrotradução viabiliza um controle monolíngue de um processo multilíngue e reforça a insegurança cultural em culturas "periféricas";
- a própria retrotradução, por sua vez, é perpetuada por sucessos passados, conforme definidos dentro do sistema de *patronage* acadêmico, dominado por publicações e citações;
- a retrotradução abnega a expertise de tradutores, negando-lhes as condições necessárias para exercitar suas habilidades e experiência e impondo o seu anonimato;
- tradutores *expert* não necessitam de retrotradução pois possuem as habilidades necessárias para identificar e solucionar problemas;
- - em culturas em que tradutores e traduções não gozem de um status inferior, o paradigma importado de retrotradução é corrosivo, pois instiga e reforça uma desconfiança tanto no tradutor como na tradução;
- - a retrotradução não é um meio aceitável para "dar voz" aos tradutores, pois comunicação entre o tradutor e seu cliente deve ser proativa, contínua e construtiva e não reativa e defensiva.

Finalmente, a seção aplicada desta tese descreve dois métodos alternativos para tradução de questionários. O primeiro é um método paralelo no qual dois tradutores *expert* produzem uma tradução inicial cada e a partir dessas trabalham em cooperação através de um processo de revisão para produzir uma versão consensual. Apesar de o processo de construção consensual não ter sido um sucesso absoluto, ambas as tradutoras voluntárias claramente demonstraram a validade da afirmação de que a retrotradução é desnecessária. Esta afirmação é justificada mais além pelos resultados de uma retrotradução de uma das traduções preliminares que não detectou nenhum problema que as tradutoras *expert* não haviam discutido e ainda deixou de detectar muitos problemas por elas discutidos, assim mostrando que a retrotradução não é um substituto para tradutores *expert*. A segunda demonstração não teve sucesso no sentido de produzir uma tradução de qualidade aceitável

porque um dos tradutores não era *expert* ou profissional. Paradoxalmente, este fato justifica a condição *sine qua non* de que tradutores *expert* são um pré-requisito para a obtenção de uma tradução bem sucedida de questionários relacionados à saúde.

Resumindo, esta tese apresenta uma reavaliação completa do processo de retrotradução, suas justificativas e bases conceituais, da perspectiva de um tradutor e pesquisador de tradução e sob a luz das grandes mudanças que ocorreram desde que a retrotradução foi introduzida em 1970. Vista sob diversas perspectivas teóricas, a conclusão invariável e sem exceções é de que a retrotradução é uma ferramenta que não serve a seus propósitos e, conforme apontou Andrew Chesterman, uma ferramenta que não cumpre sua função deve ser descartada.

Palavras-chave: retrotradução, adaptação transcultural, questionários relacionados à saúde; estudos da tradução; *expertise* em tradução

LIST OF TABLES AND FIGURES

PRELIMINARY NOTES

- In this thesis I have chosen to employ the first person voice. There are a number of reasons for this, in addition to personal preference in terms of style. The primary reason for choosing this voice is that a certain proportion of the subject matter of this thesis is directly related to my experience as a professional translator of medical texts Additionally, two recurrent themes in this thesis are the impossibility of objectivity and the inevitability of subjectivity in both writing and reading. Finally, I also note that while earlier works on translation tended to employ the first person plural (for example, Nida 1964 and Catford 1965), many prominent contemporary scholars of translation use the first person singular to refer to themselves in their own work (Lambert 1991; Lefevere 1992; Pym 1993; Toury 1995a; Venuti 1995; Nord 1997a; Hermans 1999).

- This thesis has been written to be read in electronic format and makes frequent use of internal links to cross-reference between sections that are distant in terms of page numbers, but related in terms of ideas.

- I have deliberately eschewed interdependent references (using *ibid*, for example). This is a consequence of prioritising the electronic format over the hard copy and is in order to make all references searchable by electronic means.

- This thesis employs British spelling. However, throughout this thesis I have adopted the orthography "back-translation", rather than "back translation" or "backtranslation", and "decentered", rather than "decentred", since these are the forms used by Richard Brislin, author of the classic paper on back-translation (Brislin 1970). Along the same lines, I have followed Venuti in using "foreignization", rather than "foreignization". In citations I have preserved the form used by the authors being quoted, including spelling variants such as "color" rather than "colour".

- Similarly, I have preserved italics, underlining and bold face as they appear in bibliographic sources. Where I have added such features for emphasis, I have indicated this fact in the reference.

- This is a thesis about translation. I have exerted a great deal of effort in attempting to ensure that all translations are acknowledged and all translators named, whenever I have been aware that I was working

with translations, in the first case, and where the information was available in the second case. When working with texts written in languages other than English I have attempted to locate published English translations. Where I have been unable to locate an "authorised" translation, I have translated texts myself and have duly indicated this fact.

- It was pointed out to me during my viva that this thesis would have been improved by the use of a formal definition of "expert". I have to agree. However, since I did not work from a formal definition, any attempt to shoehorn one into an already complete work would be a distortion. I therefore take this opportunity to explain that my use of the word, in the term "expert translators", is intended to denote translators who have acquired expertise. I am therefore talking about experts in translation and not necessarily experts on translation. My expert translators are at the top of their profession in terms of the practical activity of translation. This a different type of "expert" from those who are members of "expert committees" or those who provide "expert opinions". While, in the event, the two expert translators I used do also have the second type of expertise, my interest in this thesis is in their practical expertise as manifest in their translation practise.

1   INTRODUCTION

1.1 Background

At the time of writing, I have been a professional translator for more than twelve years. The motivation for my research project comes directly from this professional experience, but the approach I have taken and the methods I have employed rely heavily on my contact with the academic discipline of translation studies.

As Theo Hermans has pointed out, "translators never 'just translate'" (Hermans 2007: 60), since "a translation is never a translation per se", but "a translated tourist brochure, computing manual" etc. It is not therefore sufficient to simply describe oneself as a translator or even as a translator and translation studies scholar. As with everything translation-related, context is all important. I shall therefore elaborate a little on the context of my professional experience as a translator.

The great majority of the texts that I have translated over the last twelve years are academic in nature and a majority of these are texts for scientific periodicals on medicine and other health-related subjects. I live and work in Brazil and translate exclusively from Portuguese to English.

In the course of my work as a translator of academic articles, I have often translated papers describing *cross-cultural adaptation* projects and have not infrequently been asked to work on the *back-translation* phase of such projects. These two terms are of fundamental importance to this thesis and I shall return to them shortly.

During the course of my studies for my masters degree I was exposed to concepts and theories developed within translation studies and it became apparent that the common ground that has been established between approaches such as descriptive translation studies and functionalist approaches, not to mention more radical proposals from individuals such as Arrojo and Venuti, was at odds with the conceptual basis of back-translation. These then are the basic ingredients that make up the context in which this project has been conceived and carried out.

I am aware that this use of back-translation is not a familiar concept within translation studies, which is one of the reasons for studying it within the discipline. The fact that back-translation is an unfamiliar concept is unsurprising since, to my knowledge, this application of translation had not been studied within translation studies

when I began this project, although I have seen a single article that has been published in a translation journal since then (Ozolins 2009).

In conducting this research project and writing this thesis, I therefore hope to open a fertile new field of translation research for scholars interested in studying a widespread practical application of translation that is used in fields such as the social sciences, psychology and medicine, by private enterprises (particularly the pharmaceutical industry) and by healthcare organisations such as professional associations and by both public and private healthcare services.

The specific method in which I am interested is back-translation when used as a tool for evaluating forward translations as part of cross-cultural adaptation of health-related questionnaires and their related scoring systems. I shall now explain a little about each of the following elements: health-related questionnaires, cross-cultural adaptation, and back-translation itself.

1.2 Health-related questionnaires and cross-cultural adaptation

For the purposes of this thesis, health-related questionnaires are instruments designed to collect health-related data. The data collected may be used for a variety of purposes, ranging from the results of a one-off administration to a single patient for the purposes of diagnosis, to data from multiple repeated administrations of a questionnaire to patient cohorts for prospective epidemiological trials that can last decades and are used to model risk factors and guide public health policy.

Questionnaires may be administered by self-report, completed by an interviewer or completed on the basis of information provided by a third party (mothers about their newborn infants, for example). In addition to collecting data for research and taking snapshots of patients' health status at a given point in time, these questionnaires are often administered repeatedly to the same patient or patient sample in order to measure changes over time.

It is important to note that my interest is exclusively in questionnaires that are designed to be administered to members of the public (and filled-out by a health professional) or to be filled out by members of the public themselves, irrespective of whether they are also used by health professionals without reference to patients. I am not concerned with instruments designed solely for use by medical professionals.

Instruments designed solely for use by health professionals would fall firmly within the field of 'medical translation', since they employ specialised vocabulary that lay members of the same language community would not understand and which is acquired by learning and experience. Paradoxically, this type of instrument is actually easier to translate than the apparently "simple" language used to communicate with patients because techniques that are appropriate to "medical translation", for example, consulting lists of controlled medical vocabulary, are not of use when translating questionnaires designed for administration to the general public. Indeed, one of the criteria for judging translations of the SF-36 is "avoidance of technical or artificial terms" (Bullinger et al. 1998: 914).

The Short Form Health Survey (SF-36) is a good example of a typical questionnaire designed for administration to members of the general public. It can be used both for research and in clinical settings and measures overall health status using 36 questions which, with the exception of the first two which are simple patient identification questions, are basically a list of possible limitations to quality-of-life caused by ill health. Respondents choose from among predefined responses to indicate the degree to which ill health has compromised their daily life. A typical question from the SF-36 is shown below together with its response options and their scores.

20. During the **past 4 weeks,** to what extent has your physical health or emotional problems interfered with your normal social activities with family, friends, neighbors, or groups?

(Circle One Number)

Not at all	1
Slightly	2
Moderately	3
Quite a bit	4
Extremely	5

(Rand Health 2010)

The SF-36 was developed by the Rand Health insurance company to help them analyse the cost-effectiveness of treatments, but it has since been adopted worldwide and is now used by many other health insurance companies and the United States government's Medicare program (Rand Health 2012).

As with any specialised subject, health-related questionnaires have their own terminology. In the example above, the question "During the past 4 weeks ..." is the item while the answers, "Not at all 1", "Slightly 2", etc. are the item responses, each of which has an individual *score*. Many questionnaires provide both an overall score for the major concept measured and scores for sub-scales, known as domains. The SF-36, for example, provides a total score indicating overall health status plus two separate scores for the physical components of health and the mental components of health. Physical health is further subdivided into the following health concepts: "physical functioning", "role-physical", "bodily pain" and "general health" and mental health is similarly subdivided into the health concepts "vitality", "social functioning", "role-emotional" and "mental health".

The SF-36 is just one of many thousands of health-related questionnaires, but it is one of the most widely used and has been adapted for more than 60 different countries and been used to collect data for more than 5000 publications to date (IQOLA 2011a).

Before explaining what adaptation of a health-related questionnaires for use in a different country entails, it is worth considering why health-related questionnaires are used in countries and languages in which and for which they were not developed.

Richard Brislin formalised and popularised the back-translation technique within cross-cultural psychology during the 1970s and 1980s. He considered that the reason for producing multiple language versions of a questionnaire was to collect comparable data allowing "a literature to be built up around a commonly shared set of concepts and operational definitions" (Brislin, 1986: 138).

From the early 1990s onward, questionnaire-based techniques were increasingly adopted within healthcare in order to "determine the impact of medical intervention on quality of life" and by "public health researchers to assess the outcome of healthcare services." (Guillemin 1992: 1417). Since all the measures developed at that point were in English, but there was "nonetheless a need for measures specifically designed to be used in non English-speaking countries and also among immigrant populations", researchers were faced with a choice of creating de novo questionnaires in other languages or of translating the existing English questionnaires. However, creation of a psychometric instrument of this type "is a time consuming process", so, with the exception of a single World Health Organization (WHO) project in which multiple language versions were developed simultaneously and in

parallel in many different locations (WHOQOL Group 1993), the solution adopted has invariably been translation (Guillemin 1992: 1417).

Stated objectives for adapting such questionnaires include to ensure that "the impact of a disease or its treatment is described in a similar manner in multinational trials or outcome evaluations" (Beaton et al. 2000: 3186) and to accumulate "international comparative data on a myriad of health issues." (Ozolins 2009: 1).

I shall present concepts from translation studies theory in Chapter 3, but there are two groups of labels from translation theory that are particularly useful for describing cross-cultural adaptation projects, so I shall introduce them here. The first is Christiane Nord's concept of the initiator and/or commissioner of a translation (Nord 1997a; Nord 1997b), as distinct to its sender or author, and the second is Lawrence Venuti's construct of hegemonic versus peripheral cultures and languages. Venuti defines hegemonic languages as the standard dialect within countries and English globally (Venuti, 1998: 4, and throughout).

Here in Brazil, standardised Brazilian Portuguese is the hegemonic language and non-standard dialects and all indigenous languages, whether codified or not, are peripheral. On the global level, however, and particularly so with relation to scientific publishing, Brazilian Portuguese is peripheral with relation to English, which is both the hegemonic language of many (scientifically) hegemonic countries and also the global scientific lingua franca.

Cross-cultural adaptation projects that make use of back-translation are not a uniform group in terms of translational characteristics such as source and target languages, the people and/or organisations fulfilling the role of initiator/commissioner or even the motives for translation, but two characteristics do exhibit strong trends.

The first characteristic that translations produced by cross-cultural adaptation projects tend to share is inter-lingual direction of adaptation (in terms of a continuum from central to peripheral cultures). The vast majority of cross-cultural adaptation projects begin with original texts written in English (Beaton et al. 2000: 3186).

The second characteristic that exhibits a tendency is intercultural direction of adaptation (also in terms of central versus peripheral cultures). In other words, the characteristic of where the translation projects are initiated; whether in the source or the target culture. This factor defines whether the translation is being "imported" or

"exported",[1] from the point of view of those conducting and paying for the translation project (the commissioners and initiators). Traditionally, the great majority of these projects are also initiated by researchers or organisations based in "hegemonic" countries, although a small but growing number of them are now being initiated in "peripheral" countries.

I am only therefore dealing with the first two of the four possible combinations of direction and initiator: these translations are either imported into a peripheral country or exported from a dominant country and are almost never imported to a dominant country or exported from a peripheral country. In the next subsection I shall (very briefly) describe what back-translation is and how it fits into cross-cultural adaptation of questionnaires.

1.3 Back-translation for cross-cultural adaptation

Language teachers sometimes use a form of back-translation as an illustrative technique to reveal differences between languages. When this is the objective, a translated phrase is "back-translated" into the mother tongue of the student word-for-word, retaining the sequence of the language being acquired. For example, after teaching an English-speaking student that one way to ask someone how they are in Portuguese is to say *"Tudo bem?"*, a teacher might then back-translate this as "All well?" to illustrate the different ways the two languages accomplish a similar function.

Back-translation for cross-cultural adaptation is entirely unlike this application of back-translation. Here, the objective is not to inform an English speaker of how a given function is achieved in, for example Portuguese, but *to test whether a Portuguese translation is correct.*

As used to test translations of questionnaires, the back-translation technique consists of asking a second translator to translate a translated questionnaire back into the source language and then comparing this "back-translation" with the source questionnaire and identifying differences. In this case the back-translation is not a word-for-word translation that follows the (intermediate) source language[2] structure,

---

[1] This factor is often conflated with the factor of inter-language direction, but the two can often be distinct. For example, the EORTC questionnaires are developed in Belgium and "exported" all over the world, but the source texts are written in English. (EORTC, 2009 & 2011).
[2] In other words, the target language of the forward translation.

but a "normal" translation using the structures and conventions of the language in which it is written.

It is worth mentioning that the term "back-translation" contains the same ambiguity as the word "translation" does, since it too describes both process and product and it is difficult to separate the two either theoretically or in practical experiments.

Having very briefly described my object of study - back-translation as part of cross-cultural adaptation of health-related questionnaires - I shall now present the objectives of this thesis.

1.4 Objectives

As mentioned above, with the exception of a single article (Ozolins 2009) back-translation has never been the subject of research within translation studies. Notwithstanding, a search on PubMed for the three possible spellings ("back-translation", "backtranslation" or "back translation") returns 2958 articles (PubMed 2013), which is a clear indication that back-translation is a widespread practical application of translation and, as such, a legitimate object of translation studies. The first objective of this thesis is therefore to introduce the most important back-translation literature to the translation studies community.

However, as I have also already mentioned, many of the concepts and theories developed within translation studies are at odds with the underlying principles of back-translation. Additionally, I myself have several reservations to use of the technique.

The second objective of this thesis is therefore to present an analysis of back-translation showing why it is not necessarily the best means of achieving the objectives of cross-cultural adaptation and how it can lead to several undesirable consequences.

Having demonstrated that there are sound reasons for rejecting back-translation, the third objective I hope to achieve is to demonstrate that, in the absence of back-translation, expert professional translators are capable of translating health-related questionnaires to a high standard and that a translations process involving extensive communication between those involved can not only produce high quality translations, but can also offer opportunities for creative cross-cultural adaptation that back-translation cannot.

Finally, the overall objective that I hope to achieve by achieving the three objectives just described is to provide a source of theoretical support and supporting evidence for researchers who, whether because they have been convinced by my arguments or because they have

chosen to do so independently, wish to produce translations of research instruments without using back-translation.

This objective is linked to one of the themes that I shall explore in this thesis – patronage. I believe that it can be successfully argued that the primary function of an academic article is not to publicise the findings of research, but to be accepted for publication. This may seem a radical proposal, but, logically, publication takes precedence over everything else since it is a precondition of publicising findings, of inviting confirmation or rejection and, within the patronage system, of accruing the status of an author who has published and is cited. The result of this is that if an academic community has reached consensus on a subject, such as the need for back-translation, there are serious disincentives to attempting to use alternative techniques.

Theoretical support for this position can be found in work by Lefevere and Venuti, among others (although I have never seen the point made quite so directly as I make it here) and I shall discuss the issue in greater detail during the thesis proper. At this point in the thesis, however, I bring up this hypothesis because it is the reason for one of my motives for writing.

My position with relation to patronage in academia is not so radical as some. Indeed, I have conducted this research project with the help of a bursary from CAPES, a national Brazilian research funding agency, so I have myself been explicitly patronised. Furthermore, I do not reject peer review, nor even, *a priori*, a system linking academic status with publication. I am, however, aware that peer-review has one major weakness, which can be illustrated by considering the flat earth fallacy.

For a long time the most learned people in Europe were convinced that the earth was flat. The fallacy was hard to disprove, since many ships did sail off into the distance never to return, as though they had indeed fallen off the edge of a flat world, and nobody had successfully circumnavigated the globe (not, at least, according to the collective memory of European society at the time). The consensus belief was, however, completely erroneous.

I do not presume to claim that faith in back-translation is as deluded as belief in a flat earth. I do, however, claim that *if* back-translation is not actually necessary to achieve a good quality, functioning translation of a research instrument (and is possibly actually a hindrance to achieving this objective), *but* the consensus among peer

reviewers is that it *is* necessary, then peer review will tend to continue imposing back-translation on the research community.

I therefore consider that one of the most important objectives of this thesis is to present arguments that do not only convince my peers in the translation studies community, but are also capable of convincing the current users of back-translation, and their peers, that (i) back-translation is not always necessary, (ii) that back-translation should often not be the preferred option and that (iii) adopting a cooperative approach involving greater communication with translators is not only preferable for translators, but is also preferable for their own research objectives.

1.5   The structure of this thesis

It is customary to introduce one's object of study and research objectives before moving on to a review of the literature on the subject. In the case of this thesis, it could be considered that the medical literature on the subject of back-translation and the published results of back-translation projects are more directly the object of study than back-translation itself, at least until the end of the fourth chapter. However, by the end of the thesis, the focus of study can properly be considered to be back-translation itself.

Notwithstanding, this thesis will trace a fairly circuitous path in terms of the way I have attempted to reconcile the dilemma of addressing an audience versed in translation studies theory on subjects such as statistics and epidemiology whilst also opening a dialogue with an audience from the health sciences that draws on the arguments and terminology of translation studies.

This element of the thesis will become evident in the second and third chapters, which deal with published literature, since they cover two distinct disciplines. The first of these, Chapter 2, describes the development of back-translation within cross-cultural psychology, starting with Richard Brislin's classic article (Brislin 1970), its adoption within medicine and the developments that different cross-cultural adaptation methods employing back-translation have undergone. The second, Chapter 3, very briefly introduces concepts from translation studies that I have found useful for understanding and analysing back-translation. Chapter 3 was originally a full, in-depth literature review, but in view of the length of the thesis as a whole, it has been reduced to a bare minimum. The original unabridged version of Chapter 3 is included in an Electronic Supplement.

The chapter on back-translation is not intended to contain critical analysis, although it is of course impossible to review literature without making choices of selection and omission which are in themselves biased. Rather, the intention is to present the back-translation literature (primarily for the benefit of translation studies scholars to whom it will be a new concept) in the terms used by its proponents, adherents and users.

Conscious critical analysis will be reserved for Chapter 4, in which I shall revisit translation studies theories and the back-translation literature, this time with the intention of bringing the former to bear on the latter, and will also present arguments that are the fruit of my reflections both on projects in which I have been involved and on the published results of projects on which I have not worked.

I consider that a biased perspective would be the inevitable result of any attempt to discuss back-translation in the light of translation theory without first describing it in its "natural habitat", and since recent scientific publications on the subject take the conceptual basis of the technique for granted, the thesis proper begins in Chapter 2 with a presentation, in some detail, of the origins of the back-translation method in the work of Richard Brislin[3] and the experimental data on which assessments of its utility are founded.

Having presented the theoretical justifications and experimental evidence that were originally offered in support of back-translation, I will move on to describe contemporary versions of the technique, restricting the scope to its use within medicine, which is the area in which I have the greatest amount of professional experience, and focusing on later adaptations of Brislin's techniques and their implicit concepts and assumptions.

One of my aims in the second chapter, dealing with the literature on back-translation, is to make explicit as far as possible the many shared assumptions that underlie back-translation. These are rarely stated overtly in the literature and medical researchers acquire them through experience in the area. As a result, literature from cross-cultural psychology (where the technique was developed) and from the health sciences is first presented with no attempt at critical analysis, but with a great deal of explanation of terms, concepts and assumptions that healthcare professionals do not feel the need to make explicit when publishing for their peers.

---

[3] Brislin 1970; Brislin 1973; Brislin et al. 1980; and Brislin 1986

The classic literature is presented first, because it is indispensable to an understanding of the justifications for back-translation, since the contemporary literature takes the technique's efficacy as a given. The current state-of-the-art of back-translation is then described in a manner that will hopefully be understandable to people who were not previously familiar with the terminology and concepts involved in cross-cultural adaptation, epidemiology and/or statistics. If, on occasion, the level is inevitably too basic for some of my readers who are well-versed in these areas, I hope that they will understand the reason and accept that the intention is not to patronise, but to ensure that all information necessary to the analysis has been presented in advance.

In attempting to refrain from criticism when presenting the literature on back-translation I do not intend to conceal the fact that I am critical of the concept and unconvinced of the technique's utility. What I hope to achieve is to describe the literature in its own terms, primarily in order to introduce it to the translation studies community, but also because, if I wish to entertain any serious ambition of changing existing practices within the scientific community, I must convince its members of my arguments and to criticise each element of the back-translation paradigm immediately after introducing it, and before having covered the many caveats and modifications that have built up over more than 40 years, would probably, and justifiably, be seen as unfair. I have therefore decided to present the literature on back-translation without raising my objections and criticisms and before introducing the translation studies literature on which many of my arguments are based.

The third chapter is dedicated to work in the area of translation studies. Since the thesis will be submitted to a translation studies department, and in view of the size of the thesis, I felt it would be acceptable to summarize the translation studies literature rather than covering it in the same exhaustive manner as the back-translation literature. Indeed, it might even be considered that the large body of extant literature on the authors I shall draw upon makes a separate chapter on translation theory unnecessary. However, I came to the conclusion that a chapter that at least presented the relevant translation studies literature was indispensable because my central argument is that back-translation is often unnecessary to achieve the objectives of cross-cultural adaptation, but the users of back-translation are medical researchers who cannot be expected to be versed in translation studies theory and terminology and if I wish to have an impact on current practice, it is these medical researchers whom I must convince.

Chapter 3 therefore very briefly introduces the most important concepts on which I shall draw in the analytical chapter and the authors who developed them, providing readers who do not have a background in a language-oriented discipline with the references needed to explore the subject in greater depth.[4]

Chapter 4 presents arguments arising from analysis of the subject of back-translation in light of translation studies theories and also arguments that are the result of my own professional experience and my reflections on the subject, including an analysis of the validity of the original justifications for back-translation as offered by Brislin. I believe that a number of the theoretical and practical points I raise in Chapter 4 have not been proposed previously. These have emerged from my consideration of both the body of work on back-translation and the general medical literature, both of which I have come into contact with during the course of my career as a medical translator, but which I now interpret not only from the perspective of a translator who has worked on back-translation projects numbering into double figures, but also in the light of what I have learnt and internalised in six years of postgraduate research into translation. As a result, I consider the analysis presented in Chapter 4 to be the most important contribution this thesis makes to translation studies.

Chapter 4 presents a series of arguments, covering the evidence in favour of back-translation and presenting evidence that is unfavourable to back-translation. Each of these arguments ends with my conclusions and the chapter itself concludes by rejecting back-translation and then considering the implications of eliminating the process.

Chapter 5 marks a transition from description and theory to practical application, presenting an alternative method for translating research questionnaires that I developed to fill the gap left by back-translation on the basis of elements drawn from the cross-cultural adaptation literature, from translation studies literature and from my own professional experience as a translator. I employed this method to translate a health-related questionnaire and the process and results and also my conclusions are described in this chapter. I should, however, make it clear that the studies I have conducted of translation processes

---

[4] A very much extended version of Chapter 3 is included in the Electronic Supplement for readers who wish a more detailed explanation of the concepts involved, including extensive English translations of quotations from texts published in Portuguese.

are not strictly speaking "scientific experiments", which is why I refer to them as demonstrations.

Although the results of the process described in Chapter 5 were very satisfactory, I was concerned that the actual process was possibly still more elaborate than strictly necessary. Chapter 6 therefore presents a "streamlined" version of the process described in Chapter 5, the results of this process and my conclusions about them.

Since Chapters 4, 5 and 6 all contain their own conclusions, and given their distinct nature, there is no standalone Conclusions section. Rather, the final chapter, Chapter 7, contains final comments, including suggested avenues for future research.

The list of references covers all of the publications that I have cited in the thesis proper and also those that are only cited in the extended version of Chapter 3, since many of them contributed to the development of my arguments.

The appendices to the component of the thesis intended to be printed contain the original questionnaire translated for the demonstrations described in Chapters 5 and 6 and the most revealing translated versions, plus the translation briefs sent to the translators and translation agency. All of the translations, including intermediate versions that contained few or no alterations, are provided in the Electronic Supplement in the interests of completeness and to avoid publication bias, since future researchers could conceivably be interested in data I considered irrelevant.

2 BACK-TRANSLATION LITERATURE

2.1 Overview

My object of study is the "translation → back-translation" stage of the cross-cultural adaptation process, as used within the health sciences, particularly in epidemiology, but also for diagnosis and monitoring in health-related fields as varied as sports science, orthopaedics, psychiatry and psychology. The back-translation technique was originally developed over forty years ago and is in widespread use in the health sciences to date.

The practice of cross-cultural adaptation began within psychology and has since been adopted throughout the health sciences in general and in many social sciences, but, with a single exception that I shall discuss in subsection 2.4.2 (Ozolins 2009), it has hitherto not been the subject of discussion within translation studies. My description of the object of study itself – back-translation – will therefore take the form of a review of the most important works in the literature on the subject.

Where the work presented discusses concepts that may be unfamiliar to readers whose academic backgrounds are in language-oriented disciplines, I shall attempt to provide explanations that will make them more accessible. These elements are primarily related to concepts involved in statistical analysis and epidemiology, but where it appears to me that the assumptions behind certain conclusions have not been made explicit in the articles being discussed, I shall do my best to fill in the gaps. I trust that those who are already acquainted with these concepts will not consider me patronising and that those who already understand them in greater depth than I do will condescend to forgive my simplifications.

The remainder of this section presents the theoretical literature on cross-cultural adaptation from within the health sciences. The classic 1970 publication by Richard Brislin, "Back-translation for cross-cultural research" is dealt with in some detail (subsection 2.2.1), as is later work published by Brislin (subsection 2.2.2).

Having covered Brislin's contributions, I will then move on to three sets of cultural adaptation guidelines that have been widely adopted and which embody the most up-to-date iterations of those cross-cultural adaptation methods that utilize back-translation (subsections 2.3.1 to 2.3.3). The final part of this chapter is devoted to the very small

body of work offering critical analysis of back-translation, whether positive or negative (subsections 2.4.1 to 2.4.3).

2.2 Richard Brislin

Richard W Brislin can properly be considered the "father" of back-translation. Although, as he himself acknowledged, he was not the first person to use or recommend back-translation, it was his doctoral thesis and an article he published summarizing the findings of that thesis (Brislin: 1970) that brought back-translation to the attention of a wider audience.

Since 1970, back-translation has virtually attained the status of the "gold-standard" for translation of health-related questionnaires (Ozolins 2009: 2) and also of many other health-related documents (Grunwald and Goldfarb 2006: 1), although these are beyond the scope of this thesis, which is exclusively related to translation of questionnaires.

I shall begin by presenting the 1970 article, before moving on (in subsection 2.2.2) to three chapters that Brislin wrote for textbooks, published in 1973 (with Lonner and Thorndike), 1980 and 1986.

2.2.1   The classic study: Richard Brislin's "Back-translation for cross-cultural research" (1970)

The classic article on back-translation was published over forty years ago by Richard Brislin in the Journal of Cross-Cultural Psychology. It was entitled "Back-translation for cross-cultural research" and summarized findings from his doctoral thesis, which he had submitted to Pennsylvania State University the previous year. In the article, Brislin described the process by which he developed translation techniques and translation testing techniques on the island of Guam, with the help of students who spoke English plus one of eight Micronesian languages or Korean. The paper also describes a series of experiments using these techniques and their results and ends by recommending a seven-stage process for the translation of cross-cultural research materials.

This paper is still extensively cited to this day. The total number of citations since 1970 is both extremely large and virtually impossible to quantify, but a Google Scholar search for articles citing "Brislin: Back-translation for cross-cultural research" returned a total of 1765 articles citing Brislin's paper (Google Scholar 2011a). The true number,

including articles and book chapters that were only ever published in print and never digitized or indexed electronically, must be far greater.

The impact and longevity of this paper, in addition to the fact that the principles and concepts behind back-translation have remained virtually unchallenged ever since, mean that I shall devote a large proportion of this chapter to Brislin's initial work before going on to discuss elements of his later work that alter or expand on the 1970 paper (in the next subsection). Having described this starting-point in detail, I will then be in a position to describe more recent permutations of cross-cultural adaptation in terms of how they conform to, or differ from, Brislin's original methods.

Brislin begins by briefly describing what he calls "back-translation techniques for cross-cultural research" (1970: 185). The first step, he says, is to write an instrument that is appropriate for translation. This is a theme that recurs throughout his writings on back-translation.

Brislin quotes a book chapter written by his colleagues Werner and Campbell, but not yet in print when he was writing, entitled "Translating, working through interpreters, and the problem of decentering." in which they suggest five rules that an investigator could follow in order to arrive at "an easily translatable version" (of the source text to be translated). These rules are as follows:

> 1) simple sentences; 2) repetition of nouns rather than use of pronouns; 3) avoiding metaphor and colloquialisms; 4) avoiding English passive tense; 5) avoiding hypothetical phrasings or subjunctive mood."
>
> (Werner & Campbell, quoted in Brislin 1970: 185)

Brislin also states that "The investigator might also add context to his ideas and redundancy", but does not make it clear whether this suggestion comes from Werner and Campbell or is his own.[5]

Having explained that the process begins with a source text that is designed to be translated, he then briefly describes the back-translation process itself, saying that two bilingual people should be employed, one to translate from source to target and the other to "blindly" translate the first translation back into the source language (i.e. without reading the original source text). Having done this,

> The investigator now has two versions in the original language, which, if they are identical, suggest that the target version from

---

[5] The suggestion was actually originally made by Nida, who went into great detail on the subject, and whose major work was on Brislin's list of references (Nida, 1964:131-143).

> the middle of the process is equivalent to the source language forms.

<div align="right">(Brislin 1970: 186)</div>

It is not clear whether "forms" refers to parts of speech, conjugations of verbs, etc. or simply to standardized paper-based documents, such as questionnaires, but the distinction makes little difference to the basic concept, which is the fundamental hypothesis on which the entire back-translation technique is based. This is the theory that by comparing a source text to a back-translation, evidence can be acquired about the accuracy and quality of the forward translation that led to the back-translation, couched in terms of its equivalence to the source text.

Brislin makes it very clear that, at this point in his research, this is still only a hypothesis and stresses his earlier use of the word "suggest", listing three possible reasons why the previous statement might not prove to be true in practice. The first reason he offers is that translators may have a shared set of "rules" by which they translate the same non-equivalent pair of words in both directions – giving the example of "amigo" and "friend" (Brislin 1970: 186).

The second reason Brislin gives for why a back-translation could be identical to a source-language original without the intermediate forward translation itself being equivalent to the original is that the back-translator may have been able to "make sense out of" the forward translation despite it being of poor quality (Brislin 1970: 186).

The third possibility would be that the forward translator could have used grammatical forms from the source language, which would be no barrier to back-translation, but would render the intermediate translation "worthless for the purpose of asking questions of target-language monolinguals"(Brislin 1970: 186).

Having made it clear that there are reservations about exactly what is demonstrated when forward translation and back-translation are identical, Brislin states that these issues will be investigated in the course of the paper and goes on to introduce the second basic concept underlying the back-translation method,

> ... suppose that the researcher finds errors in comparing the two source language versions. Knowledge of the errors gives the researcher some insight into the competence of his translators. If the two source language forms are not identical, he can confer with the two bilinguals, clearing up errors.

<div align="right">(Brislin 1970: 186)</div>

Having introduced the basic justifications for using back-translation, Brislin moves on to a related, but distinct concept known as "decentering", which he defines as, "a translation process in which the source and the target language versions are equally important during the translation procedure." He explains that in this method the source text is not inviolable, but can be modified during the process. "Such modification might be based on knowledge of what terms will not translate well." (Brislin 1970: 186).

Having introduced the rationale behind back-translation and the concept of decentering, Brislin briefly describes work that in 1970 was either recently published or still unpublished and that dealt with translation problems and techniques.

He begins by listing four techniques that could be used separately or in combination, according to a paper that he had co-authored with Campbell, Stewart and Werner (Campbell et al. 1970), but which had not yet been submitted for publication when he was writing.[6] The four techniques were back-translation, the bilingual technique, the committee approach and pre-test procedures.

The first of these he had already described, the second consists of testing bilinguals in both languages and identifying discrepancies, the third is forward-translation in committee and the last, pre-testing, is field-testing of a completed translation. Brislin states that these techniques were rarely used at the time, citing the same unpublished paper, which had reviewed 80 articles before reaching that conclusion (Campbell et al. 1970, quoted in Brislin 1970: 187). Brislin was, however, able to refer to three studies that he said had reported successful use of back-translation.

The first was a 1963 study by Fink, entitled "Interviewer training and supervision in a survey in Laos", in which Fink, unable to find people competent to translate from English to Lao, translated from English to Thai to Lao, back to Thai and finally back to English again. He found that some of the back-translations differed significantly from the originals and in response to this he revised the original English and checked for problems in all three non-English versions (Fink 1963 quoted in Brislin 1970: 187).

---

[6] It appears that this paper was never actually published, despite being described in the references as "to be submitted to the International Journal Of Psychology", since, in the preface to "Cross-cultural research methods", Brislin and Lonner state that although it "circulated informally under the title 'Back-translation and other translation techniques in cross-cultural research' ... the involved schedules of the four authors have prevented a final published version" (Brislin, Lonner & Thorndike, 1973: xii).

Brislin's second example is the same unpublished paper that he co-authored with colleagues from Pennsylvania State University (Campbell et al. 1970), which he says reported success after having "Navajo subjects translate and back-translate several simple English passages".

The third paper mentioned was published in 1963 by Sinaiko who "studied French and English in governmental work simulating multi-nation conferences." Brislin tells us that "Skilled interpreters were used as subjects" and that Sinaiko claimed "There was virtually no change in meaning" (Sinaiko 1963, quoted in Brislin 1970: 187).

Brislin suggests that the first two studies were successful because the original English was open to revision and the third because English and French have a similar structure.

The next section of the paper is devoted to similarity of structure between source and target languages, but despite having expressed the opinion that Sinaiko's success had been due to structural similarity, here Brislin does not relate language similarity to translation success, merely stating that Jacobson and Bass had been successful with Indo-European languages including Danish, French, German, Italian and Spanish. (Brislin 1970: 187-8). However, the question of language effect is revisited when Brislin discusses the results of experiments he conducted on Guam.

Brislin next moves on to the bilingual technique which he suggests can, together with pre-testing, "complement back-translation", summarizing a five-step "bilingual" procedure from Prince and Mombour's 1967 publication in which an initial translation is administered to a sample of bilinguals divided in two at random. The first "group is asked the first half of the questions in language one, and the second half in language two" and then the "order is reversed for the other group". When items elicit "discrepant response frequencies", the procedure recommends that "further attempts at translation should be made" and the process continued until "comparable frequencies are obtained". Interestingly, the procedure also recommends discarding items that continue "to yield discrepant responses" (Prince & Mombour 1967: 236-237, cited in Brislin 1970: 189).

No comment is made on the success of this procedure, but it is made clear that respondents are equated by randomization, ruling out the possibility that test-retest reliability could be because of memory, since each respondent either answers the original or a translated version of each half and is tested only once.

Brislin's paper then moves on to pre-testing, which, it states, "is necessary even after careful translation since nonsensical answers can occur in response to the most carefully constructed questions." (Brislin 1970: 189) Two methods of pre-testing a questionnaire are given.

The first of these methods is the random-probe technique by which randomly-chosen respondents are asked probe questions of the type, "What do you mean?" after answering certain questions, also selected at random. The theory behind this is as follows,

> If a respondent's justification of the original answer is bizarre, then the intent of the question is not being conveyed. When the answers to this "meaning" probe are taken over many respondents, a researcher should have a good deal of insight into the quality of his questionnaire.

> (Brislin 1970: 189-90)

The second of these methods consisted of asking interviewers to rate the questions they were asking "to indicate how clear the question was to people - in effect, how adequate the translation was in formulating a well-worded question." Brislin cites an example in which the questions rated highest for clarity also scored highest for test-retest reliability with two interviewers questioning the same respondent, saying that this provided "added data in support of translation adequacy" (Brislin 1970: 189-90).

Next, Brislin describes a project that combined back-translation with an interview-based pre-test and mentions two projects that combined back-translation with a committee, but concludes by saying that,

> These and all other studies using back-translation, however, have a notable lack of emphasis on criteria for equivalence of source, target and back-translated versions.

> (Brislin 1970: 190)

which brings him to a consideration of the literature of the day on equivalence in translation:

> It would be a simple task to review and evaluate the back-translation literature if there were some criteria for translation quality or translation equivalence against which a given study could be gauged. The criteria now in the literature are only suggestions.

> (Brislin 1970: 190)

The only works by scholars of language that Brislin references are Catford's "A linguistic theory of translation" (1965) and Nida's "Towards a science of translating" (1964) and he summarizes their contributions on translation quality assessment by stating that both regarded equivalence of meaning as "the most important aspect of translation". He then returns to work by psychologists, saying that Werner and Campbell had compared "the meanings of the source and back-translated versions" on the basis of which he theorizes that, "this procedure implies that the unit of translation quality may be a unit of meaning", although he then credits the suggestion to Treisman (Brislin 1970: 191).

We are then told that Carroll used meaning as an evaluative criterion, assessing translations on the basis of meaning judgments, although "he did not actually devise a unit of translation quality that might be the unit of [a] translation quality scale." (Brislin 1970: 190).

From the theoretical consideration of the possibility of a unit of translation quality, Brislin moves on to practical attempts to demonstrate translation equivalence. He begins with Miller and Beebe-Center who

> ... reasoned that if people could perform bodily movements after having heard either a source or target language instructions,[(sic)] and if the results of the bodily movement criterion were similar across all people, then the source and its translation must be equivalent.
>
> (Brislin 1970: 191)

and then mentions Allport and Pettigrew, whose "Zulu subjects gave hand motions to indicate the presence or absence of the Trapezoidal Illusion", before highlighting the fact that the suggestion is "limited to the kinds of materials that can be examined through bodily movements." (Brislin 1970: 191).

Still in the realm of practical tests of equivalence, Brislin tells us that Miller and Beebe-Center and MacNamara shared the opinion that if answers to questions given after seeing either the source or target version of a passage are equivalent across different people, then "the source and its translation should be equivalent". He ends the subsection on evaluating quality and equivalence by saying that Spilka had noted a "lack of an ultimate criterion of translation quality", but that a suggestion by Nida, that "one of the requirements of translation is to 'produce a similar response' on the part of readers of a source and target version" could be interpreted to mean that, as is the case in "behavioristic psychology, an observable, verifiable response may be

the standard for an ultimate criterion of translation equivalence and quality."(Brislin 1970: 191).

Brislin considers factors that affect translation quality in the final preliminary passage of the article, before beginning to discuss his own methods. The section is very brief, but the three points made are indicative of Brislin's concept of translation and had an impact on his study designs.

The first point Brislin makes is that Treisman had considered information load a component of the level of difficulty of source language materials and that this could affect translation. Brislin states that "this point is of importance to cross-cultural researchers who wish to know how difficult their source language version can be before translation will be poor" (Brislin 1970: 191).

The second point, which Brislin considers similar to the first and which was made by both Nida and Spilka, is that "familiar materials should be less difficult for translators and thus should be more easily translatable." (Brislin 1970: 191).

The third and final point Brislin makes relating to factors affecting translation quality is that "Miller and Beebe-Center (1956) noticed that translation results were partially a function of the target language, with translation to French giving better results than translation to German." (Brislin 1970: 191).

This reference to the effect of varying translation languages ends Brislin's discussion of the state-of-the-art in 1970 and the next section of the article, entitled "The Present Problem", describes two research objectives that Brislin pursued at the University of Guam. The first was to examine five criteria for determination of whether source, target and back-translated forms are equivalent and the second was to examine the effects of "content area, difficulty and language on translation quality" (Brislin 1970: 192). Three of these five criteria provide the foundations for the vast majority of later incarnations of cross-cultural adaptation involving back-translation (criteria three and four are rarely used today), but Brislin's discussion of language effect and his description of the investigations of language-dependent variation are also important because they illustrate the conditions and constraints of the translation process that he conducted to investigate back-translation and shed light on his concept of translation.

Notwithstanding, there can be no doubt that it is the criteria for equivalence that are responsible for Brislin's lasting impact on the scientific community. Criterion 1 is the "original" back-translation

method and because of its lasting impact I shall quote his description of it almost verbatim:

> Monolingual raters examine the original and back-translated forms of a passage, and write down errors that they feel would lead to differences in meaning if the two forms were administered. Different raters' judgments of these 'meaning errors' can be compared in two ways. The first is the correlation between number of errors found by two or more raters over a large number of passages. The second is the percentage overlap in different raters finding <u>exactly</u> the same meaning errors.
>
> (Brislin 1970: 192)

To date, the conceptual basis of the back-translation technique is still that the forward translation is evaluated against "raters' judgments of these 'meaning errors'" in the back-translation, <u>without reference to the forward translation itself</u>. This is the essence of the back-translation method.

Criterion 2, by which "Bilingual raters look at the original and target versions and write down meaning errors as in criterion one." does not strictly speaking require a back-translation stage, but Brislin suggests analyzing the results together with those from criterion 1 and many more recent permutations of the cross-cultural adaptation process include a similar bilingual evaluation stage. Once more it is the raters who judge what constitutes a meaning error, but Brislin considers that

> This criterion is a more direct test of original-target language equivalence than that found in criterion one since criterion one tests the same language forms, and criterion two tests the different language forms.
>
> (Brislin 1970: 192)

Criterion 2 is also still being applied in current iterations of cultural adaptation processes. The number of raters vary, they may work in committee or independently and they may be working on an interim stage of a back-translation step or be rating a translation that has already completed the back-translation phase (including criterion 1 testing), but the principle of a bilingual rater comparing a forward translation with its source text and indicating errors is a component of many current forms of cross-cultural adaptation requiring translation.

Brislin's criterion 3 was as follows (abbreviated):

> Subjects should be able to answer questions about target language and back-translated passages. ... The answers to the

questions should be the same for different subjects if the versions are equivalent.

<div align="right">(Brislin 1970: 192)</div>

The passages that Brislin refers to here are essays that he used in his experiments, rather than the questionnaires to which he often refers and which are nowadays the subject of the majority of cross-cultural adaptation projects. The essay format is undoubtedly better suited than a questionnaire to a comprehension test in which respondents have to locate information in a text. In the modern setting, criterion 3 is rarely used.

Before describing criteria 4 and 5, Brislin explains that they "presuppose a satisfactory back-translation according to criterion one." and on this basis he considers them to be "tests of how sensitive criterion one is to foreseeing and preventing problems in actual test administration." (Brislin 1970: 193).

Despite introducing criteria 4 and 5 as related to "actual test administration", the description of criterion 4 that follows is once more related to the "passages", as was the case with criterion 3. There is a fundamental assumption embodied in the criterion, which is as follows:

> If the passage in English asks for a performance of some sort, the subject may be requested to perform a task with the target language version as the instructions. If he can complete the task, then the original and target versions are undoubtedly equivalent. Specifically, the original and foreign language versions are functionally (workably) equivalent. This is the 'performance criterion.'

<div align="right">(Brislin 1970: 193)</div>

With criterion 5, Brislin returns to the issue of testing the equivalence of original and translated questionnaires (since up to this point his volunteers had been translating the essays and then instructions for making a picture out of blue and green pieces of paper). The technique is "an extension of the bilingual procedure outlined by Prince and Mombour (1967)." and consists of randomizing a sample of "bilinguals" into four groups and then administering the entire source language questionnaire to the first group, the entire target questionnaire to the second, the first half of the source and second half of the target to the

third group and the opposite combination of halves to the fourth,[7] before testing the results statistically. Here, the hypothesis is as follows,

> If the versions are equal, then item frequencies should be the same, as should the total score for the entire questionnaire, across groups. Since the four groups are equated by randomization, and since the versions of the questionnaire are presumed equivalent, then the two versions of the questionnaire should elicit the same responses.
>
> (Brislin 1970: 193)

Criteria 4 and 5 are described as the "ultimate" criteria for translation because, "following Nida (1964), it is assumed that if a source version and its translation elicit the same response or pattern of responses, the most important purpose of a passage is being conveyed."(Brislin 1970: 193).

Brislin returns to the five criteria from a practical perspective later on in the paper when he describes the experiments he conducted in order to test them. The next subject he deals with from a theoretical perspective is the issue of the effect of the variables content, difficulty and language on translation quality. As mentioned above, these subjects are not directly related to back-translation, but the discussion of them throws light on Brislin's basic assumptions about translation.

The first of these three variables, content, is investigated in order "to see if a translator can look at all content areas objectively", on the basis that certain "content areas, such as facts about a piece of art, may produce fewer difficulties than content areas more involving to the translator". The hypothesis is that "if certain content areas are more difficult to translate, it might be determined why they are more difficult, and whether the problem can be corrected." (Brislin 1970: 193).

In Brislin's experiments content was varied along a "continuum from factual to attitudinal" whereas the variable difficulty of original English prose was quantified using "Thorndike-Lorge (1944) word counts and the Flesch Readability Index (Flesch 1949)". According to Brislin, "the purpose of varying difficulty is obvious. Cross-cultural researchers will want to know the upper level of difficulty of the

---

[7] Brislin's definition of "first half" and "second half" is not intuitive. The "first half" actually refers to all the odd-numbered questions while the "second half" refers to all of the even numbered questions (rather than 1-16 and 17-33 as might be expected). In other words, the split-halves questionnaires have alternate questions in two different languages and the "first half" language is used for questions 1, 3, 5, and so on, whereas the "second half" language is used for questions 2, 4, 6, and so on.

original English that can be expected to translate well." (Brislin 1970: 193-4).

The third variable investigated for its effect on translation was the target language. The research question Brislin was trying to answer by varying the target language among nine Austronesia" languages was, "whether or not some [target] languages provide better results than others when using the back-translation technique." (Brislin 1970: 194).

Brislin ends his presentation of the theoretical background to his work by mentioning that Werner and Campbell's five rules (listed on page 41 above) were followed when writing the easier texts used in his experiments.

The remainder of Brislin's landmark study is ostensibly concerned with the practical application of his theoretical framework, but there are several points in his description of methods, his presentation and discussion of results and in the final conclusions he draws that shed further light on theoretical aspects of his work, such as underlying assumptions about (and concepts of) translation, in addition to providing more detail on the context in which back-translation began to gain acceptance within the scientific community.

Brislin begins the description of his experimental methodology by briefly setting the scene with an explanation about the University of Guam, where his research was conducted. The people who Brislin refers to as "bilinguals" or "translators" throughout the article were all students at the university. These students spoke English plus one of nine Austronesian languages or Korean. The local languages were Chamorro, Kusaien, Marshallese, Palauan, Tagalog, Ponapean, Trukese, Ulithian and Yapese. Additionally, Korean was used for pre-testing. Brislin states that structurally all of these languages are very different from English and that it "was thought that using such languages would give the back-translation procedure a stringent test" (Brislin 1970: 194 & 207).

Having introduced the location and languages under investigation, Brislin describes the texts that the University of Guam students translated and which were used to test the first three criteria of equivalence (monolingual raters, bilingual raters and comprehension testing). Since these texts were also to be used to test the effect of the variables content, difficulty and language, Brislin prepared texts on three different subjects and at three different levels of difficulty which he then tested with native speakers of the nine Austronesian languages plus Korean.

The three subjects, intended to vary along a "continuum from factual to attitudinal", were on fine art, which was expected to be the "least involving", child rearing, which was considered neutral, and "the intelligence of the black versus the white race", which Brislin considered "would be very involving ... since all translators were members of dark-skinned groups". Notwithstanding, he accepted that he had no evidence for this assumption and acknowledged that "the conservative designation of the content areas would simply be 'three different content areas.'" (Brislin 1970: 193-5)

For each of these three content areas Brislin initially prepared three 300-word essays designed to be of three distinct levels of difficulty, with Flesch indices of 80, 60 and 40 (the lower the score the more difficult the text). However, he was forced to abandon the Flesch 40 texts because "bilinguals could not even start to translate these very difficult passages, and so the level was discarded." (Brislin 1970: 195).

This meant that Brislin now had 6 essays, two each on art, child rearing and racial intelligence. By asking translators who spoke English and one of two languages (Chamorro or Palauan) to translate or back-translate an essay at each of two difficulty levels in each of three subject areas, a three-level analysis of variance (ANOVA) could be performed with 3 cells for content, 2 cells for difficulty level and 2 cells for language.

For readers who are not acquainted with the technique, ANOVA basically compares multiple variables against each other simultaneously in order to eliminate the possibility that an apparently significant result from a comparison between two variables is actually the result of a third (or more) variable that has not been analyzed, but which affects one of the variables that has been analyzed, or which has a more powerful effect on the outcome than the variables that have been analyzed.

Brislin explains that even though there were University of Guam students who spoke at least one of 10 languages plus English, the work done by speakers of the other eight languages could only be analyzed in a two-level design (content against difficulty), because "translators for languages other than Chamorro and Palauan were not available in large numbers". He also takes care to point out that "a given cell entry was provided by two bilinguals, one translating from English to his language, the other back-translating from his language into English", and mentions that the order in which each translator was given each difficulty level and each content area were systematically randomized

i.e. each order was used as often as another and a given order was assigned to each translator at random (Brislin 1970: 195).

Continuing the description of his experimental methods, Brislin comes to the students he used as translators. All of them had completed at least one semester of "English Composition", on which they could only enrol after passing standardized tests of English proficiency. Brislin comments that this reduced the number of eligible translators, but increased the quality of their English and also increased the proportion of translators who were students in their final year at University. A total of ninety-four students met Brislin's criteria and took part as translators (Brislin 1970: 196).

The instructions given to the translators are of relevance here and I shall quote them verbatim (the words "Kusaien" and "English" were substituted and/or reversed depending on which languages and directions the translator in question was working on):

> I would like you to read the English essays that you have in your hands. Then I would like you to translate the English essays to Kusaien.
>
> Do not be afraid of the word 'translate'. I want you to get the ideas and meanings of the English essays into Kusaien. Don't worry about translating every English word into a Kusaien word. Just get the ideas and meanings of the English into the Kusaien language.
>
> Imagine I wanted to write a letter to a Kusaien friend. I don't read or write Kusaien, but you would surely help me write the letter. You would tell me how to write the ideas and meanings in the letter to Kusaie.
>
> I am interested in the ideas and meanings of these English essays, just as I would be interested in the ideas and meanings of a letter that I wanted to write to Kusaie. Are you ready to start putting these English essays into Kusaien? Do you have any questions?
>
> (Brislin 1970: 196-7)

The translators were given a copy of Webster's Collegiate Dictionary (monolingual American English), translations took place in a small quiet room and the translators were paid ten dollars each and asked not to discuss the content of the essays.

The next part of the paper deals with experiments that Brislin conducted, using the essays and students described above, in order to

test the five criteria of equivalence that were discussed earlier (starting on page 47).

Starting with criterion 1 (monolingual meaning errors), Brislin explains that a single female rater, bilingual in English and Spanish, examined the original and back-translated versions of all of the back-translated essays. Rater two, who was Brislin himself, examined fifteen sets of six essays (three content areas multiplied by two levels of difficulty), but at no point in the paper is it made clear which target languages were rated by Brislin.

The instructions given to the monolingual raters (rating back-translation against original) were as follows,

> Examine the second copy of the English against the original and write down any errors that you feel might affect the meaning between the two passages. That is, write down errors that might make differences in the meaning people would infer from reading only one passage. For instance, if the original is 'food' and the second copy 'hunger', that might be a meaning error, depending upon the context. Examine the two English versions, and write down errors that might affect meaning.
>
> (Brislin 1970: 197)

It is not entirely clear at this point in the paper, but reference to the results and discussion sections reveals that the reliability of criterion 1 was assessed by comparing errors found by rater one and rater two in the 90 back-translations that both rated, whereas the ANOVA design, in which each cell entry contains the number of errors found by comparing a back-translation with its original, was used to detect the effects of content, language and difficulty and this analysis was conducted using rater one's results alone.

Criterion 2 (bilingual meaning errors) was tested by asking either Palauan-English or Chamorro-English bilinguals to rate ten translations of one of the six essays – the harder essay on child rearing. Brislin states that the bilinguals found this task very difficult and time-consuming and reported "difficulties in setting criteria for themselves as to when a meaning error was made" which he attributes to the "vast differences between the two languages being rated." The raters themselves practised translating for two hours and practised rating their own language against English, also for two hours (Brislin 1970: 198).

Criterion 3 (questions about a passage) was tested using two Palauan and two Chamorro raters. Each rater was given ten translations of the harder child-rearing essay and asked to answer questions

"demanding eight facts, using only information from each essay." They were also asked to specify exactly where in each essay they had found the information they had used to answer each question. In parallel, two raters from the Pennsylvania State University answered the same questions working from back-translated versions of the essays. The theory behind this was as follows,

> The purpose was to determine if there were differences in number of correct answers from raters seeing target versions versus raters seeing back-translated versions. If differences emerge, then a designation of which step — translation or back-translation — led to difficulties might be made.
>
> (Brislin 1970: 198)

Brislin does not state which language the questions bilinguals answered were written in – but presumably they were written in English since the monolingual raters would obviously have had to answer English questions and he provides no details of any translation/back-translation process for producing Chamorro and Palauan versions of the questions.

The first three criteria were all tested using the six 300-word discursive essays in three content areas and at two levels of difficulty. These were not appropriate for the criteria 4 and 5 tests since criterion 4 demands a performance and criterion 5 was designed specifically to test questionnaires. Brislin therefore employed his own criterion 1 method (back-translation with raters identifying monolingual meaning errors) in order to produce suitable materials to test criteria 4 and 5.

The materials for the test of criterion 4 (the performance task) were instructions for assembling a picture using blue and green pieces of paper. The instructions were originally written in English, translated into three of the island languages (Chamorro, Palauan, and Kusaien) by students who "seemed to have done well with their six essays" and then back-translated into English by students "with similar qualifications" (Brislin 1970: 198-9).

Brislin reports that the "translations proved very easy" since "no Criterion One meaning errors were found" and explains that the "colors were chosen since Palauans use the same word for both colors" and that this "possible confusion was added to give the picture-makers more chances to make a mistake."

Criterion 4 was then tested by randomizing ten native speakers of the three Austronesian languages to follow instructions in English or their native language. Therefore, fifteen people followed the English instructions and five people followed instructions in each of Chamorro,

Palauan, and Kusaien. The number of errors each person made in assembling the picture was counted and the maximum number of possible errors was eight.

The experiment Brislin used to test criterion 5 (administration of a questionnaire) became the basic model for the translation phase of many cross-cultural adaptation projects. The translation/back-translation component is based on criterion 1, but criteria 2, 3 and 4 were not applied. Brislin did, however, use the decentering technique.

The questionnaire translated was the Marlowe-Crowne Social Desirability Scale (Crowne & Marlowe [1964] 2009) and the target language chosen was Chamorro. Six translators and back-translators were used who "were upperclassmen, seemed capable, and expressed interest in doing more work." After a first translation/back-translation stage, criterion 1 was applied and "a revision of the English was needed ... so that a comparable Chamorro item would be possible." Brislin does not mention how many monolingual raters were used, nor whether statistical methods were applied to the results. However, it can be presumed that because only one back-translation was produced at each stage, the only statistical treatment that could have been used would be inter-rater comparison, since there would not have been multiple translations that could be compared and the absence of criterion 2 testing means there would also be no possibility of direct comparison between the forward translation and the original or between the forward translation and the back-translation (Brislin 1970: 199).

After a second translation/back-translation step (working from the revised English original and with different translators), another "review session" and "slight revision of several more English items", the second translation was administered to ten "Chamorro-speaking college students [who] could not read their [own] language" by playing them a tape recording of the scale. This led to "a few more changes of the English version." Finally, a third translation/back-translation step was conducted with a fifth and sixth translator and monolingual rating conducted again. Brislin states that "At this point there were no Criterion One meaning errors.", but still provides no indication of how many monolingual raters were involved. This version was reviewed by all six translators who "agreed that it was as close to the revised English as possible." The process of translating and decentering took more than 200 man-hours (Brislin 1970: 200).

The actual criterion 5 test of equivalence was designed to provide results suitable for statistical analysis, as was the case with the tests of

the effect of content, difficulty and language. However, whereas those tests employed an ANOVA procedure to detect for associations between three interrelated variables, the criterion 5 results were analyzed using Pearson's correlation coefficient, which tests for statistical agreement between repeated or paired observations. Another difference between the criterion 5 test and the previous tests is that the respondents were a mixture of "80 bilingual high school and college students", which is significant because all the other testing and translation procedures had exclusively involved non-native English-speaking University students who had achieved a certain level of competence in English or native speakers of English, in the case of the monolingual raters (Brislin 1970: 200).

In order to allow for comparisons between bilinguals answering the questionnaire in different languages, the first half of Brislin's sample was randomized to answer either an English or a Chamorro version. The English questionnaire was the decentered version that had served as the original for the final version of the Chamorro questionnaire. The other 40 students were randomized to answer either a questionnaire in which odd-numbered questions were in English and even-numbered questions were in Chamorro (which Brislin termed the "first-half English" version) or a questionnaire in which odd-numbered questions were in Chamorro and even-numbered questions were in English (the "first-half Chamorro" version). This gave him four groups, which he illustrated as follows:

Table 2.1 – Bilingual split-half test groups for Brislin's criterion 5

	Second Half English	Second Half Chamorro
First Half English	Group 1 — All English	Group 2 — Half English, Half Chamorro
First Half Chamorro	Group 3 — Half Chamorro, Half English	Group 4 — All Chamorro

(Brislin 1970: 200)

Having identified the subsets of his sample of bilingual students, Brislin describes the results that he considers would confirm his hypotheses, as follows:

A high split-half reliability coefficient for groups one and four would suggest satisfactory reliability of the revised English and Chamorro scales. A high coefficient for groups two and three would suggest both adequate translation and would justify the inference that the criterion one meaning error was sensitive to

> determining translation problems in actual administration of translated materials.
>
> <div align="right">(Brislin 1970: 200-1)</div>

This statement ends the description of experimental methodology and next Brislin presents his results. Certain details of the results section throw light on features of the experimental methods and theoretical assumptions which had not been entirely elucidated earlier in the paper, but in general the results are not of great interest in isolation. What is of interest is Brislin's analysis and interpretation of these results and the conclusions he draws from them. I shall therefore close this description of Brislin's seminal work reporting the results of the tests of the five criteria for equivalence, together with Brislin's discussion of them, before moving on to the results for content, difficulty and language effects and the conclusions he drew from them.

Criteria 1, 2 and 3 were all tested at least partially using the moderately difficult essay on child-rearing and on this basis Brislin presents the results for each criterion separately, before providing a statistical analysis of their inter-relationships, but in the discussion section he deals with all three of them together.

The correlation coefficients for the level of agreement between rater one and rater two (Brislin himself) judging "monolingual meaning errors" in the back-translations of 90 essays on three subject areas and at two levels of difficulty were as follows: easy art, .92; difficult art .93; easy child rearing .84, difficult child rearing .88; easy racial intelligence .79; and difficult racial intelligence .71. (Brislin 1970: 201)

Jacob Cohen's classic text "Statistical power analysis for the behavioural sciences" has a chapter on the significance of a product moment "r" (which is another name for Pearson's "r") and in it he states that correlations between .5 and 1 and between -.5 and -1 are "large" and are

> ... around the upper end of the ... r's one encounters in those fields of behavioural science which use them extensively, e.g. differential, ... clinical and counseling psychology ...
>
> <div align="right">(Cohen 1977: 80)</div>

However, a few pages earlier (p. 78), Cohen also explains that "correlation coefficients above the .50 - .60 range [are normally encountered] only when the correlations are measurement reliability coefficients", which is exactly what Brislin's coefficients are – tests of the reliability of a measurement – and this is why Brislin himself only

describes criterion 1 reliability as "adequate" despite the fact that all of his coefficients were above .70.

The percentage overlap between the two raters was calculated by dividing the number of errors found by the rater who found most errors by the number of errors that coincided with those errors found by the rater who found fewer errors. In other words, the largest number of errors found was the denominator and the number of those errors that were also found by the other rater was the numerator in a fraction that was then converted to a percentage. (Brislin 1970: 201)

The percentage overlap between the two raters judging "monolingual meaning errors" in back-translations of the 90 essays on three subject areas and at two levels of difficulty were as follows: easy art, 64%; difficult art 60%; easy child rearing 54%, difficult child rearing 59%; easy racial intelligence 51%; and difficult racial intelligence 54%. (Brislin 1970: 201)

Having described criterion 1 reliability as adequate, Brislin states that

> Evaluative comments on the magnitude of percentage overlap are more difficult to make since standards do not exist. Considering the scoring procedure, however, the magnitude of error overlap seems sufficiently high to state that a fair amount of agreement exists between raters. This last statement is, of course, the writer's subjective judgment.
>
> (Brislin 1970: 201)

There is no discussion of the criterion 1 results in isolation other than to repeat the statement that reliability was adequate. Instead, they are considered together with criteria 2 and 3, so I shall describe Brislin's conclusions about them after presenting the results for criteria 2 and 3 and the analysis of associations between criteria 1, 2 and 3.

The results of the test of criterion 2 relate to eighteen passages translated into Chamorro – all of the moderately difficult child-rearing essays (n = 10) and eight others – and ten passages translated into Palauan – all of the moderately difficult child-rearing essays. The raters indicated the number of "bilingual meaning errors" they found in each essay and the results Brislin presents are summarised into simple averages (means) and the correlations between them. He does not provide ranges or medians and neither does he provide information on differences between different levels of difficulty and different content areas (presumably because he only had results for eight essays that were not the moderate child rearing essay). Brislin does not provide data on

"error overlap" as he did with criterion 1, even though he did use two raters for each language and could therefore have done so. Instead he chooses a rather unusual selection method, by which

> In this and all other presentations of criteria one through three, two persons rated the number of errors per passage. The mean score of either could be presented, but the scores found in the text and tables are the number of errors reported by the first rater to complete his or her judgments

> (Brislin 1970: 202)

However, it is still not clear whether the intended meaning is that all of the results come from the first rater to complete the entire dataset or whether each result for each cell comes from the first rater to finish the task providing figures for that cell.

The mean number of "bilingual meaning errors" found by Chamorro students was 1.8 and the inter-rater reliability (agreement) was r = .86. The mean number of "bilingual meaning errors" found by Palauan students was 7.6 and the inter-rater reliability (agreement) was r = .68. Brislin describes these coefficients as "adequate for the group comparison being made here" (Brislin 1970: 202). In the discussion section, Brislin mentions that Palauans had lower reliability and that this is understandable since they also found translation harder than Chamorro speakers. He points out that "Judging someone else's translation includes many difficulties of the actual translation task, e.g., remembering corresponding terms and structural rules" (Brislin 1970: 211).

The results provided for criterion 3 were also extensively summarised. Criterion 3, the comprehension test, had been tested in three different subsets. The first two were bilingual. Two Chamorro and two Palauan bilinguals had answered written questions (presumably in English, but this is not specified) about a target version of the child-rearing essay. The third subset comprised two monolingual raters.

However, Brislin collapses the bilingual results together (Palauan and Chamorro), despite the striking difference between their results for criterion 2 and informs the reader that, out of nine possible errors, 70% of the forward-translated essays rated by bilinguals were in the range from 0 to 2 errors. He provides no range or maximum value for the other six essays that were not within this range, but states that the mean number of errors in the twenty target essays was 1.7.

With relation to the monolingual raters, Brislin provides the percentage of back-translated essays falling into the range of 0 to 2

errors (75%), the mean number of errors per essay (2.0) and once more does not specify the range or maximum value for the five essays falling outside of the range of 0 to 2 errors. Finally, Brislin provides the reliability coefficients for the numbers of errors bilingual (r = .76) and monolingual readers (r = .86) made when answering questions on a set of 20 moderately difficult essays on child-rearing, 10 translated into Palauan and 10 into Chamorro, and 10 back-translated into English from each language.

Brislin next moves on to the criteria 4 and 5 results, but I shall first deal with interactions between the results for criteria 1, 2 and 3, since this is how Brislin groups them in his discussion section. I shall then consider the results and discussion relating content, difficulty and language effects, since they are also based on results from the same 300-word essays, before returning to consider criteria 4 and 5, which use different data, and Brislin's discussion of them.

Brislin presents the results for the relationships between criteria 1, 2 and 3 as a table in which the criteria are listed down the left hand side (y-axis) and also across the top of the table (x-axis) and correlations between each pair of criteria given in the cell where their x-axis and y-axis position coincide (shown below, Table 2.2). For example, the top-left cell crosses criterion 1 against itself and therefore provides the inter-criteria correlation – i.e. the reliability for criterion 1. The second cell in column two shows the correlation between criterion 1 and criterion 2, and so on. However, the criterion 3 results are subdivided into monolingual and bilingual raters on the y-axis, but presented as monolingual and combined results on the x-axis. This means that while the table shows intra-criteria reliability for questions about the target (a bilingual test), it does not provide monolingual intra-criteria reliability (back-translation against back-translation). Notwithstanding, the monolingual intra-criteria correlation was provided earlier and was the same as the correlation between monolingual rating and a combined "both translations" category including results from both forward and back-translations (0.86). Quite how a statistic that combines monolingual reliability of 0.86 and bilingual reliability of 0.74 can itself have reliability of 0.86 is not explained, but it is clear that not all of the data used can be the same (for example Brislin may have used only the results from one rater in one case, but averaged two raters' results in another – the level of detail provided is insufficient to speculate further).

Table 2.2 – Intercorrelations between Brislin's criteria 1, 2 and 3

Criterion	Criterion			
	One	Two	Three (Target)	Three (Both Translations)
One: Monolingual Meaning Errors	(.88)			
Two: Bilingual Meaning Errors	.58**	(.75)		
Three: Questions About Target	-.02	.15	(.76)	
Three: Questions About Back-Trans	.76**	.38	.30	(.86)

**p < .01

(Brislin 1970: 204, parentheses indicate intra-criteria correlations)

Only two of the inter-criteria correlations were statistically significant. According to Brislin, the "highest, r = .76, indicates that judging meaning errors in or answering questions about the back-translated versions are similar tasks" and the "correlation between monolingual and bilingual meaning errors suggests that translation quality can be judged by either or both criteria, although different aspects of the two-step back-translation procedure are involved" (Brislin 1970: 204-5).

In the discussion section Brislin returns to the intercorrelations between criteria 1, 2 and 3, saying that although they were imperfect they still, "suggest that similar aspects of translation quality were being judged." He goes on to state that the high correlation between monolingual meaning errors and the comprehension test applied to monolingual English speakers after reading the back-translated essays showed that "it was easy or difficult to answer questions about an essay, depending upon the number of meaning errors." (Brislin 1970: 211).

On the basis of the moderate correlation between monolingual and bilingual meaning errors (criteria 2 and 3), Brislin proposes that what he terms "two minor, common-sense principles" have been demonstrated. These are as follows,

> 1. The goodness of the back-translation is dependent upon the goodness of the translation, but the correlation is not perfect.
>
> 2. Therefore, poor results in the back-translation can be caused by mistakes in either the source to target step, the target to source step, or both.

(Brislin 1970: 211)

He considers that the statistically non-significant correlation between the monolingual and bilingual components of criterion 3 add further

weight to these statements, but finishes his discussion of the criteria 3 intercorrelation by observing that, "perhaps the restricted range of the two-part criterion accounted for the low correlation between parts", or possibly, "the questions about the passages were not difficult enough to differentiate good and poor translation" (Brislin 1970: 211). Certain aspects of the conclusions Brislin drew from his criteria 1, 2 and 3 results are implicit in the presentation and discussion of results from the investigation of content, difficulty and language effects and will be highlighted when these parts of his research are presented, but there is no further explicit interpretation of the results for criteria 1, 2 or 3 in the paper.

The results for criterion 4 relate to six groups of bilinguals who each spoke one of three island languages plus English. Five speakers of each language followed instructions in their own language and five speakers of each language followed instructions in English. The instructions had been translated and back-translated and no criterion 1 errors were detected in the back-translations after the first attempt. There were a total of eight possible errors, but the actual range (of all subjects, not just of the mean for each group) was zero to one. In other words none of the bilinguals made more than one error following any version of the instructions. The mean number of errors for each group of five were as follows: Chamorro speakers following English instructions, 0.0; Chamorro speakers following Chamorro instructions, 0.0; Kusaien speakers following English instructions, 0.2; Kusaien speakers following Kusaien instructions, 0.2; Palauan speakers following English instructions, 0.4; and Palauan speakers following Palauan instructions, 0.2.

Brislin states that the results were statistically equivalent between groups, which indeed they are since five groups were within the first 2.5% of possible variation (≤ 0.2 out of 8.0) and all six groups were in the first 5% of possible variation (≤ 0.4 out of 8.0). On the basis of these results, he states that, "It can be safely assumed that the translations were quite adequate for instructing the subjects to perform the task" (Brislin 1970: 201-2). Brislin discusses the results from criteria 4 and 5 together, so I shall now present the criterion 5 results before moving on to Brislin's conclusions about both sets of results.

Criterion 5 was tested by administering the Crowne-Marlowe scale to four groups and then calculating reliability coefficients from the results. The four groups answered a version of the scale entirely in English, entirely in Chamorro, with odd numbered items in Chamorro and even ones in English or with odd numbered items in English and

even ones in Chamorro. There were twenty subjects in each group and the results analyzed were the number of items for which respondents chose the "true" response. These results are best illustrated in tablular form, as follows.

Table 2.3 – Results for Brislin's criterion 5: means, standard deviations and reliability coefficients

	Second Half English			Second Half Chamorro		
First Half English	Group 1 All English			Group 2 Both languages		
	M	SD	r	M	SD	r
	14.45	5.01	.85	14.00	4.62	.83
First Half Chamorro	Group 3 Both languages			Group 4 All Chamorro		
	M	SD	r	M	SD	r
	13.95	4.95	.86	14.20	5.22	.90

(Brislin 1970: 203)

The correlations are intra-version correlations, i.e. they indicate the level of similarity between the scores for each group of twenty respondents who answered each version of the questionnaire. They do not indicate degree of overlap. Nevertheless, Brislin states that

> The means and standard deviations of the four groups are very similar to the norms reported by Crowne and Marlowe (1964) for comparable groups in the United States. The high correlations for groups one and four suggest adequate English and Chamorro scale reliability, and the correlation for groups two and three suggest satisfactory translation.

(Brislin 1970: 203)

He is therefore suggesting that the fact that twenty different people scored similarly on each of the four versions means that they understood the four versions in a similar manner, which in turn means that the target version has been well-translated.

Brislin then explains that another test had been run on the data from the two groups answering bilingual versions of the scale. He compared the number of subjects who answered "true" to the English versions of the questions with the number who answered "true" to the Chamorro versions, providing a partial view of the "overlap" between subjects. Using total numbers of subjects is not however sensitive to

which subjects answered "true" to which items. For example, if half of the respondents in each language answered true in the sequences 1, 2, 5, 6, 9, 10 etc. and the other half in the sequence 3, 4, 7, 8, 11, 12 etc., their results would appear identical if summarised in this manner, despite half of each group being exactly the opposite of the other half. Nevertheless, the correlation coefficient was r = .89, which Brislin once more says suggests good translation (Brislin 1970: 203).

In the discussion section, Brislin states that, "The success of criteria four and five support the validity of criterion one as a predictor of translations that can be successfully administered.", explaining that he "assumed that the administration of test materials with many criterion one errors would have been unsuccessful." (Brislin 1970: 211).

Specifically with relation to the criterion 4 picture-making task, Brislin states that the low number of errors was particularly interesting since Palauan does not have words for the all colours used and "the Kusaien language demands very difficult specification of spatial orientation and direction". He also considers it interesting that "the six bilinguals involved in the performance task translations had no criterion one meaning errors on the first attempt" despite the instructions being "similar to the poorly translated passages about art." It should be pointed out that the six bilinguals in question were three forward translators and three back-translators, so each absence of errors is the result of two people's work and of two translations, plus the person following the instructions. Brislin speculates on the possible reasons for the absence of errors, suggesting that it was because these were six of the best translators and because they "had already translated the 1800-word set, thus having had a good deal of practice." (Brislin 1970: 211-2). Brislin then summarizes the conclusions he draws from the criteria 4 and 5 results, stating that,

> In these two criteria, responses either to the instructions or to the questions were equivalent across languages. Workable, functionally equivalent versions, [in] English and its translation, have thus been demonstrated.
>
> (Brislin 1970: 212)

The next point he makes is that "an objective and verifiable response can be the standard for a cross-cultural research strategy." and that criterion 1 "was validated against the two response standards" (Brislin 1970: 212).

Brislin then states that he "considers the response criteria four and five the most important aspect of the present research",

acknowledging that Prince and Mombour and Spilka had suggested similar ideas previously (Brislin 1970: 212).

I shall now briefly deal with Brislin's investigation of the effects of content, difficulty and language. While this part of his work does not have a great deal of relevance to the modern practice of cross-cultural adaptation, his presentation and discussion of it throws light on attitudes and assumptions related to the three criteria that have been adopted (one, two, and five) and occasionally gives details about the project as a whole which are not provided in the section dealing with the five criteria. I shall restrict myself to those parts that are of relevance to the subject of back-translation in cultural adaptation.

As explained on page 52 above, Brislin prepared texts on three different subjects and at three different levels of difficulty, although the most difficult level later had to be discarded. This gave him data for a 3 (content areas) x 2 (levels of difficulty) x 2 (languages) ANOVA design[8] which he used to try and detect associations between the number of criterion 1 meaning errors found by a single rater and the three variables content, difficulty and language. Each cell contained the averaged results of ten forward translations and ten back-translations in a given area and language and at a given difficulty level, as judged by a single rater reading the back-translation and the source text.

Before reporting the effect results, Brislin explains that two sets of Palauan essays had to be excluded from his analysis (one set contained twelve essays: one forward and one backward translation for each combination of area and difficulty level). He explains that this was because "several of the essays were almost incomprehensible" and "it was felt that these sets would provide an unrealistic description of the capability of Palauan translators" (Brislin 1970: 205).

Brislin's ANOVA indicated a significant main effect for all three of these variables, with language and content significant to $p<0.001$ and difficulty significant to $p<0.05$. He also detected an interaction between language and content that was significant to $p<0.05$. This means that, according to the t test, there is less than a one in 1000 ($<0.1\%$) probability that the effect of language and content occurred by chance (the null hypothesis) and there is less than a one in twenty ($<5\%$) probability that the difficulty effect and the language against content interaction were due to chance.

---

[8] See page 52 above for a (brief) explanation of the purpose of ANOVA procedures.

Brislin begins his discussion of the main effects with the language effect, saying that the ANOVA data "suggest that translations to Chamorro are better than translations to Palauan." After saying that not enough is known about Palauan and Chamorro to come to a conclusion about reasons for the effect, he then suggests that there is a "plausible explanation [that] lies not in the languages but in the translators." The next few passages are highly relevant to the discussion of back-translation that follows, and so I shall quote them almost verbatim.

Brislin first provides some information about the Chamorro bilinguals from Guam and the Northern Mariana Islands,

> The Chamorro translators had lived in Guam, Saipan or Rota all their lives, being close to the English language due to nearby American military installations and due to contact with an American school system, attended from age six.
>
> (Brislin 1970: 209)

which is in contrast with the Palauans, who "had not had this degree of contact, and were probably not so competent in English as the Chamorro speakers." before suggesting that the language effect may actually be a function of English competence, as follows, "A differential command of English, one of the two languages involved in each translation set, might have been the cause of the significant language effect." (Brislin 1970: 209).

Brislin then provides further data to support this hypothesis,

> Of the 20 Chamorro translators, 13 had completed a two-semester sequence of College English Composition with at least a grade of "C," while only four of the 20 Palauan translators ... has achieved this level of competence. ... One reason why the Palauan students had not completed the English composition sequence is that they had to spend several semesters in courses preparing them for composition.
>
> (Brislin 1970: 209)

Brislin closes his discussion of the language effect by suggesting that, in addition to the effect of competence in English, there might still be an effect from language, related to differences in how much Spanish had been assimilated into each island language. He makes it clear that he himself believed "that both of these factors were operative, but can document only the first, i.e., translators differential familiarity with English."

The next main effect that Brislin discusses is the effect of content. As was explained on page 50 above, Brislin had theorized that content would be more difficult if it was more "involving" and that "facts about a piece of art" would present few difficulties.

However, the actual result was that both Chamorro and Palauan translations of the racial intelligence essays had the least number of criterion 1 errors and the essays on art had the greatest number. For Chamorro translations, both art and child-rearing had significantly more errors that the racial intelligence essays, while for the Palauans all differences were significant; they made significantly more errors in the child-rearing essays than in the essays about racial intelligence and made significantly more errors in the essays about art than in the child-rearing essays. He describes this as a content effect and a content x language interaction, since the content effect was magnified in Palauan translations (Brislin 1970: 205-6).

In the discussion section, Brislin offers two theories to replace his rejected hypothesis that "more involving" equates to harder to translate. The first suggestion is once more based on Nida's concept of "differential familiarity" and it is as follows,

> All the translators had completed an introductory psychology course where child development and the race-intelligence controversy formed units, but the translators had not necessarily had experience with the concepts present in the 'art' passages. The translators might have been more interested in working on the familiar passages.
>
> (Brislin 1970: 210)

It will be noted that this revised hypothesis that familiar passages might be more interesting (and implicitly, easier to translate) is almost the opposite of the original hypothesis that an "involving" passage would be more difficult.

The second suggestion that Brislin makes to try to explain why the art essays exhibited a greater number of errors is that they "contained more detail and were thus more difficult to translate", on the basis that "too much detail in a passage does not allow context and redundancy to ease translation." Once more, Brislin considers that both the "experience" and "detail" factors played a part (Brislin 1970: 210).

The interaction between language and content is explained as follows,

> The greater competence of Chamorro versus Palauan bilinguals
> in translating the art passages caused the content-language

interaction. The Chamorro translators may have had more competence in English, and thus could handle the detailed description in the art passage.

(Brislin 1970: 210)

The results for the third effect – difficulty as gauged by the Flesch Readability Index – are reported as follows, "Difficulty level is also important, with the level purposely written to be easier than the other causing fewer translation errors." (Brislin 1970: 205). In the discussion section, Brislin provides more detail about the third, most difficult level that had been abandoned after pre-testing because, "translators had not been able to start because they were unable to find target language equivalents for the difficult English words", stating that "different levels of difficulty can be written that will cause more or fewer errors in translation, and that rules exist to help write a relatively easy level." (Brislin 1970: 210).

Difficulty was the last of the three effects included in Brislin's original ANOVA design, but he ran an additional analysis of variance with "the six orders of presentation as one variable, content and difficulty being the others", and found that "order did not have even a marginally significant effect, nor did any of the interactions with order." (Brislin 1970: 210). This design was to detect "major effects", i.e. independent variables with an association with (impact on) the number of criterion 1 meaning errors across all other variables.

In the analysis that failed to find a main effect, the six possible orders were taken as one variable and their effect on content and difficulty was analyzed. The language variable was ignored, since no translators translated one language and then the other because no translators translated both Palauan and Chamorro. However, when the data were tabulated, a much simpler order effect was detected:

... it seemed that translating one content area's passage led to improvement on the second regardless of difficulty. That is, regardless of whether it was the more or less difficult version, the second of the two essays seemed to have fewer errors, a practice effect.

(Brislin 1970: 208)

A test based on the difference between the mean number of criterion 1 errors in the first and second translations of each content order was also significant (Brislin 1970: 208). In the discussion section, Brislin says that the "unexpected practice effect" is possibly as important as the difficulty effect and that it was "of both statistical and practical

significance, causing a decrease of two errors per 300-word essay, concluding that, "as in many tasks, translation seems to improve with practice" (Brislin 1970: 210).

Brislin dedicates the remainder of his landmark study to the methodological implications of his results and conclusions, beginning by providing two pieces of advice for achieving accurate translations which had not been previously suggested or had previously lacked supporting evidence. The first piece of advice is in two parts. Firstly, "asking for a translation of material with specific detail ... seems to cause many errors and should be avoided unless the translator is known to be excellent" and secondly, "the researcher should secure translators familiar with the content involved in the source language materials" (Brislin 1970: 212).

The second recommendation is that although "the translation should have no Criterion One meaning errors" before it is administered, "the first translation sequence ... is unlikely to fulfil this requirement", but that by using an iterative back-translation process involving "review of problems, revision of the source, and several more three-step sequences" the "no-error standard" can be met. Nevertheless, when this is achieved, "a pre-test still is necessary" (Brislin 1970: 212).

At this point, Brislin states that "it is desirable to use multiple methods in research demanding translation whenever possible" and signals his agreement with Gough's position that "a researcher cannot depend solely on the back-translation technique" (Brislin 1970: 212).

Mentioning once more that pre-testing detected errors that back-translation had not detected, and stating that this was "probably due to some back-translators being able to make good sense out of target language passages with several errors", Brislin suggests that an additional step be added to his back-translation procedure,

> in which one of the best translators could be asked to read a translation and tell the writer if 1) if the grammar was good, 2) if words are used which most native speakers would understand, and 3) if he felt that other people would have any problem reading the material and answering questions about it.
>
> (Brislin 1970: 213)

leading to a four-step procedure of "original → target → target check → original" which "assumes that the translation-checker is willing to be critical" (Brislin 1970: 213).

Brislin then addresses the claim, attributed to Phillips, that back-translation leads to poor results, saying that Phillips' experience was

possibly the result of the original English not being open to revision or because "Phillips sampled a poor set of translators", explaining that since he (Brislin) had found a range of quality in translations by 94 different translators, "a researcher employing translators from a similar population might obtain good or bad results depending upon which people he hires." (Brislin 1970: 213).

In support of the claim that "back-translation gives insight into the competence of the translators", Brislin mentions that one Palauan set of translations had "the least number of criterion two (bilingual) meaning errors, and had a low criterion three (questions about target) error score", but had a very poor back-translation, "suggesting that the final step ruined a good translation." On the basis of this finding, "the back-translator would not be asked to do any further work" (Brislin 1970: 213).

In response to the observation by Miller & Beebe-Center that there is no unit for assessing translation adequacy, Brislin proposed his criterion 1 meaning error as a unit of translation quality, with the justification that it "is reliable and possesses validity as shown by its prediction of good results with the test materials for criteria four and five" (Brislin 1970: 213-4)

Having defined this (inverse) unit of translation quality, Brislin proposes parameters for an "adequate" translation. Stating that he feels that 50 errors or less per 1800-word essay set "represents good translation", and extrapolating this to mean "eight errors per 300-word essay, or slightly more than one error per paragraph", he refers to the fact that around 50 such errors were removed from the Marlowe-Crowne scale (used to test criterion 5) to justify the claim that "the errors in other essays could also be corrected." (Brislin 1970: 214).

The very last part of the paper is a proposed seven-stage procedure "likely to provide adequate translation from English to other languages". Brislin introduces his procedure with the statement that each step "presupposes an ideal situation" and by accepting that "most research projects would have to change the procedure in response to a given problem." (Brislin 1970: 214).

The first step is to "write an English form that is likely to be translatable." Brislin suggests using "Werner and Campbell's five rules, the Flesch Readability Score, and Thorndike-Lorge word counts", in addition to adding redundancy and context. (Brislin 1970: 214)

The second step is simple, "Secure competent translators familiar with the content involved in the source language materials" (Brislin

1970: 214). No distinction is made between forward and backward translators.

The third step covers both the forward and backward translations, "Instruct one bilingual to translate from the source to the target language, and another to blindly translate back from the target to the source. Allow the translators some practice time" (Brislin 1970: 214).

The fourth step is a combination of his criteria 1, 2 and 3 tests, "Have several raters examine the original, target, and/or the back-translated versions for errors that lead to differences in meaning (meaning errors)." However, criterion 3 appears to have become optional, since the recommendation to "have other raters answer questions after having read only one of the versions" is qualified with "if possible", and an iterative element has been added; "If errors are found, repeat step three, changing the original English when necessary, the process known as 'decentering.'" The fourth step also includes a recommendation on translator employment policy, "Retain or dismiss translators based on a study of the interrelationship among the different criteria of translation adequacy." (Brislin 1970: 214-5).

The fifth step (in full) is as follows,

> When no meaning errors are found, pre-test the translated materials on target language-speaking people. Revise the translation and/or the original English in light of insights gained during the pre-test. Ask a bilingual to critically examine the translation.
>
> (Brislin 1970: 215)

It will be noted that the earlier suggestion (see page 70) by which the critical bilingual assessment was an intervening step in a four step back-translation process has been modified once more. Here, the back-translation (step 3 in the 7-step procedure) has reverted to the original → target → original format and the qualitative bilingual assessment is now an additional stage performed after pre-testing.

The sixth step is actually the final step in the translation procedure *per se*, since step seven is concerned with reporting the results with a view to comparing translation projects. Step six is a version of criterion 5, and is supposed to "finally demonstrate translation adequacy". The process is to "administer the materials to bilingual subjects, some who see the English versions, some who see the translation, and some who see both", and the test of adequacy is that "response should be similar across groups, as assessed by means, standard deviations, and correlation coefficients" (Brislin 1970: 215).

The seventh step was to "Report experience using the different criteria for equivalence." The reasoning behind this is that, after determining "the verdict of translation adequacy derived from the meaning error standard and a simple pre-test" and comparing it with "the verdict derived from the more formal and time-consuming administration to subjects", then if "the verdict is the same for many research projects, future research might only demand the simpler meaning error standard and pre-test" (Brislin 1970: 215).

Back-translation, assessment by raters and pre-testing, in combination with a "decentering" approach, was therefore originally considered to be a possible alternative to validation – the time-consuming administration to subjects.

It will have been noted during the course of this subsection that Brislin actually described several different versions of his translation method. The first version was the method used to translate the six 300-word essays and included the tests of his three criteria for equivalence (see pages 51 to 55). This first formulation is not, strictly speaking, a full translation method since the translations are not modified in the light of the findings and decentering was not employed other than the very crude techniques of discarding the most difficult source essays and the weakest forward and back-translated essays and only translating the two easier essays for each content area. The second version was the entire process of producing materials for criterion 4 and testing them (see page 55). The third version of Brislin's method was used to produce materials for criterion 5 and test them and was the first to incorporate "iterative" back-translation (see pages 56 to 58). The final version of the translation method is the seven-step approach described in the preceding paragraphs. In addition to these four different sequences of translation, back-translation and testing, Brislin also made the recommendation that a bilingual review stage should be included in the back-translation step (see page 70), but he did not describe a translation process that used the resulting four-step back-translation phase. Table 2.4 below summarizes the four different versions of Brislin's method.

Table 2.4 – Four back-translation methods described by Brislin (1970)

Criteria 1 – 3	Criterion 4	Criterion 5	7-step method
Ten bilinguals[*] produce 10 forward translations of six essays, two each on child care, race and art.[*]	One bilingual[*] produces one forward translation[*] of instructions.	One bilingual produces one forward translation of questionnaire.	An English form is written to be likely to be translatable
Ten bilinguals produce 10 back-translations of each forward translation.	One bilingual produces one back-translation.	A second bilingual produces one back-translation of questionnaire.	Competent translators, familiar with the content involved in the source language materials, are contracted.
One or two[†] monolingual raters judge errors in the back-translations with reference to the original.	One monolingual rater[**] judges errors in the back-translation with reference to the original.	The original is decentered by researcher plus translators one and two in response to any errors identified by a single[**] monolingual rater.	One bilingual translates from the source to the target language.
Inter-rater and intra-rater agreement are analyzed.	The number of errors is counted.	A third bilingual produces one forward translation of decentered questionnaire.	Another bilingual blindly translates back from the target to the source.
Two bilingual raters[*] judge errors in the forward translations with reference to the original.	No further errors identified by monolingual rater[††]	A fourth bilingual produces one back-translation of decentered questionnaire.	Several raters examine the original, target, and/or the back-translated versions for errors that lead to differences in meaning.
Inter-rater and intra-rater agreement are analyzed.	Ten speakers of the target language follow the instructions; half follow the original version and half the translation.	The decentered original is modified again by researcher plus translators three and four in response to any errors identified by monolingual rater.	Other raters answer questions after having read only one of the versions.
Two bilingual respondents[*] answer questions about the forward translation.[‡]	The number of errors made is counted for each language.	Second forward translation is pre-tested with 10 target language natives	If errors are identified, decenter original then repeat forward and backward translations and rate them.[††]
The number of errors made is counted.	The numbers of errors made following instructions in each language are analyzed.	The twice-decentered original is modified again in response to pre-test results.	Translators are retained or dismissed on the basis of the interrelationships of the different criteria
Two monolingual respondents answer questions about the back-translation.		A fifth bilingual produces one forward translation of decentered questionnaire.	Pre-test on target-language speakers.
The number of errors made is counted		A sixth bilingual produces one back-translation of questionnaire.	Revise translation and/or decenter original in light of pre-test results.
The results of each stage are analyzed for associations with each other.		No further errors identified by monolingual rater[††]	Ask bilingual to critically examine translation
		All six translators review final translation, comparing it to last decentered original, and declare it finished.[‡‡]	Administer materials to bilingual subjects – some see English versions, some translations, some both.
			Analyze means, standard deviations, and correlation coefficients.

* Per target language.
† Brislin excluded the second rater's results from the majority of the statistical tests he employed.
‡ It is presumed that the bilinguals answered questions in English, but this is not made clear in Brislin's text.
** This is not actually specified in the text, but since no mention is made of inter-rater comparisons and the only mention of Brislin himself rating was related to the essays, I am assuming only one rater was used in each case.
†† Brislin makes it clear that the process must continue until a monolingual rater finds no errors in a back-translation. In his study this occurred at the first and third iterations for criteria 4 and 5 respectively and in the 7-step method no limit is specified.
‡‡ The entire process to produce the Chamorro version of the 33-item, 551-word Crowne-Marlowe scale took "more than 200 man-hours" (Brislin 1970: 200).

(Abridged and adapted from Brislin 1970)

The differences between the four methods described in full make it clear that when studies simply state they have translated "as per Brislin (1970)", "according to Brislin (1970)" or "as recommended by Brislin (1970)", but do not specify which version, the actual translation techniques used are unknown.

2.2.2   *Later work by Brislin*

In addition to "Back-translation for cross-cultural research", Brislin wrote three chapters discussing back-translation that were published in cross-cultural psychology textbooks. All are now out of print and the preface of the last of these three books states that the first of them was already out of print in 1986 (Lonner & Berry 1986: 11).

None of these chapters alter the fundamental assumption of the back-translation technique (that comparison of a back-translation with the source text for an intermediate translation can be used to evaluate the intermediate translation). They do however contain additional information on the original experiments, some changes in emphasis, particularly in the 1980 and 1986 chapters, certain modifications to the recommended back-translation procedure and further clues to the way that their authors conceptualise translation. Perhaps even more informative than the additions is a tendency to omit or suppress an increasing proportion of the methods originally recommended, to the extent that the most recent chapter (1986) does not discuss criterion 2 (comparing source with translation), criterion 3 (comprehension testing), criterion 4 (performance instructions) or criterion 5 (split-half testing) and makes no mention of anything like a "unit of translation quality".

The first of these three chapters appeared in a book co-authored with Walter Lonner and Robert Thorndike, entitled "Cross-cultural research methods" (Brislin et al. 1973). This books' chapters are not credited to individual authors and in the text of the chapter in question the authors refer to themselves in the plural of both the first and third persons (and never in the singular), so from a strict bibliographical perspective the chapter entitled "Questionnaire wording and translation." is by Brislin et al., rather than Brislin alone. However, the content of the chapter is almost exclusively derived from Brislin's doctoral thesis and research he conducted immediately afterwards on the island of Guam.

This chapter is listed in the bibliographies of at least one of the articles describing each of the three contemporary cross-cultural adaptation methods that I shall present in subsections 2.3.1 to 2.3.3

below and Google Scholar lists it as having been cited by 1300 articles (Google Scholar 2012a).

The second textbook chapter is entitled "Translation and content analysis of oral and written material." and was published in 1980 in the textbook "Handbook of cross-cultural psychology", edited by Harry Triandis and John Berry. This chapter is in the references describing the Sperber method (section 2.3.1 below), but is not cited by the authors of either of the other two methods. Notwithstanding, Google Scholar lists 1583 citations (Google Scholar 2012b).

The most recent original work[9] Brislin published on back-translation is a chapter entitled "The wording and translation of research instruments." It was published in a textbook entitled "Field methods in cross-cultural research" (Lonner & Berry 1986).

None of the publications that cover the contemporary cross-cultural adaptation methods described in this thesis (subsections 2.3.1 to 2.3.3) list the latest chapter in their references, preferring the earlier works, but Google Scholar lists 1310 citations (Google Scholar 2012c) and I have seen it referenced in many articles describing individual cross-cultural adaptation projects, including the handbook for the EORTC method, which was used as the basic model for the translation project that Ozolins describes and which I discuss in subsection 2.4.2 below (EORTC 2009; Ozolins 2009).

I shall now describe the principle additions and modifications to the back-translation technique, as defined in the three textbook chapters, showing how certain elements (such as writing with translation in mind, decentering, collaboration with translators and the need to document results) received increasing emphasis in successive publications.

I shall start with additions and developments related to the theory and concepts behind back-translation and cross-cultural adaptation in general and in order to do so it is first necessary to introduce a concept from cross-cultural psychology. This concept is known as the "etic-emic distinction" and it is the conceptual basis for a great deal of the techniques and basic hypotheses used in the field.

---

[9] In addition to these three book chapters, and very much more recently, Brislin also co-authored an encyclopaedia entry entitled "Back-translation: a tool for cross-cultural research" (Brislin and Freimanis, 2001). However, the material contributed by Brislin is quoted almost verbatim from the earlier publications, drawing heavily from the 1980 and 1986 book chapters and contains no additional methods or recommendations. The material by Freimanis describes back-translation as applied specifically to the language pair English-Chinese and I am not qualified to comment on it.

The concept was introduced to cross-cultural psychology in 1969 by Berry and differentiates between "*etic* constructs – those that exist in identical or near identical form across a range of cultures – and *emic* constructs that are limited to one culture" (Behling & Law 2000: 3) The chapter on questionnaire translation is the second chapter in the 1973 textbook and does not define these terms. However, the first chapter, entitled "Introduction" and ostensibly co-authored by Brislin, introduces the concept using the following example of the pitfalls of mistaking emics for etics, quoting from Wesley & Karr as follows:

> If, for instance, a German and an American mother are both asked how they would punish misbehavior, both may give the same answer. However, the German mother may have interpreted "misbehavior' to mean 'being ten minutes late for dinner,' the American mother perhaps as 'not coming at all.' The mothers might have given entirely different answers had the misbehavior been specified to mean 'ten minutes late for dinner' ... Mothers of two cultures may answer one and the same question with, for instance, 'moderate punishment.' By 'moderate punishment', the mother of one culture may mean a verbal scolding, the mother in the other culture, a physical punishment such as a slap in the face.
>
> (Wesley & Karr 1966: 260, quoted in Brislin et al. 1973: 25)

Interestingly, a footnote states that the actual terms, "emic" and "etic" have their origin in linguistic terminology, where "Phon*etic* notation is meant to be a general system which can describe all sounds in all languages, while phon*emic* relates to sounds that are meaningful in a given culture." (Brislin et al. 1973: 25, footnote).

In the 1980 chapter, Brislin considered a suggestion made by Przeworski and Teune that cross-cultural questionnaires should contain a core set of etic items relevant to all cultures and a set of culture-specific items, different for each culture and designed for the emics of each (Przeworski & Teune 1966 & 1970, discussed in Brislin 1980: 393).

In such an approach, "a set of questions would be compiled to measure the specific concerns of each culture", which Brislin considered better than using "questions based on imposed or even derived etics" (Brislin 1980: 393). He points out that there would then be no statistically interpretable cross-cultural relationships between the emic components, but that significant interrelations would be expected between etic and emic components within each culture (Brislin 1980: 393).

The 1986 chapter begins with a discussion of the relative merits of translating "existing instruments"[10] and writing new instruments specifically to be translated and applied in two or more cultures. This discussion employs the emic-etic distinction.

Before dealing with the implications for translation of using existing or newly-developed instruments, Brislin considers some general points for and against using existing instruments. He begins with the advantages, which he describes as "considerable". The first of these is that using an existing measure that has been used previously means that there would be existing data available for comparison, allowing "a literature to be built up around a commonly shared set of concepts and operational definitions". The second point Brislin makes in favour of using existing instruments is to save time and expense, since "time, energy and funding" are always limited (Brislin 1986: 138).

The third point is ostensibly still part of the list of points in favour of existing instruments, but leads in to the reasons for avoiding them. The advantage here is a "sense of security" offered by using an established measure, based on the idea that "if some established researcher has used a certain measure and obtained respectable publications, then it must have merit" and "can't possibly be the target of criticism or improvement", but Brislin warns that "one should not be lulled into a false sense of security through choice of a popular existing device." (Brislin 1986: 138).

The first disadvantage is linked to the emic-etic distinction and is that researchers using existing instruments may miss important aspects of phenomena "as viewed by (and seen as important by) people in other countries" and risk imposing conclusions based on concepts that do not exist in the culture they are working in, since "existing instruments provide operational definitions of certain concepts" and it cannot be guaranteed that either the concepts or their definitions are common to both/all cultures under study (Brislin 1986: 139).

The emic-etic concept is central to the process of "decentering" (see page 43 above), which itself developed out of the attempts to define rules for writing "translatable English" that were discussed in the 1970 paper (see page 41 above). The original 1970 paper listed five rules for writing translatable English and these increased to ten in the 1973 chapter and then to 12 in the 1980 and 1986 chapters. In the 1980

---

[10] Brislin explains that he uses "existing instruments" as a "shorthand term to designate measures which were developed and standardized in one culture and can possibly be used for data gathering in another culture" (Brislin, 1986, p. 138).

chapter, the twelve rules are presented as a simple list of statements, but in the 1986 version the statements are accompanied by explanations and clarifications running to several hundred words per rule.

Below are the final twelve rules, as shown in the 1980 and 1986 chapters. Rules 6 and 11 were added in the 1980 chapter.

1. Use short, simple sentences of less than sixteen words.

2. Employ the active rather than the passive voice.

3. Repeat nouns instead of using pronouns.

4. Avoid metaphors and colloquialisms. Such phrases are less likely to have equivalents in the target language.

5. Avoid the subjunctive, for example verb forms with 'could', 'would', 'should'.

6. Add sentences to provide context for key ideas. Reword key phrases to provide redundancy.

7. Avoid adverbs and prepositions telling "where" or "when" (e.g., frequent, beyond, upper).

8. Avoid possessive forms where possible.

9. Use specific terms rather than general terms (e.g., the specific animal, such as cows, chickens, pigs, rather than the general term "livestock").

10. Avoid words which indicate vagueness regarding some event or thing (e.g., probably and frequently).

11. Use wording familiar to the translators

12. Avoid sentences with two different verbs if the verbs suggest different actions.

<div align="right">(Brislin. 1980: 432; Brislin 1986: 144-149)</div>

Although rule 6 was added in the 1980 chapter, Brislin also recommended increasing redundancy and context in 1970 and 1973, but not as part of these rules. In 1973 he quoted George Miller's statement that "redundant information is an automatic mistake catcher built into all natural language" (Miller 1953: 8, quoted in Brislin et al. 1973: 35) and Chapanis' claim that a "word is much harder to understand if it is heard in isolation than if it is heard in a sentence" (Chapanis 1965: 73, quoted in Brislin et al. 1973: 35), which had accompanied a set of rules Chapanis had written for "language to be used in adverse conditions" (1965: 75). Chapanis' rules are not relevant, but the comment that

follows is especially relevant to the subject of back-translation: "The writers consider the translation process as an example of an adverse, difficult state of affairs" (Brislin et al. 1973: 35).

The twelve rules are introduced as having been developed "on the basis of experiences preparing research instruments in over twenty languages" and Brislin states that their purpose is to ensure that the *translators* will:

1 have a clear understanding of the original language item;

2 have a high probability of finding a readily available target language equivalent so that they do not have to use convoluted or unfamiliar terms

3 be able to produce target language items readily understandable by the eventual set of respondents

<div align="right">(Brislin 1986: 143)</div>

Brislin then suggests that there are additional benefits to following his guidelines, the first two of which are that the resulting items should be (a) "readily understandable to respondents" and (b) "understandable to researchers in other cultures". Since "items on a questionnaire or survey instrument constitute the operational definitions of concepts" and "more and more research is [now] done by members of cultures who previously were only the hosts for visiting research teams", one additional application of the guidelines for writing "readily understandable English" is "for researchers to communicate their findings to widely dispersed colleagues", since "communication with a broad, world-wide audience is dependent on publication in English" (Brislin 1986: 143-144).

I shall now present the most relevant parts of Brislin's justifications for each of the twelve rules for writing translatable English, some of which extend to several hundred words.

Rule 1 was to construct sentences of no more than sixteen words. No reason is given for the specific number of words chosen, but the reason given for keeping to a limit is so that each sentence only covers "one dominant idea" because multiple ideas would be "difficult to disentangle" with subordinate clauses causing confusion. Brislin also points out the difference between items and sentences, saying that an item can contain multiple sentences (Brislin 1986: 144).

Rule 2 was to avoid the passive voice and the majority of the justification for this is based on the statement that using the active voice makes it easier for *the translator* to "identify subject, verb and object"

and "match adjectives and adverbs to the appropriate nouns and verbs" (Brislin 1986: 144).

Rule 3, to avoid pronouns, is justified because following it avoids "weak noun-pronoun links" and because many languages have "far more pronouns" than English (Brislin 1986: 145).

Rule 4 was to avoid metaphors and colloquialisms and the only justification for its inclusion is that "such phrases are unlikely to have equivalents in the target language". The remainder of the discussion of this rule is dedicated to explaining that single-country standardized tests are actually very likely to contain such features because they are "very good for communicating *within* a community" and therefore do well in statistical validation tests in their country of origin (Brislin 1986: 145). Later in the same chapter, Brislin touches on the subject again, this time with relation to translating existing questionnaires, saying that "Feeling blue", as an example of idiomatic wording, might be best translated as "the more direct 'depressed'" (Brislin 1986: 153).

The rationale for Rule 5, to avoid the subjunctive, is basically the mirror image of the reason for avoiding pronouns. Whereas other languages may have more pronouns, Brislin says they "rarely have readily available terms for the various forms of the English subjunctive." arguing that by using the subjunctive, researchers "force the translator" to guess or to approximate. The discussion of this rule ends with a very interesting statement, "Assuring clear communication is the researcher's job, and it should not be carelessly delegated to translators." (Brislin 1986: 145).

The discussion of Rule 6, adding context and redundancy, begins with context, but goes well beyond the justifications for this and amounts to a procedure for preparing a special version of a single-country instrument especially for translators. Such a version would be extended by providing additional context and redundancy, but the translator would "not necessarily have to provide target language equivalents for every word" (Brislin 1986: 145). In other words, at least some of the context and redundancy is purely for the benefit of the forward translator and is not intended to be used in the translation, or, at least, the translator is not obliged to include it in the forward translation.[11]

---

[11] Brislin does not take up the question of how the (monolingual) person rating the resulting back-translation against the source text, with all its additional context and redundancy, would be able to determine whether elements that have not been carried over to the back-translation were "redundant", and not therefore errors of ommission.

The additional information needed could be as extensive as discussing the origins of the superstition of "avoiding stepping on cracks in the sidewalk" and, for countries where sidewalks do not have cracks, for example, "the context would include the reason why this item is on the original test" in order that "a cultural equivalent of avoiding sidewalk cracks can be found" (Brislin 1986: 145).

Brislin provides another example of the type of additional context that might be needed, which is providing markers of social status when researching in Asian countries, but it is clear that in this case the extra information would have to be carried forward into the target-language version and would not have the same "disposable" status as the additional context provided exclusively for the benefit of the translators, since it would be needed by the target population (Brislin 1986: 145).

The second part of Rule 6, adding redundancy, is "to help translators catch mistakes", on the basis that "if unsure of the meaning of an item from one phrase, they may find it in the redundant information found in another phrase."(Brislin 1986: 146).

The basic justification for "avoiding adverbs and prepositions telling 'where' and 'when'" (Rule 7) is that "there are often inadequate direct equivalents of these words", which Brislin illustrates with the question "how often does an event have to occur for the word 'frequently' to be used?" (Brislin 1986: 146). However, in this case he suggests that the differences between languages are the result of differences between cultures, citing work that demonstrated how perceptions of time differ across different cultures (Hall 1959, quoted in Brislin 1986: 146).

Brislin lists three reasons for following Rule 8 (avoiding possessive forms). The first is that to assume native speakers of English will have the same "concepts concerning ownership" as native speakers of other languages is "an example of cultural imposition" (Brislin 1986: 146). The second and third reasons are that "especially in long sentences, translators may find difficulty in matching what is 'possessed' with who is 'doing the possessing'" and that the English possessive "yours" may need to be rendered as "one of three forms in many other languages (corresponding to single, dual and multiple referents)" (Brislin 1986: 146-7).

The justifications for Rule 9, using "specific rather than general terms", is that "people in various language communities do not categorize specific items in the same manner ... Rather, the items are grouped in different ways across cultures". Brislin provides two

examples of this. The first, "livestock", is an example of a general term that refers to different items in different cultures. The second is an example of how what is a single item in one language may consist of more than one item in another. Brislin states that Japanese has two terms for brother, one meaning older brother and the other meaning younger brother, saying that there is no general term for both (Brislin 1986: 147).

This issue then leads to another concern; how does one ask the same question in two countries, one with a single term and the other with two distinct terms? In contrast with the inclusion of information on social status, which Brislin suggested as possible additional context (Rule 6) for an instrument designed for "Asians", but which he did not recommend adding to an English version, here (Rule 9) the recommendation is that "researchers should then ask English-speaking respondents about interpersonal relationships with older and younger brothers," proposing that were it to then be found that "English speakers do not make major distinctions ... while Japanese do", this would be "an important finding" (Brislin 1986: 147).

The third reason given for using specific rather than general terms is related to the psychometrics of questionnaire-based research. Brislin cites work by Ajzen and Fishbein, stating that they have shown that "behavior is best predicted ... when specific rather than general questions are asked" (Ajzen and Fishbein 1977 & 1981), returning to the livestock example to justify this by saying that people "behave differently regarding treatment of chickens, cows, sheep and pigs" meaning that questions on "livestock" should be replaced by "inquiries about those specific animals" (Brislin 1986: 148).

Rule 10 is to "avoid words indicating vagueness about some event or thing" and while Brislin relates it to Rule 9 on specificity, the justification is actually closer to that for Rule 7 (avoiding adverbs and prepositions telling 'where' or 'when'). The reason for employing Rule 10 is as follows: "Even when there are seemingly equivalent terms to words like 'probably' or 'maybe', the number of times an event has to occur to be labeled 'probable' may differ from culture to culture". Here, Brislin acknowledges that the problems caused by vagueness also affect monolingual research, but in the case of cross-cultural research, he claims that the lack of "shared information and shared experiences" limits the extent to which interviewer-respondent interaction can overcome them. The solution Brislin offers is to be specific, for example by asking about, "the number of times per hour, day, month or year" (Brislin 1986: 148).

Rule 11 is the least specific of Brislin's 12 rules for writing translatable English. While other rules recommend avoiding certain parts of speech, or limit the number of verbs or total words per sentence, here the guidance is to "use wording familiar to the translators". The first reason given in support of this recommendation is that familiarity will allow translators to create a "well-worded target language version", which, in turn, will be "readily understandable" by respondents. A second advantage of this approach is that "translators are treated more like colleagues than hired help" (Brislin 1986: 148).

The passages that follow make it clear that Brislin was not recommending using language with which translators are already familiar, but suggesting that translators should be familiarised with the actual wording they will be asked to translate, stipulating that researchers "should sit down with translators and go over the materials to be translated, line by line." The element of treating translators as colleagues extends beyond affording them professional respect and Brislin says their contributions should be considered as "good data", so that if a translator has difficulty with a given phrase, the researcher should explain "the underlying purpose of [the] item" before "researcher and translator together" work to find a "culturally equivalent item" (Brislin 1986: 149).

The twelfth and final rule for writing translatable English is to avoid "sentences with two different verbs if the verbs suggest two different actions". The first justification given for this is to avoid translators having difficulty "attaching the relevant subject to the appropriate verb". The second reason applies equally to monolingual research, since Brislin suggests that using more than one verb in a sentence makes "interpretation of respondents' answers difficult" since "researchers will have difficulty ascertaining why respondents endorsed or rejected a given item" because they could be responding to "one of the verbs, the other, or a combination" (Brislin 1986: 149).

The added attention to preparation of materials for translation and the increased role given to translators in the problem-solving phases of the process are the most notable developments in the overall progression from back-translation as described in the 1970 to its incarnation in the 1986 book chapter. The original article did recommend following 5 rules for writing items for translation, but ended by translating an existing questionnaire that had not been written according to those rules. Furthermore, in the 1970 version of the translation process, the translators' status was closer to "hired help" than "colleagues".

In the 1986 chapter, Brislin returns to the translation of existing instruments and discusses other authors' work in which modifications were not made to the existing source text items, but to the translated versions of those items, meaning that his preferred technique of decentering was not used. Brislin quotes a personal communication from Gough that was also quoted in the introduction to "Cross-cultural research methods" (Brislin et al. 1973: 26). Gough had translated his California Psychological Inventory (Gough 1969) and stated that,

> Most personality assessment material has some sort of diagnostic rationale ... In translation it is this intent that must be maintained, not the content. Thus, translators must know the infrapsychology of the tools they are converting and they must know the empirical connotations of an item as well as its linguistic and literal referents.
>
> (Gough, quoted in Brislin et al. 1973: 26 and Brislin 1986: 150)

In the 1973 publication this quotation is part of the introduction, but by 1986, Brislin had adopted it as part of his arguments in favour of his translation method and had also elaborated on it, adding the following comment, "Of course, the person who instructs the translator about the 'infrapsychology of the tools' and 'empirical connotations' is the researcher in charge of the project" (Brislin 1986: 150).

Brislin then makes a general recommendation about modification of existing items (which he credits to Butcher & Garcia 1978, and Butcher & Clark 1979), stating that "the important thing to keep in mind is the detailed knowledge of the target culture necessary to modify items and to insure good, usable translations"(Brislin 1986: 151).

Another recommendation for modifying items in the translation is drawn from work by Butcher and Garcia and relates to difficulty translating items with "negative wording". Brislin gives the example, "I can read for a long time without tiring my eyes", and says that the solution was "to rephrase the item and to reverse the scoring key for the item."(Butcher & Garcia 1978, quoted in Brislin 1986: 151).

To explain for readers who are unfamiliar with questionnaire-based research, the result of this in practice would be that if, for example, the original response scores for the statement above had ranged from 1 for "agree completely" to 5 for "disagree completely", then a replacement item worded "My eyes get tired when I read for a long time" would be scored from 1 for "disagree completely" to 5 for "agree completely". The relationship of lower numbers indicating better

vision and higher numbers indicating poorer vision would therefore remain unchanged and statistical analyses would be unaffected.

The mention of item scoring leads on to a discussion of statistical analysis of the data, which "will almost always be done 'back home,'" and will be the basis for conclusions about the "adequacy of the translated instrument" (Brislin 1986: 152).

The final subsection of the 1986 chapter is ostensibly about developing new items, but the majority is actually a presentation of several different authors' contributions on the desirability, or otherwise, and the simplicity or difficulty of creating new items and instruments.

Their opinions on the ease or difficulty of creating new instruments range from the position taken by Burisch who considered "two hours and a bottle of wine" sufficient resources to construct a valid scale, to the opposing position of Gorsuch, who warned that scale development "should be left to those with advanced training" and was "not a task for the average master's level project" (Burisch 1984: 219 and Gorsuch 1984: 235, quoted in Brislin 1986: 155). As would be expected, Gorsuch defends using established scales, whereas Brislin's one-time colleague Campbell likens this type of "loyalty" to someone clinging to a "magic bundle of rituals of which he does not dare disturb any part" and dismisses it with the verdict that "this is not science" (Campbell 1968: 255, quoted in Brislin 1986: 154)

The 1986 chapter closes with a description of Brislin's final version of his translation procedure, but before presenting the 1986 iteration of Brislin's technique, I shall describe the changes that occurred between 1970/73 and 1980.

The classic 1970 paper introduced five methods for producing and/or testing translations to be used in research, which Brislin termed criteria 1 to 5 (see pages 47 to 50 above). The 1973 book chapter repeated all five of these criteria unchanged, but in the 1980 chapter the terms criteria and criterion had been abandoned and some of the criteria themselves were also omitted.

Rather than describe criteria for testing translation quality, the 1980 chapter returns to the four "basic translation methods" mentioned early on in the 1970 paper and credited to an unpublished paper by Brislin in conjunction with Campbell, Stewart and Werner (see page 43 above). The first of these, "back-translation", is described in the same way as in the 1970 and 1973 publications and is the same as criterion 1. The second, "the bilingual technique", is basically the same as criterion 5 from the 1970 paper, i.e. split-half testing of bilinguals (see pages 56

to 58 above). The third is "the committee approach", of which Brislin had always seemed mistrustful, and here he states that "the weakness of the method is that committee members may not criticize one another, and may even unify against the researcher" (Brislin 1980: 431). The final technique, pretest procedures, was also covered in the 1970 paper (see pages 45 to 45 above), but whereas it had been an integral part of the translation process in the 1970 version, the 1980 chapter (very briefly) describes field testing after the translation is complete "to insure that people will comprehend all material to which they will be expected to respond" (Brislin 1980: 431).

No mention is made in the 1980 chapter of criterion 2, in which bilinguals compare source texts with translations, of criterion 3, comprehension testing, or of criterion 4, the performance test. The methods used for criteria 2, 3 and 4 are also missing from the 1986 chapter, which mentions pretesting only very briefly, deals with bilinguals only in relation to new item creation and does not refer to the committee approach in any way. In the 1986 chapter, Brislin has abandoned his recommendation that researchers should use as large a number of different methods as possible and he advocates just one method for preparing and testing translations, which is introduced under the title "Translation: a recommended procedure" and consists of "back-translation, and its close relative, decentering." (Brislin 1986: 159).

The 1986 description of back-translation is relatively unchanged from previous versions, "one bilingual translates from the source to the target language, and another blindly translates back to the source", but the reiterative series of cycles, demanding different translators at each step, is made explicit with the words "the procedure can be repeated for several rounds" (Brislin 1986: 160) and in a diagram (Figure 2.1 below).

Figure 2.1: Back-translation in Brislin (1986)
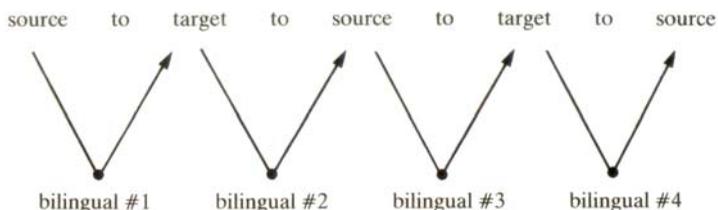
(Reproduced from Brislin 1986: 160)

The 1986 description of the actual steps involved in back-translation employs the emic-etic distinction and the implication is that items that are not translated correctly after several back-translation cycles are emic and will therefore be eliminated according to the decentering principle. This is implied in the statement that "if a concept 'survives' the decentering procedure, it is assumed to be *etic* since there must be readily available words and phrases in the two languages", but is not stated explicitly. Continuing the tendency to treat translators more as colleagues than as hired help, the 1986 version of the procedure recommends consulting them and conducting "extensive discussions" to reveal the reasons why "materials were and were not translatable" (Brislin 1986: 160).

Another new element in this description is that the "original language version which will eventually be used in data collection ... will often be the final back-translation", on the basis that it "is most likely to be equivalent to a target language version, probably the version immediately preceding the final step". However, while this had not been recommended in the text of Brislin's earlier publications, it appears that it is what he had actually been doing in the research described in 1970, since the example he provides is from the Crowne-Marlowe scale that he translated for his criterion 5 test (see page 63 above). The example is an item that read, "I have never intensely disliked anyone". After translators had problems with "intensely", the decentered version actually administered (in the source culture) was "I have never really disliked someone". Brislin cites Campbell to defend the validity of this on the basis that "if changes from 'intensely' to 'really', and from 'someone' to 'anyone', change the underlying concept so much that there is differential response, then the underlying concept is weak" (Campbell 1964, cited in Brislin 1986: 160-161).

Nevertheless, Brislin accepts that this position is not consensus and suggests that researchers who "are worried [about] noncomparability with previous studies which have used the original version ... can gather data using both versions." (Brislin 1986: 161).

The final addition to Brislin's method described in the 1986 chapter is an extra (optional) step for use when the target version includes "phrases which are unfamiliar to the sample of respondents" or if "the researcher is concerned that the back-translation is better than the target-language version." This step is to ask a monolingual "similar to the respondents to rewrite the material so that it will be clear to native speakers" (Brislin 1986: 162).

This revised model is also illustrated in a diagram
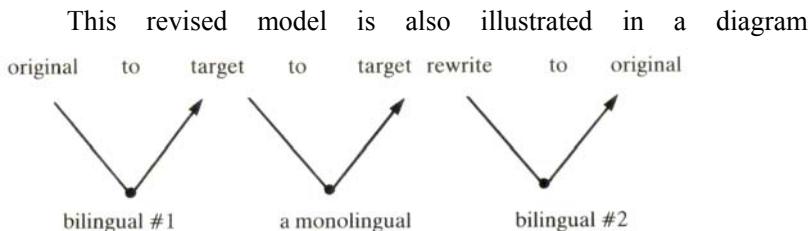
(below),
although, where the first diagram showed two "source text → forward-translation → back-translation" cycles, this diagram depicts one cycle of "source text → forward-translation → rewriting → back-translation". This is presumably because at least one "source text → forward-translation → back-translation" cycle would have had to have been completed before either of the conditions requiring a rewriting step could be detected.

Figure 2.2: Back-translation with monolingual rewriting in Brislin (1986)
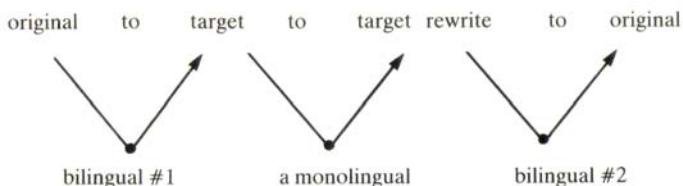
(Reproduced from Brislin 1986: 162)

In summary, the major changes in Brislin's position from 1970 to 1986 are as follows:

- The rules for writing translatable English have increased from 5 to 12, incorporating the addition of context and redundancy, and are now viewed as almost a prerequisite for preparation of the source material to be translated.

- The attempt to use meaning errors as units of translation quality has been abandoned.

- Criteria 2, 3, 4 and 5 are no longer recommended as tests of translation, although a version of criterion 5 is described.

- The translators, who in the 1970 paper and 1973 chapter had been students, paid ten dollars for participation and viewed as

experimental subjects, are now expected to provide valuable insights and should be treated as colleagues.

- Emic-etic distinctions and, as a result, decentering have become central to the entire translation method and the suggestion is made that the last back-translation should be used for data collection in the source culture.

- A new recommendation to add local emic items to tap culture-specific concepts in the target culture has been introduced.

- In contrast, items that are "not translatable", defined as not appearing in the back-translation, are considered to be emic to the source culture and are modified in the source language version.

- The bilingual approach and the committee method, which survived up to 1980, have not been included in the latest chapter.

<div align="right">(Summarised from Brislin 1970, Brislin et al. 1973, Brislin 1980 and Brislin 1986)</div>

The next three subsections of this thesis present contemporary cross-cultural adaptation methods that have been developed taking Brislin's pioneering work as a starting point. These methods have all been developed to translate health-related questionnaires and all of them include a back-translation stage.

2.3 Beyond Brislin: contemporary use of back-translation for cross-cultural adaptation of health-related questionnaires

There are a large number of contemporary protocols for cross-cultural adaptation of health-related questionnaires and they invariably employ back-translation. There are, however, three methods that have gained greatest acceptance,[12] developed by three teams of researchers. None of them are the work of a single person and all of them have been developed as part of projects designed to produce research materials for specific uses. None of them have the speculative element of Brislin's original work and two were first described in articles reporting on specific translation projects, while the third was introduced in an article that reviewed seventeen different projects to cross-culturally adapt quality-of-life measures. Two of these groups of researchers are loosely

---

[12] Based both on the number of citations they receive in articles describing cross-cultural adaptation projects and also on the basis of my experience of articles and questionnaires that I have translated myself.

related, in that they have certain members in common and both focus on quality-of-life measures.

The first of these methods is the one published by a team with no links to the other two. It was first described in an article entitled "Cross-cultural translation - methodology and validation", published in the "Journal of Cross-Cultural Psychology" - the same journal in which Brislin published in 1970 (Sperber et al. 1994). This method was further developed in "Translation and validation of study instruments for cross-cultural research", published in the journal "Gastroenterology" (Sperber 2004). The earlier of these articles makes frequent reference to Brislin's work, the second to a lesser extent. I shall refer to this method as the Sperber method hereafter.

All of the articles describing the other two methods have links to work conducted while translating the Nottingham Health Profile questionnaire (Hunt et al. 1991). Interestingly, that project did not use back-translation. Both of these methods have been developed to adapt quality-of-life measures, both involve large teams of professionals in several different research centres and both demand considerably more resources than the Sperber or Brislin methods (whichever version of the Brislin method is chosen).

The first method for translating quality-of-life measures was developed by the International Quality of Life Assessment project in order to cross-culturally adapt the SF-36 Health Survey. It is described in "Translating health status questionnaires and evaluating their quality: The IQOLA project approach" (Bullinger et al. 1998). I shall refer to this as the IQOLA method hereafter.

The second of the quality-of-life methods was first proposed in an article that did not describe a particular adaptation project, but was a review of the methodology employed in seventeen different adaptation projects. That paper, written by Guillemin, Bombardier and Beaton, provided guidelines for the cross-cultural adaptation process (Guillemin et al. 1993). In 2000, these three authors plus Ferraz published "Guidelines for the process of cross-cultural adaptation of self-report measures" (Beaton et al. 2000). This article describes the translation method adopted by the outcomes committee of the American Association of Orthopaedic Surgeons (hereafter AAOS), although at a certain point in the article, Beaton et al. state that they are describing the method used by the IQOLA. Notwithstanding, the actual method they describe has important differences from the method described by Bullinger et al. in 1998. I shall refer to this as the AAOS method.

In each case I shall present the most recent version of the process, but I shall also refer to earlier publications if relevant information is not provided or fundamental concepts and assumptions are not discussed in the latest publications and also in order to illustrate the evolution of attitudes towards back-translation.

Examples of this development can be found in both the Sperber and AAOS methods. Both were first proposed in the first half of the 1990s (1994 and 1993 respectively) and in the initial papers the authors' position was that not enough projects were using back-translation and other cross-cultural adaptation techniques such as validation. In contrast, the most recent publications (2000 and 2004 respectively) both consider back-translation to be an indispensable (and universally accepted) component of cross-cultural adaptation and do not therefore devote as much space to explaining, justifying or contextualising the technique.

I shall devote a subsection of this chapter to each method, starting with the Sperber method.

2.3.1   *The Sperber cross-cultural translation procedure*

The first article describing the Sperber method, "Cross-cultural translation: methodology and validation", was the result of a project to translate an American questionnaire on attitudes to preventative medicine (Sperber et al. 1994). The more recent paper was by Ami Sperber alone and was entitled "Translation and validation of study instruments for cross-cultural research" (Sperber 2004).

The earlier paper begins by explaining the problem that led to the development of cross-cultural adaptation. The basic assumption is that if a cross-cultural study has methodological failings that originate in translation then its validity will be compromised, but this lack of validity will not be apparent, leading to "erroneous conclusions". "The challenge", is therefore, "to adapt the instrument in a culturally relevant and comprehensible form while maintaining the meaning of the original items." In the case of cross-cultural psychology, the erroneous conclusion would be that "there are culturally different attitudes or norms" rather than the correct conclusion that "items were interpreted differently because of linguistic shortcomings and the differences were actually semantic" (Sperber et al. 1994: 501-2). The 2004 paper states the same case with fewer details.

Notwithstanding, whereas the 1994 article states that the differences between studies, "in terms of translation techniques and

rigor of post-translation validation, attests to the lack of a standardized approach to this methodological problem" (Sperber et al. 1994: 505), the 2004 article states that the "back-translation technique is preferred even though it is time consuming and can be expensive" (Sperber 2004: S126).

Having justified the need for cross-cultural adaptation, Sperber et al. mention the possibility of dispensing with translation altogether, before stating that "most researchers have used direct translation methods to conduct cross-cultural comparisons." Sperber et al. opt for translation and then, citing Brislin, state that there are basically two possible situations; either a *de novo* questionnaire will be created, in which case decentering can be used, or an existing, validated questionnaire which cannot be altered will be translated (Sperber et al. 1994: 502-3). With relation to the first case, the 2004 paper adds the following, "an assumption underlying this approach is that neither language is primary (no source language)", but Brislin is no longer cited and the term "decentering" has been dropped (Sperber 2004: S125).

In the case of a questionnaire that "cannot, itself, be changed in any way", Sperber et al. recommend writing the "questions in the original language bearing in mind the anticipated translation process" and indicate Brislin et al.'s 1973 book as a source of guidelines for this process (Sperber et al. 1994: 503). The 2004 paper does not make this suggestion, but does state that

> Most questionnaires are translated from English, so there is a potential problem of ethnocentricity or what has been termed cultural hegemony in cross-cultural research.
>
> (Sperber 2004: S125)

Sperber et al. state that "several translation and evaluation methods are used in both situations" and then go on to describe what are in essence Brislin's criteria one and five, plus an adaptation of his pre-testing method. Criterion 1 is described as follows,

> The first is the technique of back translation. In this method the original translation is translated back into the source language by a blinded, independent translator. The two source-language versions are then compared and, if necessary, revised either by individual translators or by a committee of specialists.
>
> (Sperber et al. 1994: 503)

They then state that three "critical translation problems that adversely affect many studies" have been identified (once more citing Brislin et al. 1973). The first problem is that "some translators are not sufficiently

aware of the rigorous requirements of cross-cultural translation." (Sperber et al. 1994: 503). An identical phrase appears in the 2004 article, followed by this explanation,

> They may spend time on literal translation without devoting enough attention to cultural nuances. Colloquial phrases, slang and jargon, idiomatic expressions, and emotionally evocative terms may be particularly difficult to handle.
>
> <div align="right">(Sperber 2004: S125)</div>

To support these points Sperber states that "potential cultural differences" are possible in the "interpretation of many terms", giving the examples of "family", "adolescence", "femininity" and "masculinity" as terms that are interpreted differently across different cultures (Sperber 2004: S125). Sperber then provides two examples of actual translation projects from his own personal experience of translating questionnaires. The first example is from a project to translate a scale to rate "patient concerns" into Hebrew. The original included "a question about patients' concerns relating to difficulty getting health insurance" and, according to Sperber, the "validation process ... showed that the translation was well done", but patients did not understand the question. The reason for this was that "basic health insurance is universal in Israel and is unaffected by health status" and Sperber's conclusion was that "the translation was good but the item was culturally irrelevant." (Sperber 2004: S125).

The second example was from a questionnaire designed for patients with irritable bowel syndrome, in which "the term used for bowel movement" chosen by the translators was the word used by non-Bedouin Arabs and would be incomprehensible to 80-90% of Israeli Bedouins. Sperber's comment was as follows, "the translation was literally good but confusing when applied to a different cultural group" (Sperber 2004: S125).

The second of the three "critical translation problems" listed in Sperber et al. is that

> ... translators are not always knowledgeable enough about the specific content area of the instrument. Specialized medical subjects are an example of this type of difficult content area.
>
> <div align="right">(Sperber et al. 1994: 503)</div>

Sperber says the same, and adds the following, "good professional translators are often incapable of translating medical material", citing the 1973 paper by Brislin et al. in support (Sperber 2004: S125). This is

the only mention of Brislin in the entire 2004 article, even though the great majority of what is stated can be traced directly to his work.

Sperber states that the third problem can "stem from overly competent translators who ... achieve a back translation that is similar to the source even though the original translation is not good" (Sperber 2004: S125). This problem is then itself broken down into three possible problems, which Sperber et al. explain by practically transcribing verbatim the three limitations of back-translation that Brislin himself had described at the start of his article (see page 42 above).

The 2004 article provides what Sperber describes as an example of the problem of retention of source language "grammatical form", by which a back-translator given "Do you sometimes feel that your stomach is full?" realises that it is a mistranslation of "Do you sometimes feel fed up?" and corrects the mistake in the back-translation with the result that, "researchers who are presented with 2 identical English versions can only conclude that the translation is excellent and leave the critically faulted target-language version unchanged." (Sperber 2004: S126).

The version of criterion 5 that Sperber et al. describe is not identical to Brislin's original version since the "instrument is given to bilingual persons in alternating language order and assessed accordingly." The method is therefore similar to half of Brislin's criterion 5 test (the half in which two monolingual groups each answered a questionnaire in a single language), but it does not include his split-half test. Rather, comparison of translation with source is achieved by having each person answer both versions, at different times. Sperber et al. point out that using "bilingual subjects for pre-testing can also create methodological problems", explaining that since "bilingual individuals adopt some concepts, values, attitudes, and role expectations of the culture of the second language" they may therefore, "represent a separate population whose responses cannot be generalized automatically to the monolingual population" (Sperber et al. 1994: 503). The only difference in the 2004 iteration is that there is no longer any hedging about whether bilinguals constitute a separate population, Sperber is sure they do (Sperber 2004: S126).

The third method that Sperber et al. describe is "field pre-testing", saying that Brislin et al. (1973) suggested

> ... comparing the results of the newly translated instrument with the results of a previously used and recognized scale tested in the same language, on the same topic, and on the same people.
>
> (Sperber et al. 1994: 504)

This suggestion is not taken seriously, however, on the basis that "It is rare that a suitable instrument already exists, and if so, why develop a new one?" This technique was not mentioned in the 2004 paper.

The major contribution to cross-cultural adaptation claimed by the first article is a process of "formal comparison of the original version of the questionnaire with the back-translated version (the two English versions)", by which "two measures of comparison were used to evaluate the success of the translating process." These measures were "comparability of language and similarity of interpretability" (Sperber et al. 1994: 506). The people judging comparability and similarity were

> ... 29 students and faculty members of the Department of Health Behavior and Health Education of the School of Public Health in the University of North Carolina
>
> (Sperber et al. 1994: 509)

Two reasons were given for assessing "perceived similarity of form and meaning" separately. The first was that Sperber et al. "believed that asking raters to assess these dimensions separately would enhance their distinctness" and that if no distinction had been made, "ratings may have reflected overall similarity, combining meaning and form" (Sperber et al. 1994: 507).

The second reason was that the authors

> ... believed that obtaining separate ratings for each dimension would assist us in deciding when and how to rewrite items that appeared, in back translation, to differ from their original counterparts.
>
> (Sperber et al. 1994: 507)

The 2004 paper uses the same method, but does not attempt to justify it in this way. The 29 raters actually used in the earlier study have been rounded up to "at least 30 raters who are fluent in the source language" and the stricture that "raters are independent of the investigators and do not include the translators" is made explicit. Instead of arguments to justify the twin-rating system, the 2004 paper makes a series of statements that are treated as givens. The first of these statements is as follows,

> This process enables us to identify potentially problematic items and reassess and retranslate them until we are as confident as possible that the item will be interpreted in the same manner in both languages.
>
> <div align="right">(Sperber 2004: S126)</div>

Sperber also states that "similarity of form and meaning" were assessed separately in order to "enhance the distinctness of the dimensions", but whereas in 1994 the authors "believed" this to be the case, in 2004 Sperber simply states that it is so.

The actual rating system remained the same from 1994 to 2004 and consists of two 7-point Likert scales for each item on the questionnaire. The first relates to "comparability of language" and ranges from 1 for "extremely comparable" to 7 for "not at all comparable". The second relates to "similarity of meaning" and ranges from 1 for "extremely similar" to 7 for "not at all similar" (Sperber et al. 1994: 506; Sperber 2004: S127)

In the original study, a 35-item survey of attitudes to preventative medicine was translated into Hebrew by "an experienced translator in the United States" who was "a bilingual physician", and back-translated in Israel by the "director of a company that specializes in medical and scientific translating and editing", and "both were aware that the versions would be compared." (Sperber et al. 1994: 506). When the two versions were rated by the 29 students, just 4 out of 35 items had a mean score higher than 3, and only one of these was above 3.5. The item that scored close to 5 was retranslated, but the other three were left unaltered and tracked during the validation process and "no apparent differences between them and the other items" were detected (Sperber et al. 1994: 515)

Sperber et al. do not provide inter-rater agreement as Brislin had done, but a table in the appendices provides maximum and minimum ratings (the range) for each of the thirty-five questions plus a rewritten version of the question that scored close to five. These figures give an idea of the degree of overall variation among the raters. I have summarised the ranges for each scale, by item (as presented by Sperber et al.) below.

Table 2.5 – Minimum and maximum ratings (out of 7) for comparability of language (Lang.) and similarity of interpretability (Int.), by item

Item	1		2		3		4		5		6		7		8		9	
Lang.	1	3	1	6	1	7	1	6	1	6	1	5	1	4	1	7	1	7

Int.	1 3	1 6	1 6	1 6	1 5	1 3	1 3	1 6	1 7
Item	10	11	12	13	14	15	16	17	18
Lang.	1 6	1 5	1 6	1 7	1 5	1 4	1 4	1 5	1 5
Int.	1 5	1 4	1 6	1 6	1 4	1 4	1 4	1 5	1 5
Item	19	20	21	22	23	24	25	26	27
Lang.	1 5	2 7	1 5	1 4	1 5	1 4	1 5	1 7	1 5
Int.	1 4	2 7	1 5	1 5	1 6	1 5	1 6	1 7	1 6
Item	28	29	30	31	32	33	34	35	36
Lang.	1 5	1 7	1 6	1 6	1 6	1 5	1 3	1 4	1 4
Int.	1 6	1 7	1 4	1 6	1 5	1 3	1 3	1 5	1 3

(Adapted from Sperber et al. 1994: S517-521)

In addition to presenting the actual "translation → back-translation" process used to translate the attitudes to preventive medicine questionnaire, the 1994 article also describes the stages employed to validate it. Although none of them are revolutionary, I shall nevertheless present them because certain elements of the description and discussion are of relevance to back-translation itself.

Briefly, a group of 32 bilingual (English and Hebrew) physicians answered both versions of the questionnaire with a six-week interval. Eighteen answered each language version first. Test-retest reliability was "well below the acceptable range" according to "conventional interpretation" although the English-second reliability was acceptable. The explanation provided was that "The 'Hebrew first' group was skewed toward specialists in family medicine, whereas the 'English first' group was skewed toward specialists in internal medicine and the medical subspecialties" (Sperber et al. 1994: 51-2).

Two further groups, one English-speaking and one Hebrew-speaking, were selected and given the appropriate version of the questionnaire twice, with an interval of approximately two months. Test-retest reliability was calculated as a correlation coefficient (r) and Cronbach's alpha was used to assess internal reliability of test and retest. These groups' results were then used to assemble a "mixed" test by selecting at random 18 items in English and 17 in Hebrew. Once more, correlations were calculated, although this time against the monolingual versions, and Cronbach's alpha was calculated for the mixed set. The 2004 article does not deal with validation other than to state, at the very end that "the specific validation method adopted is less

important than the recognition that the translation process must be appropriate and the validation process rigorous" (Sperber 2004: S128).

There is no discussion section in the 2004 article, but since the methods recommended in both papers are almost identical and the assumptions behind them are the same, the discussion section in the 1994 article is of relevance to both.

Sperber et al. stated that their bilingual analysis, "confirmed the effectiveness of the translation process and the reliability of the Hebrew questionnaire" and also that

> ... the fact that the first survey had a lower total mean score than the retest irrespective of whether it was in English or in Hebrew provides further evidence that the survey was stable and consistent, independent of the language in which it was administered.
>
> (Sperber et al. 1994: 515)

They then explain what their validation was designed to have demonstrated were significant differences between groups to have been detected. In the event that there had been test-retest differences within monolingual groups, the conclusion would be that a shift in attitude had taken place between tests (the questionnaire is supposed to measure attitudes), but not that there was a problem with the translation. If there had been a difference within just one of the bilingual groups (either English first or Hebrew first), then this "would have been very difficult to understand." (Sperber et al. 1994: 516).

The outcome of interest, however, would be if a "discrepancy had been found in both bilingual groups" since "the most likely explanation would have been a lack of translation validity." They acknowledge, however, that in this scenario "a change in attitude could not be ruled out" either. In the event, however, "within- and between-groups analysis revealed no significant discrepancies among any groups" which they say "[lends] further credence to the validity of the translation process" (Sperber et al. 1994: 516).

The final paragraph of the 1994 study begins with the information that "the translation and validation processes described for this study were time-consuming", before going on to state that

> ... it is impossible to achieve 100% validation, and one can always contend that significant differences in cross-cultural comparisons could be the result of methodological flaws rather than actual differences.
>
> (Sperber et al. 1994: 516)

In summary, the back-translation component of the Sperber method is basically back-translation as described by Brislin, with the back-translator blinded to the original, but with the innovation that multiple raters judge the degree of "comparability of language" and "similarity of interpretation" of questionnaire elements on a scale of 1 to 7, rather than using Brislin's criterion 1 meaning error. The scores are converted to means and individual items are reviewed if they score higher than a certain cut-off (since 1 means *most* comparable or similar). This contrasts with Brislin's method in which the cut-off was a predetermined number of total errors and the response to exceeding the cut-off was to conduct a new translation/back-translation of the entire instrument.

The Sperber method uses variants of Brislin's criterion 5 test to validate the translation and makes it clear that this process is distinct from psychometric testing of the translated instrument.

2.3.2   *The IQOLA project cross-cultural translation procedure*

The IQOLA project was initiated in 1991 with the goal of "developing validated translations of a health status questionnaire for use in multinational clinical trials and other international studies of health". Since then, translation projects to adapt the SF-36 questionnaire have been completed in more than 60 different countries (IQOLA 2011a). The SF-36 is a

> 36-item survey that measures eight domains of health [and] is a generic measure [that] has been useful in assessing the health of general and specific populations, comparing the relative burden of diseases, differentiating the health benefits produced by a wide range of treatments, and screening individual patients.
>
> (IQOLA 2011b)

The eight domains of the SF-36 are as follows, Physical functioning, Role-physical, Bodily pain, General health, Vitality, Social functioning, Role-emotional and Mental health (Perneger et al. 1999: 1042).

Below is a typical item from the SF-36

6. During the past 4 weeks, to what extent has your physical health or emotional problems interfered with your normal social activities with family, friends, neighbors, or groups?

❑ Not at all
❑ Slightly
❑ Moderately
❑ Quite a bit
❑ Extremely

(IQOLA 2011b)

The IQOLA website also states that the SF-36 has been used as a measure in research reported in more than 5,000 publications.

The scientific article describing the actual translation process was published in 1998 and describes adaptation of the SF-36 for Canada, Denmark, France, Germany, Italy, Japan, the Netherlands, Norway, Spain, and Sweden (Bullinger et al.1998: 916-917).

The IQOLA cross-cultural adaptation method includes three stages. The first is translation following a standardized procedure, the second is psychometric testing and the last stage is a series of studies to evaluate validity (IQOLA 2011b).

I shall only cover the publication describing the first of these stages, "Translating health status questionnaires and evaluating their quality: The IQOLA project approach"(Bullinger et al. 1998), since that is the stage during which the actual translation is produced and the stage in which back-translation is utilized.

The IQOLA process has an interesting element, which is a "National Principal Investigator", responsible for coordinating each translation in the target setting. The process specifies two forward translations of the original US-English SF-36, produced by native speakers of the target language who have experience of questionnaire translation. These initial translations contain one translation of each question and the instructions plus a list of "all possible translations of the response choices". Translators are asked to "place emphasis on conceptual rather than literal equivalence" when translating. Another interesting feature is that the translations are targeted at a reading age of 14, or younger in some countries (Bullinger et al. 1998: 914).

Once the forward translations are ready, and before back-translation, two exercises are conducted with them. The first is a "Thurstone scaling exercise", the first step of which is to select the two most extreme response choices (end points), for each "response continuum" (the examples given are "excellent" and "poor"). A group of 100 native speakers are then asked to position all the other candidate

response choices between the two end points, along a "100mm LASA scale" (a ruler printed on the assessment form). Bullinger et al. say the aim was to provide "additional information that would help in selecting response choices that had similar values as those in the original instrument", although additional criteria such as clarity and use of common language were also considered (Bullinger et al. 1998: 914).

The second exercise is conducted by the translators, who are asked to rate the difficulty of translating the items on the original SF-36 from 0 to 100 (where 100 is most difficult). The results from the first ten countries were analysed with the objective of using them for a sophisticated form of decentering of the original SF-36 (Bullinger et al. 1998: 914).

The two translators for each language then meet with their respective National Principal Investigators to attempt to agree on a single forward translation. A lay panel is convened if consensus cannot be reached or in order to "elicit expressions more likely to convey the concept under study" (Bullinger et al. 1998: 914).

The synthesized forward translation is then given to two more "bilinguals" who are "native speakers of the foreign language if possible, or else native English speakers with extensive knowledge of the target language". These two translators conduct a modified version of Brislin's criterion 2 (bilingual rating). They each score the "quality" of the translation on a scale from 0 to 100, where 0 is "not at all perfect" and 100 is "perfect", using the following three criteria:

> (1) clarity of the translation (i.e., the use of simple and understandable expressions); (2) common language use (i.e., avoidance of technical or artificial terms); and (3) conceptual equivalence (i.e., representation of the content of the original source instrument).
>
> (Bullinger et al. 1998: 914)

The National Principal Investigator then meets with the first two translators to discuss the ratings and "modify the translation as needed." (Bullinger et al. 1998: 915).

The forward translation is then ready for back-translation. Two back-translations are produced, by native speakers of American or British English, and "reviewed by researchers at the Health Assessment Lab for conceptual equivalence with the original source version" (Bullinger et al. 1998: 915). The Health Assessment Lab is in Boston, in the United States (Bullinger et al. 1998: 913).

No further information is provided on how the researchers rate the back-translation against the original, neither is it made clear whether they also see the forward translation, what languages other than English they understand, how many of them are involved or even whether they use qualitative or quantitative methods (or a combination of both). Nevertheless, this stage is a recognisable descendent of Brislin's criterion 1, based on the assumption that assessing a back-translation for equivalence with the first source text provides meaningful information about the intermediate forward translation.

Irrespective of how the Health Assessment Lab researchers arrive at their conclusions, items and responses that were "deemed not to be conceptually equivalent were discussed with the National Principal Investigator", but no indication is given of how many such items were actually identified in the ten translations reported on by Bullinger et al., what modifications they led to or how consensus was arrived at. The first ten translations were also discussed together at an international meeting (Bullinger et al. 1998: 915-6).

The forward translations were then pilot tested in the target setting by administering them to "focus group[s] of up to 50 respondents who differed in health status". Respondents were asked if they found any items "difficult, upsetting, or confusing." Once more, we are told that "difficulties encountered by the respondents were noted, and the translations were revised as needed", but no examples are given of these difficulties or of how many were encountered in each country (Bullinger et al. 1998: 916).

At this point in the article the description of the translation phase proper is complete, but the article continues, presenting the results of the translation ratings and describing how they were used. The objectives of these ratings were as follows,

> ... to: (1) examine the rate of agreement in difficulty and quality ratings between independent raters within each country, (2) identify SF-36 items that consistently presented problems in translation, and (3) identify differences across the countries with regard to the difficulty and quality of SF-36 item translations.
>
> (Bullinger et al. 1998: 916).

The second of these objectives is a preliminary step towards decentering the SF-36 and producing the "international" version. The reasons for wishing to achieve objectives one and three are not given, but the different cross-sections of the dataset - intra-national in the case of the first objective and international in the third – would provide insights that

would also be of use in a decentering exercise and so aided in achieving the second objective.

Analysis of the results showed that "there was little variability in ratings given by individual raters within a country and the results were highly skewed." In other words, both the difficulty and quality of a translation were dependent on the person rating it. Despite this, Bullinger et al. calculated mean ratings for each item in each country "by averaging the rating for each item and response choice over the two raters within each country" and did the same for each item across all countries and for each country across all items (Bullinger et al. 1998: 917).

In order to "better understand the translation ratings" a cut-off score was calculated at the 75th percentile for difficulty and the 25th percentile for quality, above and below which translations were "determined to be problematic" (Bullinger et al. 1998: 921). It is important to point out that the 75th percentile is not 75% of the total possible score of 100 (i.e. a score of 75), but the point on the range of all scores in a given sample at which 75% of results are below the cut-off and 25% of results are above it. In the event, scores above 25 out of 100 for difficulty, and scores of less than 90 out of 100 for quality were considered "problematic" (Bullinger et al. 1998: 920).

The article gives vague details on a few examples of some of these problematic items, saying that an item asking about physical functioning that included bowling and golf as examples of moderate activity was "difficult to translate because several ... activities are not common outside of the United States", that "the concept of social activities was viewed differently in Europe than in the United States", that an easier-to-translate "synonym for *pep* is *life*" and, finally, that "words were found to convey the concept of depression" when translators were confronted with "*feeling blue*" as an emotional state. It also reports that "standards of equivalence were set to convert one mile, several blocks, and one block into metric equivalents of one kilometer, several hundred meters, and one hundred meters" (Bullinger et al. 1998: 921).

In the discussion section, Bullinger et al. list as strengths of the IQOLA approach "independent empirical tests of translation quality and comparison of translations across countries, as well as the international comparison of response scaling values", but admit that "greater standardization of the qualifications of translators and quality raters and more input from lay groups" are needed (Bullinger et al. 1998: 922).

The article closes with a number of statements, the most important of which is that "reworking translations with low-quality ratings yielded improvements of the translations, as did the process that compared backward-translations and the original SF-36 questionnaire." (Bullinger et al. 1998: 922). However, there are no further details about the nature of the reworking or of the improvements, whether monolingual or bilingual ratings were responsible for identification of a greater or lesser proportion of translations to be reworked or whether forward or backward translators or someone else was called upon to resolve problems that had been identified.

2.3.3 *The AAOS cross-cultural translation procedure*

The American Association of Orthopaedic Surgeons was founded in 1997 to deal with orthopaedic health policy and patient advocacy (AAOS 2011a). One of its functions is to produce outcomes instruments, which are

> ... designed to collect patient-based data for use in clinical practices to assess the effectiveness of treatment regimens and in musculoskeletal research settings to study the clinical outcomes of treatment.
>
> (AAOS 2011b)

Each instrument consists of a questionnaire for administration to patients over the age of 18 that produces a score from 1 to 100, where 100 indicates least disability and 1 greatest disability. Instruments currently exist for arm, shoulder and hand (one instrument); spine; foot and ankle; hip and knee; and lower limbs; plus one for "sports knee"; and one for overall musculoskeletal function (AAOS 2011c).

The method used to adapt these instruments is described in an article entitled, "Guidelines for the process of cross-cultural adaptation of self-report measures" (Beaton et al. 2000). These guidelines are a development of recommendations made in "Cross-cultural adaptation of health-related quality of life measures: Literature review and proposed guidelines" (Guillemin et al. 1993). Indeed, three of the four authors of the 2000 article, Beaton, Guillemin and Bombardier, are also the only authors of the 2000 paper. The 2000 paper also includes contributions by Ferraz and the lead author is Beaton rather than Guillemin.

Before describing the AAOS method, it is necessary to explain the reason why I am referring to the method described in Beaton et al. as the AAOS method. In fact, the paper makes two contradictory

statements about which method it is actually describing. Early on, the paper states that "the translation process outlined in this article is the first step in the three-step process adopted by the International Society for Quality of Life Assessment (IQOLA) project", but the paper contradicts itself a few lines further down, stating that "the cross-cultural adaptation process being recommended ... is the method currently used by the American Association of Orthopaedic Surgeons (AAOS) Outcomes Committee" (Beaton et al. 2000: 3186-7).

On the basis of comparison with the methods described in Bullinger et al., which differ from the Beaton et al. procedure in several significant ways, and also on the basis that when Beaton et al. refer to sample forms that can be used as examples, the reference they give is for a paper authored by McConnell, Beaton and Bombardier describing how to use the AAOS arm, shoulder and hand instrument (McConnell et al. 1999, cited in Beaton et al. 2000: 3187), I came to the conclusion that the method being described is indeed the AAOS method, although it was presumably based on the IQOLA method. One further clue appears in the discussion section, where Beaton et al. refer to "submitting reports to a body such as the AAOS" (Beaton et al.: 3190).

As is the case with the Sperber method, the first publication that outlined what was to become the AAOS method was written from the perspective that back-translation and other cross-cultural adaptation techniques were underutilized in the field in question (in this case, health-related quality-of-life research). Guillemin et al. state that "many researchers in QOL may not be aware [of] or do not quote this methodological work developed in the psychology and sociology literature" (Guillemin et al. 1993: 1428). In 2000, the same three authors, plus Ferraz, stated, "It is now recognized that if measures are to be used across cultures, the items must not only be translated well linguistically, but also must be adapted culturally" (Beaton et al. 2000: 3186). Notwithstanding, the principal elements of the method did not undergo substantial change between 1993 and 2000, so I shall only refer to the earlier paper when it provides additional detail.

Beaton et al. say that one of the reasons for adapting health status measures is an "increase in the number of multinational and multicultural research projects". They also explain that while "most questionnaires were developed in English-speaking countries", even in English-speaking countries researchers must still "consider immigrant populations in studies of health, especially when their exclusion could lead to a systematic bias" (Beaton et al. 2000: 3186).

On this basis, Beaton et al. claim that "cross-cultural adaptation ... necessitates use of a unique method, to reach equivalence between the original source and target versions of the questionnaire" stating that the method "allows increased confidence that the impact of a disease or its treatment is described in a similar manner in multinational trials or outcome evaluations" (Beaton et al. 2000: 3186).

The most distinctive feature of the AAOS method was first described in the 1993 article. This feature is the objective of the translation stages of the adaptation process. The Brislin method had aimed for "equivalence" or "functional equivalence", the Sperber method for "equivalence of meaning" and the IQOLA method was designed to achieve "conceptual equivalence" (which meant equivalence of psychometric concepts, not equivalence of respondents' conceptualizations). In contrast, the AAOS method defines four different types of equivalence.

In the earlier publication, Guillemin et al. state that "translators aiming for conceptual equivalence should consider the following" and then, rather confusingly, lists "conceptual equivalence" as one of the components of "conceptual equivalence" (Guillemin 1993: 1424-5). The 2000 article resolves this confusion by redefining the overall objective as "to achieve equivalence between the source and target versions in four areas". The areas are Semantic Equivalence, Idiomatic Equivalence, Experiential Equivalence and Conceptual Equivalence (Beaton et al. 2000: 3188-9).

The AAOS method consists of six stages, the last of which is related to documentation and central control of the project. The six AAOS stages are as follows:

Table 2.6 – The six stages of the AAOS translation method

Stage I Initial translations (x2, in parallel)	Translator 1: a native speaker of the target language, with knowledge of the subject matter and concepts. Translator 2: a native speaker of the target language, who is ignorant of the subject matter and concepts
Stage II Synthesis of the translations	Performed by both translators together in the presence of a recording observer.
Stage III Back translation (x2, in parallel)	Carried out by two translators separately, both are native speakers of the source language and both are ignorant of the subject matter and concepts.
Stage IV Expert committee	The committee consolidates all previous versions and produces a "prefinal" version. The minimum composition comprises methodologists, health professionals, language professionals, and the translators (forward and back-translators) involved in the process up to this point. The committee will look at four aspects of equivalence.

Stage V Test of the prefinal version	Ideally, between 30 and 40 people from the target setting should be tested and interviewed about what they thought was meant by each questionnaire item and the chosen response. The distribution of responses is examined to look for a high proportion of missing items or single responses.
Stage VI Submission of documentation to the developers or coordinating committee for appraisal	In effect Stage VI is a process audit, with all the steps followed and necessary reports followed. It is not up to this body or committee to alter the content, it is assumed that by following this process a reasonable translation has been achieved.

(From Beaton et al. 2000: 3187-90 [extensively abridged])

It will be noted from Table 2.6 that the back-translations and original are not compared in isolation of the forward translation, which is a significant departure from Brislin's criterion 1. The two forward translations are synthesised to form a preliminary translation and this is reviewed by the expert committee along with the original, both initial forward translations and both back-translations.

No indication is given of what weight is afforded to each stage when the versions are "consolidated" by the expert committee or whether some members of the committee have more authority than others and no examples are given of the type of problems that back-translation in particular might be expected to detect.

Back-translation is, however, described as "highlighting gross inconsistencies or conceptual errors in the translation" and the reasons given for choosing two "naive" translators for the back-translation, who "should neither be aware nor be informed of the concepts explored, and should preferably be without medical background", are "to avoid information bias" and "to elicit unexpected meanings of the items in the translated questionnaire" thereby "increasing the likelihood of 'highlighting the imperfections'" (Beaton et al. 2000: 3188).

In common with Brislin, Beaton et al. warn that "agreement between the back translation and the original source version does not guarantee a satisfactory forward translation, because it could be incorrect; it simply assures a consistent translation." (Beaton et al. 2000: 3188).

In the AAOS method, the expert committee must "examine the source and backtranslated questionnaires for all such equivalences" and three specific examples of the types of problems that the committee is expected to solve are provided within the definitions of the different

types of equivalence. The four types of equivalence are defined below. I have underlined the three examples.

- **Semantic Equivalence.** Do the words mean the same thing? Are their [sic] multiple meanings to a given item? Are there grammatical difficulties in the translation?

- **Idiomatic Equivalence.** Colloquialisms, or idioms, are difficult to translate. The committee may have to formulate an equivalent expression in the target version. For example the term "feeling downhearted and blue" from the SF-36 has often been difficult to translate, and an item with similar meaning would have to be found by the committee.

- **Experiential Equivalence.** Items are seeking to capture and [sic] experience of daily life; however, often in a different country or culture, a given task may simply not be experienced (even if it is translatable). The questionnaire item would have to be replaced by a similar item that is in fact experienced in the target culture. An example might be in an item worded: Do you have difficulty eating with a fork? when that was not the utensil used for eating in the target country.

- **Conceptual Equivalence.** Often words hold different conceptual meaning between cultures (for instance the meaning of "seeing your family as much as you would like" would differ between cultures with different concepts of what defines "family"—nuclear versus extended family).

<div align="right">(Beaton et al. 2000: 3188-9, underlining added)</div>

Once consensus has been reached, the resulting "prefinal" version is pilot tested. The AAOS version of pre-testing involves administering the prefinal questionnaire to "between 30 and 40 persons" and then interviewing them "to probe about what he or she thought was meant by each questionnaire item and the chosen response." which, it is claimed, "ensures that the adapted version is still retaining its equivalence in an applied situation." The subjects' responses are also analyzed using descriptive statistics to test for "a high proportion of missing items or single responses." (Beaton et al. 2000: 3189).

Beaton et al. make it clear that the pre-test does not guarantee that the translated questionnaire has retained the psychometric properties of the original since it "does not address the construct validity, reliability,

or item response patterns that are also critical to describing a successful cross-cultural adaptation."(Beaton et al. 2000: 3189).

Step six of the AAOS process is self-explanatory. It is "submission of documentation to the developers or coordinating committee for appraisal of the adaptation process", which Beaton et al. describe as a "process audit" which is included "to verify that the recommended stages were followed" (Beaton et al. 2000: 3189).

Neither article describes the process of validating the psychometric properties of questionnaires, but Beaton et al. do define the scope of what such a validation project is expected to achieve:

> The new instrument should retain both the item-level characteristics such as item-to-scale correlations and internal consistency; and the score-level characteristics of reliability, construct validity, and responsiveness.
>
> (Beaton et al. 2000: 3189)

The paper also lists certain elements of the validation process, such as testing that each item in each scale is still correlated with its own scale rather than with another scale (construct validity) and assessment of "score level attributes" for validity and reliability. ." (Beaton et al. 2000: 3189).

Part of the validation process for any psychometric questionnaire is to test its validity and reliability. According to Beaton et al., when a cross-culturally adapted questionnaire is being validated it should undergo the same validation procedure as the original and the results should additionally be compared with the results of "similar tests performed in the original setting using the original instrument" and "it is expected that the adapted version would perform in a similar manner." (Beaton et al. 2000: 3189).

In summary, the most distinctive features of the back-translation component of the AAOS method are its insistence on "naive" back-translators, the subdivision of "equivalence" into Conceptual Equivalence, Experiential Equivalence, Semantic Equivalence and Idiomatic Equivalence, and the fact that back-translation is not a discrete process with unique input and output, but one of many processes leading up to the definitive expert committee.

2.4 Contemporary critical evaluation of back-translation

As mentioned earlier, very little has been published in terms of criticism, whether positive or negative, of the back-translation

technique. When I first searched the literature at the start of this project (in early 2009) I was only able to locate a single article that analyzed the effectiveness and utility of back-translation for medical instruments (Perneger et al. 1999).

However, during my qualification hearing Professor Lincoln Fernandes kindly drew my attention to an article that had been published after I had conducted the initial literature search (Ozolins 2009) and the references provided in that article led me to a third (Grunwald and Goldfarb 2006), although there are reasons why this last paper should be treated with caution, as I will show shortly (in subsection 2.4.1).

The first two of these articles present examples from projects to translate health-related questionnaires and use them to illustrate critical analyses of back-translation. The third illustrates how the back-translation process works using an excerpt from a fictional clinical drug trial Informed Consent Form and translations and back-translations of it.

However, in terms of the positions they adopt with respect to back-translation, these three publications break down differently. Grunwald and Goldfarb come to a very favourable conclusion on the utility of back-translation. Ozolins discusses certain drawbacks inherent to back-translation, but, after an analysis of both translation problems and some of the e-mail exchanges between researchers and translators that led to their resolution, he also comes to a favourable conclusion. In contrast, after detailed qualitative and statistical analysis of two translations of the same questionnaire, one produced using back-translation and the other without, Perneger et al. come to a conclusion that is not by any means a wholehearted endorsement of back-translation. In view of this, I shall not present these three publications in chronological order. Rather, I shall begin with Grunwald and Goldfarb (2006), on the basis that they agree to the greatest extent with the literature reviewed up to this point. I shall then move on to Ozolins (2009), who opens the possibility of alternatives to back-translation, but concludes in favour of it. I shall then end this section on criticism of back-translation, and the chapter itself, by describing the comparison Perneger et al. made of two different translations of the same questionnaire and the data collected with them (1999).

2.4.1   *Grunwald and Goldfarb (2006)*

As mentioned above, there are certain issues related to Grunwald and Goldfarb's article that could be considered reservations. The first of these is that this paper is based on analysis of three paragraphs from a

"fictional" text, rather than an actual back-translation project to translate a "real life" text. Secondly, the fictional text is not a questionnaire, but an Informed Consent Form for a clinical trial (so the objective would be to inform volunteers about a drugs trial and record their consent). The third and final feature of this article that should be borne in mind when considering the results and conclusions is detailed in a disclosure of interests statement provided at the end of the article.

> David Grunwald is CEO of Global Translations, a provider of translation services to the clinical research industry. ... Norman M. Goldfarb is Managing Partner of First Clinical Research, a provider of a[sic] clinical research best practices consulting, training, implementation and research services.
>
> (Grunwald & Goldfarb 2006: 6)

To contextualise this statement, the article was published by an online magazine called the Journal of Clinical Research Best Practices, which is hosted on the website of First Clinical Research of which Goldfarb is Managing Partner.

Each of these three issues has its implications, but this article makes some contributions that I could not find anywhere else and I have therefore included it on the basis that, as long as readers are fully aware of its peculiarities, they should be in a position to judge the validity of its conclusions for themselves (and also of course, the validity of the conclusions I draw about it in Chapter 4).

The first of these issues - the fact that the examples are not only fictional paragraphs, but paragraphs taken out of context - is indeed regrettable, but, to a certain extent, the fact that one of the authors runs a translation services company and the other works in clinical research could be considered to ameliorate this since they should be in a position to produce a fictional consent form that is very close to a real one. The article gives no reasons for not using a real consent form.

The second feature of the article that bears on its interpretation is also related to the text. The subject of this thesis is translation and back-translation of questionnaires, but this article uses examples from a consent form. While consent forms do not have the response options that questionnaire have, they do fulfil the same general function of enabling structured communication between health professionals and their patients or members of the public in general. This means they share certain features, chief of which is the need to use non-specialist language, while providing (and/or acquiring) the necessary information.

The information contained in the disclosure of interests statement is of course an alert to the possibility of bias, but it also reveals the reasons why this article contains two important elements that are not discussed explicitly in any of the literature covered so far.

The first of these elements is a discussion of situations in which back-translation is not merely a preferred option chosen by researchers who wish to be rigorous, but a regulatory obligation, and it can be assumed that these insights come from Goldfarb's experience.

The second is a discussion, including specific prices and estimated percentages, of how much back-translation increases the cost of translation and this can be assumed to tap into Grunwald's professional experience.

I shall now present the main points made by Grunwald and Goldfarb in support of back-translation. They begin by stating that guidelines in Europe, Japan and the United States all demand that consent forms should be written in language understandable to the people who will sign them and then point out that this demands translation if these people do not understand the language the original form is written in (Grunwald & Goldfarb 2006: 1).

They then claim that consent forms are highly technical documents and that therefore it is essential to use certified medical translators, although they do not specify a certification body or minimum level of certification (Grunwald & Goldfarb 2006: 1).

Next they provide three reasons why even certified medical translators are likely to make errors:

- Translation is more of an art than a science.

- Translators are unlikely to specialize in the specific medical condition under study.

- Time and money are usually limited.

(Grunwald & Goldfarb 2006: 1)

Grunwald and Goldfarb describe a "normal" translation process as follows: "the translator translates the document; an editor then reviews the translation and makes corrections." They then suggest that back-translation is more "robust" and briefly describe the process. There is nothing like the level of detail provided by Brislin or the three cross-cultural adaptation processes so far described, but three statements are of interest. The first is that "the back translator must have excellent command of the foreign language as well as being a native English medical translator, and perform a more literal translation than normal."

The second is a suggestion that, in addition to the original authors of the consent form, the back-translation can be shown to institutional review boards. The third statement Grunwald and Goldfarb make is that they consider that back-translation deals with the first two of the causes of error they had listed, but costs money and time, thereby exacerbating (or being prevented by) the third (Grunwald & Goldfarb 2006: 1).

Grunwald and Goldfarb provide excerpts (of three paragraphs each) from the fictional consent form in English, from two different Spanish translations of these paragraphs, from an edited version of one of the Spanish translations and from back-translations of each of the three translated versions. They then list five "errors" that were detected by back-translation (in three different back-translations of around 240 words each).

In view of the artificiality of these examples I shall not waste space presenting them, but they can be summed up as follows. Out of five changes listed, two were artefacts introduced during the back-translation and three were present in one of the forward translations, but none of these three changes were present in the other forward translation.

The first change that was present in one of the translations was "experimental 'study drug'" translated as "medicamento en estudio" and back-translated as "medication under study", the second was a clause "you will come in for two visits", translated as "deberá presentarse a dos visitas médicas", and back-translated as "you shall come for two medical visits." and the third "substantive change" was that the (non-existent) drub name "harontin" was translated as (the equally non-existent) "harontonina". Grunwald and Goldfarb do not explain why they consider this "substantive", but give the following reason, "Translators could not find this term in any of the medical dictionaries, so they substituted it with a name that had a more Spanish flavor." (Grunwald & Goldfarb 2006: 2 & 6).

On the basis of these results, Grunwald and Goldfarb say there are three reasons why back-translations may differ from originals:

- Translator changed the meaning in the forward-translation.

- Translator changed the meaning in the back-translation.

- Literal back-translation of correct forward-translation appears to be error

(Grunwald & Goldfarb 2006: 6)

They then make three suggestions:

• Back-translations are valuable tools, but cannot replace editors.

• It may be essential for ICF authors to read the back-translations.

• A larger, more definitive, experiment would be worthwhile.

<div align="right">(Grunwald & Goldfarb 2006: 2)</div>

The article then provides information that I have never found in academic work on back-translation, which is how much it costs. According to Grunwald and Goldfarb, writing in 2006, a typical $800 fee for translating an informed consent form would increase by 80%, to $1400. They also point out that additional time is needed to compare versions and correct errors. Notwithstanding, their final position is that, "given the importance of accurate translations, the additional investment is more than justified." (Grunwald & Goldfarb 2006: 2).

The article then makes two final points. The first is that some ethics committees demand back-translation (an excerpt from Boston University Medical Center's Institutional Review Board Procedures is provided by way of example) and the second is the suggestion that back-translation could be of use to quality-control personnel at translation companies as a means of "identifying qualified translators."(Grunwald & Goldfarb 2006: 2)

Returning briefly to the issues raised by the declaration of interests at the end of this article, it should be pointed out that the Journal of Clinical Research Best Practices is not a peer-reviewed journal. It is advertising driven and has "industry" partners and subscribers. The journal's guidelines for authors contain the following: "We welcome articles from product and service providers that are willing to share their expertise with our readers. However, promotional content, if any, must be subtle." (First Clinical 2012).

2.4.2   *Ozolins (2009)*

In contrast with Grunwald and Goldfarb's paper, Uldis Ozolins' 2009 article "Back translation as a means of giving translators a voice" draws on examples from a real-life translation project employing back-translation.

This article was published in "Interpreting & Translation", but it does not draw on translation studies theory, making its case on the basis of excerpts from translations of a medical questionnaire and from e-mail exchanges between the medical researchers who commissioned the

translations and some of the translators working on either forward or backward translations.

Where this article differs from all of the literature presented so far (and also from the next article covered in this subsection, which is the last dealing with back-translation) is that it starts from the premise that the validity of back-translation has been challenged from within the translation industry. This is not a position I have seen espoused in any of the literature on back-translation, but it undeniably reflects a common feeling of dissatisfaction with back-translation that I have heard expressed anecdotally by professional translators and translation project managers (and which I, as a professional translator, also share). Ozolins only provides one reference for this particular point, which is a statement from a United States normative standard entitled "Standard guide for quality assurance in translation", quoted in an article by Kim Vitnay. The quotation is as follows, "A back translation will not result in a text that is identical to the source text, and furthermore, a back translation is not necessarily a good indicator of the quality of the translation." (ASTM International 2006, quoted in Vitnay 2007).

Ozolins describes this situation as a "cleavage" between the perspectives of those who commission back-translations and those who carry them out. He suggests that this cleavage "indicates relatively hermetically sealed universes of practice and belief among different sectors of the translation profession." (Ozolins 2009: 2).

While discussing this cleavage between opposing viewpoints on the merits of back-translation, Ozolins also mentions the same point raised by Grunwald and Goldfarb, which is that "the medical field now sees this methodology as something of a gold standard" and (quoting Vitnay 2007) back-translation is often an obligatory step in fulfilling "clients' needs to meet regulatory requirements" (Ozolins 2009: 2).

Ozolins credits the back-translation technique to Brislin, quoting his first and last texts on the subject, (Brislin 1970; Brislin 1986) and identifies the WHO "and other international medical research and treatment organisations" as the main drivers of the adoption of back-translation, listing its applications as "international questionnaires and surveys, as well as diagnostic and research instruments" and the reason for doing so as "the desire to find international comparative data on a myriad of health issues." (Ozolins 2009: 1).

As the title of Ozolins' article suggests ("Back translation as a means of giving translators a voice"), he is of the opinion that back-translation offers the chance to "establish an ongoing dialogue between

a translator and client for mutual benefit" and it is this dialogue that makes his article unique since he presents extracts of exchanges between the researchers running the translation project and the forward and backward translators. To my knowledge, this is the only published account of such exchanges.

The specific translation project that Ozolins describes was conducted in Australia and (with the exception of an additional target language that was added after the project had started) the ultimate objective was not to collect international comparative data using translated questionnaires, but to collect data within Australia from non-English speaking patients using versions of a questionnaire translated into twelve different languages: Arabic, Chinese, Croatian, Greek, Italian, Macedonian, Maltese, Polish, Russian, Spanish, Turkish and Vietnamese (Ozolins 2009: 2).

The questionnaire was developed in Melbourne, in English, and is entitled the Multi-attribute Arthritis Prioritisation Tool (hereafter MAPT). It was designed to help with prioritising patients for hip or knee replacement while reducing the "number of clinical visits necessary", by asking patients "suffering from severe arthritis a series of questions about their degree of pain, discomfort or problems associated with hip or knee pain", thereby "saving valuable clinical time and expense." (Ozolins 2009: 2). Ozolins provides an example item from the MAPT:

> Q: Do you have hip or knee pain that does not get better even while you rest (for example, while sitting)
>
> 1- None or mild pain
>
> 2- Moderate pain
>
> 3- Severe pain
>
> 4- Extremely severe pain
>
> 5- The pain is so severe that I cannot bear it.
>
> (Ozolins 2009: 2)

and, as will be observed, it follows the now familiar Likert format of a question with a preset list of responses scored along a continuum. Ozolins says that "the fine gradations of patient response" determine the instrument's validity and that "it was these gradations that were the essential items to maintain in any translation." (Ozolins 2009: 2).

The research team had decided to adopt the European Organisation for Research and Treatment of Cancer (hereafter EORTC) translation procedure, so I shall briefly describe how this method

compares to the cross-cultural adaptation methods described so far. The procedure they used is the second edition of the EORTC procedure (described in Cull et al. 2002) and it has some important differences to the three methods described in subsections 2.3.1 to 2.3.3 above.

The second edition of the EORTC process requires that every forward translator "should be a native speaker of the language into which the questionnaire is being translated, with a high level of fluency in the other relevant language", and each back-translator "should be a native speaker of [the original language] and fluent in the language from which the questionnaire is being backtranslated." (Cull et al. 2002: 4).

Two forward translators produce two translations independently and the "person responsible for coordinating the translation process" compares the translations. (Cull et al. 2002: 8). Two back-translations are produced and also compared by "the person coordinating the translation process", so, in common with the IQOLA method, the EORTC process requires a bilingual person other than the translators to play the role of coordinator (Cull et al. 2002: 9). However, in the IQOLA method this role is taken by a scientific researcher, but the EORTC method uses a "project manager" who "can be an investigator or a translation agency staff member" (Dewolf et al. 2009: 7).

Ozolins comments on this project-managed translation process saying that the system "dispenses with any expert committee" and states that the method "runs the entire process at arm's length to the authors who only engage after the substantial work of securing appropriate back and forward translations has been accomplished." and is

> ... clearly based on a double-blind model of clinical trials, where in this instance authors are removed from the toing and froing of the translation process; sufficient use of back translation will, it is assumed, eventually indicate the conceptually equivalent translation.
>
> (Ozolins 2009: 3)

Ozolins explains that the agency had originally resisted the use of back-translation and proposed its own quality-control system. In addition to the difficulty of finding, for example, "Turkish or Polish or Macedonian translators whose A language was English to perform the back translation", Ozolins also points out that there was very little literature on using back-translation with "local multilingual target readers" (Ozolins 2009: 4).

None of their arguments convinced the researchers, who wanted "a methodology that would fit in with international EORTC precedents

and the demands of peer review." (Ozolins 2009: 4). The method they finally adopted to translate the MAPT was as follows:

- Briefing by the MAPT team of the forward translators and checkers

- One forward translation

- Independent checking of the forward translation

- One back translation

- Comments (by email) by the MAPT team on any noted discrepancies between the English source text and the back translation

- Comments (by email) by each of the forward translator and back translator on the MAPT team's comments

- A final teleconference between MAPT team and forward translator on any unresolved items, and a further round of back translation if no resolution.

- Documentation on translators' qualifications and experience

- Pilot testing etc by the MAPT team.

<div align="right">(Ozolins 2009: 4)</div>

Ozolins does not provide any insight into what the documentation mentioned in the penultimate item actually demonstrated. As a result, there is no way of knowing whether the translators used were indeed accredited, whether they were native speakers of their target languages (as specified by EORTC), how much experience they had or what their qualifications were. The fact that results differed greatly across languages makes this a pertinent omission. Indeed, comparison of the phases of the MAPT translation process with those specified in the EORTC procedure reveals that the MAPT process is significantly different.

The first step in the MAPT process, briefing, is not part of the EORTC process (although it could be argued that that role is performed by the manual written by Cull et al. if this were given to translators). The second and third steps, translation and back-translation, are each performed by a single translator and each produces a single translation. The forward and back-translations are compared by the researchers, who presumably do not include a speaker of each of the twelve (later thirteen) target languages, as specified by EORTC. Furthermore, the final teleconference only involves the forward translators, in contrast

with the EORTC method in which the final version sent for pretesting is arrived at through discussions with both sets of translators.

The bulk of the remainder of Ozolins' article concerns e-mails exchanged between translators and the "MAPT team". The first example given illustrates a case in which a difference identified between the source text and the back-translated version led to a modification in the forward translation. The MAPT researchers questioned a back-translation of an item asking about the frequency with which pain stopped a patient from sleeping and the forward translator accepted that their forward translation had asked about the frequency of the pain, rather than the frequency of inability to sleep (Ozolins 2009: 5). This is the only example Ozolins gives of such a case, but he provides some figures for how often differences identified in the back-translation led to changes in the forward translation. He states that the number varied greatly from language to language, with a range of 1 to 20 changes and an average of five changes per language. Comparison of back-translations with the source text had originally identified 11 to 50 discrepancies per language, with an average of 24, meaning that between a quarter and a fifth of suspected differences actually led to alterations to forward translations (Ozolins 2009: 9).

Ozolins provides a greater degree of detail on the cases in which the forward translation was not at fault, reproducing exchanges involving the research team and either or both translators. In the first such example the researchers question the use of "poorer" where the source text had "worse". The forward translator's response was that the word used did mean worse and the back-translator agreed (Ozolins 2009: 5).

In contrast with this example, Ozolins states that "the bulk of disputed instances for back translators were strongly defended by them, often with a pointer to the issue of how the back translator has to choose between a number of viable options" (Ozolins 2009: 5).

One such response from a back-translator actually defended the forward translation as much as the back-translation, explaining that although he or she had back-translated the word used for "moderate" in Arabic as "mild" that did not mean that "the Arabic word used in the FT for moderate is not the right word."[13] (Ozolins 2009: 5). In the next example the researchers question "rather difficult" as a back-translation

---

[13] In Ozolins' examples, FT can stand for "forward translation" or "forward translator" and BT may mean "back-translation" or "backward translator".

of "moderately difficult" and once more the resolution was that the forward translation did not need changing.

Ozolins says that these "exchanges properly consumed much time for MAPT team and translators as they worked through the scales of intensity" explaining that different languages divide these scales in different ways. He also states that the question of whether the distances between translated response options were the same as the distances on the English questionnaire's response scales was a "critical issue" (Ozolins 2009: 5).

This leads Ozolins to discuss what he terms "back translation noise", which is the four-fifths to three-quarters of the discrepancies that did not lead to changes in the forward translation. He introduces the subject of "back translation noise" by accepting that a "critic of back translation would ... have been able to draw on many occasions in this project where the back-translation project led to wrong identification of discrepancies" and listing six types of back-translation noise, which he considers to be "linguistic issues". These are apparent discrepancies in usage of singular or plural, with the specific example of "hip" and "knee", apparent discrepancies in use of tenses, capitalisation and contractions and a recurrent tendency in "most languages" for "hip or knee pain" to be back-translated as "pain in the knee or hip" (Ozolins 2009: 6).

Ozolins points out that these "apparent linguistic discrepancies are perfectly familiar to translators (and bilinguals in general) but often opaque to monolinguals" (Ozolins 2009: 6). This provides the foundation for his claim that the true utility of back-translation was that it "ultimately gave the MAPT team confidence in the translations." (Ozolins 2009: 10).

Ozolins provides some specific examples of "back translation noise" and the translators' explanations of discrepancies that "apparently signal non-conformity in translation where in fact this is not the case." The first example is from a back-translation from Greek in which simple present and past tenses in the English source text were back-translated as continuous tenses. The forward translator states that there is "no difference in Greek". The back-translator contributes, "the meaning in Greek is the same" (Ozolins 2009: 6-7).

The continuous tense examples are followed by two examples of how English use of "some" and "any" provoked differences between the source text and the back-translations in the absence of problems in the forward translations. In the first case the MAPT team question an insertion, requesting confirmation that the Croatian translation has

"pain" not "any pain". This was resolved during a teleconference as follows, "Croatian does not use the word any. This is only in the back translation." (Ozolins 2009: 7). In the second case, the word "some", in "there are some things I cannot do" was missing from a back-translation, which was explained by the back-translator as follows,

> In the FT the word 'some' is not used as it is not idiomatic and therefore in back translating it is not present. Not meaning any disrespect but if 'there are things that one cannot do' does that not imply that they are some and not all?
>
> (Ozolins 2009: 7)

The hint of irritation that can be detected in the second half of this response becomes more explicit in successive examples.

Ozolins suggests that back-translators, who never saw the entire original questionnaire, expressed feelings of "being ambushed", but in the next two examples, it is the forward translator's comments that betray a touch of irritation. In both cases the MAPT researchers questioned why the possessive "my" was missing in a back-translation from Macedonian. In the first case the word was missing from "enjoy my life", in the second it was missing from "my relationships" (Ozolins 2009: 7).

The forward translator's response to the first query was "The FT is the same [as the BT], following the logic that I can enjoy only my life and not somebody's else.[sic] There is no other natural way to say it in Macedonian.", whereas the back-translator's response was more neutral, "FT has enjoy life (which is more appropriate in the Macedonian language in terms of syntax)" (Ozolins 2009: 7).

In response to the query about "my relationships", the back-translator wrote,

> In Macedonian it is stylistically improper to use possessive pronouns more than once in the same sentence, unless necessary, so it is superfluous to repeat the word my. There is no other natural way to say it in Macedonian.
>
> (Ozolins 2009: 7)

Other comments from translators addressed the issue of back-translation directly. One forward translator explained that the "translation process is more than replacing words with the words in the other language." and that "often the same meaning needs to be expressed using different words, which back-translation will not reflect" (Ozolins 2009: 10).

The strongest negative reaction was from a back-translator, who expressed a very dim view of back-translation,

> Dear [MAPT team]: I strongly recommend you speak to your translators face to face. What is coming through is lack of meaning you wanted, which is exactly what is to be expected if you do not talk to either the FT or the BT. Suggest that you talk to your FT to explain the SENSE you want to convey, not try to match the words. Word matching will not work.
>
> (Ozolins 2009: 10-11)

The same back-translator also wrote,

> Once again, I would much rather the Melbourne Uni team talk to a translator to explain the flavour they want to get across, and this can be done. Using FT/BT/ blind will simply lead to the wrong nuances being imported.
>
> (Ozolins 2009: 11)

and ended with what Ozolins called a cryptic comment,

> Once again, garbage in garbage out. If you want meaning, you need to explain what you want prospectively.
>
> (Ozolins 2009: 11)

Ozolins also provides examples that he claims illustrate that the back-translation process allowed both researchers and translators to be confident in the final results. In the case of the researchers, the "aim was precision, pursued at every turn, with the ultimate thought not only of how the patients would read the questionnaire but how orthopaedic surgeons would interpret it." (Ozolins 2009: 9-10).

The example Ozolins provides to illustrate this tenacity does not include the actual words or phrases that are being questioned, but the implication is that a back-translation had asked whether a person's knee or hip condition made it "harder" to look after themselves, whereas the source text had asked whether they made it "difficult". He does not state whether this difference was an artefact of back-translation, but his point is that "hands-on engagement in the translation process ... ultimately gave the MAPT team confidence in the translations" (Ozolins 2009: 10).

Ozolins also claims that the process, in particular "the flow of communication and the ability to have their voice heard, however critically, and to explain their choices", meant that the translators also "had confidence in the final product" and states that "overwhelmingly, the translators engaged in this project", even the one who wrote the "garbage in garbage out" comment, "were willing to do further

translations for the same clients using similar methodologies." (Ozolins 2009: 11)

In addition to its role in convincing the researchers that translations were precise and in allowing translators to explain their work, Ozolins also gives two examples of cases in which the translators' explanation led to changes in the researchers' understanding. The first of these relates to a large number of cases (how many is not specified) in which "some of the time" in a source text was replaced with "sometimes" in back-translation. Ozolins describes this as a case of "unexpected author intentions", since although the researchers stated that "*Sometimes* suggests that a person might receive help occasionally but not on a regular basis" and that "Some of the time implies that a person has regular help", the forward translator's reply was "I'd bet that most English speakers would miss the meaning of "help on a regular basis" when they read 'some of the time' here", adding, "to convey the idea that the person "has regular help", in Spanish we would have to say something like: *Con alguna regularidad* (With some regularity)" (Ozolins 2009: 8).

This, as Ozolins points out, "was an explicit questioning of whether what the MAPT team meant would be understood – by an English reader." However, despite having introduced the episode as an example of communication leading to changes in the researchers' understanding, Ozolins states that the MAPT team did not accept the translator's offer "to make the implicit understanding of regularity explicit", attributing this to unwillingness to "diverge from the English original." (Ozolins 2009: 8).

Ozolins suggests that this situation, "where the authors have persuaded themselves an item has a certain implication but this implication may not be clear to a reader", may not be unusual in questionnaire construction and acknowledges the "peculiar challenge" that this lays down for a translator, since "the authors want the implication to come across in translation, but not be made explicit" (Ozolins 2009: 8).

The second example of back-translation providing opportunities for client education comes from the MAPT team's notes. After discussions with one of the Chinese translators (Ozolins does not state whether it was the forward or backward translator), the team accepted that length could be expressed as "how short" and "how long" and that "very difficult" could be expressed as "large difficulties" because the

translator had explained to them that in Chinese it is necessary "to provide pictures" (Ozolins 2009: 11).

In addition to the twelve languages spoken by immigrant patients in Australia, the MAPT questionnaire was also translated into French. Ozolins explains that French was a later addition and that the methodology used was slightly different, since the forward translation was sent to the "team of medical colleagues in France" who had requested it after the initial check had been performed, but before back-translation. This team suggested changes to the response choices for just one question. In contrast, the team in Melbourne continued with the back-translation process and the French translation that had been accepted by the "expert group" in France turned out to be the version that generated the greatest disagreement, to the extent that French was the only language for which a second back-translation cycle was conducted. This process "resulted in a final translation arguably not significantly different from the first", raising the question of whether the "more favourable response of the French readers [should] have been taken, saving effort?" Ozolins does not answer his own question, but explains that the researchers were "not prepared in this instance to rely on the opinion of an expert group (the French team)" (Ozolins 2009: 9).

In summary, Ozolins' article provides a unique insight into interactions between a medical research team and their translators during a back-translation project. He illustrates cases in which back-translation detected differences that led to changes in the forward translation, at a rate of approximately one change for every five differences detected, and describes the four out of five differences that did not lead to changes as "back translation noise". However, his main argument is that the true value of the back-translation process is in providing opportunities for dialogue between translators and clients, offering advantages for both.

The advantage Ozolins sees for translators is that, as long as the process is "based on transparent communication between translators and authors", then it can be a means of "enabling translators to have their voice heard by clients." (Ozolins 2009: 1).

The advantage Ozolins sees for the clients echoes Grunwald and Goldfarb in that it stems from regulatory requirements and/or the need to satisfy "the demands of peer review", since "the medical field now sees this methodology as something of a gold standard" (Ozolins 2009: 4 & 2). For the researchers, then, the major issue was "the degree of confidence that [they] wanted to have in the ultimate translation" (Ozolins 2009: 9).

Ozolins' final conclusion is unequivocal in its approval of the particular methodology used by the MAPT team, which, he claims, "ensured great confidence for the authors in the translations produced, while also proving an unusually rich occasion for translators to provide clients with an understanding of their tasks and challenges, providing a learning experience for both sides." (Ozolins 2009: 11).

2.4.3   *Perneger, Leplège and Etter (1999)*

The earliest of the three articles providing some type of critical analysis of back-translation was published in the Journal of Clinical Epidemiology in 1999 and is related to the IQOLA cross-cultural adaptation process (described on pages 100 to 105 above). This paper compares what is termed a "rapid translation" of the SF-36 health status survey questionnaire (from English to French) with the version resulting from what they described as a "comprehensive adaptation", produced using the full IQOLA method (Perneger et al. 1999: 1037). This was one of the initial ten translations of the SF-36 described by Bullinger et al. and so it started from the original version, rather than the decentered "international" version (Bullinger et al. 1998: 916-917).

The "rapid translation" was not a simple forward translation, but began with three independent forward translations, two by professional translators (one freelance and the other working in-house for the WHO) and the third by a team of medical researchers. These translations were then synthesized by "two independent experts", one of whom was "the head of French-English translation services at WHO", together with Perneger (who is a Doctor of public health and preventative medicine). This synthesized translation then underwent pretesting with a sample of 50 people and the entire process took 3 months (Perneger et al. 1999: 1038). Perneger et al. refer to this as the Geneva version or the rapid translation in their article.

In contrast, the IQOLA translation process took several years. The sequence was as follows: multiple independent forward translations were "checked by back-translation"; each item was rated for translation difficulty, quality and concordance with the original; a Thurstone scaling exercise involving 60 people was conducted for the response continua (see page 101 above); the original American developers took part in selecting the translation options; lay panels were set up to discuss acceptability with varying sectors of the public; meetings were held throughout the process with IQOLA teams producing other language

versions to maintain a common approach; and two different versions of the instrument underwent psychometric testing at different stages of development (Perneger et al. 1999: 1038).

Perneger et al. had access to data collected in two consecutive years (1993 and 1994) from 946 Swiss adults enrolled either on a newly established managed care health plan or on a conventional indemnity health insurance plan. The data were collected as part of an evaluation study of the managed care plan that had used the rapid translation in 1992 and 1993, but adopted the official IQOLA translation of the SF-36 for 1994, so Perneger et al. were able to compare the results from the rapid translation with results from the IQOLA version for exactly the same 946 respondents. They claimed that this meant that "differences in psychometric properties cannot be ascribed to differences in test populations", before immediately acknowledging that one uncontrolled variable remained, which was time, since 1 year had passed between answering the rapid translation and answering the IQOLA translation.

This approach is essentially a test-retest exercise since statistical tests of agreement between repeated measures can be applied treating the two sets of results as though they had been collected using the same questionnaire because they use the same scoring scales with all elements in the same order and are administered to the same people. This is preferable to assuming that two different random samples will have similar distributions of the variables of interest.

Here, because the two samples are the same people no such assumption needs to be made, so attributing differences in the results to differences in the questionnaires used to collect them is therefore done on the basis of a more robust theoretical foundation than if two randomly selected populations were used.

Perneger et al. conducted an impressive array of statistical tests on the results from the two questionnaires, but they also conducted qualitative assessments of the wording, thereby identifying specific sections of the data as of greater interest and leading them to conduct further statistical tests.

The greater part of the qualitative analysis consisted of comparing the two questionnaires' wording to identify differences. This analysis revealed that differences in wording "were too numerous to be reported in detail" since "only 2 of 36 items were worded identically, and only the yes/no response scales were translated in the same way" (Perneger et al. 1999: 1039). Only two examples are provided, and one of these is the now familiar "have you felt downhearted and blue", which was translated as "avez-vous eu le cafard" in the Geneva version

and as "vous êtes vous senti(e) triste et abattu(e)" in the IQOLA translation (Perneger et al. 1999: 1039).

On the basis of the analysis of wording discrepancies, Perneger et al. analyzed the results for the most glaring differences and they make it clear that they had expected the large degree of difference in the wording to lead to significant differences between the results of each version of the questionnaire. However, when they analyzed the results from two items for which the wording appeared to be substantively different, this was not what they found.

In the first case, the response options to the first item in the general health subscale were "excellent", "very good", "good", "fair" and "poor" and the two questionnaires had identical translations for the first three of these options. The last two, however, were translated as "médiocre" and "mauvaise" in the IQOLA version, but as "passable" and "médiocre" in the Geneva version, meaning that option 4 (the second-worst) in one questionnaire was option 5 (the worst) in the other. Despite this Perneger et al. reported that

> ... the distributions of these variables appeared to differ no more at the lower end than at the upper end of the distribution ... as if the rank of the response option, not the attached label, determined the response that was chosen.
>
> (Perneger et al. 1999: 1039).

Later in the article, they return to this finding, commenting that for the "ordinal response scales ... the rank of the response may be at least as important as its description." (Perneger et al. 1999: 1044).

The second example of apparently significantly different wording is an item asking about "walking more than one mile", which was translated as "marcher 2 kilomètres ou plus," in the rapid translation and as "marcher plus d'un km à pied" in the translation produced after more than two years' work, following the IQOLA procedure. Once more, discrepancy in distance (one question asks about a distance twice as long as the other), which might be considered very significant in a health status questionnaire, had very little effect on the results, since the difference between the percentage of respondents who reported limitations at 2km (13%) and the percentage who reported limitations at 1km (11%) could not be distinguished statistically from variations attributable to chance (Perneger et al. 1999: 1039). This is shown by the fact that the p-value was greater than 0.05, which is an arbitrary (but generally accepted) cut-off for the point beyond which findings are

considered to be significant. In other words, these changes in wording did not affect the results.

In addition to wording differences, Perneger et al. also analyzed the questionnaires to identify possible inconsistencies. An example of a possible inconsistency would be that "someone who has difficulty climbing one flight of stairs cannot have no difficulty climbing several flights". Therefore, if a respondent said they could climb several flights, but couldn't climb one flight, their dataset would be tagged as inconsistent. Perneger et al. identified 15 possible inconsistencies in the questionnaires and then applied statistical methods to compare the percentages of respondents whose data were free from all 15 inconsistencies for each questionnaire. These percentages were 94.3% for the Geneva version and 94.0% for the IQOLA version, which, in common with the results for the 1km and 2km questions, are also indistinguishable from a statistical point of view, since the likelihood that this could occur by chance was almost 70% (5% is the conventional upper limit, sometimes 1% is even adopted). Notwithstanding, the non-significant difference of 0.3% favours the rapid version, since it means that three fewer people provided inconsistent responses (Perneger et al. 1999: 1039).

In addition to the tests guided by qualitative analysis, Perneger et al. conducted several purely statistical tests, including tests of the three main groups of psychometric properties that are conventionally used to conduct initial assessments of both original and translated questionnaires of this type. These were a descriptive analysis comparing means, standard deviations, floor effects and ceiling effects; tests of convergent and discriminant validity between items and scales; and tests of scale reliability in terms of the correlation within and between the two questionnaires. In addition to these tests they also conducted factor analysis to illustrate the extent to which the results for the items in each of the SF-36's eight subscales cluster together – reflecting the extent to which they all test the same construct[14] and an analysis of differences between "known groups". Since these tests are likely to be unfamiliar to some readers, I shall briefly explain what each of them is intended to demonstrate before presenting the results.

The descriptive analysis provides a general outline of the way that the 946 respondents' answers to each of the questions are

---

[14] A quick reminder: the eight subscales of the SF-36 are designed to measure the following health "domains": Physical functioning, Role-physical, Bodily pain, General health, Vitality, Social functioning, Role-emotional and Mental health (Perneger et al.: 1999, p.1042).

distributed. Means are calculated by summing the scores from all respondents and then dividing the result by the number of respondents, thereby giving an average score for an item, scale or questionnaire. They are measures of the central tendency of the data. The standard deviation is the square root of the extent to which results deviate from the mean (the variance) and gives an impression of the spread of the data.

Floor and ceiling effects are, respectively, the percentage of results at the lowest and highest values and as such illustrate the percentage of respondents who would potentially have been even further from the mean if the scale used had been larger. In other words, the floor effect for a self-report health status questionnaire, subscale or item is the proportion of respondents who scored lowest for health status and who could potentially have reported that their health status was even worse if the measure had included even more extreme categories. Conversely, the ceiling effect gives an idea of the proportion of the sample whose health status is so good that it may be off the top of the scale used (Bartman 2012).

One important feature of floor and ceiling effects is that there is no way to determine how much further the scale would have had to extend in order for zero respondents to score maximum or minimum (or if such a point exists, some people may always choose the highest or lowest response to certain types of questions). In other words, the additional information that is "hidden" below the floor and above the ceiling is simply lost. (Bartman 2012).

Tests of convergent validity are designed to show that items designed to test constructs that are theoretically related do (or do not, if the test of validity is not passed) produce related results. Tests of divergent validity are designed to show that items that test constructs that are theoretically unrelated do not (or do) produce related results. For example, each of the SF-36's 8 subscales theoretically assesses a different aspect of health status. The tests Perneger et al. performed for convergent and divergent validity were designed to show that items correlated well with other items in the same subscale, demonstrating convergent validity, and also that they correlated better with items in the same subscale than with items in other subscales, thereby demonstrating divergent validity (Perneger et al. 1999: 1038).

Tests of internal consistency are designed to demonstrate that the items within a test such as the SF-36 all contribute in the same direction to the final assessment of the overall construct, in this case health status. Good internal consistency is one indication that a measure reliably

measures that which it purports to measure. Perneger et al. tested each questionnaire for consistency with itself using Cronbach's alpha and tested the consistency from one year to the next (i.e. between the results for the two different versions of the questionnaire) using intraclass correlation coefficients (Perneger et al. 1999: 1038).

Cronbach's alpha compares the variance in the results for each item with the variance in the results for the entire scale. The more the items agree, i.e. the greater the extent to which they all tend towards indicating better or poorer health status (in the case of the SF-36), the closer Cronbach's alpha will be to unity (1.0). If the items are entirely unrelated ("independent" in statistical terminology), then Cronbach's alpha will be zero (Bland & Altman 1997: 572).

Intraclass correlation coefficients can be considered as "an index of correlation between repeated measures, i.e. as an index of repeatability", which, "in the case of repeatability studies ... is essentially a ratio of the variability between subjects to the total variability" (Bland & Altman 1990: 337). Ideally, the time interval between two data collections should be as short as possible, if the objective is to test the measurement method rather than the sample (i.e., if the intention is to detect variation caused by unreliable measurement), so that there is as small a chance as possible that any observed variation has occurred in the sample during that interval. Perneger et al. therefore corrected the intraclass correlation coefficients using the results for Cronbach's alpha, to control for the passage of time. They also compared them with intraclass correlation coefficients calculated for two sets of data collected one year apart (1992 and 1993) using the Geneva translation both times on the basis that variation in these datasets could not be due to differences in questionnaire wording (Perneger et al. 1999: 1038).

The next test of validity that Perneger et al. conducted was factor analysis. In fact, they conducted factor analyses on two levels. The first-level units were the individual items on the questionnaire and the second-level units were the eight subscales making up health status.

The purpose of factor analysis is "to discover if the observed variables can be explained largely or entirely in terms of a much smaller number of variables called *factors*" (Darlington 1997). In the case of the SF-36, the theoretical basis behind the format is that the 35 of the 36 items should provide information on eight different components of a single construct, which is health status. The 36th question asks respondents to rate their overall health in comparison to one year

previously. There is also an intermediate level at which health is divided into two dimensions, Mental Health and Physical Health.

This subdivision of a multifaceted phenomenon is one of the reasons why this type of scale is developed in the first place. Since it is not possible (or rather, not very informative) to measure health status with a single question such as "how are you feeling?", the questionnaire compiles responses to questions about specific elements of health which together provide an assessment of health in each of the eight domains, an assessment for mental and physical health separately and an assessment for overall health.

Perneger et al. suggest that the ideal result of the item-level factor analysis would be an eight-cluster solution and the ideal result for the subscale level factor analysis would be a two-cluster solution, in terms of demonstrating that the (translated) questionnaire was "coherent with the theoretical structure of the instrument" (Perneger et al. 1999: 1041).

The "known groups" test is based on splitting the sample in various ways (16 in this case) to produce two groups so that one of the groups has a given attribute and the other does not. Examples include splitting the sample into those people who had been hospitalized during the year and those who had not and splitting the sample into those who had consulted a psychiatrist or psychologist and those who had not. The theory behind these tests is that since these groups differ in terms of one characteristic, they should also differ in terms of other related characteristics and the questionnaires should be capable of detecting these differences. This analysis is based on prior hypotheses such as, for example, if someone has been taken to hospital one might expect their physical health score to be impacted, and, if someone has consulted a psychiatrist or psychologist, then one might expect their mental health score to be affected.

The results for the descriptive statistical analysis revealed some differences that were significant from a statistical perspective. The mean scores for all eight subscales were significantly different for the two questionnaires, according to Student's *t* test for repeated observations. However, estimated standard deviations for all eight scales were similar. Perneger et al. state that this shows that "absolute scores ... were not equivalent, but differences between scores obtained with the same instrument could be interpreted similarly" (Perneger et al. 1999: 1040). This is more important than whether absolute scores are similar because, as long as ceiling and floor effects do not cut off too much data, both

scales are still differentiating between different levels of health status and recalibration can bring them back into line.

The results for floor effects were "negligible for both instruments ... reflecting the good health status of this population" (Perneger et al. 1999: 1038). In other words, very few respondents (less than 1% in almost all scales) had such poor health status that they were off the bottom of the scale. Ceiling effects "were important for five of the scales but tended to be higher for the Geneva version" (Perneger et al. 1999: 1038). The difference in ceiling effects is itself a manifestation of the difference in absolute scores, since if more of the results for the rapid version are at the maximum (the ceiling effect), the average result (the mean) will also be higher.

Perneger et al. state that the criterion for convergent validity is a correlation coefficient greater than or equal to 0.4 (Perneger et al. 1999: 1038 & 1040). This is an arbitrary cut-off and they do not cite a reference to support it. According to Professor Trochim of Cornell University, there is no consensus on the ideal cut-off for demonstrating convergent validity and so the objective is to amass as much as evidence suggesting convergent validity as possible, in order to give readers confidence in the results (Trochim 2006). The factor analysis also provides evidence of construct validity, so if the factor analysis results agree with the convergent and discriminant validity results then the value of each is enhanced.

Both the rapid translation and the IOQLA translation met the criterion for convergent validity since the average coefficient for the correlations between each item and its own subscale was in the range 0.6 to 0.7 and none of the coefficients were below 0.4 (Perneger et al. 1999: 1040).

The criterion for discriminant validity is that each item should have a higher correlation with its own scale than with other scales. Professor Trochim agrees with this position (Trochim 2006). Of 245 possible comparisons, 241 met this criterion for the rapid translation and 242 met the criterion for the IQOLA translation. Perneger et al. describe these results as excellent (Perneger et al. 1999: 1040).

The results for the internal consistency of each questionnaire and for each subscale of each were described as "satisfactory". This is all the detail that is provided in the text, but a table containing the actual results for Cronbach's alpha for each subscale of each questionnaire reveals that for two of the subscales both questionnaires had the same degree of internal consistency, but the rapid translation had better consistency for four of the six subscales in which scores differed. Table 2.7 below lists

these results with the higher score in bold type for domains in which one questionnaire scored higher than the other.

Table 2.7 – Results for tests of the internal consistency of two different translations of the SF-36 questionnaire, administered to the same sample of 946 young adults, on different occasions with a one year interval

	Cronbach's alpha*	
SF-36 Subscale	Rapid Translation (no back-translation)	IQOLA Translation (with back-translation)
Physical functioning	**0.92**	0.89
Role-physical	**0.85**	0.82
Bodily pain	0.81	0.81
General health	0.81	0.81
Vitality	**0.81**	0.78
Social functioning	0.81	**0.84**
Role-emotional	**0.79**	0.78
Mental health	0.85	**0.86**
*Figures closer to 1 indicate better internal consistency.		

(Data extracted from Perneger et al. 1999: 1042)

The rapid questionnaire did not only perform better in terms of same-scale comparisons. The range of Cronbach's alpha for the IQOLA translation was 0.78 to 0.89, whereas the range of Cronbach's alpha for the rapid translation was 0.79 to 0.92.

In addition to testing each questionnaire for internal consistency, Perneger et al. also tested the results for the two different years against each other, as though they had been produced with the same questionnaire. This would not have been possible if the questionnaires had different numbers of items or different scoring schemes, but, since they are both translations of the English SF-36 and use the same layout, structure and points system, the two datasets could be treated as test-retest results.

This test was conducted using intraclass correlation coefficients and, as Perneger et al. point out, these are not only sensitive to the reliability of the instrument, but also to changes in the sample. As they also pointed out at the start of their paper, the one-year interval between data collection with the rapid translation and data collection with the IQOLA version was the one variable they were unable to control for.

The resulting intraclass correlation coefficients ranged from 0.36 to 0.86, which Perneger et al. describe as "moderate". However, they go on to state that these figures are similar to the intraclass correlation

coefficients for a test-retest analysis using data collected with the rapid translation in 1992 and then again, also with the rapid translation, in 1993. Reference to the article cited reveals that the range of those coefficients was 0.28 to 0.70 (Perneger et al. 1996: 321). In other words, changing a small proportion of the sample, but using an identical questionnaire, had a greater effect than changing translations and using an identical sample, since reliability coefficients of 0.36 to 0.86 indicate greater reliability than coefficients of 0.28 to 0.70.

The next statistical analysis that Perneger et al. conducted on the results from the two questionnaires was factor analysis. They found results that were "remarkably coherent with the theoretical structure of the instrument" and that the results were the same for both translations (Perneger et al. 1999: 1041).

The final statistical analysis of the two different translations of the SF-36 questionnaire was an analysis of known groups. In general, these results were as Perneger et al. had hypothesized for both questionnaires. For example, people who had been hospitalized did have lower physical health scores and people who had consulted a psychiatrist or psychologist did have lower mental health scores. Perneger et al. also state that when a difference was confirmed, in 14 out of 16 cases the rapid translation actually detected a greater difference between the two groups than the IQOLA version (Perneger et al. 1999: 1041).

In the discussion section, Perneger et al. are quite candid about the fact that the overall results were not as they had expected. They begin by summing up what all these results actually mean in terms of what they say about each translation's performance in the field.

> The main message of this study can be summed up in a paradox: despite numerous, and sometimes important, differences in item wordings, the two versions of the instrument displayed almost identical psychometric properties. Indeed, for all psychometric criteria, such as the variability of the scores, the internal consistency of the scales, the factorial structure of the instrument (whether for the 35 items or for the 8 scales), and known-groups differences, results were remarkably similar for the two French versions.
>
> (Perneger et al. 1999: 1042-1043).

Before considering the wider implications of this, they discussed the differences in mean scores and ceiling effects, which they suggest demonstrate the need for calibration of translated instruments on the basis of normative samples, although they also state that the

psychometric properties of both translations are "very close to results obtained for the original instrument in the United States, implying that the two French versions were both of good quality" (Perneger et al. 1999: 1044).

They then move on to consider the wider implications of these results. The first hypothesis Perneger et al. raise to explain the similarity is related to the position Brislin adopted in his 1986 paper to defend decentering (see page 88 above), since it is the suggestion that "one possibility is that the structure and item content of the original instrument are particularly robust, so that even approximate translations failed to alter them."(Perneger et al. 1999: 1045).

Perneger et al. state that the fact that two "very different translation procedures" produced such similar results "suggests that current recommendations for translating and adapting psychometric instruments are insufficiently evidence based." One example is back-translation, which, as I have shown and as they point out, is "commonly recommended". However, they conclude that "the value added to the final product by this procedure has never been empirically demonstrated". The current situation is therefore that "expert opinion and common sense still rule the field." (Perneger et al. 1999: 1045).

Perneger et al. close their article with the following summary,

> ... two adaptations of the SF-36 health survey into French that followed different methods and selected quite distinct wordings produced instruments of almost identical reliability and validity. This surprising result underscores the current lack of empirical evidence about the effectiveness of various procedures used for translating psychometric instruments.
>
> (Perneger et al. 1999: 1045)

3   TRANSLATION STUDIES THEORY

There are a large number of different approaches to translation studies, as a quick look at the list of titles in St. Jerome Publishing's "Translation theories explained" series shows. They are as follows: "Translating as a purposeful activity: Functionalist approaches explained", "Translation and Gender: Translating in the 'era of feminism'", "Translation and language: Linguistic theories explained", "Translation and empire: Post-colonial theories explained", "Conference interpreting explained", "Translation and literary criticism: Translation as analysis", "Translation in systems: System-oriented approaches explained" and "Deconstruction and translation". In the interests of brevity, in this chapter I shall only discuss concepts and theories that are directly related to the arguments I present in the next chapter where I analyze the back-translation phenomenon. I draw on ideas from several of the "schools" in the list above, since different approaches focus on different aspects of translation.

      In general, I have found functionalist insights of greatest use with relation to the translation process itself, although arguments presented by deconstructionists and descriptive scholars are also relevant. With relation to the effect on translation of interactions between cultures and the effect of translation on cultures I have found descriptive and manipulative approaches to be most useful, since they take a very broad view. I shall also refer to Nida and, very briefly, to Catford, since these are the only scholars of translation cited by Brislin's original 1970 paper (Nida 1964; Catford 1965).

      The remainder of this chapter is divided into four sections, reflecting the divisions of Chapter 4, in which I draw on this work to support parts of my analysis of back-translation. In the first section I shall present statements, arguments and concepts related to meaning and language in translation. In the second section I shall present a number of views on the question of whether to prioritise the source text and/or culture or the target text and/or culture. In the third section I present theories that have emerged from a cultural perspective on translation and a systemic perspective on culture. The final section consists of two subsections, the first dedicated to translators' status and the second to work that has investigated translators' expertise.

3.1 Meaning and language

Discussions of meaning in translation studies tend to start from a model of communication. In 1964, Nida's model was based on a conceptualization of language as a code. Nida said that this code is made up of "symbols in context" which are imposed on a "medium". The pattern imposed on the medium he called the message (Nida 1964: 121) and he further divided the message into the signal and the content (Nida 1964: 123). Nida believed that the minimum requirement for communication were source, message and receptor (Nida 1964: 7, 9, 40, 147 & 156).

The language-as-code concept is no longer widely accepted, but the remaining elements in his communication model have been developed and elaborated on rather than abandoned. For example, Nord added the role of initiator to better relate the model to professional translation, stating that "the initiator is the person, group or institution that starts off the translation process and determines its course by defining the purpose for which the target text is needed" (Nord 2005 [1991]: 47).

Although Nida's view of translation is considered traditionalist today, he never lost sight of the implications of his three-element model, stressing that "a message which does not communicate is useless" and that "in communication the effective meaning of any message is that which gets through to the receptor" (Nida 1964: 21 & 35). He also recognised that this model also leads to the conclusion, with relation to translation, that "one cannot speak of 'accuracy' apart from comprehension by the receptor" (Nida 1964: 183). Furthermore, he also realised that since "language as a mode of action is described as a system of symbols which signal behaviour", i.e. purposeful behaviour, then a "basic ingredient in any communication is the purpose of the human source producing the message" (Nida 1964: 37).

Where Nida was out of step with theories that have been proposed since, some in explicit contradiction of his work, was to consider that what the "really competent translator" does even when translating between "closely related languages" is "to decode the meaning, transfer the content, and then generate another message in the receptor language" (Nida 1964: 68).

Catford, who was the only other translation scholar cited by Brislin, did not agree that meaning was transferable, stating that "the view that [source language] and [target language] texts 'have the same meaning' or that 'transference of meaning' occurs in translation is

untenable" (Catford 1965: 35), that "the view that translation is a 'transcoding' process" is a "manifestation of the 'same-meaning' or 'meaning-transference' fallacy" and that to believe that "there is some pre-existent 'message' with an independent meaning of its own which can be presented or expounded now in one 'code' ... now in another 'code'" is "to ignore the fact that each 'code' (i.e. each language) carries with it its own meaning, since meaning ... is 'a property of a language'" (Catford 1965: 41-42).

However, Catford based his discussion of the conditions for translation equivalence on the degree to which source and target text "are relatable to (at least some of) the same features of substance" (Catford 1965: 50) and for this he has been criticised. The deconstructionists, for example, attack this concept directly, denying the possibility of any link between language and objects, claiming that what is "signified" by a "signifier" is never more than another "signifier" (Arrojo 2003 [1991]: 11).[15]

Arrojo also contested Nida's "transfer of meaning" model, in which she claimed the original text was seen as a stable and *transportable* object, the content of which can be classified completely and objectively (Arrojo, 2997 [1986]: 12) From the deconstructionist perspective, reading is not discovery of meanings, but the production of meaning, and what is signified is not hidden or built into the text, waiting for a reader to decipher and understand it. Rather, meaning is produced by the reader from their circumstances and from conventions, including language itself (Arrojo & Rajagopalan 2003 [1992b]: 88).

Deconstructionists also claim that there is no difference between figurative (or metaphorical) meaning and literal meaning. Arrojo, for example, stated that no theory of language had ever been able to establish objective and irrefutable distinctions between the literal and the figurative (Arrojo 2003 [1992b]: 36), while Rajagopalan and Arrojo have enlisted Derrida's support to state that the very concept of meaning that is literal, *original* and decontextualised, in common with the possibility of *scientific* thought that is objective and independent of ideologies and of history (Rajagopalan & Arrojo 2003 [1992]: 54) is "white mythology" (Derrida 1972: 213).

---

[15] My translation/paraphrasing from the Portuguese. I have used "signified" where Arrojo uses "significado" and "signifier" where she used "significante", rather than "referent" and "symbol", not simply because they are cognates, but because in Portuguese these two words can also be translations of "meaning" and "meaningful" (rather than "meaner" and "meant") and neither "symbol" nor "referent" convey a sense of "meaning" in English, whereas "signified" does.

Functionalist scholars of translation also hold the view that meaning is created by the receptor of a text (reader or listener). For example, Nord maintains that the meaning of a text "is not something inherent in its linguistic signs; it cannot simply be extracted by anyone who knows the code", since a text "is made meaningful by its receiver and for its receiver", while "different receivers (or even the same receiver at different times) find different meanings in the same linguistic material offered by the text", leading to the conclusion that "a text has as many meanings as there are receivers" (Nord 1997a: 31).

However, in contrast with deconstructionists who entirely reject the concept that interpretation is an activity that aims to recover the author's *original intention* (Rajagopalan 2003 [1992b]:63), functionalists are also interested in the intended meaning of the sender. Functionalism begins from a definition of communication as a type of action and therefore foregrounds intention. Furthermore, it sidesteps issues of consciousness and subjectivity highlighted by the deconstructionists, deducing the existence of intention on the basis that a course of action must be intentional if "there was a choice to act one way or another, to refrain from acting in a particular way, or not to act at all" (Nord 1997a: 19).

Nord resolves the apparent paradox that results as follows. First, "intention is defined from the viewpoint of the sender, who wants to achieve a certain purpose with the text", but, it is "the receiver who 'completes' the communicative action by receiving (i.e. using) the text in a certain function" (Nord 2005 [1991] p. 53).

Furthermore, since in their view the use of signs (i.e. communication) "aims at a particular goal" and since "in order to obtain the intended goal, the producer and the receiver must have some kind of agreement about the meaning of the sign", they also conclude that "signs are conventional and culture-specific" (Nord 1997a: 23).

The position that meaning is created by the receiver has implications for whether to prioritise source text and/or source culture or target text and/or culture.

3.2 Source versus target orientation

The theme of whether to prioritise the source text and/or culture or the target text and/or culture has been a recurrent issue in translation studies at least since Schleiermacher suggested that translators must choose between leaving the author in peace and taking the reader to the author,

or leaving the reader in peace and taking the author to the reader (Schleiermacher 2001: 43).[16]

The reason that Nida is considered source-oriented despite his dynamic equivalence, which is target oriented, is that he believed that the answers to questions of *what* to say and *what* to mean should be sought in the source text, except in cases of "untranslatability", when a solution derived from the target language and culture becomes permissible.

His basic position is exemplified by the following: "a translator must be content to be like his author, for it is not his business to try to excel him" (Nida 1964: 151), and

> ... radical changes are not to be made merely for the sake of editorial improvement or at the translator's whim or fancy. The translator's basic task is to reproduce what he has been given, not to improve it, even when he thinks he can do so.
>
> <div align="right">(Nida 1964: 226)</div>

This was the consensus position when Nida was writing and he himself considered that the central question was the extent to which one should prioritise meaning over form, in other words questions related to *how* to say and *how* to mean. His position, deciding in favour of target language forms, was even slightly radical at the time.

However, with the notable exception of Venuti, the majority of modern translation studies scholars have overwhelmingly adopted a target-oriented approach that considers that all translation decisions (including the initial decision to translate something and the decision of what to translate) are (or should be, depending on the approach) taken on the basis of target-culture conditions and target language requirements.

I shall draw on both prescriptive and descriptive target-oriented approaches when analysing back-translation in the next chapter. Functionalist translation scholars such as Nord and Vermeer provide the prescriptive perspective and I have found their insights more useful with respect to issues of prioritisation of source or target text and matters related to the mechanics of translation, while the self-proclaimed "descriptive" branch of translation studies is more useful for questions of interaction between source and target culture.

---

[16] My paraphrase of a Portuguese translation (by Margarete von Mühlen Poll) of the original German (Entweder de Uebersezer läßt den Schriftsteller möglichst in Ruhe und bewegt den Leser ihm engegen; oder er läßt der Leser möglichst in Ruhe und bewegt den Schrifsteller ihm entgegen.)

These two approaches arrived at their target-orientedness via different routes. For the functionalists, it was Vermeer who broke the link to equivalence, abandoning a purely linguistic view and defining translation as "a type of human action", which, in turn he defined as "intentional, purposeful behaviour that takes place in a given situation at the same time as it modifies the situation" (Nord 1997a: 11). The next element was to acknowledge that "since situations are embedded in cultures, any evaluation of a particular situation, of its verbalized and non-verbalized elements, depends on the status it has in a particular culture system" (Nord 1997a: 11).

The implications of this are that "translation cannot be considered a one-to-one transfer between languages" and that "a translation theory cannot draw on a linguistic theory alone, however complex". Rather, what "is needed is a theory of culture" (Nord 1997a: 11). This realisation, in turn, inevitably means that

> ... one of the most important factors determining the purpose of the translation is the addressee, who is the intended receiver or audience of the target text with their culture-specific world-knowledge, their expectations and their communicative needs.
>
> (Nord 1997a: 11).

Paul Kussmaul explains that because Vermeer "looks at the recipients of a translation in their specific situation in the target culture", he redefines the source text merely as "information offered", meaning that

> ... the source text should no longer be seen as the 'holy original', and the purpose (Skopos) of the translation can no longer be deduced from the source text but depends on the expectations and needs of the target readers.
>
> (Kussmaul 1997: 29)

It is a common mistake to assume that target-*culture* orientation equates to a blanket rule that the target *text* takes precedence over the source text and, sure enough, Kussmaul tells us that Vermeer "was accused of advocating arbitrariness and a disregard of the value of the source text" (Kussmaul 1997: 30). However, while Kussmaul agrees that Vermeer demonstrated that "logically the target-culture conditions are superordinate to source-culture conditions" he says that this does not necessarily mean a free or domesticating translation, since "there may well be cases where target-culture conditions and target-reader expectations are such that a text remains embedded in its source culture" (Kussmaul 1997: 30). In other words, it is the target *culture* that takes precedence, not the target *text*, so if the target cultural system employs

literal or even word-for-word translation for certain uses, such as to reveal language structure in a foreign language class, then the source *text* will take precedence over the target *text*, but only because that is what the target *culture* demands.

As Nord points out, the effect is to dethrone the source text, as illustrated by Vermeer's "Skopos rule", which she translates as follows:

> ... translate/interpret/speak/write in a way that enables your text/translation to function in the situation in which it is used and with the people who want to use it and precisely in the way they want it to function
>
> (Vermeer 1989: 20, quoted in translation in Nord 1997a: 29)

This is the reason for the label "functionalist", since from this perspective, the function of the translation is the ultimate arbiter when taking translation decisions.

In contrast, descriptive translation studies arrived at its target orientation via a different route. This approach to translation widens the focus from analysis of individual translations and theorizes about translating and translations on the level of cultures and in some cases even on a global level. These theorists all share an interest in the history of literary translation and the differences in translation practices at different periods in time and in different cultures, which they analyse from a systemic point of view. This has given them a wider perspective on translation norms, since they observed that what is considered the "correct" way to translate changes diachronically within cultures as well as synchronically between cultures.

The systemic approach was pioneered by Even-Zohar (and later Toury) who adopted the concept of cultural systems and applied it to translation. Even-Zohar coined the term "polysystem" to indicate that cultural systems are not uniform and actually consist of a series of interrelated and interacting systems. In the foreword to "Papers in historical poetics", Even-Zohar explained his basic premise as follows,

> Literature is herein conceived of as a stratified whole, a polysystem ... The repertoire of components (items, models) possessed by the polysystem behaves according to certain principles. The main principle governing this behavior is assumed to be the opposition between primary and secondary patterns (activities/systems). These are hypothesized as universals of any cultural system ...
>
> (Hrushovski & Even-Zohar 1970: 7)

This disarmingly simple hypothesis that literature is a system that interacts with other cultural systems and that translated literature is a subsystem of literature had (at least) two major implications. The first is that it provided a logical justification for considering translations from the perspective of entire cultures and, specifically, from the perspective of target cultures, on the basis that "translations should be regarded as facts of target cultures; on occasion facts of a special status, sometimes even constituting identifiable (sub)systems of their own, but of the target culture in any event." (Toury 1995a: 139). The second is that it suggests that patterns in translation policy (in terms of what to translate) and translation norms (how to translate) are determined on a cultural level.

The descriptive/systemic approach to translation therefore reframed meaning as a question of culture-specific interpretation and reframed source versus target orientation as a question of cultural convention, leading many of these authors to investigate the relationships between interaction between cultures and translation.

3.3 Translation and cultures

The cross-cultural adaptation literature approaches cultural differences as barriers to translation equivalence. This perspective is exemplified by Catford, one of the two scholars of translation that Brislin referred to. He states that cultural untranslatability occurs when a "situational feature, functionally relevant for the [source language] text is completely *absent* from the culture of which the [target language] is a part" (Catford 1965, p.99).

This type of issue also concerned Nida, who distinguished between grammatical and lexical problems, saying that the former can be dealt with "more readily", since the receptor language structure makes certain changes obligatory. He identified three relevant levels of lexical relationship between cultural terms. The first was "terms for which there are readily available parallels, e.g. river, tree, stone, knife", the second was "terms which identify culturally different objects, but with somewhat similar functions" and the third level was "terms which identify cultural specialties". The first level presents few problems and the third obliges "foreign associations", but the second level demands that a decision be taken; "one must either use another term which reflects the form of the referent, though not the equivalent function, or which identifies the equivalent function at the expense of formal identity" (Nida 1964: 167). The second and third levels are what cause the "problems" that users of cross-cultural adaptation wish to solve.

However, Nida was also aware of another class of cultural issues that affect translation, although in more subtle ways. These issues relate to the status of different cultures and/or the status of their languages. It should be borne in mind that Nida's experience of translation was gained in Latin America, Africa and Asia and the materials he translated were Christian religious texts to be used by missionaries in these areas (Nida 1964: ix).

Nida dealt with the different status of cultures or societies themselves from two different perspectives. The first perspective focuses on the status of the target culture and language as perceived by translators and the initiators of translation projects. For example, he warned that one "of the most serious mistakes made is to equate adults in other cultures (especially members of primitive societies) with children in our own", because "adults acquire a great deal of information by means of formal education, and their inexperience or deficiency in literacy does not by any means place them in the category of schoolchildren in our society" (Nida 1964: 143).

The second perspective from which Nida approached the question of cultural status is the view that members of target cultures have of their own culture's status. Nida discussed the concept of populations' "cultural security" with reference to target culture preferences for greater or lesser degrees of formal adherence to the source text, giving the example of diglot publications in which an indigenous language text "is accompanied by the Spanish equivalent" and saying that "the Spanish obviously has greater prestige than the Indian language can ever have" and that the indigenous people "are likely to feel culturally quite insecure, and hence insist that their own language must conform in so far as possible to Spanish" (Nida 1964, p.182).

With the exception of a single reference by Sperber to avoiding "cultural hegemony" (Sperber 2004: S125), the cross-cultural adaptation literature does not cover the effects of interaction between cultures on translation or the effects of translation on cultures, but both subjects are foci of interest in translation studies, particularly the descriptive branch.

Toury's work is a case in point. He pointed out that "translation activities and their products not only can, but do cause changes in the target culture" (Toury 1995a: 138) and, on the basis that "cultures resort to translating as a major way of filling in gaps, whenever and wherever such gaps manifest themselves", this led him to consider the circumstances under which cultures tend to import to a greater or lesser extent using translation (Toury 1995a: 138).

Toury considered that "the starting point is always one of a certain deficiency" in the target culture and that this may be identified "in view of a corresponding non-gap in another culture", or, particularly in a colonial situation, "an alleged gap may be factually pointed out for it by a patron of sorts who also purports to 'know better' how that gap may best be filled" (Toury 1995a: 138). However, Toury claims that even when the gap is identified by comparison,

> ... the more persuasive rationale is not the existence of something in another culture/language as such, but rather the observation that something is 'missing' in the target culture which should have been there and which, luckily, already exists elsewhere and can be taken advantage of.
>
> (Toury 1995a: 138)

Lambert analysed a variety of configurations of import and export between different "cultural traditions" and deduced a number of import/export rules, tending to support Toury's view of the relationship between source and target cultures, since he found that the exporting system holds more power than the importing system (Lambert 2006 [1995]: 98-99).

The different status of cultures is also reflected in different status of translation within different cultures. Even-Zohar distinguished between what he termed 'primary' and 'secondary' positions. When translated literature occupies a primary position,

> ... it participates actively in modeling the center of the polysystem ... it is by and large an integral part of innovatory forces ... when new literary models are emerging, [it] is likely to become one of the means of elaborating these new models ...
>
> (Even-Zohar 1978: 22)

One sign of this state of affairs is that "often it is the leading writers (or members of the avant-garde who are about to become leading writers) who produce the most important translations"[17] (Even-Zohar 1978: 22).

In the inverse scenario, translated literature "maintains a *secondary* position" in which it "has no influence on major processes[,] is modelled according to norms already conventionally established by an already dominant type [and] becomes a major factor of conservatism" (Even-Zohar 1978: 24).

---

[17] Examples of this phenomenon abound. Lord Byron is a classic example from English literature and Machado de Assis, founder of the Acadamia Brasileira de Letras, is an example from Brazil.

Venuti has written extensively about the effects of the second situation, from a position within a society (the United States) that is an exporter of translations and in which translation is in a secondary position. He quantified the translation patterns of the United States and the United Kingdom and found that they "point to a trade imbalance with serious cultural ramifications" (Venuti 2004 [1995]: 14). These ramifications include the way that British and American publishing has

> ... reaped the financial benefits of successfully imposing Anglo-American cultural values on a vast foreign readership, while producing cultures in the United Kingdom and the United States that are aggressively monolingual, unreceptive to the foreign, accustomed to fluent translations that invisibly inscribe foreign texts with English-language values and provide readers with the narcissistic experience of recognizing their own culture in a cultural other.
>
> (Venuti 2004 [1995]: 14)

The fluency and invisibility mentioned in the citation above are central to Venuti's discussion of the status of translation and translators and I shall return to them in the next subsection.

However, there is one more author writing on cultural issues whose theories I have made use of in the next chapter. André Lefevere is interested in the way that cultural systems control the production and dissemination of writing, including translation, which he defines as rewriting.

He identifies two control factors. The first of these factors is the "professional" and, in the case of the literary system, "professionals are the critics, reviewers, teachers [and] translators". They "occasionally" exert control by complete repression of "certain works", but the most prevalent control mechanism is to "rewrite works of literature until they are deemed acceptable" (Lefevere 1992: 14).

This raises a question; acceptable to whom? According to Lefevere, a second control factor sets the parameters within which the first control factor operates (Lefevere 1992: 14). He calls this second control factor "patronage" and defines it as "something like the powers (persons, institutions) that can further or hinder the reading, writing, and rewriting of literature." Patronage is exercised by patrons who may be persons or groups of people (Lefevere 1992: 15).

Of course this is a systemic model and Lefevere duly defines the role of patrons as to attempt to "regulate the relationship between the literary system and the other systems, which, together, make up a

society, a culture." They do this through "institutions set up to regulate, if not the writing of literature, at least its distribution: academies, censorship bureaus, critical journals, and, by far the most important, the educational establishment." (Lefevere 1992: 15).

Furthermore,

> Once a literary system is established, it tends to try to reach and maintain a 'steady state', as all systems do ... Strictly regulated systems even appoint individuals to institutions expressly created to bring that state of affairs into being, such as the Académie Française and other academies.
>
> (Lefevere 1992: 38)

Lefevere subdivides patronage into "three elements", the first is an "ideological component, which acts as a constraint on the choice and development of both form and subject matter", the second is an economic component, since it is the patron who "sees to it that writers and rewriters are able to make a living", and the final element is status (Lefevere 1992: 15).

Lambert provides a good example of one way that two of these factors can be brought to bear on translation,

> ... in case the initiative has been taken by political authorities (e.g. for religious texts), we may also assume that the result of the translational activity as well as the book product once finished will be supervised by those who commissioned it. Those who order generally also pay, those who pay and who are in charge are generally also the ones who order.
>
> (Lambert 2006 [1995]: 95)

Lambert also points out that "institutions other than the political ones may be at least as influential (e.g. private companies, especially multinationals)" (Lambert 2006 [1995]: 96).

Lefevere further classifies patronage regimes as either differentiated or undifferentiated, saying that undifferentiated patronage exists when "its three components, the ideological, the economic, and the status components, are all dispensed by one and the same patron." (Lefevere 1992: 17). He also echoes Lambert in suggesting that the power to patronise is shifting from political to corporate entities, saying that "undifferentiated patronage need not be based mainly on ideology ... [the] economic component, the profit motive, may well lead to the re-establishment of a system with relatively undifferentiated patronage" (Lefevere 1992: 19).

The ideological, economic and status elements of patronage all affect the status of translators and translations, which is the subject of the first subsection of the next section.

3.4 Translators

As Nida stated back in 1964, "the principles and procedures of translation cannot be fully understood or objectively evaluated without recognizing the important part played by the personal involvement of the translator" (Nida 1964: 9). In this section I shall briefly summarise work related to translators from which I have drawn inspiration. The section is divided into two subsections. The first presents a little of what has been written on the subject of the status of translators and of translations and the second provides an overview of insights into the competence and expertise of translators.

3.4.1   The status of translators and translations

The general consensus on the subject of translators' status is summed up in a statement made by Nida, who said that for "the most part, the translator's lot has been one of little thanks, poor pay and plenty of abuse" (Nida 1964: 145).

        This consensus has not changed since, Hönig, for example, says that translators "are rarely seen as experts ... by their clients and other users of translation. Often the only expertise that the public appreciates is that of knowing a foreign language well" (Hönig 1997: 12). The corollary is that "many users/clients are very reluctant to see themselves as laypersons in the field of translation" and, if they know a foreign language, they feel that this "qualifies them to not only have a say in matters of translating, but also to evaluate and criticise them" (Hönig 1997: 15). Furthermore, a naive belief in symmetry in translation means that "in many cases ... clients exercise considerable pressure on translators to provide quasi-scientific proof that their translation is correct" with the result that the translator translates "defensively", following a strategy of "I know that this does not sound particularly good but I challenge you to prove that it is not correct" (Hönig 1997: 17).

        According to Vermeer, the way to change this situation is to make commissioners of translations aware that the translator, who "is the expert in transcultural communication on whom the success of the communicative act depends", is "a human and not machine and [is]

therefore to be treated as a human and respected as an expert in his field" and is "the commissioner's partner who works for him and collaborates with him ... and therefore needs the commissioner's confidence and goodwill and collaboration" (Vermeer 1994: 14).

In contrast with this ideal, what has been observed is that very often "people deny the translator the necessary information about documents", which prompted Vermeer to ask whether doctors or lawyers could "work efficiently if you hide important information from them?" and whether they would even accept work under such conditions (Vermeer 1994: 14).

Lambert has also described a similar situation, in this case drawing on research conducted into the Belgian translation market and published in Dutch (Hermans, Simeons & Jansen 1994). He investigated translators' status and found that "Job dissatisfaction was a problem for most people involved in business translation, both among employers and employees". In the case of translators, they went so far as to complain they were "being treated like a 'fax machine'." (Lambert 1996: 285).

Lambert and Hermans hypothesised that this was because

> ... knowledge of a foreign language is sometimes supposed to be God-given, the general view is that there is no reason to spend much money on someone who just happens to know the necessary languages or who enjoys language games anyway.
>
> (Lambert & Hermans 2006 [1998]: 155)

Furthermore, the prevailing attitude is that translation can "be done over the weekend or at home, when the 'real job' is over" (Lambert & Hermans 2006 [1998]: 155).

Lambert considered that the root of this attitude was that "the general language component, in business communication ... is not integrated into the corporate identity", pointing out that "many companies try to communicate multilaterally while their planning is monolingual" and that this is symptomatic of a situation in which "translators and language experts are nowadays being used by managers and customers who view language very naively", concluding that "in addition to competent translators, experts will have to train their environment, their employers and customers" (Lambert 1996: 285).

With relation to medical translation, Ruuskanen studied the attitudes of Finnish medical experts to their translators and found a paradoxical situation in which, despite "the specialist knowledge clients expect them to have, translators and editors seem to be regarded as glorified secretaries". This echoes Lambert's findings, but Ruuskanen

also pointed out that "when the x-ray technicians and the lab assistants are thanked, the translators ... are often left out of the acknowledgements" (Ruuskanen 1995: 299).

This phenomenon has been labelled the translator's invisibility by Venuti, although in relation to literary translators. He links translators' status to the status of their translations, saying that in the Anglo-American literary system translation "is stigmatized as a form of writing, discouraged by copyright law, depreciated by the academy [and] exploited by publishers and corporations, governments and religious organizations" (Venuti 1998: 1). Venuti attributes this low status to a tendency in native Anglophone cultures to value the apparently neutral discourse of science.

> The enormous economic and political power acquired by scientific research during the twentieth century [and] postwar innovations in advanced communications technologies ... have affected every medium, both print and electronic, by valorizing a purely instrumental use of language and other means of representation and thus emphasizing immediate intelligibility and the appearance of factuality
>
> (Venuti 2004 [1995]: 5)

This preferred style is also imposed on translation, leading to what Venuti has famously termed the translator's *invisibility*. With relation to "contemporary Anglo-American culture", Venuti claims that

> ... translated text, whether prose or poetry, fiction or nonfiction, is judged acceptable by most publishers, reviewers, and readers when it reads fluently, when the absence of any linguistic or stylistic peculiarities makes it seem transparent ...
>
> (Venuti 2004 [1995]: 1)

This transparency is the result of "fluent discourse, of the translator's effort to insure easy readability by adhering to current usage, maintaining continuous syntax, fixing a precise meaning" and it "conceals the numerous conditions under which the translation is made, starting with the translator's crucial intervention in the foreign text" (Venuti 2004 [1995]: 1). Furthermore, Venuti considers that the translator's invisibility is directly proportional to the extent to which their translation is fluent (Venuti 2004 [1995]: 2).

In addition to fluency, Venuti identifies a second factor in Anglo-American culture that encourages translators' invisibility. This is a conception of authorship based on a belief that

> ... the author freely expresses his thoughts and feelings in writing, which is thus viewed as an original and transparent self-representation, unmediated by transindividual determinants (linguistic, cultural, social) that might complicate authorial originality.
>
> (Venuti 2004 [1995]: 6-7)

This, in turn, has two implications for the status of translations. Firstly,

> ... translation is defined as a second-order representation: only the foreign text can be original, an authentic copy, true to the author's personality or intention, whereas the translation is derivative, fake, potentially a false copy.
>
> (Venuti 2004 [1995]: 7)

Secondly, "translation is required to efface its second-order status with transparent discourse, producing the illusion of authorial presence whereby the translated text can be taken as the original" (Venuti 2004 [1995]: 7).

Finally, Venuti points out that the norm of fluency and the expectation of transparency devalues not only the status of translations, but also of translators, even, in the case of the latter, in terms of their own self-image, stating that however "much the individualistic conception of authorship devalues translation, it is so pervasive that it shapes translators' self-presentations" (Venuti 2004 [1995]: 6), before defining the translator's invisibility as "a weird self-annihilation, a way of conceiving and practicing translation that undoubtedly reinforces its marginal status in Anglo-American culture" (Venuti 2004 [1995]: 8).

In addition to pursuing a foreignizing translation strategy, Venuti suggests that, in order to change this status quo, translators should

> ... demand contracts that define the translation as an 'original work of authorship' instead of a 'work-for-hire,' that copyright the translation in the translator's name, and that offer standard financial terms for authors, namely an advance against royalties and a share of subsidiary rights sales.
>
> (Venuti 2004 [1995]: 311)

In the next (and final) subsection I shall briefly present some of the findings of research that has investigated translators' competence from a cognitive perspective.

3.4.2 _Translation as an expert activity_

As I have shown, Hönig, Nord, Vermeer and Venuti all believe that translators should be treated with the respect accorded to professionals, but that this is not the rule. In this subsection, I shall present a little of what is currently known about expert translators' skills and abilities. The focus here is on what differentiates them from other language professionals (i.e. the actual practice of translating one text into another) and from multilingual people who are not expert translators but may engage in translation (i.e. the way they go about the practice of translating one text into another).

Researchers who have investigated the cognitive aspects of translation have arrived at a consensus: "translation has come to be seen as an expert activity involving both linguistic and cultural knowledge and competence" (Kaiser-Cooke 1995, p.135). Hurtado Albir, for example, points out that there is a difference between bilingual competence and translation competence, explaining that translation competence is specialised knowledge comprising a group of skills and knowledge that makes translators unique and differentiates them from other bilingual speakers who are not translators (Hurtado Albir 2005: 19).

She uses a definition of specialised knowledge developed within cognitive psychology to situate translation competence as a form of specialised knowledge, on the basis that it is dependent on specialization by the person who is acquiring translation competence, unlike communicative competence. She highlights three characteristics that are common to specialised knowledge of all types: dependence on a firm foundation of extensive knowledge, organisation in complex structures, and applicability to problem solving (Hurtado Albir 2005: 21).

Kaiser-Cooke explains the difference between the way that novice and expert translators operate as follows,

> Novices tend to view problems at their explicit, face value (in translation, at the word level, for example) whereas experts, whose knowledge is organised around inferences about principles and abstractions, categorise problems at a higher, more abstract level, which in turn facilitates both analogy-building and more flexible adaptation of problem-solving strategies to meet new situations
>
> (Kaiser-Cooke 1995, p.135)

In turn, the "problem-solving activity ... takes on the nature of cognitive routines which do not require reflection but occur 'automatically'", and

"at various stages in the knowledge acquisition process, qualitative leaps take place, in which knowledge is both automatised and restructured on a higher level." (Kaiser-Cooke 1995, p.135-6).

This is why novices "are unable to produce adequate translations". They have "inadequate inference and abstraction capabilities, underdeveloped holistic processing and insufficient problem representation." (Kaiser-Cooke 1995, p.136).

Furthermore, Kaiser-Cooke also states out that it is not just resolution of problems that is characteristic of experts, "problem-recognition is [also] a salient feature of expertise". As she points out, "we are all familiar with novices or laypersons who describe texts as 'easy to translate' because they are not aware of the difficulties (i.e. the nature of the problem) involved" (Kaiser-Cooke 1995, p.136).

Fabio Alves, in turn, has shown that it is not only with respect to identification and solution of problems that novice translators differ from experts. He explains that observations made by Lorenzo have shown that the composition and revision phases of translation are different from each other to the extent that the cognitive rhythm of revision is completely different to the rhythm of the composition phase (Alves 2005: 119, referring to Lorenzo 2009). Lorenzo's results had shown that novice translators very often worsen their own translations during the revision phase due to a lack of text revision skills (Alves 2005: 121, referring to Lorenzo 2009), leading him to hypothesise that novice translators struggle with revision because they lack both strategies appropriate for detecting and evaluating errors and strategies for correcting them, so the greater part of their revision effort is fruitless, or even counterproductive (Alves 2005: 119, referring to Lorenzo 2009).

Alves explains that novice translators acquire competence by means of a process of restructuring and development of incipient knowledge – a translation pre-competence – into specialized knowledge, which is translation competence itself (Alves 2005: 121). In other words, translation competence is the result of consolidation of specialized knowledge over time (Alves 2005: 121). Specialized knowledge, in turn, is the result of personal development through effort and years of practice (Alves 2005: 124).

Alves conducted experiments to investigate whether differences in level of experience differentiate experienced translators qualitatively from novices, to the extent that the former obligatorily perform better than the latter (Alves 2005: 124). Working from the assumption that (bilingual) language competence is not the same as translation competence, but is a prerequisite for translation competence, he chose

three people as experimental subjects who were all bilingual in Portuguese and German and all had at least a bachelors' degree in languages. He considered this the minimum level of language competence needed to produce a translation from German to Portuguese (Alves 2005: 126-127).

In these experiments Alves gave the same German source text to three people, all with language competence, but differing in degree of translation experience and formal education and/or research experience in translation studies. He refers to these people as T1, T2 and T3, in ascending order from least to most experienced. The first, T1 had the least translation experience and had taken some modules in translation studies as part of her language degree. The second had more experience than T1, had published some translations, but did not translate regularly, and had a doctorate in translation studies. The third, T3, also had a doctorate in translation studies, but was additionally described as a translator of undeniable experience and as a nationally renowned translator whose translations are published by high-prestige publishers and was chosen as an example of consolidated experience (Alves 2005: 127).

Each of the three subjects translated the text from German to Portuguese, following a translation brief defining, among other elements, target audience. Alves used a program called *Translog* to monitor keypresses, pauses, mouse usage and other elements. The program produces a report of a translator's computer use from the start to the end of a task.

When the logs of computer use were analyzed, Alves found that T1 translated each different lexical item literally and continuously in the order they appeared as she progressed through the translation (Alves 2005: 134). He commented that T1's linear process appears to be motivated by an excessive concern with translating lexical items in detriment to macrotextual elements (Alves 2005: 135).

His verdict was that T1's translation process and textual production revealed that she had sufficient bilingual subcompetence, but lacked mastery of other subcompetencies needed to perform the task of translation (Alves 2005: 137). He stated that she had not mastered the instrumental subcompetence, the knowledge about translation subcompetence or, most importantly, the strategic subcompetence (Alves 2005: 137). These subcompetencies will be defined at the end of this subsection, on page 121.

The second subject produced a text which Alves judged had a satisfactory degree of cohesion and some interesting collocations. He

considered the text to be fluent, cohesive and coherent, but he also identified a number of failures (Alves 2005: 139).

In contrast, T3's composition phase resulted in a draft that contained two typographical errors (corrected in the revision phase) and had certain solutions indicated as provisional using brackets, but which Alves considered was coherent and cohesive, provided the information contained in the source text and offered the target audience a fluent translation that met the specifications of the brief (Alves 2005: 143).

Alves' analysis of T3's translation process highlighted his well-balanced management of the task. Alves identified the decision to postpone solutions to certain clearly indicated items until the revision phase as evidence of hierarchical process management (Alves 2005: 143).

Alves evaluated the three subjects on the basis of these and other results as follows: T2's performance was superior to T1's and the texts produced by T3, both after composition and after revision, were indisputably superior to the texts of the other two subjects (Alves 2005: 147). Alves' overall conclusion was that specialized knowledge, in this case translation competence, increases or reduces in line with degree of experience (Alves 2005: 147).

It is clear from all of the above that translation competence is obviously a highly complex phenomenon, or group of phenomena. The translation subcompetencies referred to by Alves are elements of Hurtado Albir's model of translation competence and I shall end this chapter with her description of this model.

Hurtado Albir's model of translation competence was developed after analysis of the results of empirical-experimental studies of translators. It defines translation competence as including both specialised knowledge, comprising all of the declarative and operational knowledge needed to know how to translate, and a number of psychological and physical components. The specialised knowledge is further classified into five subcompetencies: bilingual subcompetence, extralinguistic subcompetence, translation knowledge subcompetence, instrumental subcompetence, and strategic subcompetence (Hurtado Albir 2005: 28). Below is my English translation of Alves' Portuguese translation of her description of each of the subcompetencies and components:

> The bilingual subcompetence comprises essentially operational knowledge necessary for communication in two languages: pragmatic, sociolinguistic, textual, lexical and grammatical

knowledge. The extralinguistic subcompetence comprises essentially declarative knowledge about the world in general and specific areas: (bi)cultural and encyclopaedic knowledge. The knowledge about translation subcompetence comprises essentially declarative knowledge about the principles that guide translation (translation units, types of problems, processes, methods and procedures employed) and about professional matters (types of tasks and target audiences). The instrumental subcompetence comprises essentially operational knowledge about use of documentation sources and information and communication technologies as applied to translation.

The strategic subcompetence comprises operational knowledge needed to ensure the effectiveness of the translation process. It plays a central role, since it controls the translation process and it is used to plan the process and the translation project (choosing the most appropriate method), to assess the process and the partial results in the light of the final objective sought, to activate each of the other subcompetencies and compensate for their deficiencies, and to identify translation problems and apply procedures to solve them.

Finally, the psychological and physical components that impact on the model are: cognitive components, such as memory, perception, attention and emotion; attitudinal aspects such as intellectual curiosity, perseverance, discipline, a critical spirit, knowledge of, and confidence in, one's own abilities, knowledge of one's limitations, motivation, etc.; and skills such as creativity, logical reasoning, analysis, synthesis, etc.

(Hurtado Albir 2005: 28, my translation of Fabio Alves' translation)

I am aware that this citation is longer than is conventional, but I believe it is justified in view of the fact that many customers of translators apparently still believe that what we do can be equated to what a fax machine does (i.e. "scan → encode → send", or "receive → decode → print"). I think that, taken in conjunction with Alves' analysis of translators' methods, the passage above makes it clear that what translators actually do is not only very much more complex, it is also in no way analogous to a fax machine.

In the next chapter I shall attempt to draw together the many disparate threads of this subject, analysing back-translation in the light of the major concepts discussed by the translation theorists presented in this chapter and in the light of my own experience as a translator and my own theoretical reflections on the subject.

4   ANALYSIS OF BACK-TRANSLATION

4.1 Overview and preliminary assertions

It is my intention in this chapter to present a number of arguments that, taken in conjunction, lead to a common conclusion. This conclusion is that back-translation is neither effective for its avowed objectives nor necessary to achieving those objectives and should therefore be avoided whenever specialist expert translators are available to translate from the source to the target language.

In view of the wholesale acceptance of back-translation for cross-cultural adaptation of questionnaires and given the huge body of literature that has already built up, apparently confirming the utility of back-translation, I am aware that these arguments must be very convincing if they are to have any chance of persuading researchers who currently use back-translation to abandon what they see as a tried and tested technique.

My main arguments against the use of back-translation can be grouped under three headings and I shall expound them in subsections 4.2 to 4.4 of this chapter, dealing with questions related to meaning and equivalence, source versus target orientation, cultural issues and back-translation and translators, respectively. However, before presenting these arguments, I wish to make some preliminary assertions. One of these assertions is actually a development of my main motivation for writing this thesis, which is the conviction that back-translation does not do what is claimed of it, but it could not be fully explored in the introduction, where it is more usual to justify the reasons for conducting a research product, because the back-translation literature is unknown within translation studies. The decision not to criticise the back-translation literature while presenting it, in turn, meant that these assertions and especially the arguments to support them had to be reserved for the analytical section of the thesis, since, as will become clear, the arguments are highly critical of back-translation.

The first of these assertions echoes and expands on a statement made by Perneger et al. In the conclusions to the article they wrote comparing a back-translated questionnaire with one that had been translated without the use of back-translation, they stated that "the value added to the final product by this procedure has never been empirically demonstrated" (Perneger et al. 1999 : 1045).

Considering that that article was published in 1999, 30 years after Brislin completed his doctorate in 1969, it seems startling that this could

be the case and yet back-translation still have been adopted as widely as it has. Surely back-translation's efficacy must have been proven at some point, even if only in the classic literature? However, I have analyzed the evidence available in the back-translation literature, including in the classic article that first analyzed the technique, and I have come to the same conclusion as Perneger et al.

The first of my assertions is therefore that the available evidence does not prove that back-translation tests translation quality (not even along a negative scale, as claimed by Brislin). This is the main subject of the first subsection of this section. However, I also assert that what back-translation may actually test, particularly in Brislin's experiments, is (a lack of) language competence. In order to support this assertion, in the subsection on proof of back-translation I will also demonstrate that Brislin did not test back-translation using expert, competent or even professional translators, but using students who could at best be described as amateur translators and that when, for example, Sperber et al. and Perneger et al. investigated translations performed by experienced specialist translators, there was no longer a role for back-translation since there was no substandard language competence to be detected.

My second assertion, covered in the second subsection of this section, is that not only is there no evidence that back-translation adds value to the final product, as Perneger et al. put it, there is actually convincing evidence that back-translation can reduce the value of the final product. In other words back-translation can be a hindrance to achieving the objectives of cross-cultural adaptation. I shall support this assertion in the second subsection of this section using examples from the literature. With reference to several translations of a single questionnaire into a number of different languages I will show how reliance on the back-translation technique can lead to results that range from the inappropriate to the absurd.

I believe that by providing convincing arguments to support my initial assertions that back-translation does not test translation quality (although it may indicate language competence below the level needed to begin learning to translate) and that it can actually lead to dysfunctional translations, I will also be demonstrating that the entire conceptual basis of back-translation needs to be revisited. Andrew Chesterman said that "If a tool does not serve the function for which it was designed, or any other function, we can get rid of it." (Chesterman 2005 [1999]: 96). While I accept the possibility that back-translation may have a function in ruling out the existence of language competence

(with relation to the very restricted context, genre, subject matter, register and of course language pair being tested), this is by no means the function for which it was designed, and in the four sections that follow these assertions I shall present my main arguments in favour of rejecting back-translation for that function, which was to evaluate the quality of translations.

In these sections I shall cover arguments related to meaning and equivalence, source versus target orientation, cultural issues, and back-translation and translators, in that order, although elements of each theme are inextricably related with elements of other themes, so the exact divisions are arbitrary.

The majority of the final section of this chapter will be dedicated to discussion of the gap that is left if back-translation is indeed rejected and of how best to proceed in order to fill it.

4.1.1   *The value added by back-translation has never been demonstrated*

The majority of my analysis in this subsection will be devoted to evidence in favour of back-translation presented by Brislin. The reason for this is very simple. The contemporary back-translation literature provides very little data that can be considered evidence of the efficacy of back-translation.

When Guillemin et al. published their review of cross-cultural adaptation projects and guidelines for how such projects should be conducted (the AAOS method, version one), they stated that their recommendations were "based on previous research in psychology and sociology" (Guillemin et al. 1993: 1418). They cited Brislin's 1970 article and his 1973 book chapter (Brislin 1970; Brislin et al 1973) and made no further attempt to justify the efficacy of back-translation. One of the justifications they gave for publishing, however, was that "many researchers in [Quality of Life] may not be aware [of] or do not quote this methodological work developed in the psychology and sociology literature" (Guillemin et al 1992: 1428).

When Beaton et al. published a revised version of these guidelines, they asserted that "cross-cultural adaptation of a health status self administered questionnaire for use in a new country, culture, and/or language necessitates use of a unique method" (Beaton et al. 2000: 3186, underlining added for emphasis), but they did not attempt to justify their assertion beyond citing Guillemin et al. It would be very difficult to produce any evidence from the AAOS method that could

demonstrate back-translation's contribution empirically, as Perneger et al. put it, since the back-translations are only used in a committee meeting, together with a synthesized forward translation. Indeed, the fact that alternative translations of problematic items are produced and back-translated during this meeting suggests that, in this version of cross-cultural adaptation, back-translation may actually be a device for the inclusion of monolingual members of the committee.

In the article describing the IQOLA method, published by Bullinger et al., there is also no attempt to justify the value of back-translation beyond citing previous literature, in this case Brislin et al. 1973 and Guillemin et al. 1993. The article presents extensive tabulated data, but these are the results of forward translators' ratings of how difficult to translate they found items and of other translators' ratings of the quality of the forward translations, on the basis of bilingual comparison with the source text. There is no numerical data on the results of the back-translation phase and the only reference to them in the text is as follows,

> Reworking translations with low-quality ratings yielded improvements of the translations, as did the process that compared backward-translations and the original SF-36 questionnaire. The cross-cultural discussion about the translations of items and responses also enhanced the quality of the final translations.
>
> (Bullinger et al. 1998: 922)

The quality ratings mentioned in the quotation above were derived from bilingual comparison (i.e. not back-translation) and the cross-cultural discussions involved the translators and the National Principle Investigators, all of whom were at least bilingual in their native language plus English.

In the first article describing the Sperber method (Sperber et al. 1994), the need for back-translation is justified with reference to a study that found unreasonable results and attributed the failings to direct translation and Brislin's 1973 and 1980 book chapters are cited both as sources of methodology and as support for the need for back-translation. As with Guillemin et al., the basic assumption is that not enough researchers use back-translation and one of the objectives of the article was to show that they should.

However, in contrast with the other contemporary back-translation methods, Sperber et al. do provide some data that relates specifically to the results of the back-translation phase (rather than to

bilingual phases, combined phases and/or target population validation of completed questionnaires). This data includes the partial results of monolingual ratings of back-translated questionnaire items against the corresponding items from a source language questionnaire, both in English. I shall analyze this data together with the data Brislin provided on monolingual ratings.

In the article written by Sperber alone and published ten years later, the exhortations to use back-translation are apparently no longer necessary since he states simply that "The back-translation technique is preferred" (Sperber 2004: S125).

Considering that there has been almost no attempt to prove the value added by back-translation since Brislin,[18] it may be assumed that much of the faith in its efficacy has accrued from cumulative citations, none of which have challenged the original assumptions and all of which are traceable back to Brislin. This is a positive-feedback loop and I shall return to this phenomenon in section 4.3, when I deal with the effects of patronage. Notwithstanding the reasons for the lack of evidence, it makes it necessary to return to the data that Brislin presented when he first recommended back-translation (since even his own later works do not provide additional evidence).

It will be remembered that Brislin had hoped that the results of his back-translation method, or, as he put it, "the verdict of translation adequacy derived from the meaning error standard and a simple pre-test" (Brislin 1970: 215), could replace target culture validation once its own merit had been demonstrated in several studies (see page 73 above). As is clear from the extensive validation processes recommended by all of the contemporary cross-cultural adaptation methods, this obviously did not occur. What actually happened was that back-translation was added to the list of existing methods.

In addition to suggesting that back-translation could replace statistical validation with target populations Brislin also said that the "correlation between monolingual and bilingual meaning errors suggests that translation quality can be judged by either or both criteria, although different aspects of the two-step back-translation procedure are involved" (Brislin 1970: 204-5) and that *bilingual* error detection was "a more direct test of original-target language equivalence than that found

---

[18] There is of course Ozolins' study, but he was not claiming the same utility for back-translation – i.e. that it could be used to assess translation quality. Rather, he claimed that back-translation could be useful for "giving translators a voice" and I shall address this claim in subsection 4.4.2 as part of my discussion of back-translation and translators.

in criterion one since criterion one tests the same language forms, and criterion two tests the different language forms" (Brislin 1970: 192). Bearing this in mind, and considering that all of the contemporary cross-cultural adaptation methods include some type of bilingual stage, it is inevitable that the question be raised of what value the monolingual component (rating back-translations against source language texts) is adding to an inherently bilingual (or multilingual) process such as translation. However, before seeking the answer to this question in the available evidence, I would like to mention some details related to the production of this evidence.

Towards the start of his landmark 1970 article, Brislin discussed previous experiences with back-translation. He was attempting to demonstrate that back-translation works best with some type of decentering, i.e. when the original English text is open to revision. However, he also mentions a case of successful back-translation in which the original English was not open to revision. This was a (US) government-funded research project conducted by Sinaiko into what was in 1963 extremely cutting-edge technology – teleconferencing. Sinaiko had said that "There was virtually no change in meaning" and Brislin attributed this to the relative structural similarity of English and French. However, bearing in mind that both Nida and Ozolins said that closely related languages can be more problematic than distant ones, giving the example of English and French, this explanation is questionable.

Brislin also describes two other reports of successful use of back-translation in detail, both of which revised the original English. The first of the two described in greater detail, Fink (1963) had been forced to use back-translation because "no interpreters competent in both languages could be found". In the second, Werner and Campbell had "Navajo subjects translate and back-translate several simple English passages". Sinaiko's work, in contrast, was "simulating multi-nation conferences." and "skilled interpreters were used as subjects." (Brislin 1970: 187).

This contrast between two situations in which professional translators were not available, leading to the use of sub-competent translators and resulting in back-translation identifying significant numbers of items needing attention, and one situation in which skilled interpreters were used and back-translation found nothing, raises the question of what level of competence Brislin's own translators had acquired. Before analyzing the evidence Brislin presented in favour of back-translation (and also data from Sperber et al.) I will therefore digress a little from the subject of proof and make some points about the

translators used to produce the evidence he provided in support of back-translation.

Brislin began his experiments with translations produced by university students who were speakers of English plus one of nine different Austronesian languages because he believed "using such languages would give the back-translation procedure a stringent test" (Brislin 1970: 194 & 207). He called these people bilinguals, but there are a number of clues in the text that indicate that, with the exception of some Chamorro speakers, they all had fairly weak English, measured on the scale of competence required for translation. The criteria for selection were, in addition to being a native speaker of one of the island languages, to have passed a test of English proficiency and taken a one-semester course in English composition, for which the prerequisite was passing the proficiency test. All of the translations used for the meaning error ratings, the picture task and the Crowne-Marlow scale, in short all of the translations presented in the classic 1970 article, were translated by these students.

I would first like to contrast these characteristics with what was recommended by Nida, one of only two language specialists listed in Brislin's bibliographic references (the other was Catford, who hardly mentioned translators). Nida said that the first requirement of a translator was "a satisfactory knowledge of the source language", but that a second requirement, "a complete control of the receptor language", was "even more important", because "dictionaries, commentaries and technical treatises" can provide information on the source message, but "there is no substitute for thorough mastery of the receptor language" (Nida 1964: 150). Brislin's students fail to meet one of these requirements for all of the translations they undertook, irrespective of whether they were forward translations (i.e. into the mother tongue) or back-translations (i.e. into the acquired language), since in the first case they were lacking in source language competence and in the second they were lacking target language competence.

I would also like to contrast these language qualifications with those of the *weakest* of the three translators investigated by Alves (discussed in subsection 3.4.2 of the previous chapter). She had a bachelor's degree in her acquired language and had been introduced to the basic principles of translation studies at degree level and yet was not considered to have sufficient language competence for part of the task and was considered to have no translation competence.

Brislin's original study design was intended to test translations of nine essays in English on three different subjects and at three different

levels of difficulty. However, he was only able to use the two easier levels because his "bilinguals could not even start to translate" the most difficult level and so it had to be discarded. (Brislin 1970: 195). It should be pointed out that the easier of the two levels that remained equated to "third grade English" (eight-year-olds) and the more difficult to "junior high school" (eleven to thirteen-year-olds). Furthermore, Brislin was only able to find sufficient bilinguals, even according to these criteria, to test these essays for two of the nine languages, Palauan and Chamorro (Brislin 1970: 195).

I maintain that if level of difficulty above that expected of a thirteen-year-old made one third of the test essays untranslatable for these students, it is clear that they did not have the necessary language competence in the source language, when forward translating, and in the target language when back-translating.

I would also point out that if degree of difficulty was an absolute barrier to effective translation then some of the greatest works of literature and science would never have been translated, or would have been rendered worthless in the process. Level of difficulty is a parameter that is relevant to discussions of language acquisition, not to translation, at least not to translation by competent professionals. In my experience, professional translators do not judge texts in terms of how difficult they are to understand, but in terms of how much effort must be expended in research and reading around the subject in order to understand them and whether this equates to a worthwhile cost-benefit relationship, taking into account factors such as current workload, other work available and the possibility of winning new clients or branching out into new areas.

Furthermore, as pointed out in particular by Hurtado Albir, Kaiser-Cooke and Alves, but also by Nida and Nord, translation competence is much more than just language competence in more than one language. Translation competence is primarily acquired through experience translating and the fact that Brislin felt the need to include the phrase "don't be afraid of the word translate" (Brislin 1970: 197) in his instructions to the students is only the first of many signs that they had little or no experience of translation.

In addition to excluding all of the translations for seven languages due to a lack of sufficient students who met the language criteria, Brislin also had to exclude two sets of English-Palauan-English translations, a total of 24, because "several of the essays were almost incomprehensible" and "it was felt that these sets would provide an unrealistic description of the capability of Palauan translators" (Brislin 1970: 205). However, it is clear that they did not in fact throw an

unrealistic light on Palauan translators' abilities since when Brislin recalculated the results including these sets and discarding two other sets at random, the results were unchanged (Brislin 1970: 205). In other words, these four translators followed the same trend as all the other translators, they just exhibited the tendency to a greater extent.

When two of the best Chamorro students and two of the best Palauan students were asked to perform bilingual rating, i.e. comparing forward translations with source texts and identifying meaning errors, they had "difficulties in setting criteria for themselves as to when a meaning error was made" (Brislin 1970: 198) and Palauans, who found translation harder than Chamorro speakers, also had greater difficulties judging translations. Brislin himself pointed out that "Judging someone else's translation includes many difficulties of the actual translation task" (Brislin 1970: 211).

Brislin identified a discernable practice effect, by which translators tended to make fewer errors in each successive translation, and concluded that "as in many tasks, translation seems to improve with practice" (Brislin, 1970: 210). This may appear to echo Alves and others, but I would point out that the improvement Brislin detected was over a maximum of 1800 words of translation, which was the most any single student translated before these results were calculated. If 1800 words' practice is enough to noticeably affect a translator's performance, it seems to me highly likely that their total previous experience (if any) was itself not much greater than 1800 words.

Finally, Brislin's definition of good translation was "eight errors per 300-word essay, or slightly more than one error per paragraph" (Brislin 1970: 214).

On the basis of all of the above, I therefore believe that it is fair to state that Brislin did not in fact test back-translation using translators (which is how he refers to them), or even novice translators, but using language students who had not yet acquired the necessary language competence to begin to develop translation competence.

Having contextualised his experiments in terms of the human resources employed, I am now in a position to return to the evidence Brislin presented in favour of back-translation. As explained in detail in subsection 2.2.1, Brislin developed five criteria for demonstrating translation equivalence and the numerical data he presented relate to these criteria. His criterion 1 was tested by two raters who were Brislin himself and a female bilingual (not bilingual in any of the experimental languages). They followed instructions that requested them to compare the six original English essays with 90 English back-translations and

"write down any errors that you feel might affect the meaning", giving the example of "if the original is 'food' and the second copy is 'hunger', that might be a meaning error, depending on context." (Brislin 1970: 197).

The first test Brislin applied to the results was simply to compare the two raters' scores and analyze the degree of agreement. I shall discuss the method used to do this shortly, but first it is necessary to point out such a test *cannot* demonstrate. This test in no way shows whether the "errors" identified actually were errors. It does not show whether any differences that actually were errors also existed in the forward translations. It does not show that all errors in the back-translations were detected. It does not show that all errors in the forward translation were detected.

Notwithstanding, the correlation coefficients for inter-rater agreement (of 0.71 to 0.93) between monolingual raters appear excellent as shown on page 58 above. However, the percentage overlap figures (of 51 to 64%) show that one rater detected many fewer errors than the other, finding between two-thirds and one half of the number of errors that the other rater detected. These correlations therefore demonstrate that two native English-speaking raters found differences, some of which they agreed on, although one found a greater number than the other, between an English translation of a document that the raters themselves could not read and which had been translated by a non-native speaker of English, who was not an expert in translation, on the basis of comparison with another English text, that was itself written by a native speaker of English and had been translated into the text that the raters could not read, also by a non-expert translator.

This is by no means the same as empirically demonstrating the value added by back-translation. Furthermore, since these errors were not followed up and the iterative technique was not used, there is also no data to demonstrate that back-translation would have contributed to correcting any errors that did indeed exist in the forward translations.

Furthermore, the term "reliability" coefficients is misleading because these coefficients do not necessarily compare the same errors across the two raters, they simply compare the total number of errors found.

I shall explain this in greater detail. Together with the reliability coefficients, Brislin provided figures for percentage overlap. These figures show that the rater who found fewer differences (hereafter Rater F) "missed" between 36% and 49% (depending on subject and difficulty) of the differences found by the rater who detected the larger

number of differences (hereafter Rater L). Additionally, the way the figures were calculated means that only those differences found by Rater F that coincided with differences found by Rater L were taken into account for the overlap calculation. In other words, in addition to the 36% to 49% failure of overlap in one direction, there is an unspecified failure of overlap in the other direction, which means that some of the errors found by Rater F may not have been detected by Rater L and that the number of such errors is unknown. Despite this, Brislin used the total number of errors found, irrespective of overlap, to calculate his coefficients of reliability of 0.71 to 0.93 and does not provide the data needed to calculate the degree of mismatch.

As an example of what I mean by mismatch, imagine a source text that contained the words

"The child was wearing a blue hat"

and a back-translation of an unseen intermediate forward translation containing the words

"The baby was wearing a green hat".

I do not believe that it would be valid, for example, to say that two raters' opinions correlated if one rater identified a meaning error because the back-translation had "baby" where the source text had "child", but the second rater identified a meaning error because the back-translation had "green", where the source text had "blue". According to Brislin's system, such a situation would score 100% agreement because both raters found one error. I believe it would be 100% disagreement because neither found the same error. I also maintain that finding these differences would not prove that the intervening forward translation was incorrect. Brislin himself said that Palauans use the same word for blue and green, to give just one of countless possible alternative explanations (Brislin 1970: 199).

Brislin's reliability coefficients would treat the imaginary case above as equal to a case in which both raters identified the same difference, despite the 36% to 49% known mismatch where Rater L did not identify the differences identified by Rater L and an unknown (but smaller) mismatch in the other direction.

Brislin said that "a fair amount of agreement exists between raters." and qualified it with "this last statement is, of course, the

writer's subjective judgment." (Brislin 1970: 201). I believe that not only is this judgment of the degree of agreement subjective, but also that the level of *disagreement* between raters shows that the ratings themselves are also highly subjective judgements.

Sperber et al. provide the only other data specifically related to the back-translation step and they used a similar system of rating back-translations against source texts. In this case however, there were 29 raters and, rather than indicating an absolute number of meaning errors, each rater evaluated each item from a 35-item questionnaire (plus one retranslated item) on two scales from 1 to 7. On the first scale they rated "comparability of language" and on the second "similarity of interpretability". Counterintuitively, a score of 1 represented greatest comparability/similarity and a score of 7 stood for least comparability/similarity.

Sperber et al. did not provide statistics for the level of agreement between their raters, but they did provide a table including the maximum and minimum ratings for each item on each scale (summarised in Table 2.5 on page 97 above). Sperber et al. presented these data by item, with both scores for each item together and with minimum and maximum scores side by side, which made it difficult to identify patterns in maximum and minimum ratings. I have therefore re-tabulated these data for each scale separately, so that all minimum and all maximum ratings can be viewed aligned as arrays. The results are interesting.

Table 4.1 below lists the ranges of scores for comparability of language.

Table 4.1 – Minimum and maximum ratings (out of 7) for comparability of language

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Min	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Max	3	6	7	6	6	5	4	7	7	6	5	6	7	5	4	4	5	5

Item	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
Min	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Max	5	7	5	4	5	4	5	7	5	5	7	6	6	6	5	3	4	4

(Adapted from Sperber et al. 1994: S517-521)

The most striking characteristic of these results is that, with the exception of item 20, every item was rated at 1 by at least one rater (there is no way of knowing how many raters chose the minimum and maximum values for each item, only that at least one of them did).

Although 1 is the lowest score numerically, it actually equates to the highest level of comparability possible. This is a ceiling effect for 35 out of 36 items (ceiling in terms of the concept comparability, it is a floor effect in terms of absolute scores).

Furthermore, although at least one rater scored items, 3, 8, 9, 13, 20, 26 and 29, as 1 for comparability, at least one other rater scored each of the same items as 7. In other words, at least two people had the opposite opinion on comparability of language for 19% of the items and they expressed this opinion to the greatest extent possible within a 7-point scale. In other words, for these items the range of *disagreement* was 100% of the possible range. This also shows both floor and ceiling effects on each of these items. Additionally, the mean scores of 3.43 and 3.41 for items 9 and 13 (and also, to a lesser extent, of 3.14 for item 26), show that this was not invariably a case of a single outlier choosing 1 or 7 when most raters were at the opposite extreme.

It can also be observed from this table that only items 1 and 34, with a range of 3 points out of a total possible variation of 7 points, have ranges that cover less than half the total possible variation. This means that for 94% of the items the range of disagreement between raters was more than half of the total magnitude of disagreement that can possibly be indicated on a 7-point scale. It is possibly unsurprising that Sperber et al. did not present coefficients for the agreement between their raters.

Table 4.2 below shows the ranges of scores for similarity of interpretability for the same 36 items, also with minimum and maximum scores aligned.

Table 4.2 – Minimum and maximum ratings (out of 7) for similarity interpretability,

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Min	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Max	3	6	6	6	5	3	3	6	7	5	4	6	6	4	4	4	5	5

Item	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
Min	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Max	4	7	5	5	6	5	6	7	6	6	7	4	6	5	3	3	5	3

(Adapted from Sperber et al. 1994: S517-521)

In this table the degree of variation is a little less exaggerated, but it is still greater than the degree of agreement. A total of six items (17%) were rated within 50% of total variation, but none were rated within a range of less than 3 points. In other words, 100% of items provoked at

least 43% of the possible range of disagreement. Of course this also means that there were zero items with total agreement (i.e. items for which all raters chose the same score).

Furthermore, in common with comparability, at least one rater also scored every item except item 20 as 1 for similarity of interpretability, which, despite being the minimum score, is the highest possible level of similarity of interpretability. This is evidence of a ceiling effect affecting all items but one (assuming the measure is similarity rather than total score). There is once more no way of knowing from the statistics provided how many raters chose 1 for each item, nor how large the scale would have to be to eliminate the effect.

This large range of opinions among 29 people on the comparability and similarity of two versions of 36 items, none of which contained more than twenty words, is unsurprising from a translation-studies perspective. Even Nida, supposedly a traditionalist, accepted that the meaning of a text is dependent on the receiver and none of the other authors I have referred to in the previous chapter would disagree.

Nord explains very clearly how the phenomena works, saying that a text "is made meaningful by its receiver and for its receiver", while "different receivers (or even the same receiver at different times) find different meanings in the same linguistic material offered by the text", with the result that "a text has as many meanings as there are receivers" (Nord 1997a: p. 31).

The significance of this fact for questionnaires should not be underestimated. Sperber et al. only used 29 raters and only rated 36 items (giving 72 ratings) and yet for 15.28% of the ratings, their raters disagreed as much as the scale was able to express. Questionnaires exist that are administered to sizable proportions of entire populations.

Despite the very significant degree of *disagreement* between their raters, Sperber et al. averaged the results of all raters before using them to choose items for review. It should be borne in mind that there were 20 items (out of 72) with a rating range of 5 points, 21 with a range of 6 rating points and 11 items with a rating range of 7 out of a maximum possible range of 7, which in turn means that 77.22% of items required more than half of the magnitude of the scale to express the extent to which raters *disagreed*. I consider this (i) to be evidence that there was a greater range of disagreement than there was agreement and (ii) that simple averages are an entirely inappropriate way to present these figures.

When choosing items that required review, Sperber et al. used a cut-off of 3 points on the scales of (not) similarity and (not)

comparability and found that four items were above this cut-off (i.e. less similar and less comparable). However, only a single item, with an average score of close to 5 was actually changed. This was also the only item that none of the raters scored at 1, i.e. that none of them thought was as comparable or as similar as possible. Notwithstanding, at least one rater scored the item at 2 on both scales, which presumably would be *almost* as similar and comparable as possible.

I will comment on the fact that this laborious process only led to the alteration of one item later on in this subsection when I discuss the effect of translators' expertise (or lack of it) on the results of back-translation in more detail, but here, in addition to remarking on the astounding lack of sensitivity of this method of aggregating ratings, I wish to point out the implications for the question of whether the value of back-translation has been proven.

The first significant point is that the only item that failed both back-translation and review was never validated, which means that there is no basis for comparing it with the substitute item and therefore there is no basis for deciding whether back-translation did indeed identify an inadequate translation, whether the original would have been satisfactory too, or even whether back-translation forced the substitution of a more creative translation that would actually have performed better in validation. The back-translation is shown and does indeed give the impression of an error, but the back-translation is not the same as the item that did not undergo validation nor is the back-translation of the substitute item the same as the substitute item itself that did undergo validation, since both were both in Hebrew and neither are presented.

Furthermore, three out of the four items identified by back-translation and the 3.0 cut-off were not substituted after review and Sperber et al. stated that they tracked and compared them with the items that had not been rejected by back-translation and "found no apparent differences between them and the other items." (Sperber at al. 1999: 515).

In other words, the value of having changed the item that was changed cannot be determined and three items that back-translation suggested should be changed were not changed, but were apparently no worse than items that back-translation did not identify as lacking similarity or comparability.

Far from demonstrating that back-translation has added value to the final product, this evidence appears to show that back-translation simply increased the workload involved dramatically, only to be ignored in 75% of cases, with psychometric results that the authors considered

satisfactory. The data from the raters in no way proves the utility of back-translation, but it does illustrate one of the reasons why it is problematic.

Quite apart from the subjectivity introduced by the fact that raters are individuals, all interaction with language is subjective and context-bound. The fact that someone is reading an item in order to rate its similarity to another item is part of that context. It is also entirely different to the context in which, for example, patients answer questionnaires in order to help their doctors acquire information, which will, in turn, hopefully help their doctors to help them. In the first case there is an incentive to split hairs and detect differences, since that is part of the task. In the second case there is an incentive to provide the best information possible, encouraging a constructive reading of the question, and very often an opportunity to ask for guidance, reassurance and explanation from somebody who understands why the questionnaire is being administered.

I shall now return to Brislin's evidence in favour of back-translation. As mentioned above, the only figures given that related exclusively to the results of back-translation were the coefficients for agreement between two raters and the percentage overlap between them. As explained in much greater detail on pages 58 to 63, the data provided on criteria 1, 2 and 3 are not by any means exhaustive. For criterion 1 reliability, Brislin rated fifteen sets of essays, which means that there were essays in more than one language, since the maximum number of essays was 12 for any language. Which languages or how many of each language was not specified.

When Brislin introduces the results for bilingual rating (two Chamorro and two Palauan speakers rating forward translations against source texts), he says there were a total of 18 essays in Chamorro with a mean of 1.8 errors per essay and 10 in Palauan with a mean of 7.6 errors. Unfortunately no figures are given for the magnitude of errors found in monolingual rating, so the figures cannot be compared. Notwithstanding the very significant difference between the results for the two languages, Brislin combines these results when calculating the coefficient of agreement with monolingual ratings, which he also averages. In the case of the bilingual results, this would tend to bring the aggregate results more in line with the Chamorro results, since there are more data points for Chamorro. In the case of the monolingual results, no information was given on how many essays in each language were

rated, which languages they were or how many errors were found per language.[19]

At no point is there any comparison of whether the bilingual raters found the same errors as the monolingual raters (possibly because that would entail trusting the bilinguals to identify the errors as the same in two languages, since Brislin himself was monolingual as far as these languages are concerned). Furthermore, despite having averaged the two raters' results for criterion 1 (back-translation), Brislin takes the first rater to finish the task for bilingual raters without justifying the decision.

For all of these reasons, it does not appear to me appropriate to then calculate correlation coefficients as though all these data were comparable. Furthermore, even with this arbitrary selection of data, the correlations between different criteria were only significant (i.e. not attributable to chance) for coefficients between criteria 1 and 2 (meaning errors in back-translations and forward translations, respectively), at a rate of 0.58, and between ratings of back-translations and questions about back-translations (criterion 1 against criterion 3) at a rate of 0.76.

The first figure, 0.58, merely signifies that the extensively summarised results of judging meaning errors (in different essays) tended to co-vary (in magnitude alone, not in terms of actual errors identified) a little more than half of the time. In addition to the fact that the comparisons were made with results averaged across two languages and that one rater's results were arbitrarily excluded (the second rater to finish), it must also be remembered that the raters who compared forward translations against English source essays (criterion 2) were themselves drawn from the same pool of students whose language competence was so limited that their translations were considered to have had as many as 148 errors per 1800 words, by the very people rating the criterion 1 errors in the back-translations (Brislin 1970: 207).

The second of these figures, a correlation of 0.76 for criterion 1 against half of criterion 3 (just the questions about the back-translations) leads Brislin to say that "judging meaning errors in or answering questions about the back-translated versions are similar tasks" (Brislin 1970: 204).

This is not necessarily true. All that this figure in particular shows is that the results of these tasks, as summarised into averages by Brislin, tended to co-vary, in this case at a rate of about 3 in 4. Furthermore, if

---

[19] To be strictly accurate, all essays rated were of course in English, being source texts and back-translations, the languages I refer to are the languages of the intermediate forward translations, which are the translations whose quality is supposed to be being evaluated.

the back-translations are supposedly a reflection of the forward translations, why is the correlation between the number of errors found in the back-translations and the number of errors made answering questions about the essays from which those back-translations were translated not also correlated? Not only is this correlation not significant and extremely small, at *minus* .02 it actually shows that the relationship that did exist was a tendency *not* to covary. Similarly, the correlation between the number of errors made answering questions about forward translations and the number made answering questions about back-translations of the same essays is non-significant and, at just 0.15, negligible.

Neither of these findings comes close to demonstrating the value of back-translation to a finished translation, particularly since the iterative method was not employed and the errors were never corrected. Notwithstanding, these figures do show that there was something (or some things) in the essays that these people tended to identify at rates that tended to increase or decrease in conjunction. I believe that it can be shown that this "something" is not translation quality but sub-competent language ability.

Brislin stated that the results for analysis of the effects of content (art, child rearing or racial intelligence), difficulty (third grade or junior high) and language (Palauan or Chamorro), all calculated using the criterion 1 results, showed that Palauan "caused" more errors than Chamorro, that art and child rearing caused more errors than racial intelligence and that the more difficult level caused more errors than the easier level. In addition to this he identified the practice effect mentioned above. His conclusion was that all of these factors affected translation quality, although he accepted that with relation to the language effect this could be due to "differential command of English" (Brislin 1970: 209).

I maintain that there are good reasons for considering that this differential command of English may have been responsible for all of these effects. Brislin himself points out that the Chamorro students had significantly better English than the Palauans, which means that the language effect could be dependent on differential command of English, rather than having an independent effect on the number of errors. It also seems clear to me that if difficulty increased the number of errors irrespective of language, then it was not only the Palauans who lacked English competence, but the Chamorro students were also deficient in this respect, although to a lesser degree. Finally, what Brislin describes as a content effect is also actually part of the practice effect, since

Brislin points out that the students had completed courses covering the subjects in which they made fewer errors. The practice effect within content areas reinforces this hypothesis, since the added familiarity would include the vocabulary used to describe the subject in question. In other words, it is a manifestation of increasing language competence, and, more specifically, of increasing specialised language competence.

I therefore propose an alternative hypothesis to explain these observations: all of the effects Brislin detected were actually dependent of an independent variable he did not measure – language competence. Brislin's boldly stated at the end of his 1970 article that "The Criterion One meaning error is here suggested as a unit of translation quality" (Brislin 1970: 213). I shall discuss this simplistic attitude to translation quality in the next section (4.2), but here I would like to point out that the fact that removing or adding the results for 24 essays that were considered close to incomprehensible had no effect on the statistical analysis adds weight to the competing hypothesis that criterion 1 meaning errors are not the units of a(n) (inverse) scale of translation quality, or even the units on any scale, but the results of the subjective decisions of human raters made in a particular context (having been asked to detect errors) and reflecting, with these translators and in an indirect manner, English language competence below the level needed to translate.

During the course of his experiments, Brislin appears to have felt the need for more competent translators, since when he concludes his 1970 article with the seven-step procedure (described on page 74 above), one of his recommendations is to "Secure competent translators familiar with the content involved in the source language materials" (Brislin 1970: 214) and in 1980 he recommended using "wording familiar to the translators" (Brislin, 1980: 432). It appears that finding translators already familiar with the content was a constant problem, however, since by 1986 he was no longer recommending securing translators familiar with the language, or choosing language they were already familiar with, but was suggesting that translators should be familiarised with the actual wording they would be asked to translate, stipulating that "researchers should sit down with translators and go over the materials to be translated, line by line" (Brislin 1986: 149). Without going into the question of quite how one can go over the wording to be translated with the back-translators and still have them translate blind, it is interesting to note that despite changing from amateur to experienced translators and, in later publications, from treating them as "hired help" to treating them as "colleagues" (Brislin

1986: 148), Brislin never reconsidered the use of back-translation, which he had developed with the students.

I should now like to contrast Brislin's results using amateur, inexperienced translators with results from two translation projects that used expert translators who specialised in the subjects they were translating. The first of these was the project conducted by Sperber et al. When discussing possible reasons for translation problems, they commented that "translators are not always knowledgeable enough about the specific content area of the instrument", adding that "specialised medical subjects are an example of this type of difficult content area" and stated that even "good professional translators are often incapable of translating medical material" (Sperber et al. 2004: S125). I would suggest that good professional *medical* translators are actually perfectly capable of translating medical material and, with reference to Sperber et al.'s own results, I believe that it can be shown that they are capable of doing so to the extent that back-translation, which I maintain detects little beyond language incompetence, detects nothing of use.

In the translation studies literature, going right back to Nida, there are frequent references to the fact that it is not enough to have bilingual language competence, nor even to acquire translation competence, but, additionally, "translators must specialize, e.g. in commercial, literary, legal, religious, or technical subjects" (Nida 1964: 242).

As I have shown earlier in this subsection, Sperber et al. were unable to demonstrate even the tendency towards agreement between raters seen in Brislin's results and I believe that this is exactly because they *did* use experienced specialist translators (for the 36 items discussed above). They stated that their 35-item survey of attitudes to preventative medicine was translated into Hebrew by "an experienced translator in the United States" who was "a bilingual physician", and back-translated in Israel by "the director of a company that specializes in medical and scientific translating and editing" (Sperber et al. 1994: 506). I do not therefore find it at all surprising that back-translation detected very little, nor that 75% of what it did detect was considered not to be problematic and subsequently shown to be statistically equivalent to what was not considered problematic on the basis of back-translation results. I have already mentioned that there is no way of knowing if the one item that was changed actually contained an error. There is also no way of knowing whether the bilingual test stage, which came after back-translation, would have detected an error if one did

exist or would have shown that the rejected item was acceptable from a psychometric perspective.

Returning to the evidence provided in favour of back-translation by Brislin, I shall now deal with the results for criterion 4. Criterion 4 was the performance criterion, which has not been adopted in contemporary cross-cultural adaptation protocols, probably because it is rarely practical since the "performance" in question is a physical reaction to a text. As Brislin put it "This suggestion is, of necessity, limited to the kinds of materials that can be examined through bodily movements." (Brislin 1970: 191). Notwithstanding its narrow range of applications, here I am interested in whether the results of the test of criterion 4 provide evidence in favour of back-translation, not whether criterion 4 itself is a valid or practical test of a translation.

Brislin wrote instructions for making a picture from pieces of coloured paper and had them translated into three languages and then back-translated. The back-translations were not used in the test, but a precondition of accepting the translated instructions was that the back-translated instructions had been rated against the source text and no meaning errors had been detected. On this basis Brislin claimed that the results of the test supported back-translation, since back-translation had been used to prepare the test materials.

Brislin gave these instructions to six groups of five people, two groups each of Chamorro, Kusaien and Palauan speakers. For each language, one group followed instructions translated into their own language while one group followed the original English instructions and the number of errors each subject made was recorded.

The results were 0 errors for all Chamorro speakers, 0.2 errors for the group of Kusaien speakers following either set of instructions and 0.2 for Palauan speakers following translated instructions and 0.4 for Palauan speakers following English instructions. Brislin calculated that all groups were statistically equivalent on the basis of eight possible errors. Assuming 8 possible errors, five groups were within the first 2.5% of possible variation (≤ 0.2 out of 8.0) and all six groups were in the first 5% of possible variation (≤ 0.4 out of 8.0).

There are, however, other ways to consider this result. Considering that the maximum number of errors made was one, these results also mean that no Chamorro speakers made any errors, that one Kusaien subject following English instructions and one Kusaien following Kusaien instructions each made one error, while one Palauan speaker made a single error following Palauan instructions and two Palauan subjects each made one error following English instructions. In

other words, three errors were made following English instructions, while one error was made following Kusaien instructions and one error was made following Palauan instructions.

Since no subject actually made more than one error, the range of errors Brislin considered possible seems of little relevance and it appears to me more reasonable to suggest that the basis for comparison should be error versus no error and the range should be the number of people in each group. Seen in that way, zero Chamorro speakers made an error, but 20% of all Kusaien speakers made an error, while 20% of the Palauan speakers made an error following Palauan instructions and 40% of Palauan subjects made an error following English instructions.

Notwithstanding, in the final analysis 25 people made no errors and 5 people made an error, but back-translation detected no meaning errors in the instructions. It is also interesting that three errors were made following English instructions whereas just one error was made following the translation into Kusaien and one error was made following the translation into Palauan. Since Brislin considered the groups statistically equivalent he did not raise the question of why the original instructions caused more errors than the translations. However, if viewed as a ratio of three errors following English instructions to one error each following Kusaien or Palauan (and a ratio of three to zero for Chamorro), these results seem to suggest that the forward translation had actually improved the instructions! It should be remembered that since back-translation found no errors it cannot have been responsible for improving the forward translations.

My interpretation is somewhat different. I suggest that, in contrast with what was seen with the 300-word essays, the instructions were relatively simple and the majority of these subjects and translators had sufficient competence in both languages to either understand the instructions or to translate them, as appropriate. However, one Kusaien subject answering in English and one Kusaien translator had poorer English, as did one Palauan translator and two Palauan subjects. The reasons I suggest that the translators were responsible for the errors resulting from following instructions in Kusaien and Palauan are that (i) there is ample evidence of their less than ideal competence in English, (ii) there is no reason to believe that subjects were sub-proficient in their own languages, and (iii) this explanation requires fewer variables, since the English instructions caused more errors than the translated instructions and the translators and the people following the instructions had a similar level of competence (in English), meaning that any difficulty or ambiguity causing errors when following the English

instructions might also cause a translator with a similar level of English competence problems, which would, in turn, be reflected in their translation.

At face value, the results for Brislin's criterion 5 (administration to bilinguals) appear to finally provide some firm proof that back-translation is of use. However, Brislin used his iterative back-translation method with decentering to produce the Chamorro version of the Crowne-Marlowe scale. As described in detail on page 56 and summarised in Table 2.4, the process consisted of a first "translation → back-translation" step, followed by rating of criterion 1 errors and then review. In this review stage, the two translators and Brislin modified the source text (decentered it) to simplify or eliminate items that had led to criterion 1 errors. The process was then repeated with two different translators with the addition of a pretest with 10 Chamorro speakers, which found errors not found by back-translation. The twice-decentered scale was then translated and back-translated once more and then all six translators and Brislin revised the final version. The revised version and the decentered English version (not the original questionnaire) were then administered to 80 students in the split-half design described on pages 49 and 57 above.

The first point I would like to make is that the fact that the statistical testing occurred after three review sessions involving bilinguals and the fact that pretesting detected problems after two rounds of back-translation mean that the criterion 5 results do not reflect the effect of back-translation alone and, as such, cannot provide the evidence that Perneger et al. identified as missing, i.e. proof of the value of back-translation itself.

However, I do accept that Brislin's criterion 5 results, in common with the final results in Sperber et al. 1994; the results for both questionnaires analyzed in Perneger et al. 1999; the results for all ten versions of the SF-36 summarised in Bullinger et al 1998; and the anecdotal evidence of the MAPT team's satisfaction with their translated questionnaire, provided in Ozolins 2009; all demonstrate successful translation of questionnaires in the terms demanded by the professionals who use such instruments.

Notwithstanding, this is not the same as demonstrating that back-translation is responsible for that success nor is it the same as demonstrating that success is not possible without back-translation. I would now like to go through some possible explanations for the success of Brislin's criterion 5 test.

My first point is that, in common with most of the results Brislin provided, the results for criterion 5 are heavily summarised. Although each group only contained 20 subjects, Brislin averaged the number of true and false answers in each group before calculating his correlation coefficients. In addition to assuming a degree of uniformity in the sample, this decision seems to ignore the fact that the Crowe-Marlow scale has two domains. This means that it is testing two different domains of the social desirability concept and it is inappropriate to analyze them together. For example, Perneger et al. analyzed their results separately by domain and Bullinger et al. provide tabulated results item by item.

This decision appears even less appropriate when it is known that answers to the questions 1, 2, 4, 7, 8, 13, 16, 17, 18, 20, 21, 24, 25, 26, 27, 29, 31 and 33 are used to score the "Attribution" domain and answers to questions 3, 5, 6, 9, 10, 11, 12, 14, 15, 19, 22, 23, 28, 30 and 32 are used to calculate a "Denial" score. Not only do these divisions not coincide with the split-half groups, which were answering odd numbered questions in one language and even numbered questions in the second language, but the two domains have *reverse polarity* with respect to the overall construct being measured - social desirability (Crowne-Marlow 1964, in Tatman et al. 2009: appendix).

Brislin did provide one correlation coefficient for item-by-item comparison, which was for the split-half groups and the result was good, at 0.89. The question therefore arises as to why he did not provide similar statistics for the other comparisons possible. Although this question must remain unanswered, there are a number of possible reasons why the students answering the split-half questionnaires may have been able to make sense of it even if the translation had not been good.

I would first like to return to the issue of the subjective nature of interpretation. The subjects answering the Crowe-Marlowe scale were all students and therefore, presumably, accustomed to taking tests. They were all also studying English. I would suggest that this characteristic of the respondents, in combination with the context of being asked to answer a questionnaire in which alternate questions are written in two different languages, one of which is English, may well have had the effect that the subjects treated the experience as though it were a language test. If so, given the fact that they were all students and so accustomed to trying to score high marks, they would have an incentive to try and make sense of any items that were not immediately clear,

whether because they were not well translated or because they were in English.

Furthermore, I believe that the split-half format actually makes it easier for respondents to answer the language that is causing problems rather than harder, since half of the items would be in a language in which the respondent is proficient and could be used to help deduce the content of the other half. The responses are of course obvious in this context since the only options are true or false for all items. I shall now provide a quick example of the types of clues that are available in the items themselves.

In all of his later articles on back-translation Brislin provides a series of rules for writing easily translatable English. One of these rules is to avoid "adverbs and prepositions telling "where" or "when" (e.g., frequent, beyond, upper)" (Brislin et al. 1973: 33; Brislin 1980: 432) and the reason given is that "there are often inadequate direct equivalents of these words" (Brislin 1986: 146). The Crowne-Marlowe scale already existed and so could not be written according to the rules and actually contains expressions of frequency in 27 of the 33 items. Interestingly, all of these expressions occur in both even and odd-numbered items, meaning that split-half respondents would encounter them in both languages. These expressions (with the numbers of the items in which they appear in parentheses) are as follows: "never" (items 2, 4, 24, 25, 26, 27, 29, 31 & 33), "sometimes" (items 3, 6, 19, 30 & 32), "on occasion/occasions" (items 5, 10, 15 & 23). "at times/there have been times" (items 11, 12, 22 & 28) and "always" (items (13, 16, 17 & 21).

Given that Brislin did not provide coefficients for item-by-item comparison of single-language groups against each other or against the split-half groups (it is fairly safe to assume that he would have done so if they had been better than the split-half coefficient), taking account of the similarity to a language test and the fact that respondents were students and bearing in mind that expressions of frequency are just one example of the type of help that having both languages can provide, I believe that one possibility that cannot be ruled out is that the split-half respondents deduced what was meant in unclear English items using the questions written in their own language to help them and vice-versa. Brislin's claim that errors found in pretesting were "probably due to some back-translators being able to make good sense out of target language passages with several errors" (Brislin 1970: 213) is itself an acknowledgment that language is robust and language users are resourceful.

Notwithstanding these statistical considerations, it must be acknowledged that Brislin did indeed manage to produce a translation that was adequate according to his own criteria. He claimed that this success was due to the application of back-translation, but I would point out that the six translators used were among the best (and Chamorros had the best English) and had already had experience with Brislin's methods translating the 1800 word sets of essays. Furthermore, irrespective of whether errors were identified by back-translation of pretesting, they had to be solved by the bilinguals, who were gaining experience. I believe that the success of the translation is attributable not to the use of back-translation, but to this all-important communication process in which the bilingual translators attempt to fulfil the needs of their monolingual client (although Brislin's translators had not attained competence, there is a client-translator relationship since they were paid to translate and these translators were re-contracted because their earlier work had satisfied their client, who was Brislin himself).

Of course there is an important argument in favour of using back-translation which is that it selects errors that need to be corrected. As Grunwald and Goldfarb put it, "back-translations ... can be used to identify possible problems for review." (Grunwald & Goldfarb 2006: 2).

This is true, but there are two counter arguments to this. The first is that the Crowne-Marlowe scale has 33 items and comprises a total of 551 words. Brislin needed three cycles of "source → forward translation → back-translation → decentering" to arrive at a satisfactory translation of 551 words. I would suggest that after three cycles and more than 200 man hours there were therefore very few words and absolutely no items that had not been discussed with the bilinguals. I also suggest that if my productivity as a professional translator were 2¾ words per hour, I would starve or be forced to change professions.

Notwithstanding, the more convincing argument is that this process of error identification is only necessary because Brislin was using amateur translators. An expert translator is perfectly capable of identifying points that need discussion without back-translation and a translator who is unable to identify potential problems has not yet acquired translation expertise, nor even full translation competence. As Kaiser-Cooke put it "problem-recognition is a salient feature of expertise" (Kaiser-Cooke 1995, p.136).

I believe that I have been able to demonstrate in this subsection that the little evidence that is available does not conclusively prove the utility of back-translation, nor that back-translation detects translation quality, and that a more convincing explanation for Brislin's results is

the competing hypothesis that back-translation was detecting deficient language proficiency in his subjects – deficient for the purposes of translation.

I would now like to return to the evidence presented by Perneger et al. and explain why I consider that, while it does not prove the superiority of one method over another, it does show that a translation method that does not include back-translation, but does employ experienced, specialised translators, is capable of producing a questionnaire that is satisfactory, in the terms demanded by the user of such questionnaires.

To recap, the questionnaire that was translated without using back-translation was translated three times by "an independent professional translator specialized in medical subjects, a translator at the [WHO] translation service, and a team of researchers at the Institute of Social and Preventive Medicine" and then these three translations were synthesized by "the head of French–English translation services at WHO, and a bilingual health survey expert, also at WHO" (Perneger et al. 1999: 1038).

Results from this questionnaire were compared with results from another translation of the same source questionnaire prepared using the full IQOLA method, which of course includes back-translation. The most important element of this comparison was that all of the subjects who answered both questionnaires were the same people, meaning this is the only study that can claim to have compared a questionnaire prepared using back-translation with another prepared from the same source text without using back-translation under anything approaching equal conditions.

The conclusions that Perneger et al. came to were startling, but also unavoidable. They found that

> ... despite numerous, and sometimes important, differences in item wordings, the two versions of the instrument displayed almost identical psychometric properties. Indeed, for all psychometric criteria, such as the variability of the scores, the internal consistency of the scales, the factorial structure of the instrument (whether for the 35 items or for the 8 scales), and known-groups differences, results were remarkably similar for the two French versions.
>
> (Perneger et al. 1999: 1042-1043)

and on the basis of this, they pronounced that the fact that two "very different translation procedures" produced such similar results "suggests

that current recommendations for translating and adapting psychometric instruments are insufficiently evidence based." (Perneger et al. 1999: 1045).

I believe that in this subsection I have demonstrated that the evidence presented in favour of back-translation does not amount to proof of value added to the final product, but this is not, of course, the same as demonstrating that back-translation has no function. Indeed, since the test of the questionnaire prepared without back-translation was that it should have similar properties to one that had been prepared with back-translation, the question can justifiably be raised as to what is the problem with back-translation, even in the absence of evidence of an absolute need for it. It may also be questioned why, if back-translation is a bad thing, the questionnaire translated without it was not better than the one translated with back-translation (in fact, for several psychometric properties it was indeed better, but without statistical significance).

It will require the next four sections to fully answer the question of what is wrong with using back-translation if the final result appears satisfactory regardless, but in the next subsection I shall provide some examples of how this is not always the case and show that the results can be less than satisfactory. With regard to why the back-translation method did not produce a (statistically) worse questionnaire in this particular case, I would point to clues in the article describing the IQOLA method used to translate the questionnaire. In this article Bullinger et al. say that the translators' own ratings of their work, the bilingual ratings given by reviewing translators, the meetings between translators and national principal investigators and the multi-language meetings all helped to improve the translations (Bullinger et al. 1998: 922). In addition to these elements, the IQOLA method also included standardization of the item response scales using members of the target culture and meetings between forward translators and national principal investigators (all of whom are native speakers of the target language), all before back-translation was conducted. All of these methods, in contrast with back-translation, involve the target language in the evaluation process and, regardless of whether possible problems are identified in these stages or by back-translation, it is only a process involving the target language that can ever provide a solution in the target language of the translation. I therefore suggest that such an effort-intensive, collaborative, international and *multilingual* effort would have worked without back-translation and may even have worked in this case despite back-translation.

Ultimately, as mentioned with relation to Brislin's criterion 5 questionnaire, the translation solutions that are actually adopted have to be provided by people who speak the target language, irrespective of whether the need for a solution is identified by back-translation, bilingual evaluation, pretesting or any of the other steps in the Brislin, Sperber, AAOS or IQOLA methods, or simply by competent, experienced translators, as in the "rapid" version tested by Perneger et al. This may seem an extremely obvious point to make, but there is an implicit assumption in much of the back-translation literature that back-translation ensures a culturally appropriate translation. Back-translation is limited to identifying supposed problems and, logically, if there is nobody available who has the local cultural knowledge and translational competence necessary to provide solutions, these problems will remain unsolved.

I shall cover the reasons that I believe that back-translation can often prove more of a hindrance than a help in achieving successful cross-cultural adaptation in the next three sections of this chapter (4.2 to 4.4), but, before I discuss *why* this occurs, it is first necessary to demonstrate that it *does* indeed occur. Therefore, in the next subsection (4.1.2), I shall present several translations of an English language health-related questionnaire into different languages and demonstrate that blind faith in back-translation can lead to dysfunctional translations.

4.1.2   *Back-translation can lead to dysfunctional translations: an example with race, skin colour and ethnicity categories*

It is not my intention to claim that back-translation always leads to dysfunctional translations. However, in this subsection I will show, with reference to published questionnaires, that if back-translation is strictly applied without regard for the final function of the translation, it can result in a translation that is not fit for its purpose.

Vermeer exhorted translators to "translate/interpret/speak/write in a way that enables your text/translation to function in the situation in which it is used and with the people who want to use it and precisely in the way they want it to function" (Vermeer 1989: 20, quoted in translation in Nord 1997a: 29) and it is exactly this approach to translation that can be severely hampered by back-translation.

I will illustrate one of the pitfalls of using back-translation with examples of two items from several translations of a single questionnaire into a number of different languages. The questionnaire is the patient history form for the "Research diagnostic criteria for

temporomandibular disorders", hereafter RDC-TMD (Dworkin & LeResche 1992; Dworkin & LeResche 2012). The temporomandibular joints are the connection between the jaw and the skull and disorders in or around these joints can cause problems with chewing, talking and even yawning (WebMD 2013). In addition to the patient questionnaire, the RDC-TMD also includes a clinical examination checklist, to be completed by physicians, and instructions for the examination and scoring system, also directed at a professional medical audience

The majority of the patient questionnaire consists of a series of questions designed to elicit information of use in diagnosing temporomandibular disorders. However, since the questionnaire is intended for use in research, there is also a section to collect sociodemographic data on the patients. The information requested ranges from date of birth, sex and race or national origin, to details such as recent employment history and average earnings. My interest is in the race and origin categories, although I have used the currency employed in the earnings items and the translation of "zip code" (plus extratextual sources of information) as indications of the intended target populations.

The RDC-TMD has been translated into nineteen different languages at the time of writing (2013) following guidelines laid out in a document entitled "Guidelines for cultural equivalency of instruments" (Ohrbach 2009). The guidelines cite Brislin 1970, Brislin et al. 1973, Bullinger et al. 1998, Beaton et al. 2000, Guillemin et al. 1993 and, interestingly, Perneger et al. 1998, and they prescribe back-translation as one of 11 stages in the translation process.

The guidelines state that "Quality control of instrument development is maintained by blinded independent back-translation ... by a translator whose native language is the source language and whose second language is the target language" (Ohrbach 2009: 10). Additionally, in common with the AAOS method, they specify that the "back-translator should ideally be naïve to subject content" and claim that "particularly when a source instrument will be translated into multiple other languages, back-translation helps to insure maximal similarity in item meaning" (Ohrbach 2009: 10).

The RDC-TMD guidelines also follow the AAOS approach with regard to the forward translators, specifying that one should be "knowledgeable of the instrument content, while the other translator should be naïve to instrument intent and hence culturally representative of the subject population who would be using the instrument" (Ohrbach

2009: 9).[20] The guidelines also recommend forward translators with "in depth knowledge of the setting" on the basis that they may be more capable of achieving "more nuanced translations" (Ohrbach 2009: 9).

The original questionnaire was developed in the United States and the format used to enquire about race and ethnic origin echoes the US national census, in which a race item is followed by an item on "national origin or ancestry" (Travassos and Williams 2004). The original United States versions of these two items are given below.

25. Which of the following groups best represent your race?

Aleut, Eskimo or American Indian	1
Asian or Pacific Islander	2
Black	3
White	4
Other	5
(please specify) _____	

26. Are any of these groups your national origin or ancestry?

Puerto Rican	1	Chicano	5
Cuban	2	Other Latin American	6
Mexican/Mexicano	3	Other Spanish	7
Mexican American	4	None of the above	8

(Dworkin & LeResche 2012: 9)

The first translation I would like to discuss is the Spanish version. The currency item in the Spanish version uses a dollar sign (rather than Euros) and "zip code" has been translated as "código postal" (González et al. 2002: 22). The dollar sign would tend to indicate that it is aimed at Spanish speaking populations in the United States, although it does not rule out many South American countries that also use the dollar sign. The translation of zip code is possibly an indication of an international audience, since even very recent immigrants to the United States would know what a zip code is.

While there is not therefore any conclusive indication of whether or not the audience originally intended included people outside of the United States, the download page on the RDC-TMD website states that "This translated instrument has been used across Latin America" (RDC-

---

[20] There is no discussion of the effect a naïve translator might have when translating the specialist material intended for researchers and physicians which is actually more extensive than the portion intended for laypeople.

TMD 2012). Irrespective of whether the possibility of an international audience was originally envisaged, the two race questions are actually translated as follows:

25. ¿Cuál de los siguientes grupos representa su raza?

Aleut, Esquimal o Indígena Americano	1
Asiático o de las Islas del Pacífico	2
Negro	3
Blanco	4
Otro	5
Especifique _____	

26. ¿Cuál es su nacionalidad u origen ancestral?

Puerto Riqueno	1
Cubano	2
Mexicano	3
Mexicano-Americano	4
Chicano	5
Latinoamericano	6
Hispano	7
Otro	8

(González et al. 2002: 22)

While this solution may be an excellent example of the "maximal similarity in item meaning" that the guidelines claim back-translation guarantees (and will therefore cause no coding problems if data are aggregated with data from the English version), it is not a functional translation outside of the United States.

The first response to the first item includes three different indigenous populations from North America, but many residents of South America would possibly classify themselves as Indigenous Americans. This might be convenient for coding purposes when statistics are analyzed in conjunction (although other versions break the coding structure of the original, as I shall show shortly), but it is hard to see a basis, whether genetic or sociocultural, on which Alaskan Eskimos can be considered part of the same group as Patagonian or Andean indigenous peoples, for example.

The second response is intended for all Asians, plus Hawaiians and natives of United States overseas territories in the Pacific (which includes Guam, where Brislin conducted his experiments). The third and fourth categories can be considered relevant to Latin America, but the

absence of a mixed-race category would exclude large portions of the population.

Finally, the "other" category, usually expected to be a catch-all for minorities, would have to be used by all other residents of Latin America and the fact that there is a blank space for respondents to specify their "other" race has implications for coding. By coding I mean the way that data collected with these instruments would be entered into a computer for statistical analysis. It will be noted that each response has a number next to it. These are the response codes. When the results are input to a database it is these numbers that will be input, not the words describing the categories. If these data on race and ethnicity are to be compared between populations sampled using two or more different language versions of the questionnaire, then a system must be devised by which each different option is allocated a different code and all "same" categories are given the same code. I shall comment further on the subject of coding after presenting more versions of the questionnaire.

The response categories in the second of these items are of even less relevance outside of the United States, Porto Rico, Mexico or Cuba. For example, the word Chicano has no meaning outside of the United States and, confusingly, most of South America's population would possibly self-identify as Latin American, but cannot a Mexican or a Cuban also be Latin American? Furthermore, many respondents would find it strange that a question using the word "nacionalidad" only offers options for three of the very large number of territories in which Spanish is spoken: Cuba, Porto-Rico or Mexico. What about the Argentineans, Uruguayans, Peruvians and Bolivians, to name a small handful? They would all be "others", which is of little use for epidemiology. On the basis that there is no European Spanish version, it would be assumed that Spaniards would be forced to choose "Hispano" (rather than Espanhol) or "Otro". In common with the English version, the "other" option in this item does not invite respondents to specify.

Incidentally, this item appears to have "slipped through the net" of back-translation, since the English questionnaire had "national origin or ancestry", whereas my back-translation (!) of the Spanish would be "nationality or ancestral origin", which actually makes the options available less appropriate still.

I must of course accept that although these translations are unsuitable for a majority of Spanish speakers worldwide, within the United States they would be functional translations. The same cannot, however, be claimed for the Italian version with respect to Italy.

The reasons I surmise that the intended target population for the Italian version included the residents of Italy are as follows. The earnings responses are in Euros, "Zip code" has been translated as "codice di avviamento postale (C.A.P.)", which is the name used by the Italian postal service (Poste Italiane, 2013), and the translation was conducted at the university of Naples (Michelotti et al. 2002). The Italian versions of these items are shown below.

25. Quale dei seguenti gruppi rappresenta meglio la sua razza?

Aleutino, Eschimese, Indiano Americano	1
Asiatico o delle isole del Pacifico	2
Nero	3
Caucasico	4
Altro	5
(specificare _____)	

26. A quale di questi gruppi etnici appartengono le sue origini nazionali o i suoi antenati?

Portoricano	1	Chicano	5
Cubano	2	Altro Latino Americano	6
Messicano	3	Altro Spagnolo	7
Messicano Americano	4	Nessuno di questi	8

(Michelotti et al. 2002: 9)

I shall begin my analysis with item 25. The Aleut are a people who live in the extreme North of the Northern hemisphere and, according to Wikipedia, they are "the indigenous people of the Aleutian Islands of Alaska, United States and Kamchatka Krai, Russia", although of the total global population of less than 18,000, only 700 live in Russia, the remainder are United States citizens (Wikipedia 2012a). Also according to Wikipedia, there were 4,570,317 foreigners living in Italy in 2011, accounting for 7.5% of the population (Wikipedia 2012b).

While I accept that these figures, in combination with geography, do not absolutely exclude the possibility that some of the world's 18,000 Aleut are living in Italy, they do show that even if a large proportion of them were there, it would not be epidemiologically significant with relation to Italy's population of just over sixty million inhabitants.

The full absurdity of including the Aleut as, not an additional, but the first category on an Italian questionnaire becomes clearer if one remembers that this questionnaire is only for people who have or are suspected of having temporomandibular disorders. Of course many

Italians could choose "Caucasico", some would choose "Asiatico" or "Nero" and the remainder would be included in the "Altro" designation, so it could be argued that they are at least included in item 25.

Notwithstanding, the only justification for maintaining the first two categories would be to preserve coding. However, as I shall demonstrate using other, more creative versions, it appears that coding considerations are not paramount, since not all language versions follow the same scoring system for these two items.

Unfortunately, it appears to me that the most likely explanation for the inclusion of the Aleut, the Eskimos and the Pacific islanders is that the forward translator simply translated what was included in the English version, the back-translator translated it back, the reviewer compared the two and found them to coincide, and nobody took responsibility for checking whether the result was actually appropriate to the target setting. Even so, it could still be argued that the psychometric results may not have been any different from other possible solutions, since the options white, black and Asian are offered.

There is a counter-argument based on the effect of context, and particularly on the effect of word sequence, that draws on the finding reported by Perneger et al. with relation to response scale labels and I shall present that argument after discussing additional versions of the questionnaire, but with reference to the Italian version, analysis of item 26 is enough to show that this translation solution of simple transposition would indeed have perceptible effects.

To put it simply, item 26 does not offer any option other than "Nessuno di questi" for the 92.5% of Italy's population who are *not* immigrants. Unless the respondents have ancestors from a Spanish-speaking country (I include the United States in that designation), they are obliged to choose the "none of these" category. Unless there is some connection between speaking Spanish and temporomandibular disorders (and there is not),[21] there is no valid reason for including this selection of items in a questionnaire destined for use in Italy (other than the fact that they were included on the United States version, were translated into Italian, back-translated into English and passed the back-translation test of comparison with the United States version).

---

[21] Data on the National Institute of Dental and Craniofacial Research website shows that temporomandibular disorder symptoms were less common among the Spanish population than in samples from Sweden, England and Hong Kong and similar to rates in samples from Canada and the United States (NIDCR 2013).

In contrast with the Spanish questionnaire which managed to smuggle a few differences past back-translation,[22] there is nothing in the Italian translations of items 25 and 26 that could be identified as a discrepancy by back-translation. They would back-translate in exemplary fashion, which surely means that they are exemplary translations (doesn't it)? Of course I believe exactly the opposite: these are dysfunctional translations and the fact that they passed back-translation does not reflect well on them, rather, it reflects very badly on back-translation.

Interestingly, there is a second Italian translation of the RDC-TMD, published eight years later (Macri et al. 2008). This is not the official version of the RDC consortium, which is the version I have just discussed, and it was not translated using the official RDC methodology described by Ohrbach.

However the article describing the process states that back-translation was used, "to make sure that the translated version reflects the same item content as the original version" (Macri et al. 2008: 166). Unfortunately, this questionnaire does not offer an alternative solution since, despite saying that the RDC-TMD history questionnaire "includes 31 questions covering information such as to [(sic)] demographics and psychosocial assessment", the Italian questionnaire presented in the appendix only has 24 questions, with the race, ancestry and all successive questions excised (Macri et al. 2008: 166; -171-174). With the exception of age and sex, this means that all of the demographic questions are missing from the questionnaire.

The clue to the reason is possibly contained in the title (although the truncation is not even acknowledged in the text, much less explained). The article is entitled "Development of a reliable and clinically useful Italian version of the Axis II of the Research Diagnostic Criteria for Temporomandibular Disorders (RDC-TMD)" and it is the words "clinically useful" that may provide the explanation. It is possible that the authors reasoned that demographic questions were not needed for clinical applications, only for collecting data for research.

However, the possibility that the race questions simply posed too great a challenge to the translation process cannot be ruled out (since the authors make no mention of the omission). One clue that this may be the case is that the appendix also contains the original questionnaire (or at least part of it), but the English version has been truncated at question

---

[22] "Hispano" for "Other Spanish", "Lationoamericano" for "Other Latin American" and "nacionalidad u origen ancestral" for "national origin or ancestry".

24 too (Macri et al. 2008: 175-177), once more without acknowledgment beyond having at one point described it as containing 31 items. This Procrustean approach does of course avoid any problems with back-translation!

I shall now present the same items in three further versions of the RDC-TMD questionnaire, starting with the Brazilian Portuguese version, which, although a great improvement on the Italian solution is still not an exemplary translation. In contrast, the Dutch and Swedish versions which I shall present after the Brazilian Portuguese version, are indeed very good examples of creative translation solutions that take full account of the intended function of the instrument and its intended setting.

The Brazilian Portuguese translation employs a more creative solution than the Spanish or Italian versions, but it still shows signs of having been *anchored* to the source text (by back-translation). The two items in question are as follows:

25. Qual a sua cor ou raça?
1   Aleútas, Esquimó ou Índio Americano
2   Asiático ou Insulano Pacífico
3   Preta
4   Branca
5   Outra [Se sua resposta foi outra, PASSE para as próximas alternativas sobre sua cor ou raça]
6   Parda
7   Amarela
8   Indígena

26. Qual a sua origem ou de seus familiares?
1   Porto Riquenho
2   Cubano
3   Mexicano
4   Mexicano Americano
5   Chicano
6   Outro Latino Americano
7   Outro Espanhol
8   Nenhuma acima [Se sua resposta foi nenhuma acima, PASSE para as próximas alternativas sobre sua origem ou de seus familiares]
9   Índio

10	Português	16	Japonês
11	Francês	17	Alemão
12	Holandês	18	Árabe
13	Espanhol	19	Outra, favor especificar
14	Africano		_____
15	Italiano	20	Não sabe especificar

(Parreira et al. 2009 [2004]: 7)

This solution is superior to that used in the Italian questionnaire, since it includes the categories from the Brazilian national census in the race item and covers the majority of the origins of Brazil's non-indigenous population in the ancestry question, plus "índio" for the indigenous part. However, it has also preserved the original United States categories, which leads to certain problems. The first problem is simply that a significant proportion of Brazil's population would have no basis on which to decide between two options in the response options to the race/colour question.

As Travassos and Williams and also Parra et al. have shown, Brazilians use a combination of terms referring to skin colour (preta, branca, moreno, parda and amarela, for example) and terms referring to geographic or national origins (galego, alemão, japonês and indígena, for example) to describe race/ethnicity (Travassos and Williams 2004; Parra et al. 2003). By "importing" the United States divisions, Parreira et al. have created a dilemma for Brazilians who feel their ancestry is Asian (the majority of whom would be of Japanese descent) or indigenous Brazilian.

The first category response translated from the English includes American Indian, which bearing in mind that Brazil is in South *America*, includes many Brazilians. However, the Portuguese term "Indígena" is also meant to designate American native peoples indigenous to Brazil. Similarly, the term "Asiático" must include Japanese, Chinese and

Koreans, but here in Brazil these races are collected under the term "Amarela".

This may seem a minor problem, as long as the racial categories are not going to be used as variables in statistical analyses, but if they are not, what is the point in including them at all? Assuming that the racial data does have something to offer, it must be considered whether the finding that the sequence in which words are presented can be more important than the actual words chosen (as demonstrated by Perneger et al. and explained on page 128 above) might not have an effect here too? If "Indígena" is the last item on the list, might not some respondents feel that this is a sign of lower status and choose "Índio Americano" instead? Also, would people who normally answer "amarelo" actually bother to read on past "Asiático" if they felt that they fit this category? The same question of which category to choose applies to "Outro Latino Americano" in the second item. Are not all Brazilians also Latin Americans, irrespective of whether they are of Portuguese, Italian, Japanese, (etc.) descent?

These are of course speculations. Without another version of the questionnaire that didn't include the terms used in the United States with which to compare this version, there is no way of testing them.

Notwithstanding, this translation has evidently been designed to make Brazilian race/colour/ancestry designations fit the original coding scheme without requiring recoding of the original questionnaire, in order to simplify data analysis. However, if the coding is indeed so important that it cannot be changed, it should still be possible to avoid possible sequence effects by simply eliminating the imported items but retaining the codes as above, i.e. with numbering starting after the original English numbering. If the numbers themselves were considered to have an effect on responses, then the printed version could use sequences starting from 1 and the instructions to researchers could include a conversion guide to recode the answers for aggregation with international data. Of course such a procedure would have to be coordinated across all language versions, but the RDC-TMD consortium has an international steering committee to take care of exactly this type of issue.

One question that intrigues me and cannot be answered from the published data, is how the additional items got past the back-translation stage and into the final version. They would obviously be identified as discrepancies if they had been back-translated, since the English version has no such items. The fact that the items are added, rather than substituted for the United States' categories, might suggest that they

were included at a later stage, possibly after pilot testing elicited complaints, but the question remains of why, if the additional items were allowed to flout black-translation, the original items could not also be removed.

This question becomes harder to answer after perusal of the Dutch and Swedish versions. Neither has retained the United States' categories or made any attempt to preserve the coding structure. Indeed, the Dutch questionnaire has added an extra item, but numbered the three resulting items as 25a, 25b and 25c, meaning that item 26 is the question that was previously 27 (on educational level) and all subsequent item numbers are altered accordingly, showing that coding parity was not their primary concern. The Swedish version retains the two-item configuration, but item 25 is a simple yes/no response and item 26 requires the respondent to specify, without giving any predefined options.

Neither the results from the Dutch translation nor the results from the Swedish translation could be conveniently input to the same database as the results from the English version, which was true of the Spanish version. The Brazilian version would demand additional categories, but no alteration of existing categories would be required. However, in return for paying a price in coding compatibility, the Dutch and Swedish solutions achieve something that the Spanish, Brazilian Portuguese and, of course, Italian versions failed to achieve: they are relevant to their target populations.

I shall present the Dutch version first.

25.a. In welke van de hierna genoemde landen of regio's bent u geboren?

Nederland	1	Europa (*excl. Nederland*	7
Indonesië	2	Noord-Amerika	8
Turkije	3	Latijns-Amerika (*excl. Suriname, Ned. Antillen, Aruba*)	9
Suriname	4	Afrika (*excl. Marokko*)	10
Marokko	5	Azië (*excl. Turkije, Indonesië*)	11
Ned. Antillen, Aruba	6	Oceanië (*incl. Australië, Nieuw-Zeeland*)	12

25.b. In welke van de hierna genoemde landen of regio's is uw vader geboren?

Nederland	1	Europa (*excl. Nederland*	7
Indonesië	2	Noord-Amerika	8
Turkije	3	Latijns-Amerika (*excl. Suriname, Ned. Antillen, Aruba*)	9
Suriname	4	Afrika (*excl. Marokko*)	10
Marokko	5	Azië (*excl. Turkije, Indonesië*)	11
Ned. Antillen, Aruba	6	Oceanië (*incl. Australië, Nieuw-Zeeland*)	12

25.c. In welke van de hierna genoemde landen of regio's is <u>uw moeder</u> geboren?

Nederland	1	Europa (*excl. Nederland*	7
Indonesië	2	Noord-Amerika	8
Turkije	3	Latijns-Amerika (*excl. Suriname, Ned. Antillen, Aruba*)	9
Suriname	4	Afrika (*excl. Marokko*)	10
Marokko	5	Azië (*excl. Turkije, Indonesië*)	11
Ned. Antillen, Aruba	6	Oceanië (*incl. Australië, Nieuw-Zeeland*)	12

(Lobbezoo et al. 2004: 9)

The three questions in the Dutch version can be roughly translated as follows, 25a: "In which of the following countries or regions were you born?", 25b: "In which of the following countries or regions was your father born?" and 25c: "In which of the following countries or regions was your mother born?".

All of the response choices are the same for all three items. They list, in addition to the Netherlands, six territories from which many people have migrated to the Netherlands and then the six continents, with exceptions in brackets for each of the territories already listed and a hint that Oceania includes Australia and New Zealand.

In other words, these are not race questions but geographic origin questions. They only relate to place of birth, not to nationality, citizenship, skin colour, race or ethnicity. However, they provide an extra level of detail compared to the other questionnaires presented so far, in that they ask about the geographical origins of two generations and allow for parents to be from different places. There are a number of reasons why this is not just a very creative solution that caters for anybody who happens to be in the Netherlands (and speaks Dutch), while concentrating on the people most likely to be there, but is also a sound approach from the perspectives of anthropology (in its modern, non-colonial incarnation), genetics, epidemiology and public health. I shall discuss these reasons after presenting another creative solution which is also based on sound scientific principles.

In the Swedish translation of the RDC-TMD, there are only two items, so the number sequence has been maintained. In common with the Dutch translation, the question is no longer related to race/ethnicity or skin colour, but to geographical origin. The two items are as follows:

25. Är Du eller Dina föräldrar inflyttade till Sverige från annat land?   Ja   1
                                                                                Nej   2

26. Om Ja, från vilket land?   Land _____

(List T et al. 2003: 7)

These two questions can be roughly translated as follows, "25. Have you or your parents moved to Sweden from another country?" and "26: If yes, from which country?". As with the Dutch version, the Swedish version basically divides respondents into people who have lived in the target setting for two generations or people who are immigrants or the children of immigrants. It does not provide the level of familial detail that the Dutch questionnaire offers, since there is no distinction between the three people who may or may not have been born in Sweden (respondent and respondent's mother and father) whereas the Dutch version requests information for each person separately. It does, however, offer a blank space for specification of the country of origin, meaning there are no predefined categories. Whereas the Dutch questionnaire offered seven possible named territories (including the Netherlands, but not, interestingly, Belgium) and offered continents for people who were not from those territories, the Swedish solution allows for complete self-selection of place of origin. Once more, there is no possibility of coding compatibility with the United States version because, even if every respondent were consigned to the "other" category, there is no way of knowing whether those who provided a country other than Sweden were themselves immigrants or the children of immigrants. Since there is no way of telling whether both parents were immigrants either, it is also incompatible with the Dutch solution.

I shall now take a little time to explain why this is not a bad thing and is in fact the most sensible decision to take (disregarding coding compatibility for race) on the basis of current consensus on the utility and desirability of measuring race in healthcare surveys.

The starting point for this discussion is the consensus view on race as a biological entity, in other words on the question of whether races actually exist from a biological, medical and genetic perspective. This consensus is possibly surprising, since it is as follows.

> There is wide agreement among anthropologists and human geneticists that, from a biological standpoint, human races do not exist
>
> (Parra et al. 2003: 177)

This position is based on findings from genetic studies that have shown that "most of human genetic diversity exists as differences between individuals within populations" (Travassos & Williams 2004: 662), which in turn means that races are not statistically informative variables to use when attempting to determine genetically-mediated health risks.

Genetic findings have also led to the rejection of the so-called candelabra model of human races. What this means is that races cannot be considered to have descended in isolation from a single common ancestor, separated by distinct ruptures at certain points in the past (resulting in a "family tree" that resembles a candelabra). In fact, genetic research has shown that there has always been continuous intermixing between human populations, resulting in the "trellis" model of human evolution, illustrating "recurrent genetic interchange among Old World human populations in such a way that there was no separation into evolutionary lineages and as a result there is no such thing as human subspecies or races." (Travassos & Williams 2004: 662).

So if race cannot be used as a marker of genetic susceptibility to disease, as these findings show, why do health surveys still contain race questions? In many cases the answer is simply that the "magic bundle of rituals" (Campbell 1968: 255, quoted in Brislin 1986: 154) has always included a race question, but there are also sound medical reasons for including such questions in certain types of survey.

While race does not exist from a genetic or biological perspective, races undeniably do exist as social constructs (Parra et al. 2003: 177). This means that health differences can be detected when data are divided along racial lines. For example, in the United States, large variations can be detected between Cubans, Mexicans, and Puerto Ricans (Travassos & Williams 2003: 666), which is why they are differentiated on a questionnaire to be used in the United States.

There are three major reasons for this. The first is that socioeconomic status also often differs along racial lines and low socioeconomic status is independently linked with risk of many diseases. The second is that discrimination exists within healthcare, as within all facets of society, and different racial groups can be discriminated against in terms of access to or provision of healthcare. The third is that different ethnic groups have different lifestyles and lifestyle is linked with many health problems.

Notwithstanding, these are all reasons for recording race or ethnicity when collecting data that may be used for taking public health decisions or to influence political decision making. When it comes to diagnosis, however, this data is of less use, because "... racial disparities in health and health care cannot be assumed as the existence of a causal relationship between race and health. Race as a risk marker is not synonymous with race as a risk factor" (Travassos & Williams 2003: 675).

There is, therefore, a sound basis for recording race or ethnicity in certain circumstances. However, when it is considered that social factors such as racial classification, racial discrimination and lifestyle are culture-dependent, the utility of such information in international comparisons is called into question.

Analysis of the different concepts of race in the United States and Brazil, for example, has shown that "definition of race is not consistent across societies" (Travassos & Williams 2003: 675). Travassos and Williams traced the history of racial classification in the United States and Brazil and found that in the United States, the decision of whether to classify a person as white or black is still based on the "One Drop of non-White Blood" rule, still explicitly stated in census instructions as recently as 1930, as follows, "a person of mixed White and Negro blood should be returned a Negro, no matter how small the percentage of Negro blood" (Travassos & Williams 2003: 664). In contrast, the system in Brazil, both popular and official, is based on skin colour and "Brazilians, when inquired about their color/race in an open-ended question, may answer with 135 to 500 different race-color terms" (Travassos & Williams 2003: 664), although the national census only offers the options used by Perreira et al. (branca, preta, parda, amarela or índio).

Furthermore, Brazilians do not see ethnicity in the same was as people in the United States, despite also being a country populated with the descendents of migrants. For example, when "exposed to an ethnicity question with 12 categories of origin, 86% of the respondents identified themselves as Brazilians" (Travassos & Williams 2003: 664), suggesting that the additional categories for origin chosen by Perreira et al. (German, Italian, Portuguese, etc.) were included in order to complement the United States ethnicity item, rather than because Brazilian's use such categories.

As a result of these factors, not even the apparently equivalent categories of white/branca and black/preta are actually comparable between Brazil and the United States. As Travassos & Williams point out,

> Some people that refer to themselves as 'Whites' in the US are not similar (in regard to ancestry or skin color) to 'Whites' in Brazil. For example, because of the 'One-Drop' rule, individuals with 'White' skin color but African ancestry are likely to identify themselves as 'Black' in the US, but might regard themselves as 'White' in Brazil.

> (Travassos & Williams 2003: 671)

This is not merely a problem when comparing Brazilian data with data from the United States. It also applies to any attempt to compare racial categories from one country to another, to greater or lesser degrees. Furthermore, the discrepancies are not limited to countries that speak different languages, since the root of the differences is cultural, not linguistic. For example,

> 'Whites' in the US are not comparable to 'Whites' in the UK. The term 'White' in the UK never considered Asian Indians, Middle Eastern, and North Africans, but until recently people from India were considered 'Whites' by the official US racial classification, and Middle Easterners still are.

> (Travassos & Williams 2003: 671)

Finally, to complicate the issue further, "the proportion of the Brazilian population in each color/racial category also depends on how such categories are presented and interpreted" (Travassos & Williams 2003: 671). For example, "the proportion of 'Whites' and 'Blacks' diminishes significantly when 'Moreno' is used instead of 'Pardo'" (Travassos & Williams 2003: 672), which adds further weight to my argument that retention of the United States categories will affect reception of the inserted Brazilian ones.

I believe that in view of all of the above, I have shown that the solutions chosen for the Dutch and Swedish questionnaires were not simply creative ways of dealing with the problem posed by incompatible categories, they are also soundly grounded in consideration of the extent to which racial data can be useful, which is almost exclusively restricted to contexts within cultures. In other words, they collect data that is of relevance nationally and do not attempt to achieve comparability with international data, which would be futile.

Furthermore, when specialist data, such as the RDC-TMD data, is analysed it is often compared with census or survey data to provide a denominator for calculations. If the categories used are not the same, then the results will be skewed. This means that, unfortunately, despite all the care taken with the Brazilian version to maintain the original coding, in addition to the fact that the categories are not comparable with the United States data, they are also no longer comparable with national Brazilian census data which only offers five race/colour options and has no data on "national origin".

At this point it must be acknowledged that for the Dutch and Swedish versions to have arrived at sensible solutions, there must have been a point at which back-translation was either waived for these items

(possibly because they were changed after an original version had transposed the categories, passed back-translation and then been altered subsequently) or someone took the decision to ignore the back-translation results. However the decision was taken it was the correct one.

It may be considered harsh to blame back-translation for the absurd Italian version and the odd Brazilian version of the racial items in the RDC-TMD since it cannot be known that the translators had offered more creative, functional, solutions which were eliminated in response to back-translation. This may even have been the case, but it is not necessary for this to have happened for back-translation to be responsible nonetheless. My argument does not rely on such a crude mechanism, although this does often happen.

The simple fact that forward translators know that their work will be subjected to back-translation is in itself enough to achieve this undesirable result. There are two possible mechanisms of action. The first is that knowledge of the impending back-translation makes translators adhere more closely to the source text, in the belief that any deviation will be considered an error after back-translation and do not propose any such creative solution in the first place, merely transposing the inappropriate categories. The second is that, knowing they are involved in an 11-stage translation process, including the "gold standard" back-translation, the translators abdicate responsibility for the final text, feeling, possibly with justification, that the stages that follow are responsible for "cultural adaptation" (which would surely include adapting race categories to the races of the people who actually live in the target culture!) whereas their task is simply to translate what they are given.

By appropriating responsibility for the finished product, cross-cultural adaptation guidelines simultaneously absolve translators from what under normal circumstances they would consider their responsibility – ensuring that the finished product is functional in its intended application. The result of relieving translators of their professional responsibility, and with it their status as responsible experts (Nord 1997a: 30), is situations like the ones I have presented in this subsection, where a questionnaire in Italian has a category for an Alaskan population numbering 18,000 worldwide, but does not have a category for Italians, despite (or rather, because of) having passed through 11 stages of a cross-cultural adaptation process designed to "produce valid data in another setting" and including back-translation (Ohrbach 2009: 2).

An obvious question at this point is whether researchers requesting translations could reasonably expect their translators to pay attention to details such as the appropriateness of racial categories, even if freed from the yoke of back-translation. I can only answer anecdotally on the basis of my own professional experience, but, with that proviso, the answer is an unequivocal yes.

I have researched racial categories on a number of occasions in order to formulate possible solutions for my clients. I read the articles by Travassos and Williams, Pena et al. and Parra et al. for the first time in a professional capacity while trying to decide how to translate "pardo" and "amarelo" for a client and I have referred to them frequently since for other clients. The appropriate solution, unsurprisingly, depends on the function of each translation.

If I am translating a scientific article reporting on research conducted exclusively in Brazil and in Portuguese, which is the normal state of affairs for me, then I will employ explicitation or even footnotes to explain the different categories and very often use the Portuguese words for terms such as "moreno", "pardo" and "amarelo" thereafter.

In contrast, if the translation is of an instrument, for example, and is part of an international collaboration, then I will raise issues such as those discussed by Pena et al. Parra et al. and Travassos and Williams, with relation to non-comparability of racial categories, and ask questions such as whether these categories will be used for international comparisons or are only for producing national statistics.

The important point is not the actual solution adopted, as this will always be tailored to the need, but that clients can indeed trust me to (i) detect the existence of a possible problem, (ii) proactively investigate whether the problem is relevant to the task in hand, and (iii) present them with a series of possible solutions or a request for further information, followed by a series of possible solutions.

Furthermore, they can also rely on me to accept their decision on which solution to adopt. This is not the same thing as abdicating responsibility, however. Even if a client were to ultimately ignore my advice and choose the same option that would pass back-translation, I would not have surrendered my professional responsibility to a process, rather I would have advised my client of what my experience and the results of my research tell me is the best course of action and would therefore have discharged my professional duties by ensuring that the decision had not been taken in ignorance.

Of course, if initiators insist on contracting "naive" translators for highly specialised projects, then it is clearly to be expected that, not only

will they lack the research skills necessary to find solutions for these types of problems, but they will even lack the specialised background knowledge necessary to detect the fact that there is a problem to be solved in the first place.

I shall return to the subject of how back-translation undermines translators' status as experts towards the end of this chapter (section 4.4), but in the next section I wish to explain why I believe that, particularly with relation to issues of meaning and equivalence, the back-translation method is grounded in a simplistic view of language, which in turn makes it an inherently "pessimistic" approach to translation.

4.2 Meaning and equivalence: Back-translation is founded on a simplistic concept of language

In this section I will show that the basic assumptions underlying back-translation form a simplistic, even naive, view of language and translation, that this in turn leads to a "pessimistic" attitude to translation and that flaws inherent in the back-translation process make it a self-fulfilling prophecy, in that it provokes and introduces many of the "errors" that it detects. The section is divided into three subsections.

In the first subsection I shall demonstrate the simplistic conception of language on which Brislin based back-translation. The main focus is on Brislin, since the back-translation technique itself has not changed since he wrote, but I shall also show that the contemporary incarnations are based on the same fundamental assumptions. I will then show how the simplistic concept of meaning as unchanging, reproducible and measurable, in combination with the naive assumption that translation is bi-directional, reversible and, as Venuti puts it, "transparent", leads to unrealistic expectations of what can be achieved with translation.

In the second subsection I shall demonstrate how the impossibility of meeting the overly optimistic expectations of symmetry and photocopier-like reproduction of meaning results in a fundamentally "pessimistic" view of translators and translations: it is believed that a perfect copy is possible, but it is assumed that translators will fail to achieve that goal and so need to be controlled at every step.

In the final subsection I will argue that, in combination, the two contradictory assumptions on which back-translation is founded mean that using back-translation enacts a type of self-fulfilling prophecy in which it provokes the very errors it detects. The simplistic view of

language is manifest in unattainable objectives such as those set by Beaton et al., for example. Semantic equivalence must often be sacrificed in order to achieve idiomatic equivalence, experiential equivalence or conceptual equivalence, but Beaton et al. demand all four types. Mistrust of translations and translators is manifest in statistical analyses that assume a certain proportion of errors and in blinding and other control mechanisms. The result is that researchers employ back-translation expecting to detect errors and back-translation duly detects those errors, seemingly proving its own worth. However, I maintain that it is actually back-translation and the prescriptions surrounding it that are responsible for the errors in the first place.

4.2.1 Back-translation is based on a simplistic view of language and of translation

In this subsection I shall present and support the argument that back-translation is based on a fundamentally simplistic view of language and of translation which imposes unrealistic and unattainable objectives.

My starting point is Brislin's theoretical justification for his criterion 1 meaning error test. This was his statement that both Nida and Catford considered meaning to be "the most important aspect of translation" (Brislin 1970: 191). On the basis of this statement, which is debateable, as I shall show shortly, Brislin hypothesized that "the unit of translation quality may be a unit of meaning" (Brislin 1970: 191). By this he meant "a unit of translation quality that might be the unit of [a] translation quality scale." After analyzing the results of his experiments rating the students' back-translations against English source texts, he boldly concluded that the "Criterion One meaning error is here suggested as a unit of translation quality." (Brislin 1970: 213).

This conclusion rests on a number of assumptions, including, but not limited to, the following (i) in order for the meaning error to truly be a unit on a scale of measurement, it must be reproducible; (ii) for that scale to be a scale of translation quality, translation quality must be defined exclusively in terms of meaning, and within meaning, in terms of error versus no error; and (iii) if the number of errors in a back-translation is to be considered a measurement of the translation quality of an unmeasured forward translation, there must be a relationship (ideally one-to-one) between meaning errors in the back-translation and meaning-errors in the forward translation. Unfortunately for the theory that a meaning error is a unit of translation quality, none of these assumptions are true.

The assumption of reproducibility had already been ruled out by Nida in 1964. Brislin's idea of measuring translation quality in terms of errors requires reproducibility, but Nida unequivocally stated that "no word ever has precisely the same meaning twice" (Nida 1964: 48), that "each source and each receptor differ from all others", and that "no two persons ever mean exactly the same thing by the use of the same language symbols" (Nida 1964: 52). I have demonstrated the result of this subjectivity in the subsection on proof, showing that Brislin's raters only exhibited a general tendency to agreement on the errors in back-translations and those described by Sperber et al. often chose opposite ends of the scale.

This is because meaning is not, as back-translation assumes, a fixed entity that is always the same for any given combination of words. As Venuti puts it,

> Meaning is a plural and contingent relation, not an unchanging unified essence, and therefore a translation cannot be judged according to mathematics-based concepts of semantic equivalence or one-to-one correspondence.
>
> (Venuti 2004 [1995]: 17-18)

In contrast with this nuanced concept of meaning, Brislin's statistical treatment of his raters' results, calculating correlation coefficients for agreement on the basis of total number of errors found, without correlating individual errors, shows that he considered one error to be as good (or bad) as another. In other words, as though they really were units – i.e. of equal value.

In addition to the lack of consistency in measurement of the meaning error unit, we need look no further than Nida, who Brislin claimed had considered meaning to be paramount, to find that translation quality is a great deal more complex than meaning alone and is not amenable to quantitative approaches,

> It would be convenient if we could construct some formula which would assign numerical values to these different factors and provide some more or less mechanical means by which we might rate translations. This, however, would be impossible for the diverse factors are too complex, too multidimensional, and to a large extent incommensurate
>
> (Nida 1964: 191)

What Brislin hoped could be a unit on a scale of translation quality is in fact nothing more than a count of the number of times a rater identified what they considered to be an error. Given what I have demonstrated

with relation to Brislin's translators, it is fair to assume that the errors his raters identified were language errors.

Irrespective of whether the meaning errors identified in the back-translation can be considered to constitute a measurement of translation quality, there are two further assumptions implicit in the use of these meaning errors as a measure of translation quality of the forward translation. The first assumption is that translation provides a "transparent" reproduction of the source text. The second is that translation is reversible. The first assumption is necessary for the back-translation to be used to judge the forward translation. The second assumption is that a perfect back-translation, i.e. one with no meaning errors, would be a reproduction of the source text used to produce the forward translation, in other words that the back-translation had "successfully" reversed all of the changes made by the forward translation. At the root of these assumptions is the erroneous belief identified by Hönig, that "source and target texts are held together by a relation of symmetry" (Hönig 1997: 16).

To a certain extent Catford's translation shifts relied on an ideal rank-bound translation against which to define the changes identified in the translated text. However, there are some very important differences between Brislin's assumption that translation quality is an absence of errors (and that this leads to reversibility) and Catford's positing of an unshifted translation. Catford described the unshifted translation as a bad translation and he did not classify differences from it as errors, but as shifted translation equivalents. His use of an idealised translation with one-to-one correspondence is as part of a theoretical model, designed to describe translation. In back-translation processes the ideal of one-to-one equivalence is the unattainable goal set for every translation.

Also related to this simplistic view of translation is an implicit assumption that there is only one translation equivalent for any given source text. To a certain extent Catford's work appears to support such an assumption and it may be that this led Brislin to do the same. Catford stated that "discovery of translation equivalents is based on the authority of a bilingual informant or translator" (Catford 1965: 27), apparently making no distinction between their relative merits. It seems that Catford considered any bilingual to be a suitable informant, irrespective of translation experience. Brislin certainly made no distinction.

The assumption that meaning is not dependent on the receiver, but a property of the text, is at the root of many of the misconceptions that are built into back-translation. In addition to assuming that there is a single correct interpretation of the source text and that this can be

conveyed unchanged in the forward translation, back-translation processes also assume that the back-translator will interpret the forward translation in the same way that the intended target population would interpret it.

However, as Nord points out, "different receivers (or even the same receiver at different times) find different meanings in the same linguistic material offered by the text", and "a text has as many meanings as there are receivers" (Nord 1997a, p. 31).

The forward-translator is invariably not the intended addressee of the original questionnaire, the back-translator may or may not be one of the intended addressees of the forward translation (depending on whether or not the forward translator has been informed that their work will be back-translated), but because the back-translator and forward translator are both bilingual, even if they are one of the addressees they are not comparable with the final target public of the questionnaire (since if the target population were bilingual, the questionnaire would not need translating in the first place). This means that each and every relationship to the translation, from the links between source and target, through the way meaning is extrapolated from text, to the element of response, will all be different for the translator than for their monolingual clients and final addressees. Furthermore, as I have pointed out before, context is extremely important. The final questionnaire will be administered to people who have an interest in making sense out of it and providing the information requested. In contrast, when source text and back-translation are compared, the person rating them has been specifically requested to identify problems.

In addition to the assumptions underlying the back-translation process itself, certain other misconceptions about the nature of language can be identified in Brislin's work and also in more recent literature.

When Brislin introduce his test of criterion 4 (the response criterion), he justified it on the basis that Nida had listed "producing a similar response" (Nida 1964: 164) as one of the requirements a translation should meet. Brislin maintained that a verifiable response could be used to prove that a translation was equivalent to its source text. However, Nida was referring to an emotional or spiritual response to passages in scripture and had very flexible criteria for the necessary level of similarity. In contrast, Brislin reduced the concept to a simplistic requirement for an identical physical response. In addition to the fact that this limits the technique to texts demanding some kind of physical response, it is also logically unsound to deduce equivalence of meaning from an equivalent response.

There is a perfect example of why the deduction is unsound in the text book in which Brislin's 1973 chapter was published and of which he is a co-author. As quoted in full on page 77 above, Brislin et al. gave an example of two mothers from different cultures who responded in the same way to a question about misbehaviour, but had arrived at their responses by interpreting the question in very different ways and provided identical answers, by which they meant very different punishments. In other words, an observed response alone is not sufficient data from which to deduce the cause of that response.

In addition to the unfounded assumption that an identical response indicates identical interpretation, the response criterion also assumes that the text alone is responsible for the response, ignoring contextual elements that may also contribute.

Brislin requested Palauan speakers to make pictures with green and blue pieces of paper after following translated instructions because "Palauans use the same word for blue and green" and he thought it would be hard for them to distinguish between the two. In fact the task was too easy for the majority of subjects. This expectation that the colours would cause problems is evidence of a naive concept of language. The assumption is that if, in a given language words do not exist that are parallel to the ones "we" use to describe a certain phenomenon, then that language is unable to describe it. However, as Nida pointed out, in the same book that Brislin cited in support of his response criterion,

> One of the principal reasons why some persons have supposed that some languages (never their own, of course) could not be used to speak about certain aspects of experience is because they have not understood adequately the diverse ways that different languages segment experience.
>
> (Nida 1964: 50)

In other words, although they may not have words that coincide with the colours "green" and "blue" in English and seem to use the same word at both times, they will inevitably have some means of distinguishing between the two shades we call blue and green – they might use analogues of "sea blue" and "leaf blue" or "sky green" and "plant green", for example, but their language will have a mechanism for distinguishing, as the results of the task clearly showed.

By 1986, Brislin had abandoned the response criterion and recognised that "people in various language communities do not categorize specific items in the same manner ... Rather, the items are

grouped in different ways across cultures" (Brislin 1986: 147), but even then he did not appear to see that this is one of many reasons why translation is irreversible.

In a similar vein to assuming that Palauans would have difficulties with blue and green pieces of paper, Brislin also stated that members of some cultures would not criticise when asked to appraise translations and that this was a cultural characteristic. This is also a naive attitude. All cultures must have systems for appraisal and evaluation and socially-acceptable ways of indicating when there is room for improvement. However, what may appear to a member of one culture as two slightly different levels of praise, may in another culture be obvious expressions of dissatisfaction, couched in a manner that is not offensive in that culture.

In addition to the misconception that some languages are not equipped to deal with certain concepts, Brislin's work also betrays signs of the baseless division of language into literal or figurative that has been the target of great criticism by the deconstructionists.

On several occasions Brislin recommends avoiding idiomatic expressions and metaphorical language. A common example is the phrase "feeling blue", which has echoed down the decades in a whole variety of questionnaires, appearing in the SF-36, for example, as "downhearted and blue" (Perneger et al. 1999: 1039). Brislin suggested that "feeling blue" might be changed for "depressed" (in the source text, prior to translating), apparently considering "depressed" as more "objective", "literal" or "scientific" than "feeling blue" (Brislin 1986: 153). However, "depressed" is itself a figurative term, since it has been borrowed from its physical usage for application to an emotional state, which has no physical properties. "Feeling blue" could actually be considered less metaphorical than "depressed" since it includes the verb "feeling", which is associated with subjectivity. Of course there is no such thing as a "literal" meaning that is not in some way figurative, since words are not simply labels for objects.

Later incarnations of back-translation have corrected the most obvious misconception in Brislins's work. Sperber asked his raters to judge comparability of language and similarity of interpretability and the AAOS and IQOLA processes use back-translation to identify errors but do not attempt to treat the number of errors as variables in statistical analyses. However, the underlying misconceptions about the nature of language and translation are still very much present in contemporary incarnations of back-translation.

As Hermans points out, the "more closely one looks at what constitutes 'equivalence' in translation the more problematic the notion becomes", since "strict application of the concept as ... in mathematics, is obviously unworkable" because this "would imply reversibility and interchangeability, and we all know that translation is a one-directional event involving asymmetric linguistic and cultural worlds." (Hermans 1997: 48). Notwithstanding the near wholesale rejection of equivalence within translations studies in the years since Nida, contemporary cross-cultural adaptation processes are still founded on the principle of equivalence and depend on a belief in reversibility.

The contemporary cross-cultural adaptation processes described in Chapter 2 all aim for equivalence, but the ways in which this is defined and the methods used to achieve it differ. Sperber et al., for example, considered that similarity of language form was a valid criteria for judging back-translations against the source texts, which demonstrates a very naive view of both language and translation.

The AAOS method defines four types of equivalence, semantic, idiomatic, experiential and conceptual, but does not acknowledge that achieving all four types is very often impossible. For example, semantic equivalence is explained with the phrase "Do the words mean the same thing?", but the example for experiential equivalence is an item worded "Do you have difficulty eating with a fork?" which the guidelines say would cause difficulties if "that was not the utensil used for eating in the target country", with the result that the "questionnaire item would have to be replaced by a similar item that is in fact experienced in the target culture." (Beaton et al. 2000: 3188). Whatever item is chosen, semantic equivalence must be sacrificed.

The IQOLA method provides very little detail on how the back-translations are evaluated, but in terms of its demand for equivalence, it is the least naive of the methods. The forward translations are evaluated for "clarity of the translation", for "common language use" and for "conceptual equivalence" (Bullinger et al. 1998: 914). At face value, conceptual equivalence may seem an unrealistic goal, if conceived of in the way that Beaton et al. describe it:

> Often words hold different conceptual meaning between cultures (for instance the meaning of 'seeing your family as much as you would like' would differ between cultures with different concepts of what defines 'family'—nuclear versus extended family)

> (Beaton et al. 2000: 3188)

If conceptual equivalence is achieved as defined in these terms, by, for example, specifying the members of "family" to restrict it to a nuclear rather than an extended family, then the target instrument risks no longer being appropriate for the target setting because the source culture concept would have been imposed.

This is indeed a naive approach, but in the IQOLA method, "conceptual equivalence" does not relate to conceptualization by the users of the instrument, but to the concepts that its creators wish to measure. In this case the concept of "family" would have to be defined in terms of the data that the instrument is designed to collect.

In order to define how to translate "family" it would first therefore be necessary to determine what elements of "family" are of interest. It could be the fact that family members live in a shared household, for epidemiological and public health purposes, for example, or the element of interest might be social interaction, possibly to measure perceived quality of life.

In the IQOLA method, conceptual equivalence is discussed between the translators and the national principal investigators, who are medical researchers who speak the target language, meaning that the translators can ask for explanation of concepts. Since conceptual equivalence is framed in these terms, rather than in terms of language, and since it is evaluated bilingually in the forward translations, rather than monolingually in back-translations, the IQOLA approach to evaluation cannot be accused of the same degree of linguistic naivety seen in Sperber et al. and Beaton et al. (with the proviso that the evaluation of back-translations is not described, so the degree of linguistic sophistication involved cannot be judged).

There is, however, still a tendency to see words in the same way as numbers in the IQOLA method. The Thurstone scaling exercise is a clear example of the misconception that words can be assigned precise values that are unchanged by context. To recap, the scaling exercise was performed in order to select appropriate item response options that would replicate the "ordinal and interval properties" of the original. For example, the response options for question 20 on the SF-36 are "Not at all", "Slightly", "Moderately", "Quite a bit" and "Extremely" (Rand Health 2010). The ordinal properties are the sequence of the words and the rank of each one, while the interval properties are the degrees of difference between each word. The Thurstone scaling exercise is supposed to identify the target language response options that best replicate the sequence and spacing of the source text options.

The IQOLA method identifies candidate response options by asking forward translators to provide all possible translations and then the research team chooses what it considers to be the most extreme option at each end of the scale, "excellent and poor" are the examples given (Bullinger et al. 1998: 914). An unspecified number of native speakers are then asked to position all the other response options in order along a scale between these two end points.

Unfortunately this method ignores the fact that words change their meaning depending on context. By providing many more options than the normal five, the scaling exercise inevitably presents the subjects with several near synonyms, whereas a response scale is intended to have evenly spaced intervals. The natural response to being presented with a series of near synonyms, such as those below, for example, is to identify points of differentiation and this is of course what the subjects are asked to do.

slightly,	not much,	very little,	a bit
moderately,	somewhat,	sometimes,	a fair amount
quite a bit,	significantly,	a great deal,	quite a lot
extremely,	very much,	intensely,	to a huge extent

This means that each option in each group above would be interpreted differently (on a scale with smaller gradations) than if one option from each group had been presented together, as shown below:

slightly,	moderately,	quite a bit,	extremely
not much,	somewhat,	significantly,	very much
very little,	sometimes,	a great deal,	intensely
a bit,	a fair amount,	quite a lot,	to a huge extent

In the first case there is very little difference between each option in each group, but if invited to locate each along a scale, subjects would be obliged to make fine distinctions. In the second case, the relative position of each option is not in doubt. If, however, a subject were presented with all sixteen items, they would not necessarily make the same distinctions as when presented with any given group of four.

The underlying basis for such a scaling exercise is a conviction that words, at least words such as these that qualify other words, retain their meaning irrespective of context. However, there is clear proof that

this is not the case, specifically with relation to response items in questionnaires, one of which was translated using the IQOLA system, including the Thurstone scaling exercise.

This proof is, once again, to be found in the article by Perneger et al. They compared a French translation of the SF-36 that had been prepared without back-translation and without the scaling exercise with the official translation by the full IQOLA method. Their experiment is unique because it administered the two questionnaires to exactly the same people with a one-year interval.

Perneger et al. analyzed the effect that the response options had on the results and found that, despite the fact that the last two options (out of five) were "médiocre" and "mauvaise" in the IQOLA version and "passable" and "médiocre" in the Geneva version, meaning that option 4 in one questionnaire was option 5 in the other, this made no difference to the results, "... as if the rank of the response option, not the attached label, determined the response that was chosen." (Perneger et al. 1999: 1039).

This example brings me to the MAPT translation described by Ozolins. In contrast with the literature on the Brislin Sperber, IQOLA and AAOS methods, Ozolins does not provide a blueprint for a method, but describes a project that adopted the EORTC method (although as I explained on page 119 above, the method they eventually used was significantly different). This project is of interest here because it shows that, whereas Brislin in particular pointed to a number of caveats relating to back-translation, most of which have been repeated by his successors, the second-order users of back-translation tend to forget these warnings and exhibit an even more simplistic view of language, assuming that every difference between a back-translation and the source text is evidence of a translation error.

At several points in the article Ozolins stresses the importance of the response option scales. When he introduces the MAPT questionnaire that was translated, he stated that "Trials had shown that patients' responses to this questionnaire matched very closely actual clinical observations, making it a powerful diagnostic tool" (Ozolins 2009: 2) and that "The validity of the instruments rested on the fine gradations of patient response, and it was these gradations that were the essential items to maintain in any translation." (Ozolins 2009: 3).

While the fine gradations of *patient response* may well be of importance to many questionnaires, the evidence reported by Perneger et al. proves that it is extremely naive to assume that only one form of words can achieve that. What is important is the patients, not the words.

In addition to finding that rank was more important than which word was used for which option, Perneger et al. also found that for the entire questionnaire, the fact that 34 out of 36 items were different between the two translations had no statistically significant effect on the results, showing that the fact that the questionnaire is administered at all is more important than the form of words used to ask the questions on it.

Ozolins says that the scales "consumed much time", in particular the "critical issue of whether there was an equivalent distance between the options to the distance between the options in English" (Ozolins 2009: 5). He provides few actual examples, but one is very telling, The MAPT team complained about a back-translation as follows, "The word *rather* is too vague. Does it mean slightly or moderately? It is too subjective. The original is *moderately difficult*."(Ozolins 2009: 5), but "slightly" or "moderately" are no less subjective, since, depending on context they too will change meaning.

In addition to the conviction that the response grading is somehow sacrosanct, there is also evidence that the MAPT team also fostered the illusion of univocity in popular language. They appear to believe that there is only one way to interpret any given phrase and that, with relation to their questionnaire, that way coincides with the way they understand it.

The team complained about a back-translation in which "some of the time" had been replaced by "sometimes" in back-translation, saying, "Sometimes suggests that a person might receive help occasionally but not on a regular basis" and that "Some of the time implies that a person has regular help" (Ozolins 2009: 8). This resulted in the translator questioning whether an English reader would understand what the MAPT team thought their phase meant.

Interestingly, it appears that this translator may have been correct about the ambiguity of the MAPT since at least one research project failed to confirm the trials that had shown that its results matched clinical observations. This study had investigated

> ... the relationship between the MAPT scores and the clinician's assessment of severity of [osteoarthritis], with respect to the radiographic assessment and the surgical waiting list category assigned to each patient.
>
> <div align="right">(David et al. 2011: 543)</div>

In other words, it tested whether the MAPT really did help prioritise patients and predict the final decision on whether to operate or not, as it was designed to do.

However, in contradiction of the trial findings, this study found "no significant correlation between the MAPT scores and the radiographic severity of [osteoarthritis] of the hip or knee" and so the study concluded that "the assigned surgical waitlist category does not correlate to the patient's MAPT score." (David et al. 2011: 546).

Notwithstanding, when the translator offered "to make the implicit understanding of regularity explicit", they refused because of unwillingness to "diverge from the English original." (Ozolins 2009: 8).

This is an illogical attitude that is directly traceable to the back-translation method. The researchers thought they knew what they meant, but all they could really know is what they had intended to mean. When the translator pointed out that what they thought they meant would be better expressed explicitly, they refused, because that would mean that the translation was "different" from the original and the original had worked in trials. This is tantamount to clinging to a "magic bundle of rituals" (Campbell 1968: 255, quoted in Brislin 1986: 154).

In view of what I have presented in this subsection, I believe it is fair to say that José Lambert's statement on the users of business translation services in Belgium is also applicable to many users of back-translation: "translators and language experts are nowadays being used by managers and customers who view language very naively" (Lambert 1996: 285).

Whereas the MAPT team's approach is based on a belief that the exact form of their questionnaire was responsible for its success in predicting clinical diagnosis, the incontrovertible results published by Perneger et al. demonstrate that a given form of words is not the only way to accomplish a given function.

The questionnaires that I am dealing with are all designed for administration to members of the public and, in contrast with scientific and technical terminology which is intended to be univocal and unambiguous and is often controlled through standards organisations and consensus documents, popular language is highly varied and has many different ways of saying the same thing.

In the next subsection I intend to show that the unrealistic expectation of symmetry described by Hönig and inherent to the simplistic view of language and translation on which back-translation is founded, coupled with the use of inexperienced translators, leads to pessimistic expectations of the abilities of translators. In the final subsection of this section I shall then show how back-translation methods tend to perpetuate this situation.

4.2.2   *Back-translation is inherently pessimistic*

In this subsection I will show that the result of the simplistic view of language is a "pessimistic" view of the abilities of translators and of what can be achieved in translation. This apparently paradoxical relationship should be less of a surprise if the characteristics of the people Brislin actually used as translators is considered as part of the equation.

The back-translation paradigm is based on the conviction that equivalence is achievable through translation, but Brislin's translators were unable to achieve it. Since his translators had such obviously deficient language competence, then it is assumed that a greater degree of vigilance to their "errors" will achieve the sought-after equivalence.

In fact, equivalence is only ever partially achievable and the best way of achieving satisfactory equivalence is to define the types of equivalence relationships that are important in each given situation and, through a process involving communication between translator and client/initiator/user of the translation, arrive at solutions that are appropriate for each situation and use.

Back-translation has taken the opposite route. Rather than tapping the intercultural and language experience of translators, back-translation is founded on an assumption that translators will make errors, that translation is an inherently difficult process and that ever-more rigorous controls are needed to avoid these errors. In the next subsection I shall show how these controls are themselves self-perpetuating since they are the cause of the great majority of the very errors that back-translation was conceived to avoid, but here I wish to present some examples that illustrate the pessimistic attitude to translation that underpins back-translation.

It will be remembered that the IQOLA method involved quality rating of translations. While these ratings were not based on back-translation, but on bilingual comparison, which is positive, the ratings were judged against cut-off scores, calculated at the 75th percentile for difficulty and the 25th percentile for quality, above and below which translations were "determined to be problematic" (Bullinger et al. 1998: 921). This means that, irrespective of how low they scored for difficulty, 25% of all items would be classed as problematic, since the 75th percentile is not 75% of the total possible score of 100, but the point on the range of all scores at which 75% of results are below the cut-off and 25% of results are above it. In other words, if there had been 75 ratings of 1 and 25 ratings of 2, out of 100 (where lower is easier), the 25 higher

items scored at 2/100 for difficulty would still be classed as "problematic". The other side of the same coin is that if 75 items received a quality rating of 100 then even if the other 25 received ratings of 99, they would still be defined as "problematic" under this system. Setting a fixed cut-off in this way is an indication of an initial assumptions of translation failure.

This assumption can be traced back to the 1973 book chapter by Brislin et al., who stated "The writers consider the translation process as an example of an <u>adverse</u>, <u>difficult</u> state of affairs" (Brislin et al. 1973: 35, underlining added for emphasis). I believe that this statement, when considered in conjunction with the characteristics of the translators Brislin used in his experiments, reveals that his underlying attitude to translation was pessimistic and back-translation has perpetuated his attitude.

All four of Brislin's publications contain a version of his rules for writing "translatable" English, increasing from five rules in 1970, through ten rules in 1973, to 12 rules in 1980 and 1986 (presented in detail from page 80 to page 84 above). The use of the adjective "translatable" presupposes the possibility of untranslatability. On this basis, the rules and the justifications provided can be considered to indicate features of language or translation situations that Brislin believed tend to produce untranslatability.

When introducing these rules, Brislin states that their purpose is to ensure that the *translators* will:

1 have a clear understanding of the original language item;

2 have a high probability of finding a readily available target language equivalent so that they do not have to use convoluted or unfamiliar terms

3 be able to produce target language items readily understandable by the eventual set of respondents who are part of the data-gathering stage of the research project.

<div align="right">(Brislin 1986: 143)</div>

I have classified the reasons Brislin gives for each rule into two categories. The first of these contains problems caused by English alone, i.e. problems related to understanding English, and which are independent of whether or not the text in question will be translated. This is my classification, since Brislin did not make any such distinction, and neither did he acknowledge that the majority of issues actually fall into this category. Examples are Rules 1 and 2, which are

justified entirely on the basis that they make the resultant English easier to understand. The second category lists problems that are related to the fact that more than one language or culture is involved and can therefore be considered (at least potential) sources of translation difficulties. An example of this is the lack of equivalents for metaphors and colloquialisms used to justify Rule 4.

Table 4.3 below summarises the different issues covered in relation to each of the Rules, broken down into problems classed as either "Understanding English" or "Translation problems".

Table 4.3 – Analysis of Brislin's rules for writing translatable English

	Rules	Problems each rule is designed to avoid or solve	
		Understanding English	Translation problems
1	Sentences of less than 16 words	Difficulty untangling ideas and relating subordinate clauses to ideas.	
2	Use active not passive	Difficulty identifying subject, verb, object and matching adjectives and adverbs to nouns and verbs	
3	Repeat nouns instead of pronouns	Unclear references due to vague noun-pronoun links.	Other languages have far more pronouns than English
4	Avoid colloquialisms		Such terms are very difficult if not impossible to translate. Such items are common in standardized tests from one country and survive "item purification".
5	Avoid subjunctive	Translator is forced to guess or approximate.	... other languages rarely have readily available terms for the various forms of the English subjunctive.
6	Add context ...		Translators may need more context than was needed in a single-language instrument. Translators' inability to catch mistakes without help.
	and redundancy	If there is no redundancy they may be unsure of the meaning of items from one phrase that redundant information in another phrase would have provided	

7	Avoid adverbs and prepositions telling "where" or "when"		There are often inadequate direct equivalents, changing the meaning of entire sentences.
8	Avoid possessive forms	Difficulty in matching what is 'possessed' with who is 'doing the possessing'.	the English possessive "yours" presents what is presented as one of three forms in many other languages
9	Use specific rather than general terms		General terms may not exist and where they do they are likely to refer to collections of items that are not equivalent. People do not categorize specific items in the same manner.
10	Avoid words indicating vagueness (e.g. probably, maybe, perhaps)		... the number of times an event has to occur to be labeled 'probable' may differ from culture to culture.
11	Use wording familiar to the translators	Researchers should sit down with translators and go over the materials to be translated, line by line.	If wording is familiar to translators ... they can create a well-worded target language version. ... translators are treated more like colleagues than hired help.
12	Avoid sentences with two different verbs	Translators sometimes have a difficult time attaching the relevant subject to the appropriate verb.	The difficulties of translation are extensive enough without ... problems which can be prevented

(Summarised from Brislin 1986: 143-150)

The first point I wish to make with relation to these rules, and the low opinion of translators' abilities that they reveal, is the extent to which they attempt to control and simplify use of English.

There is a stricture to limit sentence length to 16 words and use one verb per sentence (Rules 1 and 12), passives, subjunctives, possessives and pronouns are all banned (Rules 2, 5, 8 and 3), adverbs or prepositions of place or time are also prohibited (Rule 7), as is any expression of "vagueness" (Rule 10) or even colloquialisms (Rule 4). It would of course be extremely hard to say anything complex within these terms.

All of the elements listed above are part of normal English and all of them have their uses. In assuming that translators will be unable to understand English that employs these devices, Brislin is giving a very clear indication of the level of language competence he expects from translators and it is far from an optimistic outlook.

The three rules that do not recommend removing certain elements of English, suggest adding material to the text (Rules 6 and 9) or preparing the translators in advance (Rule 11). These three approaches are closer to what an expert translator would request if translating in a scenario that allowed communication with the client.

However, it is a very different prospect to add context and redundancy to the source text and make explicit the members of collective terms (and according to Brislin's decentering principle, therefore to do the same to the instrument used in the source culture too), than to allow a translator to request such information as and when needed. In the first case there is an assumption that the translator needs more information than the source culture population, since the context and redundancy is recommended for "translatability". In the second case the translator requests additional information as and when the target language, target culture or target text function demands more information than is available in the source text. I will discuss the ways that back-translation makes this impossible in the next subsection, but here my point is that there is an assumption built in to the very concept of "translatable English" which is that translators cannot understand what native English speakers can understand.

The recommendation to familiarise translators with the material in advance is the only rule that does not assume inferior language competence, just ignorance of specialist subject matter. Unfortunately, the blinding inherent in all back-translation and the demand for "naive" translators in many contemporary systems means that this is never an option for the back-translators and rarely an option for the forward translators.

The points listed under "Translation problems" in table 4.3 above could also be divided into two categories. While all of the problems tend to disparage translators' abilities, a number of them do nothing more than this, while others also provide examples of situations that Brislin considered problematic.

The purely "pessimistic" points are that colloquialisms are "difficult if not impossible to translate", that translators' "inability to catch mistakes without help" demands greater redundancy and that the

"difficulties of translation are extensive enough without ... problems which can be prevented".

Taking each point in turn, I would argue that, on the contrary, colloquialisms are not necessarily difficult and never impossible to translate. Brislin himself pointed out that such phrases are "very good at communicating *within* a community" (Brislin et al. 1973: 33). Good translators will not only be able to understand these devices in the source text, they will also be able to find suitable colloquialisms in the target language to ensure that the target text also communicates well within the target community. This type of technique is covered by Nida's "dynamic equivalence" and Nord's functional approach to translating. A competent translator should be able to select an appropriate register and conform to it in the target language. For questionnaires that are directed at lay people, the most effective register should use colloquialisms if these are a part of the way that health professionals communicate with members of the public in the target culture.

Translators' inability to catch mistakes is of course one of the major justifications of back-translation (and one I contest), but here my interest is the pessimistic view it embodies, and the same is true of the "extensive" difficulties of translation, since both demonstrate very little confidence in translators. These *a priori* assumptions of failure are at the foundations of back-translation.

In addition to these purely negative comments, the "Translation problems" column also contains descriptions of some situations Brislin considered problematic. None of these problems should actually present a problem for an expert translator with cultural knowledge of both source and target cultures *and* access to the researchers in order to request additional information as necessary. In fact the opposite is true.

These include the statements that "other languages have far more pronouns than English" that "other languages rarely have readily available terms for the various forms of the English subjunctive", that there are "inadequate direct equivalents" of adverbs and prepositions of time and place, that "the English possessive 'yours' presents what is presented as one of three forms in many other languages" that "general terms may not exist and where they do they are likely to refer to collections of items that are not equivalent", that "people do not categorize specific items in the same manner" and that "the number of times an event has to occur to be labelled 'probable' may differ from culture to culture."

Quite apart from the difficulties of writing English without using passives, subjunctives, possessives and pronouns, adverbs or prepositions of place or time, expressions of vagueness, colloquialisms or even categories, a text that had had all these elements expunged would actually be harder to translate rather than easier (harder to translate well, for a competent translator that is, it might be easier to translate in a rudimentary manner by a novice). This is because it would be too simplistic. By removing polysemy, polyphony, ambiguity and allusion, the number of options open to a translator is reduced and the range of response is restricted. In other words, by restricting the linguistic devices available, the rules hold translation to a standard that no other use of language is expected to meet. In their own language, scientists see no problem with phrases such as "feeling blue", but in the target language instrument "objectivity" is imposed, despite the finding that the original source texts are ambiguous.

While all of these phenomena are surmountable by a competent translator with a well-defined translation brief and access to the initiators of the translation, many of the solutions to these kinds of "problems" are reasons why translation is not reversible, as envisaged by back-translation. In response, Brislin's method removes elements of English that have proven problematic for his amateur translators as measured by the back-translation test of reversibility and symmetry. It is not a coincidence that the number of rules increased in successive publications.

In short, Brislin's rules for writing translatable English comprise a group of assumptions that treat translators as incompetent receivers of English and another group of assumptions that show his underlying expectation was of translation failure. Application of these rules reduces decentering to "dumbing down" or to finding the lowest common denominator, but would probably make "successful" back-translation more likely.

Although the decentering method has been abandoned by all medical back-translation processes (with the exception of the IQOLA team, which decentered after translation), the expectation of symmetry and reversibility and the resulting pessimistic attitude to translation have been carried forward into contemporary cross-cultural adaptation methods, even those that do not insist on naive translators.

In the next subsection I shall explain why I believe that the adoption of back-translation sets up a vicious circle in which back-translation is used because of an unrealistic expectation of symmetry and lack of confidence in translators, in an attempt to achieve that

symmetry, but the very methods involved in back-translation lead to the creation or imposition of "errors" which are duly detected by back-translation, reinforcing the belief in symmetry and the mistrust of translation that spawned back-translation in the first place.

4.2.3   *Back-translation enacts a self-fulfilling prophecy*

The unrealistic expectations bred of the simplistic view of translation as symmetrical equivalence and the inherently pessimistic approach to translation that assumes "errors" are inevitable (illustrated in the previous two subsections) both contribute to the number of errors that back-translation will detect. In this subsection I shall argue that, in addition to these factors, a series of fundamental flaws in the back-translation processes recommended and applied as part of cross-cultural adaptation also tend to guarantee that back-translation will detect errors.

These flaws range from artificial constraints on the translation process and deliberate selection of unsuitable translators to the definition of unattainable objectives, such as the AAOS demand for four types of equivalence that are often mutually incompatible. The result is that the decision to employ back-translation becomes a self-fulfilling prophecy.

The assumption underlying back-translation is that translators will make errors and back-translation will detect them, while the flaws inherent in the process almost guarantee that there will be errors to be detected. The result is one element in a vicious circle in which the more back-translation is used, the more it appears to justify its own existence, by consistently identifying errors in translations. Paradoxically, on the few occasions when back-translation fails to detect errors, as in the experiment described by Sperber et al., this too is held up as proof of the efficacy of back-translation.

A series of caveats acknowledged by the proponents of back-translation appear repeatedly throughout the back-translation literature. They are invariably presented as minor details that should be borne in mind, but, taken together, they are actually serious barriers to the success of back-translation and are the reason why Brislin's hope that back-translation could supplant other methods of evaluating translated questionnaires was never realised.

Brislin was, of course, the first to point out that backtranslation is not a panacea. When pretesting found errors after multiple rounds of back-translation, Brislin suggested that this was "probably due to some back-translators being able to make good sense out of target language

passages with several errors" since they "could compensate for these errors and write down a back-translated version equivalent to the original" (Brislin 1970: 213).

Sperber makes a similar point, with the following example. A back-translator realises that the question "Do you sometimes feel that your stomach is full?" is a mistranslation of "Do you sometimes feel fed up?" and corrects the mistake in the back-translation, so "researchers who are presented with 2 identical English versions can only conclude that the translation is excellent and leave the critically faulted target-language version unchanged." (Sperber 2004: S126).

Beaton et al. also warned that "agreement between the back translation and the original source version does not guarantee a satisfactory forward translation, because it could be incorrect" (Beaton et al. 2000: 3188).

Sperber et al. listed three limitations, all along the same lines, in that a poor forward translation is masked by the back-translation. They stated that "overly competent translators" can "achieve a back translation that is similar to the source even though the original translation is not good", because translators may have a "shared sets of rules", or they may correct "poorly written language" or, if "the grammatical form" of the original is mirrored in the forward translation, then the "two source language versions appear similar" but "critical differences between the two versions" will be hidden (Sperber et al. 1994: 503).

On this basis, it can be concluded that if the forward translator is incompetent, a competent back-translator may conceal this fact.

Conversely, another problem acknowledged by Brislin was that one set of (forward) translated essays that performed well on bilingual tests was rejected because the back-translation "was very poor, suggesting that the final step ruined a good translation" (Brislin 1970: 213).

It can therefore also be concluded that if the forward translator is competent, but the back-translator is incompetent, back-translation will also fail.

In Brislin's experiments with translated essays, the majority of both forward and backward translators had suboptimal language competence. This resulted in detection of many errors. However, Brislin did not use his iterative technique to attempt to correct these errors. He only corrected errors in the Crowne Marlowe scale which he translated into Chamorro using the six best translators. This raises a question. If both forward and backward translator are less than ideally competent, to

whom do the researchers turn for solutions to the errors that have been detected, and in what language do they discuss these solutions?

Returning to the caveats mentioned in the back-translation literature, Grunwald and Goldfarb point out another possible reason why back-translation may detect false positives, or create back-translation noise as Ozolins puts it. This is if a "literal back-translation of [a] correct forward-translation appears to be [an] error"(Grunwald & Goldfarb 2006: 6).

There is a wider point here. It is clear that it is not necessary for either translation to be "incorrect" for back-translation to detect differences. It is enough for one translator to "place emphasis on conceptual rather than literal equivalence", as requested of the forward translators in the IQOLA protocol (Bullinger et al. 1998: 914) and the other to follow Grunwald and Goldfarb's recommendation to "perform a more literal translation than normal" (Grunwald and Goldfarb 2006: 1). In both these cases, only the instructions for one translation direction included a directive such as this on how to interpret the source text. The IQOLA process does not specify how the back-translator should approach their task, while Grunwald and Goldfarb do not describe any instructions for forward translators. If one translation is closer to word-for-word "literal' translation and the other is a "freer" interpretation, then differences will inevitably emerge.

Generalizing, if the forward and backward translators take different approaches to the translation task, along any lines, then there will be differences. The forward translator may decide to take a "domesticating" approach, changing, for example, elements that do not exist in the target culture for local elements, as would inevitably be necessary at some point in the cross-cultural adaptation process for certain target cultures for certain elements. The examples that always crop up are using a fork to eat, or bowling or golf as forms of moderate exercise (Beaton et al. 2000: 3189; Bullinger et al. 1998: 921).

If the back-translator then translates, for example, "chopsticks", from a Japanese version the back-translation is obviously not going to coincide with the original English. On the other hand, if the forward translator takes a foreignizing approach and uses "fork" or "bowling" or "golf" in Japanese, the back-translation will not detect anything, despite the fact that these may not be appropriate items for measuring the construct in question in Japan, whether manual dexterity or physical activity level. Notwithstanding, using chopsticks, taekwondo and sumo wrestling would also possibly not be equivalent in terms of the physical activities involved.

The solution to this type of problem, caused by cultural differences, is to find out what exactly eating with a fork, bowling and playing golf are supposed to be testing and then formulate a question in the target setting that tests the same motor abilities, social insertion, enthusiasm for life, or whatever the underlying concept may be. However, this solution is entirely ruled out by the blinding involved in back-translation and the lack of a communication channel with the initiators or authors. In some cases communication is only constrained until after the translation is complete, but often translators will have no communication whatsoever from start to end. This is more common in smaller scale projects when the researchers compare versions and make corrections themselves or by sending small fragments for retranslation, but also when iterative back-translation is used since translators are changed after each stage and receive no feedback.

Before discussing the reasons for blinding the translators and running the translation process "at arm's length to the authors" (Ozolins 2009: 3), I would like to discuss the effect this has on the process of translation, in particular on the ability to solve problems caused by "terms which identify culturally different objects, but with somewhat similar functions" and "terms which identify cultural specialties" (Nida 1964: 167).

As theorised by functionalist translation scholars, the ideal approach to translating is to

> ... translate/interpret/speak/write in a way that enables your text/translation to function in the situation in which it is used and with the people who want to use it and precisely in the way they want it to function
>
> (Vermeer 1989: 20, quoted in translation in Nord 1997a: 29)

Back-translation processes erect a large number of barriers that prevent a translator from conducting either the forward or the backward translation in this manner.

Starting with the forward translation, Nord states that "the translator needs as much knowledge as possible about the communicative purposes the target text is supposed to achieve for the addressees in their communicative situation" (Nord 1997b: 44). The ideal source for this information, and particularly so with questionnaire translation, is the client, who should "give as many details as possible about the purpose, explaining the addressees, time, place, occasion and medium of the intended communication and the function the text is intended to have" (Nord 1997a: 30).

Considering these elements, the entire translation process becomes very ambiguous. What is the function of the first translation? Is it supposed to be a functioning research instrument despite the fact that it will be changed drastically during the stages to come? Is it, as with the IQOLA response options, merely a series of provisional suggested options, bearing in mind that there is almost always more than one forward translation?

Chesterman has pointed out that

> In the translator's head, the image of the potential effect/reception of the translation forms part of the translator's expectations about the target audience and their expectations. This image then plays a part in regulating the translator's decisions.
>
> <div align="right">(Chesterman 2007[1999]: 95)</div>

Is it not therefore possible that forward translators will see their true task as being to second-guess the back-translators and produce as few "errors" as possible by using "reversible" solutions? If this occurs and forward translators see the back-translator as their true target audience, then back-translation has distorted the process even further.

Furthermore, the background information that would normally allow a translator to define function is not provided and the communication channel with the client is often closed. In the case of the back-translator, the lack of translation brief and communication channel are absolute and non-negotiable.

Notwithstanding, the forward translator is in a more favourable position than the back-translator. Vermeer says that the translator "is expected to do research in order to make himself acquainted with all the necessary details of his commission and texts involved" (Vermeer 1994: 14), but the back-translator is blinded and cannot do this. In processes in which the translators are "naive", they will not have the necessary research skills to do so either.

Which brings me to the back-translators' brief. Is the brief to produce a functioning translation that will never be used? It appears more likely that they will see their role as to find as many potential differences as possible, not only to justify their fees, but also because they will have to face (even if only virtually) the medical professionals and the forward translators as part of an "expert committee".

On every level, all of the context, background information and interaction that translators usually use to situate themselves with relation to their projects are denied them with back-translation. There is

normally no briefing of any translators and there can never be briefing of the back-translators, there is no access to parallel texts and usually no definition of target audience and there is absolutely no opportunity for asking questions and exchanging draft solutions and alternative options. Even the very best and most experienced specialist translators would struggle under these conditions and yet Beaton et al., for example, specify that three out of four translators must not even have experience to fall back on.

I would like to contrast these conditions with my own experience of specialised translating without back-translation. When I translate other types of medical texts I can expect to receive supplementary material such as copies of publications on the same subject, references for articles describing other stages in the same research or similar work by other investigators and probably a link to the journal website, often directly to the instructions for authors page. If such materials are not forthcoming I can invariably locate them using the internet.

In addition to this, I maintain a constant stream of communication with my private clients, starting with preliminary questions ranging from target audience to word count limitations. During the actual translation process I often send four or five intermediate versions (for an average sized academic article) with explanations, questions and suggestions.

When my client is an agency I either do the same as for private clients, via a project manager at the agency, or I produce two final versions, one with the comments and suggestions, including alternative options for certain parts of the text, and a second version in which I take all the decisions myself, but indicate where other choices would have made significant differences.

This communicative element of modern translation practice has been made possible primarily by e-mail, but also by instant messaging, internet phone services and internet file transfer services that have for all practical purposes removed the limiting factor of file size. We can now electronically exchange different versions of documents that only ten years ago would have required a motorbike delivery.

Back-translation demands that back-translators be blinded to the original questionnaires, which in turn means that not only are they obliged to forgo any attempt at a communicative or cooperative approach to translation, but they will also be deprived of what is rapidly becoming the only reference source that a medical translator requires – the internet.

Any competent translator who uses, for example, the substitution technique to search for vocabulary to use in a questionnaire item will almost inevitably find the source questionnaire from which the forward translation they are back-translating was itself translated. The only precondition for this to occur is for the original questionnaire to have been published electronically, which, if it has been validated and has been adopted widely enough for it to be considered of use in a different culture, will almost certainly be the case.

By disconnecting from the internet in order to avoid stumbling over the source text, the back-translator is not only deprived of access to the scientific articles describing the research for which the questionnaire was designed and the results of that research, but also of access to "parallel texts", and the chance to "model their translations according to the patterns they find there" (Nord 1997b: 51).

Ozolins says that the back-translation process is "clearly based on a double-blind model of clinical trials" (Ozolins 2009: 3) and it is to be assumed that this is the reason that communication with translators is restricted at least until after the translation proper. However, I do not believe that a double-blind clinical trial is an appropriate model for evaluation of translations for a number of reasons. The first of these reasons is connected to the concept of evidence-based medicine.

Evidence-based medicine is an approach to healthcare that consists of making decisions, designing processes and formulating policy on the basis of the best available evidence accumulated through application of the scientific method. It has become the cornerstone of decision-making in most modern health systems and is the policy of the Conselho Federal de Medicina in Brazil, the American College of Physicians in the United States, the National Institute for Clinical excellence in the United Kingdom and the World Health Authority globally.

In order to be able to use the "best available evidence" it is first necessary to evaluate evidence. In the health sciences this is done using "evidence levels". Below are the evidence levels adopted by the relevant medical authorities in Brazil, the United States and the United Kingdom (all underlining added):

Projeto Diretrizes da Associação Médica Brasileira e Conselho Federal de Medicina

A   Estudos experimentais ou observacionais de melhor consistência.
B   Estudos experimentais ou observacionais de menor consistência.
C   Relatos de casos estudos não controlados.
D   Opinião desprovida de avaliação crítica, baseada em consensos, estudos
         fisiológicos ou modelos animais.

(Associação Médica Brasileira, Conselho Federal de Medicina 2008: iii)

US Preventive Services Task Force

Level I: Evidence obtained from at least one properly randomized controlled trial.
Level II-1:   Evidence obtained from well-designed controlled trials without
              randomization.
Level II-2:   Evidence obtained from well-designed cohort or case-control analytic
              studies, preferably from more than one center or research group.
Level II-3:   Evidence obtained from multiple time series with or without the
              intervention. Dramatic results in uncontrolled trials might also be regarded
              as this type of evidence.
Level III: Opinions of respected authorities, based on clinical experience, descriptive
              studies and case reports, or reports of expert committees.

(US Preventive Services Task Force 1996: 862)

National Health Service (UK)

Level A: Consistent Randomised Controlled Clinical Trial, cohort study, all or none,
            clinical decision rule validated in different populations.
Level B: Consistent Retrospective Cohort, Exploratory Cohort, Ecological Study,
            Outcomes Research, case-control study; or extrapolations from level A
            studies.
Level C: Case-series study or extrapolations from level B studies.
Level D: Expert opinion without explicit critical appraisal, or based on physiology,
            bench research or first principles.

(National Health Service [UK] 2010)

It will be observed that in all cases "expert opinion is the lowest level of acceptable evidence" (National Health Service [UK] 2010). Despite being based on the experimental method, back-translation does not fit modern evidence-based medicine. Not only is there no evidence that the technique works in the final analysis (as I have shown in subsection 4.1.1), the only "evidence" that the back-translation can provide about the forward translation is the back-translator's expert opinion of how to translate it back into the source language.

So, if all we have is two or more experts' opinions, where do we go for evidence in favour of one or the other if they disagree? I use the word disagree, since the perspective of back-translation is that a difference is a disagreement, rather than the more likely explanation that

it is evidence of asymmetry. For some reason, back-translation trusts the back-translator, which is an assumption that the back-translator will have made a "transparent" reproduction of the forward translation, even though the back-translation has only been commissioned because of suspicions that the forward translation was not a transparent copy of the source text.

The back-translators operate under a large number of constraints to produce what is basically an indirect translation. They work in the dark, with no briefing on the underlying concepts and, if the AAOS recommendation for naive translators is followed, they have no experience of their subject.

Despite this, the researchers trust the back-translation to reveal problems in the forward translation. This is because of what Lefevere has pointed out: "for readers who cannot check the translation against the original, the translation quite simply is the original (Lefevere 1992: 111). They therefore believe that the back-translation is a "transparent" rendering of the forward translation. If it were not, it would not be possible to use it to evaluate the forward translation.

However, if the back-translation is a transparent image of the forward translation and yet the forward translators had more information and better resources than the back-translators, there no longer appears to be a need for a back-translation in the first place, since surely a forward translation produced under such favourable conditions should also be a perfect image of the source text (shouldn't it)?

Furthermore, from the point of view of levels of evidence, in the AAOS and RDC-TMD methods only one of the forward translators is an "expert" at this type of translation, so his or her opinion should be preferred to that of the naive translators, since "lay" opinion is not worthy even of the lowest evidence level.

There is still a problem even experienced, professional translators are used for both directions, since if they are all equally experienced there is still no basis on which someone who does not see the target version (or does not understand the target language) can tell where a discrepancy between original and back-translation has originated. Even if a rater does see the forward translation and does understand the target language (in which case they wouldn't appear to need a back-translation anyway) and the rater decides that it is the forward translation that is at fault (which is the only result that is of any practical impact, since the back-translation will never be used), then the result is still no more than a difference of opinion (expert or otherwise), in this case between rater and translator.

To put it another way, within an evidence-based medicine scenario, in which a research instrument is being translated specifically in order to provide a tool for acquiring data, with which to conduct evidence-based medicine, it is at the very least incongruous to "test" an intermediate translations, which is the result of expert opinion, against another intermediate translation, albeit in the opposite direction, which is also the result of expert opinion, only to finally take the resulting translation and test it "properly".

By "properly" I mean from the perspective of the proponents of cross-cultural adaptation themselves. They all warn that only evidence from properly designed experiments with members of the public can prove the validity of the final instrument (Sperber 2004, Bullinger et al. 1998, Beaton et al. 2000). For them testing "properly" is testing with a sample target population using statistical tests of validity. In addition to statistical validation of the completed questionnaire for internal and external reliability, the IOQLA and AAOS methods recommend that scales should also be tested statistically. The AAOS guidelines do not specify the method, but the IOQLA method specifies accumulating sufficient data on the target population to recalibrate the scale to fit target population norms.

The restrictions placed on the translators, particularly the back-translators, are theoretically imposed in order to avoid bias, as with the techniques of randomization and blinding in clinical trials. However, in the case of back-translation, this actually pseudo-science because cross-cultural adaptation is not a scientific experiment, but a process for creating a translation which will then itself be the subject of at least one true experiment – the validation study – and then, if approved, become the data-collection instrument for any number of future experiments. This means that making translators work blind and choosing "naive" translators are unnecessary barriers to achieving a functioning translation and do not in any way improve the validity of a given translation solution, they just make it much harder for the translators to choose their solutions.

Furthermore, there are many "problems" that back-translation detects that are not translation problems at all. They are simply differences between cultures. These "problems" cannot be solved by back-translation either, but many could be solved by the client-translator communication that back-translation blocks.

By 1986, Brislin had acknowledged that "translators must know the infrapsychology of the tools they are converting and they must know the empirical connotations of an item as well as its linguistic and literal

referents" and considered that "the researcher in charge of the project" was responsible for doing making sure this was the case (Brislin 1986: 150)

If similar efforts at translator-preparation were undertaken in cross-cultural adaptation projects, then many "problems" which are detected by back-translation, but which back-translation cannot solve, could be discussed in advance. For example, in the article describing the AAOS method, Beaton et al. state that "the meaning of 'seeing your family as much as you would like' would differ between cultures with different concepts of what defines "family"—nuclear versus extended family" (Beaton et al. 2000: 3189).

In order to choose an appropriate translation, it is necessary to explain why the question has been asked. If it is because the patient may have some condition that physically prevents them from visiting or being visited, then it shouldn't really matter which members are included. If it is because the respondent might be working too hard, and the question is probing emotional self-deprivation, then maybe the original question's "family" could be "decentered" to "loved ones", etc. As Brislin pointed out, it is better to be specific, for example, "wife and children", "wife, children and parents", etc. However, for this to work the researchers need to know what they themselves mean first. If they do not know why their questionnaire works, if it is actually a magic bundle of rituals, they will not be able to provide the necessary explanations.

However, assuming that the researchers do understand the theory underlying their questionnaires, then, in a communicative process in which a translator can request this type of information and suggest solutions for appraisal, an appropriate translation can invariably be found. If, however, the translator is working blind, with the additional pressure of knowing that the translation will be "tested", the natural reaction is to translate defensively, as Hönig puts it, according to the following rule: "I know that this does not sound particularly good but I challenge you to prove that it is not correct" (Hönig 1997: 17).

In addition to cultural issues that require additional information, there are also a series of techniques by which translators could improve translations, but which need to be agreed with initiators. Brislin mentioned reversing the polarity of scoring if negatives make a translation sound awkward, but a translator obviously cannot make such a decision without consultation. The same is true of changes in layout and format and, as in the example of racial categories, of elimination of items that are culturally incompatible.

If a translator is working in isolation, blinded and with foreknowledge of impending back-translation, they are not in a position to do anything other than find a translation for the items they are presented with. If, however, they are translating in a communicative and cooperative manner, they can point out when an item appears to be inappropriate or irrelevant.

The differences that back-translation detects can basically be divided into linguistic differences and cultural differences. Back-translation processes often cause the first type by selecting inappropriate translators, by constraining translation techniques and by restricting information available to the translator. Even when these constraints are not involved, the asymmetry between languages practically guarantees that some discrepancies of this type will be detected.

However, the effect of back-translation on the second type of difference is more significant, since it prevents translators from contributing to the solutions. Firstly by restricting the (parallel, source culture) information they would use to guide their choices, secondly by rejecting any creative and appropriate solutions such as the Dutch and Swedish RDC-TMD race questions because they do not pass through the reversibility filter[23] and thirdly, and most importantly, by cutting the communication channel with the initiators of the translation.

Vermeer has noted out that people often "deny the translator the necessary information about documents" and this led him to ask two rhetorical questions: "Can a lawyer or doctor work efficiently if you hide important information from them? Would they work under such conditions?" (Vermeer 1992: 14).

By anchoring the translation to the source text, back-translation negates the skill and experience of expert translators, so it is maybe unsurprising that the AAOS and RDC-TMD methods actually prefer inexperienced translators.

While the source-anchoring effect of striving for symmetry and reversibility can lead to absurd translations such as the Aleut in Italy, it is actually the cutting of the communication channel that is more damaging, since the effect of this is to make the translator into a mere "bilingual informant" (Catford 1965: 27).

Rather than being given the freedom, as the "responsible expert" (Nord 1997a: 30), to *proactively* approach the entire questionnaire as an

---

[23] As I have mentioned before, in order for the Dutch and Swedish solutions to have been used, either someone overruled back-translation or the solutions were introduced after the back-translation step.

assignment, preparing the ground by reading the scientific literature on the original instrument and the parallel, target language, literature, before requesting explanation of key terms and of the intended function of the entire instrument and of specific items, the translator is reduced to a "walking dictionary" (Hönig 1997: 17) who *reactively* justifies translation decisions or provides alternatives on demand.

I shall take up the effects of this on translators in section 4.4 below. In the next section I shall deal with the effect that back-translation processes have on the source or target oriented nature of translation and the issue of cultural hegemony before discussing how patronage affects the use of back-translation.

4.3 Cultural issues

Whereas the previous section was concerned with the simplistic conception of meaning and equivalence on which back-translation is based and the procedural flaws that result, this section covers issues related to the intercultural nature of translation and the systemic nature of culture.

The section is divided into four subsections. In the first subsection I shall highlight some significant differences between contemporary applications of back-translation in medicine and the way that Brislin employed back-translation within cross-cultural psychology, using decentering, and in the second subsection I show that the result of rejecting decentering while maintaining back-translation is to *anchor* the translation to the source text and source culture.

In the third subsection I shall discuss issues related to cultural hegemony, arguing that back-translation enacts unequal power relations leading to cultural colonialism, which can be academic and intellectual, but can also be organisational and corporate driven.

In the final subsection, I make use of Lefevere's concept of patronage to illustrate some of the reasons why back-translation is so widely used, including the citation "snowball effect" and publication-driven academic career paths.

As Christiane Nord has pointed out, when dealing with cultures, "there can be no neutral standpoint for comparison" and "culture can only be perceived by means of comparison with our own culture" (Nord 1997a: 34). However, accepting subjectivity does not mean that one can just say anything, on the basis that everything is shaded by opinions, attitudes and ideologies anyway. Rather it imposes the responsibility to frame and contextualise statements in such a way that these opinions,

attitudes and ideologies are acknowledged, since it is not subjectivity that is the problem, but the illusion of objectivity and impartiality.

In view of this, it is necessary to state that, with the exception of a single translation project, all of my experience of translation has been gained in the Brazilian culture. My perspective on back-translation is therefore that of a translator working in the target culture, although back-translating into the source language. Furthermore, the projects conducted here in Brazil are not exclusively local subsidiaries of projects to "export" questionnaires. By this I mean that while Brazil does host large multinational projects such as ISAAC and IQOLA that export a version of their questionnaires to many different countries, many cross-cultural adaptation projects are also initiated here when a researcher or team of researchers "import" a questionnaire that they wish to use in Brazil.

4.3.1 *Brislin studied cultural differences, but cross-cultural adaptation views them as obstacles*

There is an important difference between the way that Brislin viewed cultural differences and the way they are approached in contemporary applications of back-translation for health-related questionnaires. In the early back-translation literature (which was developed within cross-cultural psychology), the primary research interest was to observe differences between cultures, hence the need for translation. In the later back-translation literature (for healthcare applications), the objective is to collect comparable data in different cultures and cultural differences are seen as barriers to be overcome, rather than phenomena of interest in themselves.

This basic difference in objectives has meant that a central element of Brislin's approach to cross-cultural research has been abandoned in contemporary cross-cultural adaptation methods designed for health-related questionnaires.

It will be remembered that Brislin classified culturally-specific concepts as "emic" and universal concepts as "etic" (discussed on pages 76 to 78 above) and that this distinction was central to his "decentering" method, since, after each step in the iterative back-translation process the *source* text was rewritten to eliminate elements that caused discrepancies when back-translation and source text were compared.

This is the method Brislin used to translate the Crowne-Marlowe scale. It is significant that this method, unlike the contemporary cross-cultural adaptation methods, does not attempt to "correct" the "errors" in

forward translations that are detected by back-translation. Rather the process of decentering the source text is repeated until successive back-translation cycles have led to a back-translation with zero criterion 1 meaning errors, when compared with the decentered source text used for the forward translation from which the back-translation with no errors had been translated. Brislin recommended administering a decentered version of the original source text to a source culture population, specifically this last version in the series, on the basis that it "is most likely to be equivalent to a target language version, probably the version immediately preceding the final step" (Brislin 1986: 160).

Brislin saw this method as eliminating elements that were culturally specific to the source culture, resulting in an entirely "etic" questionnaire, on the basis that "if a concept 'survives' the decentering procedure, it is assumed to be etic since there must be readily available words and phrases in the two languages" (Brislin 1986: 160).

Before considering the implications of this for health-related questionnaires, I would like to point out that this is not the same as evaluating translation quality. In fact, if back-translation is used to decenter the original, then it is cannot be eliminating "errors" in the forward translations since these are discarded after every step except the last one, nor is it assessing the forward translation's quality, for the reasons explained in previous sections. What it is actually doing (and this is probably the reason why only one cross-cultural adaptation protocol has used decentering of any type), is framing the "translatability" of the source text with reference to the specific translators used and the specific languages involved. More correctly, it is framing the apparent translatability of the source text as conceived by a monolingual source-language rater who has not actually read the translation. If one agrees with Campbell, it is also testing how robust the concepts are, which is a point to which I shall return in this subsection.

Notwithstanding, the use of a decentered version of the original for data collection would appear to negate some of the main reasons for translating an existing questionnaire in the first place, which were to collect comparable data allowing "a literature to be built up around a commonly shared set of concepts and operational definitions" (Brislin, 1986: 138), to ensure that "the impact of a disease or its treatment is described in a similar manner in multinational trials or outcome evaluations" (Beaton et al. 2000: 3186) and to accumulate "international comparative data on a myriad of health issues." (Ozolins 2009: 1).

Brislin only translated the Crowne Marlowe scale into one language, so he was not forced to consider the implications of using

decentered versions to collect data in the source culture when the source text has been translated into several languages.

Since decentering takes place during translation, a project such as the EORTC project which had 83 different language versions listed in 2009 (Dewolf 2009: 29-29), would be faced with having an equal number of decentered English versions in which, for each target language, items that could not be back-translated had been eliminated by decentering (item purification, in Brislin's terms). Even without following the most radical suggestion, which was to also add target culture emics to target language versions, not only would 83 different target language versions have been produced, but the 83 "decentered" versions of the original questionnaire would also be different, since the types of differences that back-translation detects are very often simply a function of the different surface structures of the source and target languages and translation solutions to such asymmetries are very often irreversible.

If the intention, as Brislin recommended (Brislin 1986, p 160-161), were to collect data with both original and decentered versions in the source culture, then some solution to the existence of multiple source language versions must be found.

They could presumably be "decentered" against each other to arrive at a "centerless" synthesis, but, this version would then have to be tested against each of the translations using back-translation once more, generating a further cascade of decentered versions.

In common with the scenario illustrated by the diagram depicting a potentially endless cycle of translation, back-translation, rejection and repetition (Figure 2.1 on page 87 above), this is highly impractical. Brislin himself said that existing instruments were used to save time and expense, since "time, energy and funding" are always limited and "cost/benefit considerations ... should become part of the researcher's planning" (1986: 138).

The result has been that, with the exception of the SF-36, back-translation has been adopted in medical settings without decentering. As the EORTC manual puts it, "It is not the purpose of the translation procedure to modify the original questionnaire." (Cull et al. 2002: 3; Dewolf et al. 2009: 6).

This means that a core element in the process of cross-cultural adaptation does not serve to adapt the original instrument to the target culture. Instead, back-translation is used to modify the target-culture instrument until, in the view of members of the source culture, often

with no knowledge of the target culture or language, it is "equivalent" to the source language instrument.

The only exception is the international version of the SF-36, which was created after 14 target language translations had been produced. While it fits the basic concept of decentering, this was a very much more sophisticated form than Brislin's Procrustean approach of eliminating problematic items on the basis of *monolingual* comparison of back-translation and source text. The IQOLA decentering process was based on translators' ratings of the source text and ratings of translations against the source text performed by bilingual raters who were also translators. These ratings were discussed by translators and national principal investigators (also bilingual) at an international meeting (Bullinger et al. 1998: 914-916). The SF-36 was therefore modified on the basis of bilingual evaluations and multilingual communication not on the basis of back-translation results.

While the IQOLA approach is undoubtedly the most enlightened and this form of decentering is more truly deserving of the name, since it decenters between 15 different languages, whereas Brislin had just two poles, the fact that it was conducted after translation means that it cannot ameliorate the source-anchoring effect of back-translation.

The abandonment of decentering is a manifestation of a key difference in approach between Brislin's cross-cultural psychology perspective and the medical perspective driving cross-cultural adaptation. For example, Brislin's reaction to the discovery that Japanese has different words for older and younger brothers was that this could lead to "fruitful hypotheses" and that, if there were also differences in attitudes, this would be "an important finding" (Brislin, 1986: 147). This is because Brislin was interested in differences between cultures as an object of study. His approach was therefore to "include items aimed at both etic and emic aspects", despite the fact that there would be "no statistically interpretable cross-cultural relationships between the emic components" (Brislin, 1980: 393).

In contrast, cross-cultural adaptation demands statistically comparable categories, to the extent that, if due attention is not paid to how categories are defined in different cultures, it can provide an illusion of comparability where none exists. To take the example of the RDC-TMD questionnaires presented in subsection 4.1.2 earlier in this chapter, the Italian and English questionnaires have the same race and ethnic origin categories, but are obviously not comparable if one considers their relative relationships to the populations they are aimed at, since the Italian questionnaire has no category for Italians, whereas

all of the categories on the United States questionnaire are descriptions of United States residents.

It is not, however, necessary to resort to such a glaringly inappropriate choice of categories to illustrate this illusory comparability. The Brazilian version includes all of the categories used by the Brazilian national census, while preserving the coding from the English version. However, as I have explained, even the categories that apparently coincide do not describe similar groups of people and so any statistical relationships that were detected on the basis of these categories would be based on the false assumptions that people who are classed as "black" in the United States are similar to people classed as "preta" in Brazil and that people classed as "white" in the United States are similar to people classed as "branca" in Brazil.

Some categories cannot be comparable across all cultures because the elements they categorise do not exist in some cultures. Brislin's solution of decentering and then adding culture specific elements would have eliminated the "Aleut" from translations of the RDC-TMD questionnaire, but possibly at a cost to the comparability of other items.

This issue of the comparability of items brings me back to the questions of to what extent the operationalization of a concept (its definition) can be altered without altering the concept itself and whether this is a reflection of how robust the concept itself is.

When Brislin decentered the Crowne-Marlow scale (without adding local culture-specific items) he gave an example of an item which had read "I have never intensely disliked anyone." with two response choices, "true" or "false". He stated that after translators had found "intensely" difficult to translate the statement was decentered and the version actually used for his criterion 5 test was "I have never really disliked someone." (Brislin 1986: 160-161).

This is by no means the same statement any more. The original statement was fairly unambiguous. For someone to answer true they had to state that they had *never* had an *intense* dislike of *anyone*. However, at least in part of the world that I come from, "really" in the second version of the statement could have several shades of meaning.

I often use the substitution technique with my clients when eliciting their preferred solutions. Were I to be asked to translate the second statement, "I have never really disliked someone.", I would wish to ask my client whether any of the following substitutions would accurately represent their intended meaning, or whether they actually preferred the ambiguity of leaving all of these meanings open:

> I have never *truly* disliked someone.
> I have never *actually* disliked someone.
> I have never disliked someone *very* much.
> I have never *extremely* disliked someone.

On the basis that I only consider the last of these options to have a meaning that is specific in the same way that "I have never intensely disliked anyone" is specific, then I would maintain that if the degree of "precision" that, for example, Ozolins' clients "pursued at every turn" (Ozolins 2009: 9) is truly necessary to maintain the useful properties of a questionnaire, then data collected using the decentered version of the Crowne Marlow scale cannot be statistically comparable with data collected using the original version.

However, the results reported by Perneger et al. suggest that the MAPT team are incorrect, since they showed that very different wording can produce what is, statistically speaking, the same response. On two French versions of the SF-36, one produced using a centrally controlled process involving back-translation and the other produced in the target setting without back-translation, thirty-four out of thirty-six items were different and all responses except yes/no were different, but the results demonstrated "instruments of almost identical reliability and validity" (Perneger et al. 1999: 1037; 1045). It must therefore be accepted that, at least in some circumstances, it is possible to acquire statistically equivalent data using questionnaires with radically different wording, meaning that back-translation and the source-text anchoring it imposes are not always necessary.

Returning to the Crowne-Marlow example, Campbell said that "if changes from 'intensely' to 'really', and from 'someone' to 'anyone', change the underlying concept so much that there is differential response, then the underlying concept is weak" (Brislin 1986: 161). In this case, I agree.

Although I have shown that, in terms of the situations described the two phrases are not equivalent, it is necessary to consider the *function* of the Crowne-Marlowe Social Desirability scale before evaluating whether or not a true or false answer would measure the same concept in each case, since Perneger et al. have shown that semantic equivalence is not always necessary for equivalence of psychometric concepts and I have shown, with relation to the RDC-TMD race categories, that semantic equivalence can actually be undesirable.

The purpose of the Crowne-Marlowe questionnaire is to measure the extent to which respondents exaggerate their good points and deny

their weak points, i.e. the extent to which they make their answers more socially desirable than is normal. The scale scores the number of times respondents answer an improbably altruistic statement in the affirmative or deny an undesirable quality, thereby providing an indication of the extent to which their answers are influenced by a need for approval (Marlowe 1999).

In this context, Campbell would appear to be correct that the intensity of the dislike and the number of people (never) disliked is irrelevant to the concept since a change in wording would not change the fact that everybody has disliked someone else at some point and to deny this with respect to oneself is to deny an unpleasant trait that is human nature, irrespective of whether the dislike is "intensely" or "really". In other words, although "I have never really disliked someone." does not mean the same as "I have never intensely disliked anyone.", for the purposes of detecting the approval motive, which is the only concept tapped by the Crowne-Marlowe scale, they are interchangeable.

Perneger et al. also suggested that their results were because the "the structure and item content of the original instrument are particularly robust", but did not relate this to an additional factor: the fact that the translation prepared without using back-translation had been translated by expert medical translators working with medical researchers who understood the underlying concepts and the reasons for each item's inclusion.

In contrast, back-translation without decentering demands near word for word equivalence with no consideration of the underlying concept nor of the way it is being measured. For example, back-translation has no mechanism for judging whether intensity of dislike is relevant in a question with a true or false response and is entirely unable to deal with underlying concepts.

This brings me to the subject of the next subsection, in which I shall explain what I mean when I say that back-translation *anchors* the target text to the source text and culture.

4.3.2   Contemporary back-translation methods are source-anchored

Introducing the IQOLA method, Bullinger et al. explain that health status measures must be culturally adapted (rather than simply translated) because they "can not be assumed to be *a priori* invariant to cultural diversity" (Bullinger et al. 1998 913). Introducing the revised AAOS method, Beaton et al. define "cross-cultural adaptation" as "a

process that looks at both language (translation) and cultural adaptation issues in the process of preparing a questionnaire for use in another setting." (Beaton et al. 2000: 3186) and Sperber et al. define the objective of cross-cultural adaptation as "to adapt the instrument in a culturally relevant and comprehensible form" (Sperber et al. 1994: 501-2).

This would suggest that cross-cultural adaptation is a target-oriented endeavour, since cultural relevance and comprehensibility can only be defined in terms of the receiving culture and, in order to test whether the underlying psychometric properties have indeed remained invariant after adaptation, testing must be conducted with the target population. Indeed, the majority of cross-cultural adaptation procedures, from bilingual rating, through to validation in the target setting are indeed target-oriented, particularly those that involve national principal investigators.

In contrast, back-translations are evaluated by monolingual members of the source culture using the source language. This has the effect of enforcing "equivalence" with source culture elements and source language features that are irrelevant or inappropriate in the target setting.

In Brislin's criterion 1 experiment, translations were conducted on Guam, but back-translations were rated against source texts in the United States and, of course, in English (Brislin 1970: 198). Sperber et al. carried on this tradition and had their questionnaire translated in the United States, back-translated in Israel and then rated by "29 students and faculty members of the Department of Health Behavior and Health Education of the School of Public Health in the University of North Carolina" (Sperber et al. 1994: 505; 509). This is the most strongly source-anchored version of the three contemporary methods, since it rated "comparability of language and similarity of interpretability" (Sperber et al. 1994: 506) and the objective was to adapt the questionnaire "while maintaining the meaning of the original items" (Sperber et al. 1994: 502.).

In addition to the fact that the objective is unattainable, since to translate it is obligatory to change meaning (or, more correctly, to create a different meaning in a different language), the fact that similarity of form is a criterion means that the more creative the translator, the more their translation will be dragged back to the source text. The Swedish and Dutch translations of the RDC-TMD race categories, for example, could not meet criteria of similarity of form or of meaning, but are very good translation solutions.

Although the IQOLA process with its national principal investigators and extensive multilingual communication is undoubtedly the most target-oriented of the three processes presented in detail in Chapter 2, it does employ a back-translation stage and this stage is monolingual. In this method the back-translations were "reviewed by researchers at the Health Assessment Lab for conceptual equivalence with the original source version" (Bullinger et al. 1998: 915). The Health Assessment Lab is in Boston, in the United States (Bullinger et al. 1998: 913). Additionally, Perneger et al. state that during an earlier stage, "the developers of the original American instrument participated in selecting the translation options" (Perneger et al. 1999: 1038). This phenomenon of source text authors exerting control over translations is also a source-anchoring element.

In the AAOS method, the back-translations are not rated during a discrete step as in other protocols. Rather they are discussed together with the forward translations at an "expert committee", comprising a specialists in questionnaire methodology, the questionnaire developer, a language professional and the translators (Beaton et al. 2000: 3187).

This configuration has the advantage that back-translation is not a type of hurdle that a translation must jump over to reach the next stage. However, there are still elements to the process that tend to lead to source-anchoring. The first is the committee's objective. They are expected to achieve "semantic equivalence", "idiomatic equivalence", "experiential equivalence" and "conceptual equivalence" and "must examine the source and backtranslated questionnaires for all such equivalences" (Beaton et al. 2000: 3188-3189).

I have already discussed the impossibility of maintaining semantic equivalence, defined as "Do the words mean the same thing?" (Beaton et al. 2000: 3188), while also maintaining the other three types. Here, however, I would like to point out that Beaton et al. state that "Consensus should be reached on the items, and if necessary, the translation and back-translation processes should be repeated to clarify how another wording of an item would work." (Beaton et al. 2000: 3189). This, together with the fact that at least the original developers, probably the methodologist and possibly the language professional (depending on target language) cannot speak the target language, suggests that the committee meeting will be at least predominantly conducted in English, meaning that back-translation is a tool to enable English to be used as the language of decision-making, despite the fact that those decisions affect the target language version exclusively (there is no decentering in the AAOS method).

The MAPT translation described by Ozolins is a slightly different case since their adaptations were all for use within the same country as the original instrument - Australia. It might be considered that there would therefore be no cultural differences, but since the target population is described as "non-English speaking patients" (Ozolins 2009: 3) then, if they are immigrants they are presumably first or second generation, and therefore do not have the same culture as their fellow Australian residents who do speak English, and if they are tourists or other visitors there will also be cultural differences.

Notwithstanding, even if the culture is considered to be the same, the back-translation process still anchors evaluation of the translations and discussion of them to the source language. The MAPT team did not have bilingual consultants to help them rate the translations, as in the EORTC method on which they modelled their project (Ozolins 2009: 3). Rather it was the authors themselves who compared back-translations with their source questionnaire and "involved themselves directly with the translators". Ozolins claimed that this meant that "every nuance and variation could be explained to the authors' satisfaction" (Ozolins 2009: 3).

However, this view of the process ignores the fact that this communication will be in English. In addition to the fact that one cannot express, for example, Chinese meaning in English (for this one needs to use Chinese), the forward translator is forced to back-translate or paraphrase their own examples using English and justify their decisions in English, for the very simple reason that the authors don't understand the target language.

In addition to the fact that back-translations are invariably rated against the source text by people who are monolingual in the source language and come from the source culture, the context in which these raters examine back-translations encourages them to find differences.

For example, one difference identified by the MAPT team in a back-translation from Macedonian was as follows, "The [back-translation] has *enjoy life* rather than *enjoy my life* (original). [Back-translation] is less personal." (Ozolins 2009: 7). In response, the forward translator quite reasonably pointed out that "The [forward translation] is the same, following the logic that I can enjoy only my life and not

somebody's[(sic)] else.[24] There is no other natural way to say it in Macedonian." (Ozolins 2009: 7).

As was the case with the Crowne-Marlowe item, the difference would not make any difference to probing the underlying concept. However, when engaged in a difference-identification task, it is natural that people will identify differences and that is what the MAPT authors did. In cases such as these the MAPT team accepted that what had seemed to be errors were "back-translation noise", meaning that the target text was not ultimately anchored to the source text in these cases. However, the effect is to anchor the translators to the source text, forcing them to justify everything in terms of the source text and in the source language. I shall discuss the effects of this on translators in section 4.4.

However, in other cases the authors did insist on changes to guarantee equivalence with the source text. The response options in particular were seen as especially important and they spent "much time" on "the critical issue of whether there was an equivalent distance between the options to the distance between the options in English." (Ozolins 2009: 5). Ozolins claimed that "the validity of the instruments rested on the fine gradations of patient response, and it was these gradations that were the essential items to maintain in any translation." (Ozolins 2009: 2)

However, Perneger et al. showed that *rank* is more important than the labels chosen for the response options and the "fine gradations" Ozolins talks about are simply five options on a scale of one to five. In other words the "patient response" is controlled in advance by the scale and, following Campbell's theory, *if* the MAPT questionnaire's underlying concepts are not weak, then minor changes to wording should not affect its validity.

To extend this reasoning, there is an alternative hypothesis to the conviction that strict adherence to an ideal equivalence is necessary. The first step is to assume that the bundle of rituals is not in fact magic at all and to acknowledge that the data collected with health questionnaires relates to real people and these people do or do not have any number of conditions, characteristics etc. The questionnaire merely probes these details, but they are facts, these people exist and they do or do not have these conditions and characteristics.

---

[24] Note the effect of forcing the forward translator, whose "saleable" language is Macedonian, to justify his/her decisions in English. While doing so, she/he commits a minor error of syntax, thereby giving the impression of pidginization and undermining his/her status.

The alternative hypothesis is therefore that questionnaires work, not because they have hit upon a "magic" formula of words, but because they are part of an initiative to investigate these conditions and characteristics in the first place. In other words, it is actually the very fact that a medical professional or researcher has created a questionnaire and administered it in order to collect a specific type of data about a patient or population that is the important element.   After all, populations and patients have health problems so there is indeed data to collect and while one obviously wishes to collect data that truly relates to the phenomena of interest, in contrast with scientific terminology which is intended to be univocal and unambiguous, popular language has many different ways of saying the same thing.

Furthermore, it must be remembered that the original questionnaires have all been through rigorous processes of validation, often involving comparing series of results with results of clinical or experimental tests that probe the same concepts. Additionally, the more rigorous cross-cultural adaptation processes, such as the IQOLA include "examination of the validity of the scales and the accumulation of normative data and other interpretation guidelines" (Bullinger et al. 1998: 914).

The normative data in particular is important because it means that any differences in scaling can be recalibrated. Since the target population is assumed to be different from the source population then it cannot be assumed that the results of source and target questionnaires should be similar. If populations did not differ there would be no need to use questionnaires in different cultures since data from one population could be extrapolated to all others. However, less sophisticated cross-cultural adaptation processes do what Brislin did with his Crowne-Marlowe results and compare them with results from the source population. Until sufficient data has been collected on the target population to be able to trace a profile of "normality", there is no basis for true comparison. Once normative data has been accumulated then the "interpretation guidelines", including the cut-off points for scale outcomes, can and should be adapted to the target population.

In the following subsection I shall discuss issues related to cultural domination, covering the relationship between translation and cultural colonialism, unequal power relations between cultures and the conditions of cultural import and export, before considering Venuti's suggestion that foreignizing translation strategies could be used to subvert the cultural hegemony of English.

4.3.3   Cultural colonisation and asymmetrical power relations

The relationship between translation and colonisation is a recurrent theme in translation studies. Venuti summarises the "inevitable reliance" of colonisation on translation as follows,

> Christian missionaries and colonial administrators, with the help of educationalists and anthropologists, typically composed dictionaries, grammars, and orthographies for indigenous languages and then set about translating religious and legal texts into them. ... Translation enabled conversion and colonization simultaneously
>
> (Venuti 1998: 165)

Brislin conducted his experiments with back-translation on the island of Guam. Guam has a long history of colonisation, starting in 1668 when Spain established a colony, and continuing to the present day, since its status as an "unincorporated territory of the United States" means it is listed as one of only sixteen "non-self-governing territories" by the United Nations' Special Committee on Decolonisation (Wikipedia 2013e). Guam has significant strategic value due to its position in the western Pacific and the BBC describes it as a "keystone of American military strategy in the region" (BBC News 2011c). Shortly after Brislin completed his research there, the United States used its airbases on Guam during the Vietnam war (BBC News 2011c).

Unsurprisingly, given the setting, Brislin was aware of the dangers of cultural imposition, stating that by using existing instruments researchers may both miss important aspects of "a phenomenon as viewed by (and seen as important by) people in other countries" and also run the risk of imposing conclusions based on concepts that do not exist in the culture they are working in, since "existing instruments provide operational definitions of certain concepts" and it cannot be guaranteed that either the concepts or their definitions are common to both/all cultures under study (Brislin, 1986: 139).

Furthermore, Brislin was a Western researcher working in a colonial setting, but he recognised that the situation was changing, stating, in 1980, that "more and more research is [now] done by members of cultures who previously were only the hosts for visiting research teams" (Brislin, 1986: 144).

However, there are also traces of the traditional colonial model, starting with the very fact that the researcher was an American and the islanders were the experimental subjects. This configuration remained the norm, despite the reference to hosts becoming colleagues. For

example, with regard to problems caused by vagueness, Brislin states that a lack of "shared information and shared experiences" limits the extent to which interviewer-respondent interaction can overcome them (Brislin 1986: 148), which, combined with the mention of data analysis always being conducted "back home" (Brislin 1986: 152), is a clear indication that the basic assumption is that the researcher is not just a foreigner, but a Westerner.

However, what makes back-translation a technique for the perpetuation of colonial asymmetrical power relations is its propensity to be used as a means of control. In 1922 the United States government banned the Chamorro language in schools on Guam and "collected and burned all Chamorro dictionaries" (Wikipedia 2013f). This is a manifestation of what Venuti described as an aggressively monolingual stance (Venuti 2004 [1995]: 14) and while the use of back-translation as part of a process for translation into Chamorro would appear to have an inverse effect to banning the language, the fact that evaluation is monolingual and in English, sidelines Chamorro in the very process of deciding whether a translation into Chamorro is adequate.

One of Brislin's arguments for using back-translation rather than a bilingual committee was as follows, "the weakness of the method is that committee members may not criticize one another, and may even unify against the researcher" (Brislin 1980: 431). I have already commented on the naivety revealed by assuming that a given language may not have the means to criticise. Here I am interested in the possibility that Brislin had conducted such meetings, disagreed with non-native English speakers on some point or another, and then felt that they had combined against him. In other words I am interested in the way that back-translation can function to assuage monolingual insecurity, reinforcing a polar "them versus us" attitude.

By restricting all discussion to English, in which the Western researcher is fluent, but in which non-native speakers reveal the acquired nature of their language competence (as shown in the "somebody's else" example on page 249 above), back-translation imposes a relationship of superiority/inferiority, which, as Arrojo points out, is a fundamental element of colonialism,

> ... if translation is a form of cultural transport, the same can be said of the processes of colonization. In the same way that the dominant ethic of translation preaches the ideal of transparency and of unconditional respect for the powerful and sacred 'original', colonization, in all of its forms, has always been inspired by the colonizer's supremacy and supposed superiority

as pretexts for dominating the culture and identity of the colonised

<div align="right">(Arrojo 1996: 64, my translation)</div>

It must of course be remembered that "the borderlines between cultural systems or sub-systems ... are notoriously hard to define" and a "culture cannot simply be equated with a language area" (Nord 1997a: 42)

Furthermore, colonisation is no longer predominantly a matter of one country colonising another, often with the excuse that they are "civilising", "converting" or even "saving" the people of the colonised country by bringing them the word of God. Rather, as Arrojo also points out, divine authority has been substituted by the authority of reason and of science (Arrojo 1996: 55).

As Venuti has identified, "the translation practices enlisted by transnational corporations ... function in the same fundamental ways as those that underwrote European colonialism" and "translation now serves corporate capital instead of a nation state, a trading company, or an evangelical program"(Venuti 1998: 165).

Lambert echoes this point, saying,

> Modern multinationals offer a clear illustration of the eclectic way of colonizing: international trade and international export networks do not necessarily require political power (although it may help), but economic markets are generally and rather inevitably linked with moral, linguistic, even artistic and social import.

<div align="right">(Lambert 2006 [1995]: 100)</div>

One example of this is the United Kingdom and United States publishing industries which Venuti says have imposed "Anglo-American cultural values on a vast foreign readership, while producing cultures in the United Kingdom and the United States that are aggressively monolingual" (Venuti 2004 [1995]: 14).

It is not, therefore, only populations that develop their own cultures. Organizations also develop their own cultures and this is not only true of large multinational corporations with their corporate cultures, but also of non-governmental organisations such as the WHO and of academic institutions such as universities. Ozolins, for example, identified the WHO "and other international medical research and treatment organisations" as the main drivers of the adoption of back-translation (Ozolins 2009: 1).

This brings up an interesting question. For example, the EORTC is an international organisation and, it will be remembered, is very

defensive of its "original" English-language source documents, stating, "It is not the purpose of the translation procedure to modify the original questionnaire." (Cull et al. 2002: 3; Dewolf et al. 2009: 6). However, the EORTC is based in *Belgium*. What, in this case, is the source *culture*? It is evidently not Belgian, since Belgian culture is enacted in Dutch, French or German, but it is also not English, since the EORTC is made up of members from fifteen European countries plus Australia, Canada and the United States (EORTC 2013). I believe that the answer can only be that when discussing translation of EORTC instruments the source culture is the culture of the EORTC organisation. This reframes the issue of source versus target culture (which is itself already a reframing of the issue of source versus target text) as an issue of source versus target organisations, which fits the systemic view of translation.

Hermans has pointed out the "the whole process of cultural contact and transmission of which translation forms part is governed by ... power, hierarchy [and] non-equality" since "relations between communities and cultures are never relations between equals" (Hermans 2007 [1999]: 60).

Lambert has observed that "it is quite hard to evaluate colonial strategies in terms of 'good' and 'bad'", since the correct question is "for whom it is good and for whom it is bad at a precise moment and in a specific situation" (Lambert 2006[1995]: 102) and so, once more "there can be no neutral standpoint" (Nord 1997a: 34). Once neutrality is acknowledge as an illusion, it is necessary to take sides.

Venuti is one scholar of translation who takes an explicit position on translation's role in enforcing unequal power relations. He was writing about the effect on the foreign text and culture of importing and domesticating translation from the perspective of someone inserted in a hegemonic importing culture – the United States. From this standpoint he defined translation as "the forcible replacement of the linguistic and cultural difference of the foreign text with a text that will be intelligible to the target-language reader" that "serves an appropriation of foreign cultures for domestic agendas, cultural, economic, political" (Venuti 2004 [1995]: 17-18).

My position is different. I live and work in Brazil and, with relation to cross-cultural adaptation, Brazil is also an importing culture. However, whereas the United States is a "central" culture, Brazil has historically been "peripheral" (although this is changing rapidly) and, in common with countless other cultural products, questionnaires have been imported via translation in the manner described by Toury, where "the starting point is always one of a certain deficiency" in the target

culture, identified "in view of a corresponding non-gap in another culture", which may or may not have been "pointed out for it by a patron of sorts who also purports to 'know better' how that gap may best be filled" (Toury 1995a: 138).

The very fact that in Brazilian medicine importation by translation is a norm, places Brazil on the weaker side of the unequal power relations described by Hermans. Lambert has said the following on the subject of power relationships revealed by importation and exportation:

> – exporting or 'active' systems are in a power position from the point of view of the importing or 'passive' systems; ...

> – the more a given society imports from one and the same neighbour, the more it is in a position of dependence;

> – the more the receiving system is in a unidirectional relationship in matters of import / export, the more it depends on its 'big brother';

> – the more the receiving system is part of a group of receiving systems that borrow their cultural products from one and the same exporting system, the more they are subordinated to a coherent network ...

> <div align="right">(Lambert 2006 [1995], p.98-99)</div>

It will be noted that, with the exception of the second point above, Lambert consistently uses the term "system" rather than culture. The fifteen European countries plus Australia, Canada and the United States that are members of the EORTC can be considered form an exporting system in the case of health-related questionnaires.

Therefore, from the Brazilian perspective, the major back-translation projects such as ISAAC and IQOLA are not bringing *back* a cultural other, in Venuti's terms, but exporting a cultural product and "standardizing" the foreign language versions against it. Although back-translation projects initiated here in Brazil might be considered to be appropriating foreign cultural products for domestic agendas, cultural insecurity, as described by Nida (see page 145 above), in combination with the source-anchoring effect of back-translation itself, means that the results of such projects are still controlled from the source culture, even if no people from the source culture are involved, since back-translation enforces the source culture model (and is itself a source-culture model) and the choice to adapt an existing instrument rather than

developing one *de novo* is an appeal to the "sense of security offered by using an established measure" (Brislin, 1986: 138).

Furthermore, back-translation serves to preserve certain elements of the source against domestication, such as Aleut, bowling, golf, but if cross-cultural adaptation is truly to produce culturally relevant instruments then what is needed is exactly domestication of the foreign text, so that it serves Brazilian agendas.

Venuti's decision to champion foreign cultures by foreignizing is a reaction to his situation living in the hegemonic country. He also recommended foreignization when importing into subordinate cultures, justifying this on the basis that "no "culture should be considered immune to self-criticism, whether hegemonic or subordinate, colonizer or colonized" (Venuti 2008 (1995): 19).

However, while this is an acceptable strategy to raise awareness among a complacent intellectual elite reading literature for pleasure or in an academic setting, I do not believe it is appropriate for translation of questionnaires. While it might be interesting to Italians that the Aleut have a category in United States racial classifications, it does not help them choose a category nor does it help their doctors treat them or their public health departments make plans. In scenarios such as a 30-minute follow-up appointment at a doctor's surgery, during which a quality-of-life questionnaire is administered to check a post-op patient's progress, administration of a breastfeeding questionnaire to mothers of infants waiting in a vaccination queue at a health centre or almost any other practical application of a health-related questionnaire, it is hard to see how a foreignizing translation could ever be anything other than time-consuming and confusing.

With relation to questionnaires designed for the general public, I cannot agree that foreignization of a questionnaire, is in any way appropriate for the objectives of cross-cultural adaptation, nor for the intended function of the translation.

On this point, I feel that the functionalists are better qualified to help and Nord's instrumental translation type is exactly what is required. The translated questionnaire must function in the target setting and, while I do not deny that no culture should be immune to self-criticism, it is the source culture that is at risk of criticism if the source-anchoring element of back-translation is abandoned. It is source text authors who will be confronted with the inconsistencies in their texts and it is source text authors who will have to formulate coherent justifications for each item if their translators ask for explanations of their questionnaires' "infrapsychometrics".

Furthermore, Venuti's description of "translation practices that establish a hierarchical relationship between the major and minor languages, between the hegemonic and subordinate cultures" (Venuti 1998: 165) is exactly how back-translation works in the current global scenario. It is the source-linked nature of back-translation that sets up the hierarchical relationship - with source text pre-eminent – since the target text is subjected to evaluation on the basis of a comparison of two texts in the source language. This makes all discrepancies appear to be the "fault" of the translation, which is in the target language, and subordinates the translation to the source text, even though the translation is for use with people who do not speak the source language nor share the source culture. The Brazilian RDC-TMD questionnaire, for example, relegates all of the racial categories that are intended to include Brazilians (with a total population of around 190 million) below a category including an ethnic minority from Alaska (with a total population of around 18 thousand). I must therefore reject Venuti's call for foreignizing translation in subordinate countries, at least with respect to health-related questionnaires

In addition to the comments above on the power relationships revealed by importation and exportation patterns, Lambert also stated that

> – any kind of explicit discourse on the import (translation) phenomenon is likely to be produced on the side of the exporter rather than on the receiving end, at least as long as the moment of decolonization has not started.
>
> (Lambert 2006 [1995], p.98-99)

It would be tempting to claim that this thesis is evidence that the moment of decolonization has started, since it is an explicit examination of an importation translation phenomenon and it has been produced on the target side.

However, as is so often the case, the lines are not so clear-cut as all that. As Lefevere has pointed out, post-colonialism started in ex-colonial societies, not in the ex-colonies. I cannot claim any different, since I was born and educated to degree level in the United Kingdom. However, I have lived and worked in Brazil, for Brazilian clients, for twelve years, I am a recipient of a Brazilian government grant and I will submit this thesis to a Brazilian University, meaning that my patrons are Brazilian.

Furthermore, while Brazil is undeniably post-colonial, it would be difficult to imagine a point at which it could ever be considered

"decolonized". In common with Australia and the United States, Brazil is an ex-colony, but the dominant culture has developed from that of the colonists. In India and many African countries, for example, local indigenous populations did not suffer near eradication and now have political control. In contrast, there is no likelihood that Brazil will be "returned" to the indigenous population. Indeed, the proportions of geographical origins in the genomes of Brazilians tested by Pena et al. and Parra et al., showed an almost perfect balance between African, European and American origins. In other words, it is no longer possible to distinguish the colonisers from the colonised (Pena et al. 2009; Parra et al. 2002: 179-181)

Notwithstanding, Brazil is a post-colonial country and a peripheral one in the terms used by Venuti, although it is rapidly becoming less so. While this thesis does not prove that "the moment of decolonization" has started, it is an explicit attempt to raise awareness here in Brazil of the issues involved in blanket importation, particularly when the process of importation includes a method that anchors translations to the source culture, i.e. back-translation.

I would like to end this subsection by pointing out that the responsibility for cultural colonisation does not rest exclusively with the hegemonic nations exporting their cultural (in this case scientific) products, but also with the members of the importing cultures who adopt those products.

Lambert says that

> what is accepted or refused as 'equivalence' or as 'translation' is basically dependent on cultural (literary, etc.) agreements, not in connection with the source text, but in connection with the receiving cultural (literary) system ...

> (Lambert 2006 [1988]: 51)

In other words, although back-translation is an imported paradigm and although many questionnaires are initiated outside of Brazil, (i.e. they are exported to here, not imported from here), in the final analysis it is their adoption and use by individual Brazilian healthcare professionals and by Brazilian health services, whether public or private, that determines their acceptance.

This, in turn, means that the receiving culture, in this case Brazil, has the necessary tools to change the situation. These tools are the series of constraints and incentives that Lefevere has termed "patronage", and their relationship with back-translation is the subject of the next subsection.

4.3.4   *Patronage and use of back-translation*

If recognition of the fact that there can be no neutral standpoint on culture means it is necessary to specify one's standpoint when writing about culture, then when a writer claims that writing is affected by patronage, that writer should also make explicit their own position within systems of patronage. I shall therefore contextualise my own position with respect to forces of patronage that I perceive have acted on the production of this thesis (it can be assumed that there are also many forms of patronage that those patronised are not aware of).

To put it in Lefevere's terms, I am a professional rewriter and writer, the first in my function as a translator and the second as a doctoral student in receipt of a bursary.

In my capacity as a translator, the effect of economic patronage is to encourage me to take part in back-translation by offering remuneration. However, this is only one element in a highly differentiated range of sources of patronage, since I am in the lucky position of having more clients than time to serve them and the majority of them do not request back-translation. I could choose not to accept back-translation projects, albeit with unpredictable results in terms of my relationships with agencies and those private clients who do request back-translation.

I do not feel pressure to accept or to refuse to engage in back-translation because of status, since I never do forward translations and so it is never my translation being evaluated. However, in my capacity as a translation studies scholar who is also a translator, I am influenced both by my academic reading and by feelings of solidarity towards those whose work is evaluated by back-translation, so I accept that ideological pressures are among the motivations for choosing the subject for this thesis and, of course, the status of holding a doctorate is one of the incentives for writing it, but that does not affect the choice of subject.

However, I had already written the research project summary for this doctorate before I was awarded my bursary and analysis of the project was part of the selection process. It could therefore be considered that I am being economically patronised by the Brazilian government research funding agency CAPES to write this thesis, which is critical of back-translation.

Notwithstanding this disclosure of interests, the type of patronage that I am interested in this subsection is the collection of incentives and constraints that contribute to researchers deciding to employ back-translation.

The results published by Perneger et al. in 1998 directly contradicted the consensus belief in the efficacy and necessity of back-translation and it might be expected that they would have led to a wide-scale re-evaluation of the technique. In fact the article appears to have had little or no impact in this respect. The article is listed on Google Scholar as having been cited 110 times (Google Scholar 2013). This figure can be contrasted with 2807 for "Back-translation for cross-cultural research" (Brislin 1970), 1584 for "Questionnaire wording and translation" (Brislin et al. 1973), 1859 for "Translation and content analysis of oral and written material" (Brislin 1980), and 1490 for "The wording and translation of research instruments" (Brislin 1986) (Google Scholar 2011a; 2012a; 2012b; 2012c). Of course, I acknowledge that Google Scholar only lists a small proportion of all citations and Brislin's work has of course had longer to accumulate citations. However, as a rough guide the numbers are still instructive and if it is considered that Perneger et al. published when the internet was still in its infancy and Brislin's work will have been cited many times in work that was never digitised, the contrast is even more striking.

Furthermore, the AAOS method articles (Guillemin et al. 1993; Beaton et al 2000) have 2195 and 1350 citations respectively (Google Scholar 2011c; 2011d). In other words, while evidence showing that back-translation is not necessary has been ignored, publications recommending it continue to be cited.

Table 4.4 below lists the major publications describing the Sperber, IQOLA, EORTC and AAOS cross-cultural adaptation methods. Additionally, the first reference in the table describes adaptation of the Nottingham Health Profile and, although it did not employ back-translation, all of the publications describing the IQOLA, EORTC and AAOS methods can be traced back to this publication by links of authorship.

Metaphorically speaking, Table 4.4 traces the genealogy of back-translation publications. To extend the metaphor, the family tree has two branches. The IQOLA, EORTC and AAOS branch of the family tree can be traced back to Brislin and Hunt et al. in terms of citations and to Guillemin et al., Aaronson et al. and Bullinger et al. in terms of shared authorship. The Sperber branch is a separate offshoot which has no direct author connections to more recent work.

Table 4.4 – Author and citation links between back-translation publications

Reference	Method described	Lead author	Other authors*	Publications cited
Hunt et al. 1991	Nottingham	Hunt SM	Alonso J	
Guillemin et al. 1993	AAOS (I)	Guillemin F	Bombardier C Beaton DE	Brislin 1970 Brislin et al. 1973 Hunt et al. 1991 Aaronson et al. 1991
Aaronson et al. 1993	EORTC (I)	Aaronson N	Bullinger M Cull A	†
Sperber et al. 1994	Sperber (I)	Sperber A		Brislin et al. 1973 Brislin 1980
Bullinger et al. 1998	IQOLA	Bullinger M	Alonso J Aaronson N	Brislin et al. 1973 Hunt et al 1993 Guillemin et al. 1993
Beaton et al. 2000	AAOS (II)		Beaton DE Bombardier C Guillemin F	Guillemin et al. 1993 Bullinger et al. 1998
Cull et al. 2002	EORTC (II)	Cull A	Aaronson N	‡
Sperber 2004	Sperber (II)	Sperber A		Brislin et al. 1973 Sperber et al. 1994
Koller et al. 2007	EORTC (II)	Koller M	Aaronson N Bottomley A Dewolf L	Brislin 1986 Aaronson et al. 1993 Bullinger et al. 1998 Cull et al. 2002
Dewolf et al. 2009	EORTC (III)	Dewolf L	Koller M Bottomley A	Brislin 1970 Cull et al. 2002 Koller et al. 2007

* Only secondary authors who have links to other publications recommending back-translation are listed.
† Aaronson et al. did not reference any previous publication with relation to back-translation, simply stating that "standard 'forward-backward' translation techniques" were employed (Aaronson et al. 1992: 366)
‡ Cull et al. provide no justification for using back-translation and only list two works in the references, both EORTC module guides.

The genealogical effect discernable in the three related methods can be attributed to the way that the academic system works. When young researchers are choosing research subjects and, later, when they are

choosing research methodologies to investigate these subjects they are often guided by the research history of the academic departments they are at and more specifically by the research interests of their supervisors. If a research area grows, lead authors of early papers, who were graduate students or research assistants, become supervisors themselves and appear further down the list of citations in their students' publications.

However, the table above only demonstrates this effect, which influences the content of research papers (an ideological constraint in Lefevere's terms), with regard to the papers advocating back-translation and while it might help to explain why the results published by Perneger et al. have not had a greater impact, it does not explain why so many other researchers employ back-translation.

I should make it clear that I am not trying to trace the reasons for importing questionnaires via translation; this can be found in the perceived gap described by Toury. In the case of Brazil, the gap perceived was between health indicators, in particular mortality, in Brazil and in the "developed" world, and it was a very real gap. What I am interested in is the reason why so many researchers have specifically used back-translation to accomplish this, despite all of the drawbacks I have described so far.

One reason for this is another systemic factor, this time more explicit. Ozolins points out that medical field now sees this methodology as something of a gold standard" and back-translation is often an obligatory step in fulfilling "clients' needs to meet regulatory requirements" (Ozolins 2009: 2). Grunwald and Goldfarb echo this, stating that for clinical trial submissions "Some Institutional Review Boards (IRBs) require back-translation" (Grunwald & Goldfarb 2005: 2).

Clinical trials are indeed responsible for a great deal of back-translation use. Koller et al. explain that the "primary objective of EORTC QLQ questionnaires is their usage as outcome measures in international cancer clinical trials." (Koller et al. 2007: 1811) and state that "pharmaceutical companies requiring translations are requested to pay for them." (Koller et al. 2007:1813) The 2009 EORTC manual defines the "Requester/Sponsor" of a translation as follows: "Individual or organisation (most likely pharmaceutical company) requesting and usually also financing the translation process." (Dewolf et al. 2009: 7)

It is of course to be expected that a pharmaceutical company that has obtained regulatory approval for a product in one country on the basis of data collected with questionnaire X, would not want the results

of a translation of into language Y to be "worse" (in terms of drug licensing) than the results of the original questionnaire. By controlling the translation process using back-translation and comparisons of pretest results with source culture results, a company (who in this case would be paying for the translation as well as everything else involved in the trial) can be more confident that they will get licensed in the culture that speaks language Y.

Seen from one point of view, this is perfectly reasonable. If the drug is safe, but trial data suggests otherwise because a questionnaire had been worded incorrectly, the drug company would be negatively affected. Seen from a different point of view, however, it is ominous. The point is that different populations do not differ exclusively in terms of language. They also differ in terms of culture, and culture includes acquired habits such as hygiene, diet and even road-safety awareness to name just three that can impact health significantly, plus a myriad of other behaviours and environmental factors that are significant to health. Furthermore, populations also differ in terms of susceptibility to disease.

If back-translation is used in combination with pretest data[25] to make the results of a translated questionnaire conform to those of the source questionnaire it is of course possible that what has been changed is limited to "meaning errors" – i.e. the translated questionnaire did not probe the same concepts as the source questionnaire until corrected after back-translation.

However, it is also possible that the difference actually exists in the population, whether because of cultural or physical reasons, and that the translation that produced results different from the source questionnaire did so because it was tested with a population that was genuinely different. This would then mean that the altered questionnaire, which produced similar results to the source questionnaire, had actually distorted rather than corrected the data. When dealing with data on drug safety and efficacy, this is a possibility that should not be ignored.

I am in no way accusing drug companies of deliberately using back-translation to mask non-uniform results. However, if one considers the ways that the economic and status elements of patronage work within institutions (success is rewarded with pay rises, bonuses, promotions and other incentives) and between institutions (success is rewarded by repeat contracts, increased responsibility, an enhanced

---

[25] The important factor in the use of pre-test data is that it is done during development of the questionnaire and influences the final wording. In contrast with collecting normative data after translation, the target population is judged against source population statistics.

reputation in the market and other positive results of client satisfaction), it is easy to imagine how successive initiators of translation projects would naturally tend to prefer a system that resulted in more uniform results to one that had an awkward habit of detecting differences between populations.

However, clinical trials and drug companies demanding back-translation still do not account for the entirety of the large volume of publications describing its use. To account for the majority, I believe one must look to what Lefevere described as "institutions set up to regulate, if not the writing of literature, then its distribution" (Lefevere 1992: 15).

The majority of medical literature of the type that describes back-translation is written by graduate students, university professors, medical school professors and professional researchers, with a preponderance of the first three categories. This literature is published in a series of scientific journals, some of which are owned by publishers and are profit-making, others are run by universities and medical boards and are not. Invariably they are peer reviewed. The editorial boards and peer reviewers are themselves senior researchers and professors at universities and medical schools.

These people all have careers that are in some way or another affected by patronage dispensed by each other, in the case of access to publication, and by governments and institutions, in the case of teaching and research posts, grants and funding. This is a mixed patronage system, but one in which all elements are interrelated.

In order to gain continued funding, departments must have members who publish. In order to gain posts in departments, researchers and teachers must publish. In this way status and remuneration are both dependent on publication.

Publication, in turn, is controlled by peer review. This is highly conducive to a citation snowball effect in which papers using methods that have been used in previous papers that have achieved publication success are more likely to be accepted for publication. In addition to the simple fact that a method needs to be accepted for publication before it can become widely accepted, there is also the effect of publication bias in itself. For a number of reasons, studies reporting positive results are more likely to be published – both submitted for publication and accepted once submitted.

Following on from this, even when there is no regulatory requirement for back-translation, the sheer weight of prior literature that has used it, has been published and has been cited, is a very strong

incentive for authors of future publications involving translated questionnaires to have their questionnaires back-translated if they wish the articles describing them to be accepted for publication.

Over the course of my career as a translator of scientific articles, particularly medical texts, and inspired by Nord's exhortations to translate with focus on the function or purpose of the target text, I have come to conclusion that the primary purpose of an academic article is not to report on the findings of research, to open intellectual debate with peers or to increase the sum of human knowledge. I do not in any way deny that academic articles are written with these purposes in mind, but, I have become convinced that, in analogy to Vermeer's demonstration that "logically the target-culture conditions are superordinate to source-culture conditions" (Kussmaul 1997: 30), on the basis of temporal precedence, the true primary function of academic articles is not to publicise the findings of research, but to be accepted for publication.

This may seem a radical proposal, but, logically, publication takes precedence over everything else since it is a precondition of publicising findings, of inviting confirmation or rejection and, within the patronage system, of accruing the status of an author who has published and is cited. If an article does not pass peer review and is not accepted for publication, then none of the more laudable objectives that are conventionally considered to be the primary purposes of academic texts can be achieved. The same logic can be applied to books and chapters, since, although in this case it is publishers' editors who make the decision on whether to publish, they base their decisions on authors' academic status and publication histories, both of which are themselves dependent on publication in peer-reviewed publications.

Furthermore, when back-translation is used to prepare instruments that are described in academic articles, I believe that the same logic applies. In this case, the primary purpose of the back-translation step is not to confirm or achieve an accurate translation, but to convince peer reviewers that an accurate translation has been achieved, thereby enhancing the chances of the article itself being published.

When translating articles for my clients I communicate with them by email and also by sending drafts of the translation at various stages with questions about preferred options and intended meanings at relevant points. I have often received comments in return that rather than clarifying an intended meaning ask which option I think has the best chance of passing peer review, whether with reference to complying with authors' instructions, or with a more direct relation to a journal's

academic preferences. I am also fairly frequently contracted to rewrite an article that has been translated by someone else, possibly the author, or written in English by a non-native speaker, specifically in order to pass peer review. Usually such articles have been rejected with comments. In such cases the authors are usually willing to grant what I consider to be a great degree of editorial freedom. They are quite happy for me to alter their structure and often even arguments and conclusions, as long as they believe the changes will lead to acceptance.

I believe that the use of back-translation is now being driven, to a great extent, by researchers' belief that, given the weight of literature apparently supporting the need for back-translation, any article describing a cross-cultural adaptation process that does not employ back-translation is at risk of being declined for publication on the grounds of methodological weakness.

Summing up, patronage therefore affects back-translation as follows:

- Academics need to publish.

- They often select topics suggested by supervisors/more senior colleagues.

- Publishing is controlled by journals, which use peer review.

- Within a given field, peers often hold a consensus view – especially if they are all linked to the founders of the field.

- Remuneration is partly controlled by universities, but university careers and the status of universities are controlled by publication (among other things) and by research.

- Drug companies fund a great deal of research and drug companies are multinational.

- Drug companies need to convince institutional review boards and drug licensing authorities.

- Back-translation is a convenient solution for all of these issues and a natural result of the mixed patronage system that exists.

John Pilger has made a very valid point with reference to academic complicity with governments that is equally applicable here, stating that in order for effects such as these to occur, there need be "no conspiracy, and that should be emphasised. It is simply how the system works,

ensuring 'access' and 'credibility' in an academic hierarchy ..." (Pilger 2002: 163). It is no more than the effect of people doing what is in their interests (because, at the end of the day, companies are groups of people too).

If, however, it could be shown that for researchers in the peripheral countries, patronage could actually accrue more quickly and effectively if, instead of adopting ready-made methods and assertions, such as back-translation, they looked for solutions that were appropriate to their own settings, then patronage could be used to change the situation it currently perpetuates.

Here in Brazil, as I mentioned at the end of the last subsection, the tools to change this situation are to hand. There are a large number of journals in Brazil that are owned by educational institutions or medical associations. There are also two major funding Brazilian government agencies. If the agencies, CAPES and CNPq, began accepting or even preferring adaptation projects that used domesticating strategies or projects that created de novo questionnaires then the number of articles describing such projects would increase.

If journals then began accepting such articles and the resources were provided to demonstrate the effectiveness of such strategies, then the situation in which back-translation is considered a prerequisite for projects involving translation could be changed relatively rapidly.

In addition to these measures, I believe it is necessary to name the translators used when preparing materials for research. One of Venuti's justifications for foreignizing translation is to eliminate the invisibility effect afflicting translators. I am fully in favour of the objective, but I agree with Chesterman when he says that there are often more appropriate places to do this than within translated texts (Chesterman 2005 [1999]: 93). The questionnaire itself does not need to bear intratextual or even metatextual clues to its status as a translation if the translators are clearly named at least in the articles describing it, since, although the people who will be interviewed are one set of end users of the translation, they do not directly control any of the elements of patronage affecting its creation.

Rather than trying to create a foreignized questionnaire, which does not appear advantageous, translators' prestige can be enhanced by giving them credit for their expertise (which, to my view, necessitates abandoning back-translation) and naming them in all publications related to the questionnaires they have translated. Under the current system, initiators often do not even know who has translated their own questionnaires so even if they wanted to they could not name them in

publications or re-contract them for future projects, both of which are routes to professional prestige (assuming the translator is competent).

It is not only the translators who would gain by such a system. Researchers who were familiar with previous translations in their area could request the same translators for future work, thereby creating a pool of ever more specialised translators. Conversely, a translator who repeatedly demonstrated a lack of competence would soon see their work dry up and be forced from the pool by economic imperatives.

In addition to better-adapted questionnaires for the Brazilian population, the Brazilian academic community would therefore also gain a pool of specialist questionnaire translators and developers and the academic prestige (another form of patronage) that currently accrues to researchers living in hegemonic cultures would be accorded to Brazilian researchers.

4.4 Back-translation and translators

I have been a professional translator for 12 years. The greater part of my work is translation of academic papers and the majority of these are submitted to medical journals (or have already been accepted by medical journals). Some of these articles describe cross-cultural adaptation projects involving back-translation, which is how I first became interested in the subject. Additionally, at a rate of about three or four times a year, I also receive requests to back-translate questionnaires as part of cross-cultural adaptation projects, sometimes from clients for whom I have translated academic papers and sometimes via a translation agency.

This subsection is therefore about subjects in which I have a close personal interest, of which I have personal experience and on which I hold personal opinions. It is perhaps unconventional for doctoral students to explicitly refer to themselves and their own subjectivity in their theses, but, as the deconstructionists have demonstrated, any attempt at "scientific impartiality" is futile, since subjectivity is an inevitable element of the human condition. I therefore conclude that not only is use of the first person singular acceptable, and a more intellectually honest approach, but that by acknowledging my own interference as researcher in the results of my research, an opportunity to contribute my own experience and insights to the discussion is opened.

Translation scholars who are interested in translation as interaction between cultures contrast different cultures with one another in terms of oppositions such as coloniser versus colonised, hegemonic

versus subordinate, Western/Northern versus Eastern/Southern, developed versus underdeveloped (or developing), and colonised versus decolonised (or postcolonial).

I live and work in my adopted country, Brazil, whereas, with the exception of some of the authors who discuss translators' expertise, the translation theorists I have drawn on live and work in developed, post-colonial, Western countries.

Brazil is an interesting case, since it does not easily fit into the categories listed above. It was colonised by Portugal, but has been independent since 1822 (Wikipedia 2013a). It cannot, however, be considered decolonised in the way that countries such as India, Zimbabwe, Vietnam and countless others are decolonised, with political power devolved to the descendents of the pre-colonial inhabitants. Rather, Brazil is similar to countries such as the United States and Australia, since it has become independent of its old-world colonial power, but the descendents of the pre-colonial inhabitants are marginalised or subsumed and the dominant culture is a development of the colonial culture.

Furthermore, Brazil is no longer an underdeveloped or even a developing country, but one of the fastest "emerging nations", together with India and China. In 2011 the Brazilian economy briefly surpassed that of the United Kingdom in terms of GDP, making it the sixth largest economy in the world. It is currently (2013) ranked seventh once more (Wikipedia 2013c). Of course this by no means makes Brazil the seventh richest country in the world. The population is very large and there is a historical deficit to overcome.

Brazil's cultural history includes long periods during which translation played the role of filling perceived cultural gaps described by Toury (see page 145 above), but contemporary Brazil exports knowledge as well as importing it, which is the reason that I have been able to pursue a career here translating (mostly) scientific literature exclusively from Portuguese to English.

This is a very different situation from the scenario described by Venuti, for example, in which American and British translators whose target language is English translate for target cultures that are native-English-speaking. Whereas they are attempting to import to (and get paid in) countries whose populations appear to have little interest in anything foreign,[26] I am working in (and getting paid in) a country that,

---

[26] Consider Venuti's verdict that the United Kingdom and United States literary cultures are "aggressively monolingual, unreceptive to the foreign" (Venuti, 2004 [1995], p. 14)

if anything, values the foreign above the domestic, and, at the very least, does not in any way disdain the foreign.

I believe that these factors go some way towards explaining why my experience of the way translators are treated is very different from the picture painted by Venuti and other translations scholars, although there is one very important exception to this trend, which is when I am working on back-translation, as I shall explain in the next subsection.

4.4.1   *Back-translation procedures do not treat translators as professionals; this is corrosive*

In this subsection I shall discuss the way that back-translation procedures negate translators' expertise, have a negative impact on the way they work and create division and mistrust between translators and clients. I shall then explain why treating translators as responsible experts and partners (and rejecting back-translation) would be beneficial both to translators and to researchers. I shall end the section by contrasting the way translators appear to be treated in the United States and Europe with my experience here in Brazil, before arguing that, just as back-translation is an imported paradigm, the mistrust of translation and translators on which it is based is also foreign to Brazil and should be treated in the same way as an unwelcome invasive species.

I have shown that Brislin's classic study was conducted entirely with amateur translators, but by 1986 he was using "translators [who] are treated more like colleagues than hired help" (Brislin, 1986: 148). Nevertheless, the basic assumption that translators cannot be trusted to communicate effectively on someone else's behalf remained since in the same publication he stated that "Assuring clear communication is the researcher's job, and it should not be carelessly delegated to translators." (Brislin, 1986: 145).

This is a very long way from the ideal situation described by Vermeer, who stated that the translator "is made co-responsible for the success of a communicative act, because he, the translation expert, is the crucial factor in it in as far as it is a transcultural act." (Vermeer 1994: 13).

Unfortunately, the situation with back-translation is as described by Hönig, "Translators are rarely seen as experts ... by their clients and other users of translation. Often the only expertise that the public appreciates is that of knowing a foreign language well" (Hönig 1997: 12).

Of course, "merely having a knowledge of two languages is no guarantee that a person can function as a translator" (Nida 1964: 145-146) and as I have shown in subsection 3.4.2 above, translation expertise is very much more than knowing a foreign language well. Indeed translation expertise can only begin to be accrued once someone knows a foreign language well.

Furthermore, since delimitations of meaning are determined by conventionalization of possible meanings in "interpretative communities" (Grigoletto 1992: 95), translators need to be members or become members of the interpretative communities in which their translations will function. This entails specialization, as pointed out by Nida (Nida 1964: 242), or a great deal of research when entering a new subject area.

Even when translating within a known subject area, a certain amount of background information is needed in order that the translator can perform research. Nord has termed these initial instructions as the "translation brief", stating that the comparison "implicitly compares the translator with a barrister who has received the basic information and instructions but is then free (as the responsible expert) to carry out those instructions as they see fit. (Nord 1997a: 30).

Unfortunately, where back-translation is concerned, the situation is very different from this ideal. As we saw in the only account of the preliminary contracting stage of a back-translation project, the company All Graduates that had been approached to translate the MAPT questionnaire specifically advised against using back-translation (Ozolins 2009: 3), but the researchers simply ignored their professional expertise and pressed ahead with back-translation.

One of the translators on that project, clearly frustrated, wrote the following:

> Dear [MAPT team]: I strongly recommend you speak to your translators face to face. What is coming through is lack of meaning you wanted, which is exactly what is to be expected if you do not talk to either the FT or the BT. Suggest that you talk to your FT to explain the SENSE you want to convey, not try to match the words. Word matching will not work.
>
> <div align="right">(Ozolins 2009: 10-11)</div>

The same translator also made the following plea,

> Once again, I would much rather the Melbourne Uni team talk to a translator to explain the flavour they want to get across, and

> this can be done. Using FT/BT/ blind will simply lead to the wrong nuances being imported
>
> (Ozolins 2009: 11)

In contrast with Ozolins' claim that BT is giving translators a voice, which I shall deal with in detail in the next subsection, this translator was complaining that the researchers weren't talking to the translators. This is an almost inevitable result of using back-translation. As the translator went on to point out to the researchers "If you want meaning, you need to explain what you want prospectively." (Ozolins 2009: 11).

Of course, without the strictures imposed by back-translation, translators could elicit the meaning required. Unless, that is, the MAPT team did not actually know what "sense", "meaning", "nuances" or "flavour" they wished to convey, but having hit on a bundle of rituals that seems to work they didn't want to change it.

Back-translation imposes the situation described by Vermeer, in which "people deny the translator the necessary information" (Vermeer 1994: 14) to do the job. He questioned whether doctors or lawyers could "work efficiently if you hide important information from them?" and whether they would even work under such conditions (Vermeer 1994: 14).

Evidently translators do accept such conditions, but the result of this is not simply that they work in the dark in terms of preliminary background information and research possibilities. By accepting the constraints imposed by the back-translation process, they are also accepting a whole series of other consequences.

Returning to Vermeer's comparison of translators with doctors, an experienced doctor does not feel the need to resort to explicit diagnostic criteria, algorithmic flowcharts or medical association consensus statements when faced with commonly recurring day-to-day presentations such as the common cold, tonsillitis or a broken arm. Rather they rely on the internalised knowledge that they have built up through training and experience and proceed directly to diagnosis and prescription. Indeed, any other approach would be prohibitively time-consuming. When, however, an experienced doctor is faced with a more serious or a rarer condition, they will return to the literature and consult with colleagues.

A good translator's mode of operation is comparable. When faced with commonly occurring (in their professional experience) constructions, concepts and idioms a good translator will simply proceed using tried and tested frameworks while adjusting to each text's

idiosyncrasies. When faced with an entire commission or merely a subcomponent of a translation project that presents a novel or particularly difficult challenge, a good translator will also return to the literature, will also consult, (although not necessarily with colleagues, more often with clients and/or authors) and will also generally reacquaint themselves with the state-of-the-art.

While doing this translators do not (normally, in the absence of back-translation) think in terms of equivalence when they are working. For large stretches of time they do not consciously "think" about their translation at all. The way I work fits in with Kaiser-Cooke's description of expert activity.

She stated that "problem-solving activity ... takes on the nature of cognitive routines which do not require reflection but occur 'automatically'" (Kaiser-Cooke 1995: 135). When I translate, a large proportion of my first draft is created semi-automatically. If the subject is familiar to me I will simply "render" the Portuguese text directly into English. This automatic process includes some fairly complex structural rearrangements that have become preferred solutions and which I no longer need to consider consciously. When the solutions do not come automatically, I do not think to myself "what is the equivalent in English", but "what do they mean by X" (when I am in doubt about the author's intended meaning) or "how am I going to say X?" (when an appropriate English construction does not occur to me spontaneously). Back-translation interferes directly with this process because the forward translator has an additional question constantly present, which is "how will X back-translate?", while the back-translator is saddled with its corollary: "I wonder what X is a translation of?"

In other words, back-translation skews the way that translators work, in the same way that rewarding teachers and schools on the basis of their students' test results can skew teaching and make teachers "teach for tests" and in extreme cases even cheat at tests on their students behalf.[27]

This may seem an extreme comparison and, although the possibility that a back-translator may "cheat" by locating the source questionnaire cannot be ruled out, I do not believe that many translators do cheat. However, knowledge that one's translation will be tested by back-translation, in the case of the forward translator, and that one is supposed to be testing another translation, in the case of the back-translator, cannot fail to affect the way they translate.

---

[27] See Robinson 2013 for a case that rocked the United States educational establishment.

It will be remembered that Brislin recommended researchers to "retain or dismiss translators based on a study of the interrelationship among the different criteria of translation adequacy." (Brislin, 1970: 214-5) and Grunwald and Goldfarb said that back-translation could be of use to quality-control personnel at translation companies as a means of "identifying qualified translators" (Grunwald & Goldfarb 2006: 2).

I maintain that it would only identify those who did not yet have language competence and a translation agency should not leave assessment of their most valuable asset to back-translation, particularly since they would still need a method of testing the competence of their back-translators. However, irrespective of the advisability of using back-translation to judge the competence of translators, it is evident that this occurs and so translators cannot ignore the implications of not second-guessing (or being second-guessed).

Ozolins said that "overwhelmingly, the translators engaged in this project were willing to do further translations for the same clients using similar methodologies." (Ozolins 2009: 11) and considered it as proof that they had been convinced by back-translation.

I disagree. I would point out that these translators were working for an agency and could not compromise their professional reputations (status and economic patronage) by rejecting their clients' chosen methodology. That does not mean they were convinced. I still accept back-translation work myself because if I pick and choose the jobs I will accept my agency can't rely on me and they have the same problem with their clients. If a client demands back-translation, an agency cannot only refuse to do back-translation and expect to retain the client for all other types of work. As Ozolins himself showed, the agency argued against using back-translation, but eventually agreed to conduct the project as their clients wished.

Furthermore, it is not only with respect to accepting entire projects that the customer's preferences carry weight. If a translator is requested by their client to make a change to a translation and there is no reason why the change should not be made, translators have no interest in defending the solution that has been rejected, even if it was perfectly valid.

Ozolins describes an exchange between the MAPT researchers and a translator. The researchers had questioned a back-translation of an item asking about the frequency with which pain stopped a patient from sleeping and Ozolins states that the forward translator accepted that the forward translation had asked about the frequency of the pain, rather than the frequency of inability to sleep (Ozolins 2009: 5).

If one considers the logic of the situation, it does not actually make any difference which of the two items the word "frequency" is attached to since "pain stopping a patient from sleeping", logically means that both must be present for a patient to answer in the positive. In other words, if there is only pain, but no insomnia, the answer is never, if there is only insomnia and no pain, the answer is never, if there is ever pain and insomnia and the additional condition that the first causes the second is also met, then the answer is the frequency with which this condition is met, as defined by the response options. It is entirely irrelevant to which item, pain or insomnia, the frequency applies because they are both preconditions of a positive answer – in other words there was no substantive error but the translator admitted fault.

It is probable that while translating "on auto-pilot" the translator had realised this logical irrelevance and so the decision on which item "frequency" should refer to was a matter of style or fluency. However, when challenged, the translator accepted the change rather than argue the case.

Ozolins said that the number of discrepancies originally identified per language was from 11 to 50, with an average of 24, but between 1 and 20 changes, with an average of 5, were actually made to the corresponding forward translations, meaning that between a quarter and a fifth of suspected differences actually led to alterations to forward translations (Ozolins 2009: 9).

Put in the technical language of "translation noise" the 4/5 of "errors" that turned out not to be errors seem like an "occupational hazard" and, if the 1/5 that are changed really are errors, a price worth paying, in terms of the extra work, for the confidence it affords. However, seen in the light of translators' status, this means that, even assuming all the items changed were actually errors, the "best" translator who only had one item changed endured something from ten to forty-nine unwarranted requests for clarification and the "average" translator, who was asked about twenty-four differences to arrive at five changes, endured nineteen pointless questions.

I have already referred to problem recognition as a feature of translation competence and I believe that an expert translator working unfettered and with access to communication with the initiators would identify all these "errors" before they happened or during revision and discuss them with the researchers. I also believe that many of these "errors" are accepted in a spirit of compromise anyway, since translators

are empathetic by nature – we are professional intermediaries, we work in a service industry and it is natural that we aim to please.

However, the true price of "back-translation noise" is that it undermines the translators, it grinds them down and demands explanations for semi-automatic processes that are not consciously monitored and so are difficult to justify and, as such, it is another part of the self-fulfilling prophecy that is the back-translation paradigm.

I shall try to elucidate. "Back-translation noise" is not random interference as the analogy suggests, it is naive insistence on "apparent linguistic discrepancies [that] are perfectly familiar to translators" (Ozolins 2009: 6). It is well-intentioned, but misguided (by back-translation) and the ultimate effect is harmful. It is harmful to translators' self-respect and self-image and it generates resentment and frustration, as Ozolins puts it, "the feeling of being 'ambushed'" (Ozolins 2009: 10).

This is bad enough, but it is most harmful because it confirms the users of back-translation in their belief of its utility. Every time a translator accepts a change in the interests of harmony, the researcher is provided with apparent confirmation of the effectiveness of back-translation. This is one of the mechanisms through which back-translation has propagated and thrived, despite its utility never having been proven, as I showed at the start of this chapter.

The effect of back-translation is to legitimise mistrust of translators by incorporating it into a standardised protocol which rarely fails to detect discrepancies that can be held up as justification for mistrusting the translators in the first place. Since the translators know that they are going to have their work judged (by people who are not experts in intercultural communication), back-translation also institutionalises the defensive translation strategy identified by Hönig (Hönig 1997: 17).

This attitude to translators and translations is a direct consequence of the illusion that translation can be evaluated objectively, which in turn is a consequence of the misconception of translations as symmetrical replicas of source texts. Hönig describes it as "destructive" and says it "relegates translators to the status of a walking dictionary without any true translatory competence" (Hönig 1997: 17).

I would call the effect "corrosive" because eats away not just at the translator's self-confidence, but also at the client-translator trust-based relationship. The translator is led to think "If they trust me so little for this questionnaire, do they also treat my other work with this degree of suspicion?" The client, who sees lists of discrepancies identified

between source text and back-translation, but cannot check them in the forward translation, quite naturally wonders "Does this translator make a similar number of mistakes in other work?"

This is because discrepancies are considered synonymous with "meaning errors". The only insight into this process available has once more been provided by Ozolins. His translators had to answer such naive queries that it is not surprising that the effect is to sound sarcastic or irritated. Consider what must have led to the following comments:

> Croatian does not use the word any. This is only in the back translation.

> In the [forward translation] the word 'some' is not used as it is not idiomatic and therefore in back translating it is not present. Not meaning any disrespect but if 'there are things that one cannot do' does that not imply that they are some and not all?

> In this particular case the logic is that my hip can cause difficulties only with my relationships and not with somebody else's

> ... translation process is more than replacing words with the words in the other language.

> ... often the same meaning needs to be expressed using different words, which back-translation will not reflect.

> (Ozolins 2009: 7; 10).

Translators are used to working to make a text that is intelligible and functional in the target language, but back-translation is constantly dragging them back to the source language, including in terms of the language they must use to discuss their translations. Irrespective of whether translators are working into or out of the original source language, they must always express themselves in that language when communicating with their employers because their employers don't understand the target language, which is precisely why they have employed translators.

This is the normal state of affairs between translators and their clients. However, when said clients are using the source language to query or simply reject a target language solution, which they have never read and could not understand if they tried, the source language, in which the translator will have to justify themselves, often does not have the mechanisms to express the target language meaning, which is exactly why back-translation detected a discrepancy, but not necessarily a reason why the target language solution was "incorrect".

Put simply, the entire back-translation process undermines translators on many different levels. First, the attempt to "test" the work of one translator using the work of another undermines trust between client and translator. For translators, the fact that the client considers it necessary to test their work indicates that they are not trusted and means that they are likely to translate defensively. As Hönig points out, "How can any translator trust his associations and linguistic reflexes if he knows that his client does not trust him?" (Hönig 1997: 17).

For clients, the fact that back-translation is a self-fulfilling prophecy means that it will invariably detect discrepancies, leading them to doubt the competence of their translators, even if they had only requested back-translation in the first place because that is what is recommended in the literature of their subject area.

The fact that language is flexible and robust, combined with the fact that there is a service relationship between client and translator, encouraging the translator to agree with the client whenever possible, means that some of the suggestions for overcoming "discrepancies" will be accepted by the translator, further reinforcing the impression that the original translation was deficient. Since wording that is rejected during back-translation is never tested in validation stages, there is no way to show whether the original wording would have performed as well as (or better than) the wording actually selected, further reinforcing the impression of back-translation's efficacy and further eroding trust in translators.

The fact that back-translation precludes communication between the translators and the initiators with relation to the text of the questionnaire to be translated means that the opportunity for translators to demonstrate their sensitivity to the needs of questionnaire translation and to the needs of the specific questionnaire being translated is lost. They must wait until back-translation identifies problems and then, if they are ever consulted again, *react* to them, rather than *proactively* seeking solutions in cooperation with their clients during (or in advance of) translation.

All of these effects are what I would call "corrosive", since they do not destroy the relationship between translators and their clients in a single catastrophic event, rather, they gradually erode away trust and mutual professional respect as translators feel their expertise is being repeatedly questioned and disrespected and clients feel that they are being confronted with sufficient examples of discrepancies between source text and back-translation to suggest that the translators are less than fully competent and even, after repeated use, that all translators

lack the competent to successfully translate such delicately balanced instruments.

Nord suggested that gaining clients' confidence would, "strengthen the translator's social prestige as a responsible and trustworthy partner." (Nord 1997a: 125). In contrast, because clients think that back-translation is the reason that their translations work, when they do, they will never build up trust in the translators and instead will come to see translators as interchangeable tools, like a thermometer or a tape measure.

In fact, translators are humans and have different skills and aptitudes. In the same way that the researchers themselves gain reputations among their peers on the strength of their work, the best way to develop a pool of respected and competent questionnaire translators would be by giving them both the responsibility and freedom of action expected by any professional expert. In other words, translators should be encouraged to take control of the translation phase of cross-cultural adaptation, not in order to impose their own vision on the translations, but in order to have the liberty necessary to produce translations that function exactly as their clients wish them to function.

I believe that for this to happen, the translators of questionnaires must be named in the publications describing the translation process. In this way they would be "made co-responsible for the success of a communicative act" (Vermeer 1994: 13). Currently this never happens.

In fact, as shown by Ruuskanen's comment that "when the x-ray technicians and the lab assistants are thanked, the translators ... are often left out of the acknowledgements" (Ruuskanen 1995: 299), this often applies to the translators of medical articles too, not just to the translators of questionnaires.

The forward and backward translators of medical instruments are often described, as in Sperber et al. 1994 and Perneger et al. 1999. However, they are never named. This is possibly related to the conception that cross-cultural adaptation should model itself on double-blind trials and/or to the belief that translators are interchangeable, which is itself a result of the assumption that there is only one correct translation for any given text, hence all "competent" translators will produce the same translation, as manifest in the complaint, recorded by Lambert, that translators are treated like fax machines (Lambert 1996: 285).

Notwithstanding, even if translators are seen in the same light as laboratory equipment, researchers generally feel obliged to name the manufacturers of their experimental equipment, in the interests of

reproducibility, and to acknowledge everybody involved in preparation of their manuscripts, including professional writers, editors, proof-readers and even translators, in the interests of intellectual honesty. It seems to me to be a strange anomaly that researchers do not name the producers of their textual research instruments in the same way that they name the producers of their laboratory instruments.

This is a matter of status for translators and the current situation means that it is impossible to build a reputation as a competent translator of questionnaires. However, it is also a less-than-ideal situation for the researchers, although they appear to currently be unaware of this. By omitting the names of the translators, the normal process of peer approval is sidestepped. Rather than a situation in which Questionnaires A, B, C and D are known to have been translated by Translators W, X, Y and Z, meaning that Researcher E wishing to translate a questionnaire that is similar to, for example Questionnaire B, could examine it and any published results and either choose or reject Translator X on the basis of what they found, Researcher E can copy the methods used, but cannot use the same materials.

If translators of questionnaires were named, not only would researchers be able to review examples of a translator's previous work, after a certain period they would also be able to identify translators with experience in their areas of specialization.

If, however, the translator is seen as an experimental subject who should be protected by anonymity, then the back-translation process is itself experimental, but the literature on back-translation considers it to be a tried and tested method and, with the exception of Brislin's 1970 paper and Perneger et al.'s serendipitous experiment, none of it claims to be testing back-translation. Rather, back-translation is used to prepare instruments that will be used to investigate a wide range of different subjects. The translators are not, therefore, experimental subjects who must be accorded confidentiality. They are (or should be) paid professionals who take responsibility for their work, in return for the chance to build reputations as competent expert translators. With the exception of volunteering experimental subjects, I do not believe that the "invisibility" of anyone involved in a scientific endeavour is healthy for science. Anonymity during peer review is not the same as invisibility, since once accepted for publications the authors' names are once more appended to the manuscript.

Of course the authors of validation studies of translated questionnaires do not deny or ignore the existence of translators (although their existence must very often be deduced from a passive

phrase such as "the questionnaire was translated and then back-translated"). They do not, however, ever provide the translators' names. Normally they just say that the questionnaire has been translated and the translation tested using back-translation.

From the perspective of the scientific method, this is not enough information for at least three reasons. Firstly, back-translation is not a uniform procedure and even if "as per Brislin (1970)" is added, there were many different back-translation methods described in that article. The MAPT team, for example, claimed to be using the EORTC method, but, as I showed on page 119 above, the differences between the two methods outweigh the similarities.

Secondly, the translators used are at least as important as the technique. Nida, Arrojo, Nord, Hermans, Venuti and many others all agree that the receiver creates the meaning of the text. In the case of translation, they create the meaning of the source text and then formulate a new text for other receivers to use to create meaning. Their cultural backgrounds, native and acquired languages and degree of translation expertise are all of fundamental importance to the translations they will produce.

Thirdly, as mentioned, if translators were named then peer approval (or rejection) of the resulting questionnaires could be used to guide future contracting decisions and translators' prior experience with questionnaires would be easily verifiable. After all, how can a translator gain a reputation as a "responsible and trustworthy partner" (Nord 1997a: 125) if nobody knows their name nor what they have translated? I shall return to this point in the final section of this chapter, in which I address the issue of how to fill the gap left by back-translation if it is rejected.

However, I would like to end this subsection by contrasting the corrosive back-translation process with my own experiences when translating without the constraints of back-translation and asserting that the mistrust of translation and translators identified by Venuti in Anglo-American cultures and by Lambert and Hönig in non-native English speaking European cultures is by no means the rule here in Brazil.

I do not think it is an exaggeration to state that back-translation imposes a suspicion of translation and translators that is not natural to Brazil. I do not feel mistreated or mistrusted. I consider myself well paid considering the local economic scenario and I have only ever had three late paying clients in 12 years, one of which was in Portugal. I have never had a client who did not pay eventually. Clients rarely challenge my translation solutions, probably because I work in close cooperation

with my clients during the translation process, but when this has happened it has invariably been when I have accepted work in areas in which I was inexperienced.

Furthermore, I translate into English from Portuguese for well-educated clients. Almost invariably my clients can read English and are competent to judge for themselves, for example, which of the many options that I often offer during this communication process is best suited to their purposes.

In addition to this I frequently receive statements of praise and gratitude from my clients. They often ask permission to add my name to the list of authors of their work, which I refuse, requesting instead that they list me as the translator, if possible with the author information, but if necessary in the acknowledgments section. I communicate with my clients throughout their translation projects, defining their needs and researching their areas of knowledge before accepting a contract. I will often read a large proportion of the work listed in their references and I do not start translating until I believe I have understood the concepts involved. Notwithstanding, I respect my clients' expertise in their own areas and make it clear that, after discussion and explanation of possible choices, their decision will always be final. In turn, they respect my expertise in my profession. The result of all this is that I enjoy my work and my clients return.

Back-translation is an insurmountable barrier to these and many more translation best practices and, having restricted the translators' working methods and communication channels, it adds insult to injury by triumphantly holding up "errors" that are entirely the result of being treated like an experimental subject rather than a competent professional.

On an abstract level, I do not believe that back-translation is the ideal solution in any situation. On a concrete level, here in Brazil, in the market of which I have experience, back-translation is much worse than a less-than ideal solution. It is a threat to the harmonious relationship between translators and their clients that currently exists and which, judging by what Nord, Hönig, Vermeer, Lambert and Venuti have to say on the subject of the translator's lot, is an enviably rare state of affairs and one that it is well worth making an effort to preserve.

Translation may no longer occupy an entirely "primary" position as defined by Even-Zohar, in Brazil, but it is by no means "stigmatized as a form of writing" (Venuti 1998: 1). With relation to medicine, and to epidemiology and public health in particular, there is no doubt that translated literature is used to fill gaps, as Toury puts it (Toury 1995a:

138) and to a certain extent, imported things are still often of higher status than their domestic equivalents in Brazil, indicating a degree of the cultural insecurity described by Nida (Nida 1964: 173).

It is possible that this explains why back-translation is often used here even when the translation project is initiated in Brazil and is conducted by people who speak both the target language and English (and do not therefore need back-translation). The underlying assumption would be that the foreign instrument is an ideal model to be imported. In fact, as the IQOLA team tacitly acknowledged by rewriting the SF-36 after experience gained in translation, the source text can often be improved and translation is an ideal opportunity for incorporating improvements.

Notwithstanding, the questionnaires that are imported are undeniably useful. In Toury's terms, there is still a gap to be filled. On a public health level, the data that is collected with them is used to plan healthcare interventions and to target spending and on the level of clinical care and research applications they are used for diagnosis and follow-up in all specialties.

However, by importing the back-translation technique together with the questionnaires, Brazilian researchers are undervaluing their own expertise, their native language and the expertise of Brazilian translators and are foregoing the opportunity for improvement that translation offers.

They are also importing the attitudes on which back-translation is founded. As Venuti puts it, in the Anglophone tradition,

> ... translation is defined as a second-order representation: only the foreign text can be original, an authentic copy, true to the author's personality or intention, whereas the translation is derivative, fake, potentially a false copy.
>
> (Venuti 2004 [1995]: 7)

It is this suspicion of the "false copy" that spawned back-translation and continues to feed it.

In addition to the negative impact it has on inter-professional respect, using back-translation in Brazil for instruments that will be administered here is paradoxical because it restricts evaluation of the text that will actually be used, and which is written in the native language of the culture in which it will be used (Brazilian Portuguese), to the language of the source text (English).

This is a throwback to the monolingual export model in which the people evaluating the translations do not understand the target language

of translation, because they come from societies that are "aggressively monolingual" (Venuti 2004 [1995]: 14)

In contrast, not only are the initiators here in Brazil themselves Brazilians and therefore native speakers of the language of the forward translation, they also invariably have a good working command of English, at least for reading within their own subject areas, since it is almost impossible to pursue an academic or professional career in Brazil in a medical subject (or indeed any science) without being able to read in English and it is very likely that they will have learned of the existence of the questionnaire they have decided to translate by reading an article in English.

Back-translation is a technique developed for use when researchers don't speak the target language and it imposes monolingual evaluation on translation, which is by its very nature at least bilingual. It is therefore entirely inappropriate to adopt back-translation in a scenario in which the researchers not only speak the target language, but are also capable of accessing the questionnaire in its source language too.

Furthermore, and more worryingly in view of the points I have made about the corrosive effect of back-translation on the translator-client relationship and on translators' confidence and sense of responsibility, there is a real danger of back-translation being adopted for uses other than questionnaire adaptation, particularly for uses in which Portuguese is the source language rather than the target language. This is a very troubling prospect indeed and one which I believe the translation community, both those who translate and those who theorise about translation, have an interest in taking a proactive stand against.

The best way of ensuring that back-translation does not spread to other forms of translation is to eradicate it within questionnaire translation and in the last section of this chapter I shall consider the consequences of rejecting back-translation, before demonstrating two possible alternative methods in the next two chapters.

However, before discussing the gap left if back-translation is abandoned, there is one last argument in favour of back-translation that I have yet to address. This is Ozolins' suggestion that, while back-translation may not be the most effective tool for evaluation of translation quality, it does at least provide an opportunity for translators to communicate with their clients. I present my reasons for rejecting this argument in the next subsection.

4.4.2   Back-translation as a means of giving translators a voice?

As I acknowledged in subsection 4.1.1 on proof of the utility of back-translation, there is one argument in favour of using back-translation that I did not address in that subsection. This is primarily because this argument has not been advanced in any of the back-translation literature proper and is not related to the question of whether back-translation does what Brislin, Sperber et al., Guillemin et al., Beaton et al. and Bullinger et al. claim it does. Rather, it is related to the question of translators' status and clients' trust in translations.

Of course, the argument in favour of back-translation put forward by Ozolins cannot be ignored, especially not in a thesis that advocates raising the status of translators, so I shall deal with it here. To recap, Ozolins accepted that there were a number of arguments against using back-translation and, although he did maintain its usefulness for error checking, his main argument in favour of back-translation was that, as long as the process is "based on transparent communication between translators and authors", then it can be a means of "enabling translators to have their voice heard by clients." (Ozolins 2009: 1).

Ozolins had defined the types of false positive errors that I have described in subsection 4.2.3 above as "back translation noise", admitting that from four fifths to three quarters of "discrepancies" did not lead to changes in the forward translation. He accepted that a "critic of back translation would ... have been able to draw on many occasions in this project where the back-translation project led to wrong identification of discrepancies" and lists six types of back-translation noise, which he considers to be "linguistic issues". These were apparent discrepancies in usage of singular or plural, apparent discrepancies in use of tenses, capitalisation and contractions and a tendency for "hip or knee pain" to be back-translated as "pain in the knee or hip" (Ozolins 2009: 6).

Ozolins pointed out that these "apparent linguistic discrepancies are perfectly familiar to translators (and bilinguals in general) but often opaque to monolinguals" (Ozolins 2009: 6) and claimed that the true utility of back-translation was that it "ultimately gave the MAPT team confidence in the translations." (Ozolins 2009: 10).

This idea that back-translation can be a trigger for communication between translators and clients is undeniably seductive. However, there are a number of reasons why I remain unconvinced that it is sufficient justification for continuing to use back-translation.

My first point in reaction to Ozolins' argument that back-translation can be a means of "giving translators a voice" may seem petulant, but it is undeniable:

Translators already have voices.

We are professional writers, we get paid for creating texts. In short, we are specialists in communication and we do not need to be given a voice. In fact, the assertion can and should be extended further:

Translators already have <u>two</u> voices (at least).

The problem is not that translators do not have voices and need to be coaxed out of our shyness. The problem is actually that the researchers who employ us to conduct back-translation often do not understand (at least) one of our voices. Furthermore, the voice they do not understand is precisely the one that speaks the target language of the translation they want. In other words, the voice they are paying us to use on their behalf.

Furthermore, Ozolins stated that the process must be "based on transparent communication between translators and authors" (Ozolins 2009: 1), but this is a manifestation of the same naive view of translation as "transparent" that led to the corrosive back-translation paradigm in the first place. In fact there can be no "transparent" communication in English about a translation in a different language. There are at least two reasons for this. The first is that in order to express a meaning in language X one must use language X, not language Y because in language Y one can only express language Y meanings. The second reason is that the questions that initiate each bout of communication are asked in the erroneous belief that the back-translation is a "transparent" window onto the forward translation (and of course, if it were there would be no need for back-translation, since the forward translation would also be a transparent window onto the source text).

Ozolins claimed that the "hands-on engagement in the translation process ... ultimately gave the MAPT team confidence in the translations" (Ozolins 2009: 10). Without going back over the cost of this confidence in terms of the man hours expended or, more importantly, translators' professional status undermined, it is pertinent to state that this engagement is by no means "hands on", since the fact that the researchers cannot understand the forward translations means

they can only ever have confidence in translators' descriptions of forward translations and of the solutions employed in them. In other words this engagement is entirely mediated by the translators.

I do not deny that the researchers need to have confidence in the translations they commission, nor that they are perfectly justified in wishing for reassurance. However, as I shall discuss in the final section, I believe that this can be achieved without antagonising, undermining or devaluing translators and without using back-translation.

Notwithstanding, if researchers (or translation agency staff) find themselves in a position in which they cannot find a qualified bilingual to assess translations or their clients demand back-translation and cannot be dissuaded, as was the case with the MAPT team, then, in order to achieve Ozolins' highly laudable objective of encouraging communication between clients and translators, I propose that the back-translation should be a documentary translation, as defined by Nord.

She states that

> ... an instrumental translation is legitimate only if the intention of the sender or author is not directed exclusively at [a source culture] audience but can also be transferred to [target culture] receivers, so that the information offer of the [target text] is included in the information offer of the [source text]. If this is not the case, the translation must be realized in document function ...
>
> (Nord 2005, [1991] p. 81)

Since the forward translator *should* be culturally adapting the questionnaire to the target culture population, the original source culture audience is *not* included in the offer of information of the forward translation and neither is the back-translator. To fulfil Ozolins' purpose, the back-translation, should therefore be a documentary translation illustrating how the forward translator has accomplished the cross-cultural adaptation, and not an attempt to produce a functional questionnaire for the original source culture, even less to try to *re*produce an existing source-text questionnaire blind.

In a documentary back-translation, the back-translator would not translate blind, but would translate with reference to both source text and forward translation. The evaluation phase would not therefore consist of comparing two instrumental translations, but of evaluating the implications of the information provided in the documentary back-translation.

This approach would make more sense than pretending that the back-translation is an instrument that will one day be administered, that

has a target audience and that must function in the target culture (which is of course the source culture of the original questionnaire), while blinding the translator, thereby curtailing their range of translation techniques.

Furthermore, the monolingual researchers would not be presented with a back-translation and a source text that must somehow overcome the unidirectional nature of translation and be identical, despite the intervening steps. Rather, they would receive something between a description of the forward translation and a report on it, which they could then discuss with both translators, having sidestepped the phase in which it appears that every discrepancy is an accusation of translation failure.

As a native speaker of English living in Brazil and working for Brazilians, I am never called upon to do the forward translation, only ever the back-translation. On several occasions I have felt myself subject to ethical dilemmas when back-translating. As many back-translation advocates recognise, it is often easy to work out what the source text form of words was (without resorting to breaking the stricture to translate blind). However, it is often the case that I would not choose that form of words myself if I were producing an instrumental translation. Should I therefore "help" the forward translator by selecting what I believe to have been her or his source? If I could produce a documentary translation I would be at liberty to explain the forward translator's solution and analyze other options.

Even so, this method would still entail a certain degree of self-justification on the part of the forward translator, in cases where the researchers come to the conclusion that the forward translation does not say what they wish it to say, but rather than forward and backward translators being set against each other, the back-translator's role would be to explain the forward translator's decisions and methods to the monolingual researchers. A type of interpretation, explicitly acknowledged as interpretation and without "translator's invisibility" to lure researchers into the false sense of security engendered by the erroneous belief that an instrumental, fluent, back-translation provides a "transparent window" onto the forward translation.

Such a process would thus be a method for client education, leading to increased understanding of the complex ways in which translation achieves its objectives. It would hopefully raise awareness, to the extent that researchers who had been exposed to such a process would achieve the necessary degree of trust in their forward translators that in future projects they would have the confidence to dispense with

even a documentary back-translation phase (and would also provide them with the arguments needed to convince their peers that such a decision was justified).

Notwithstanding, in a target setting such as Brazil, and particularly when the cross-cultural adaptation process has been initiated in Brazil, I see no need for any kind of a back-translation element, not even an "enlightened" back-translation step such as a "documentary" back-translation performed "unblinded". Here in Brazil, academic researchers invariably read English and many speak it well too. Since the target language is Brazilian Portuguese, the forward translator(s) can communicate with their clients in the same language that they are translating into and when their clients understand English they can also provide explanations and examples in the source language. This is the way I conduct all other types of translation for my academic clients. It is only when back-translation is used that communication is cut off.

Ozolins' suggestion is indeed seductive when compared with the more mechanistic incarnations of back-translation, but, in the final analysis, it is only necessary to give translators their voice back because they have been muted by the back-translation process. Removing blinding and approaching translation as a collaborative effort involving a great deal of communication between translator and initiator is an incomparably better way of encouraging clients and translators to talk to each other.

4.5 Conclusions and discussion

Andrew Chesterman says that "if a tool does not serve the function for which it was designed, or any other function, we can get rid of it."(Chesterman 2005 [1999]: 96)

I believe that I have shown that back-translation does not serve the purpose for which it is recommended in the cross-cultural adaptation literature, nor Ozolins' substitute purpose of giving translators a voice. While I have not shown that it does not serve any other function, I believe I have shown that the functions it does serve are (i) unnecessary; in the case of testing for language competence, when expert translators are available, (ii) counterproductive; in the case of its effect on the translation process, and (iii) corrosive; in the case of its effects on translators' status, self-confidence and professional reputations and relationships.

In view of everything that I have argued in the preceding sections of this chapter, I therefore believe that the inevitable conclusion is that

back-translation should be eliminated from cross-cultural adaptation processes whenever specialist expert translators are available to translate from the source to the target language and particularly when the project is initiated within the target culture.

However, to do so leaves a gap that must be filled to the satisfaction of both translators and those researchers who currently put their faith in, and trust their professional reputations to, back-translation.

Put simply, rejection of back-translation poses the question of how the translation stage of cross-cultural adaptation should be conducted, if not with back-translation. When a researcher or research team wishes to take part in an international collaboration involving questionnaires or needs a questionnaire for national research and knows of a suitable questionnaire in another language, there is still a need for questionnaire translation.

Additionally, researchers are unavoidably bound up in the systems of patronage that I described in subsection 4.3.4 above and their need to pass peer review in order to publish is very real. It is one of the most important factors deciding career advancement and academic status.

As Arrojo points out, acceptance within any given cultural community is dependent on meeting its demands and conforming to them (Arrojo 2003 [1992f]: 109) and publication is one of the demands of the academic community.

The solution must therefore be convincing if it is to persuade researchers that they will not be thwarting their own career aspirations by trusting translators without policing them through back-translation. Brislin said that "assuring clear communication is the researcher's job" (Brislin, 1986: 145) and he was correct about the responsibility, just not correct that clear communication can be ensured by back-translation.

Since Brislin developed the back-translation technique, the world has changed greatly. In particular, translation practice, translators and the tools they have at their disposal have all changed. The "peripheral" countries have also developed greatly. Furthermore, the way language and translation are viewed has changed. It is now understood that there is no ideal or perfect translation, because every decision precludes others. There are only translations that are better suited to certain functions and contexts than others. This is not, however, a limitation that is specific to translation. It is true of all writing.

As soon as a writer chooses the first word of a text, however short, they have ruled out the possibility of starting with any of the other words in the language they are using. This then has consequences for

every subsequent word they choose, not only because of grammatical considerations, but also because of the more powerful and cumulative consideration of coherence with what has gone before.

Once the inherent subjectivity not only of translation, but of reading, listening, writing and speaking in general, are acknowledged, it becomes clear that back-translation is an attempt to make translation conform to a standard that no other use of language is held to.

The only solution is to accept the subjectivity inherent in all language use and trust translators. Or, more completely, find a mechanism through which trustworthy translators can be identified and allowed to exercise their professional expertise.

The first part of the solution is the more difficult – identifying trustworthy and competent translators and this can only be done in the same way that other trustworthy professionals are identified, which is by reputation, curriculum and, basically, trial and error. The process of identifying a pool of trustworthy translators must involve a type of virtuous circle by which translators who have proven competent can be contracted for similar work, which implies making their names known.

Furthermore, by advising translators that their names will be published during the contract negotiations, translators who know they are not really qualified will be deterred from accepting the work offered. This is in stark contrast with the back-translation technique, which provides a cloak of anonymity and, on one hand, offers the safety net of back-translation to less than competent forward translators, while, on the other hand, gifts poor back-translators with the knowledge that their translations will never be used as data collection instruments.

It is not only (competent) translators who would gain by such a system. Researchers who were familiar with previous translations in their area could request the same translators for future work, thereby creating a pool of ever more specialised translators. Conversely, a translator who demonstrated a lack of competence would soon see their work dry up and be forced from the pool by economic imperatives.

As I have mentioned repeatedly, translators are not interchangeable tools, but human beings and every one is different. The key to achieving a suitable translation, from the perspective of the initiator, is to choose the right translator for the translation in question.

Nida pointed out that translators are motivated by a large number of factors, one of which is "a sincere humanitarian purpose, namely to convey an important message in an intelligible form", giving the example of "many present-day translators who strive to translate scientific texts" (Nida 1964: 145).

A professional translator of medical texts will understand the importance of the questionnaire they are being asked to translate and treat it accordingly, if allowed the freedom to do so. A (good) professional translator can be trusted to translate in the best interests of their clients, but they cannot be expected to work blind and they must be able to request collaboration from their clients.

In the next two chapters I shall demonstrate two proposals for the translation stage of cross-cultural adaptation of questionnaires that do not employ back-translation, but rely on expert translators communicating. They are intended to "slot in" to existing procedures before the expert committee review stage, at which both translators should also be present and at which solutions to cultural incompatibilities should be discussed.

I am of course aware that starting off a virtuous circle is dependent upon acceptance of these methods in the medical research community. As Ozolins showed, attempts at client education were ineffective with the MAPT team. The translation agency advised against using back-translation but none of their arguments convinced the researchers, who wanted "a methodology that would fit in with international EORTC precedents and the demands of peer review." (Ozolins 2009: 4).

The sticking point is therefore the extant back-translation literature which has not provided proof of back-translation alone but has shown that cross-cultural adaptation processes, with all their steps, work to the standard demanded by researchers in the area.

Brazil, however, offers an opportunity to begin the virtuous circle by starting with importation projects initiated here. Many dispensers of patronage, including scientific journals, are controlled by governments,[28] universities and professional associations. Furthermore, the Brazilian research community has high levels of English reading competence.

If the Brazilian research funding agencies CAPES and CNPq were to provide financial support for cross-cultural adaptation projects that were initiated here by Brazilian researchers, and those studies did not use back-translation, but did publish translators' names and both source and target questionnaires, then the Brazilian research community that is able to read both source and target text would then be able to judge the results for themselves.

---

[28] Brazil has a three-tier governmental system, with Federal, State and Municipal governements.

In the next two chapters I shall present and discuss some proposals for a translation phase for cross-cultural adaptation processes based on the conviction that by employing experienced and talented translators and encouraging them to cooperate in the production of a functioning translation, using all the tools and resources they are usually able to count on, a forward translation can be produced that is ready for the committee stage without the need for back-translation.

The first method is a parallel, cooperative, consensus-based method and I shall demonstrate it using two expert translators, both with well over ten years' experience. The second method is a streamlined version of the first, incorporating small modifications in response to the results of the first method. The second method is demonstrated using two translators sourced from a translation agency previously unknown to me and selected from the first five results returned by a Google search on the basis of their responses to my enquiries. This was done in order to gauge whether a researcher without my privileged access to translation professionals would be able to contract competent translators.

The methods, the questionnaire translated to demonstrate them, the results of those translations and my conclusions and reflections about them are all presented in the next two chapters.

5 THE PARALLEL, COOPERATIVE, CONSENSUS-BASED TRANSLATION METHOD FOR QUESTIONNAIRES

As an alternative to the translation phase of current cultural adaptation methods that employ back-translation, I propose a parallel, cooperative and consensus-based method which does not involve back-translation, but instead relies on the competence and experience of two expert translators.

In this chapter I shall describe the underlying principles on which the method is based, the prerequisites for using the method and the actual steps involved before presenting a demonstration of the method, comparing it with back-translation and discussing the results.

5.1 Basic principles of the Parallel, Cooperative, Consensus-based Translation Method for Questionnaires

This method is based on a series of basic principles that, taken together, define the form that the method takes. The justifications for these principles have all been discussed earlier in the thesis (mostly in the previous chapter), but I shall summarise the principles themselves here.

- A minimum of two **expert** forward translators must be contracted. At least 7 – 10 years' experience or more than one million words of material translated is recommended. It is also recommended that researchers who do not already have a working relationship with such translators use a reputable translation agency to source them.

- If such translators are unavailable, the method **cannot** be used, since it is based on the assumption that **expert** translators working cooperatively are capable of producing robust, functional translations and of detecting and eliminating any problems that exist in initial drafts of their forward translations with reference to the original text only, in addition to improving the wording of translation choices over a series of revisions.

- Translators and agencies should be named in publications describing the translation process, in the same way that the suppliers and manufacturers of laboratory equipment and materials are named; in the interests of reproducibility and in order to allow the mechanisms of peer approval and professional reputation to function in the same way as in other professions.

- Translators should translate into their native languages and prepare instruments for use in their native or adopted cultures (e.g. a Brazilian living in Brazil would not be contracted to translate a questionnaire for use in Portugal, but a Brazilian living in Portugal could be).

- Translation projects should be coordinated in the target culture (even when initiated in the source culture).

- Translators should be provided with a Translation Brief (see Appendix B for an example) explaining the purpose of the questionnaire they will be translating and providing (as a minimum) a reference for the article describing the original validation study for the source-language version of the scale, in addition to covering the objectives and procedures of the Parallel, Cooperative, Consensus-based Translation Method.

- Translators **must** be allowed to work using their usual methods, including translation tools and software. Techniques such as blinding translators and insisting on naive translators are responsible for many of the "errors" detected by back-translation.

- Translators should be given access to the author of the original document if possible, or another researcher knowledgeable about the questionnaire if not, in order to request clarification when needed.

- During the correction/suggestion and consensus-building stages, translators should be discouraged from changing items "just for the sake of it", in order to avoid appearing lax, or for any reason other than the fact that they believe the alteration improves the translation.

- During the correction/suggestion and consensus-building stages, translators should also be encouraged not to suppress their own feelings simply to reach a consensus. If consensus is not possible between two translators, a third can be contracted to break the impasse, but must translate the entire questionnaire before being asked to adjudicate and must be at least as experienced as the two translators already contracted and preferably should have even more experience.

- The research team **must not** interfere with the translation process (before the committee stage), but the translators may and should request from the research team any information that they believe

could help them to produce the best possible translation, including terminology and translations suggestions.

- Translations should be pre-tested in the target culture and any problems detected should be solved in the target-culture by the research team. Retranslation is not necessary, since the assumption is that problems detected at this point will not be translation problems, but cultural differences. If there is conflict between the target culture and the source language instrument, the target culture must always take priority, even if this means sacrificing comparability of data on specific items.

- Full psychometric analysis of the translated scale can be conducted once pilot testing has indicated an acceptable translation.

The translation method that I am proposing is an attempt to adhere to these basic principles. The major feature of the method is of course the absence of a back-translation stage, but all aspects of the method have been designed with the intention of making the best possible use of the experience and skill of expert translators.

Certain elements of the method, such as the decision to prepare multiple initial translations, have been adapted from existing cultural adaptation methods, while other elements, such as a decision to give each translator the final say on whether to accept or reject alterations in their own translations, are examples of "best-practice" that I have observed in my own work as a professional translator.

At this point it is necessary to make a distinction between the "ideal" version of the Parallel, Cooperative, Consensus-based Translation Method outlined above and the procedures that I have used to demonstrate the method. In common with Brislin's original work, the objective is to investigate the translation process, not to produce a translation. Although a translation must be produced as part of the process, it is essentially a by-product. The result is that not all of the stages of the Parallel, Cooperative, Consensus-based Translation Method will be employed.

Conversely, for the purposes of this demonstration, I have added certain extra stages to the translation process itself. These stages are designed to facilitate the task of evaluating the translation process and to demonstrate that the method is capable of producing a forward translation that cannot be improved further using back-translation. The extra stages include a phase in which each translator scores different

translation solutions and, ironically, a back-translation of a first draft of the translation!

The back-translation will be compared with the source version of the questionnaire translated, as advocated by Brislin. The rationale for using a draft version is that it is likely to contain elements that the translators decided to modify in later versions.

The assumption here is that, if my hypothesis is ***incorrect*** and back-translation is indeed necessary, then the back-translation should identify problems in the translation that two expert translators were unable to identify by working with just the original and the translation and therefore remain in the final version. If, however, my hypothesis is ***correct***, the back-translation should only identify "problems" that are actually false-positives, i.e. problems that have been introduced by the back-translation and do not exist in the forward translation, or that have already been discussed by the translators.

5.2 Demonstration of the Parallel, Cooperative, Consensus-based Translation Method for Questionnaires

This demonstration has two objectives. The first is to show the feasibility of the translation method in practice, using professional translators to produce a Brazilian Portuguese translation of an existing English questionnaire. Since the method has been developed on the basis of a combination of theory and professional experience, the primary aim of the demonstration is to show that the method is applicable in practice. The second objective is to demonstrate that back-translation is unnecessary when expert translators are available.

In his 1970 study, Brislin started from the position that back-translation would improve translation and did not begin to correct translations until their back-translations had been compared with their originals. In contrast, I hope to demonstrate that if expert translators are used to produce, correct and revise the original translation, there will be no problems remaining that back-translation is capable of correcting. This is not quite the same thing as to state that there will be no problems left whatsoever, since there may be problems of which the translators are already aware and there may be very minor problems of a subtlety that back-translation does not have the sensitivity to detect, but it is the same thing as to state that back-translation no longer has anything to offer and is unnecessary.

Once the exercise is complete, the method will be adapted in response to the findings and the same questionnaire will be translated

once more. This is an idea that I owe to Professor Lincoln Fernandes who suggested it during my qualification hearing. He considered that it would be preferable to translating a different questionnaire since it would increase the detail of analysis on a specific sample, rather than widening the focus and possibly losing detail.

5.2.1 *The instrument*

The instrument I have chosen for the demonstration is the "Social and Physical Environment Survey", which is part of the "Environmental Supports for Physical Activity Questionnaire" (SIP 4-99 Research Group 2002). The lead author, Barbara Ainsworth, and the University of South Carolina's Prevention Research Center kindly gave their permission for me to translate the questionnaire. Barbara Ainsworth also gave me permission to give her e-mail address to the translators in case they wished to request clarification. The questionnaire is reproduced in Appendix A.

5.2.2 *The translators*

The translation services employed for the demonstration were generously donated by a translation agency for which I have worked as a freelance translator. The agency is *Scientific Linguagem* of Porto Alegre, Rio Grande do Sul, Brazil. *Scientific* is a specialist provider of language services offering, in addition to translation, proofreading, text correction, editing, formatting and publishing.

The Parallel, Cooperative, Consensus-based Translation Method requires at least two *expert* translators. After discussing the requirements of the method with a project manager at *Scientific*, two translators were chosen.

Both translators are employed full time by *Scientific Linguagem*, both have more than ten years' experience translating medical and scientific texts from English into Portuguese and both are native speakers of Brazilian Portuguese, living and working in Porto Alegre, in the South of Brazil.

Denise Arend has a bachelor's degree in Portuguese and English from the Universidade Luterana do Brasil, Canoas, and a postgraduate diploma in English from Unilasalle and in 2008 she completed a Masters degree in Translation Studies at the University of Aston in the United Kingdom.

Janisa Antoniazzi has a bachelor's degree in English and Portuguese from the Universidade Federal de Rio Grande de Sul, awarded in 1996, and a postgraduate diploma in Text Correction and Linguistic Consulting from UNIRITTER, awarded in 2007.

I had initially considered using translators from two different agencies, on the basis that this might provide differing perspectives, but working on the hypothesis that two translators who have worked together at the same company for more than ten years should find it easier to reach a consensus than two translators who do not know each other, I decided that Janisa and Denise would demonstrate the Parallel, Cooperative, Consensus-based Translation Method.

5.2.3   *The translation brief and translation protocol*

As has been discussed in detail, my theory holds that those elements of cross-cultural adaptation guidelines that place restrictions on the methods that translators can use, such as the practices of blinding translators, of insisting on naive translators and of requiring translators to translate under test conditions are responsible for introducing "errors" that are then "detected" by back-translation. My back-translation-free method therefore attempts to do exactly the opposite: it attempts to empower the translators to produce the best possible translation using whatever tools and techniques they wish. In the same vein, the method also attempts to provide them with background information that will enable them to contextualise their task.

Christiane Nord was writing about trainee translators when she defined the basic elements of the "translation brief" (Nord 1997, 44-5). Notwithstanding, I believe that a translation brief can be of use to even the most experienced translator since it can provide terms of reference against which competing translation solutions can be judged. Of course, an experienced translator does not need a *formal* translation brief and will invariably define the basic parameters of a project with the client even when the client has not specified them. Nevertheless, a clear definition of the objectives and context of a translation project, from the perspective of the client, can streamline the process of producing a translation that will meet that client's objectives.

Nord lists four elements that should be included in a translation brief. These are, 1, "the target-text addressee(s)", 2, "the prospective time and place of text reception", 3 "the medium over which the text will be transmitted" and, 4, "the motive for the production or reception

of the text" (Nord 1997, 45-6). I therefore prepared a translation brief for the demonstration, which can be consulted in Appendix B.

Certain elements of the demonstration have been included precisely because it is a demonstration and are not part of the translation method being proposed. The translation protocol that was given to the translators together with the translation brief includes five steps. The fourth of these steps was included for analytical reasons and is not needed to produce a translation. The translation protocol is also included in Appendix B. Nevertheless, I shall now provide a brief description of the five steps involved:

I. Each translator produces an initial translation, working alone. Translators are asked to concentrate on clarity and on consistency of terminology, particularly for the Likert scales.

II. Each translator makes two types of alteration to the other translator's first version, marking translation or grammatical "errors" in blue and suggestions for improvement in green. They may include comments to explain their suggestions, but do not have to do so.

III. Each translator receives her altered translation back and rejects or accepts the alterations.

IV. Each translator compares her version with the other translator's version and scores any items for which she "feels strongly" her solution is better than the other translator's solution.

V. The translators work together to produce a final version. They are at liberty to do this in whatever manner they believe produces the best translation. They could, for example, use one of the intermediate translations in its entirety or could go to the opposite extreme and produce an entirely new translation. The only demand is that neither translator suppresses their feelings in order to reach a consensus.

Step four is the stage that has been included purely for the purposes of the demonstration. It has been included as a check for the possibility that one translator could dominate the other and also to provide a basis for comparison with the actual consensus reached and would be unnecessary in a cultural adaptation project.

5.2.4   Analysis of translations

Since this is a demonstration that is being run to investigate the feasibility of the proposed translation method, the final translation will not undergo psychometric validation, but will be subjected to qualitative analysis of each stage leading to the final version and of the final version itself. I shall conduct the qualitative analysis.

As was pointed out with reference to evidence-based medicine and to Brislin's "practice effect", (good) translation is carried out by experts and expertise is related to both ability and experience.

In common with the two demonstration translators, Denise and Janisa, I have been a professional translator for more than ten years and I have also done large quantities of freelance translation for *Scientific Linguagem* (more than three million words to date). In contrast with Denise and Janisa, I am a native speaker of British English and I only translate from Portuguese to English. I therefore feel able to state that Denise, Janisa and I are all capable of providing expert opinion, although our areas of expertise differ.

In my qualitative analysis of the stages of the Parallel, Cooperative, Consensus-based Translation Method, I will discuss some of the translation solutions that Denise and Janisa employed and analyze the suggestions each translator made, the decisions on whether to accept or reject them and the resulting effect on the translation before going on to consider their final consensus-based version. I shall also discuss the results that could be expected from comparison of the back-translation with the source text.

5.2.5   Forward translations

The first draft produced by Denise is shown in Appendix C and the first draft produced by Janisa is shown in Appendix D. As was found by Perneger et al., the great majority of items and responses are worded differently. However, unlike Perneger et al., I do not automatically assume that these differences would make any difference when administered. Furthermore, these two translations are an initial step towards a final translation.

These initial translations were then exchanged and each translator made alterations to the other's translation, with comments where necessary. As professional translators, both Janisa and Denise understand that there are many valid ways to translate most texts and they did not concentrate on differences, as back-translation evaluation

instructions would have requested them to do, but restricted themselves to points where they identified room for improvement, within the other's approach.

This was what I had hoped for, since each translation has its own consistency as a whole and in a great many cases the differences between the two translations are related to minor differences in style and word choice. Some examples of the types of differences that Denise and Janisa did not comment upon, and which I believe they were quite correct to ignore, are given in Table 5.1 below.

Table 5.1 – Examples of different wording not mentioned in comments

Location	Denise	Janisa
Running head	Ambientes Sociais e Físicos	Ambientes Físicos e Sociais
Introduction	Farei algumas perguntas ...	vou fazer algumas perguntas
Introduction	Em primeiro lugar, algumas perguntas	Vamos começar com algumas perguntas
3.1	Há quanto tempo você vive no seu endereço atual?	Há quanto tempo você mora no seu endereço atual?
3.1	Nota ao entrevistador:	Nota para o entrevistador:
3.1	mais de 12 meses deve ser informado apenas em número de anos inteiros	mais de doze meses deve ser registrado somente como anos inteiros
3.5	Não (Passe diretamente para a questão 3.7)	Não (Pule para a pergunta 3.7)
3.10	De forma geral,	De um modo geral,
3.12	Em geral, como você classifica as condições dessas	3.12 Em geral, como você classificaria a condição dessas
3.13	qual das afirmações a seguir você considera mais adequada?	qual das afirmações a seguir é a mais verdadeira...
3.15	OBS.: ENTENDEMOS POR TRILHAS PARA CAMINHADA TRILHAS PÚBLICAS	DICA: TRILHAS PARA CAMINHADA SIGNIFICAM TRILHAS PÚBLICAS
3.16	Sim - UTILIZA CENTROS PÚBLICOS	Sim – USA CENTROS PÚBLICOS

In addition to using significantly different ways to basically "say the same thing" at many points, Denise and Janisa also used wording that coincided exactly at certain other points. These were all response options and many are recurrent. Some examples are shown in table 5.2 below.

Table 5.2 – Examples of identical wording choices

English	Denise & Janisa
Number of months (twelve months or less)	Número de meses (doze meses ou menos)
Number of years (one year or more)	Número de anos (um ano ou mais)
Refused	Recusou-se a responder
Very pleasant	Muito agradável
Not very pleasant	Não muito agradável
Not at all pleasant	Nem um pouco agradável
Heavy	Pesado
Moderate, OR	Moderado, OU
Light?	Leve?

In addition to identical response options, there are many response options that only differ because of grammatical agreement with another element. For example, in question 3.7 on lighting, Janisa used the phrase "a iluminação das RUAS", which is feminine, for "the STREET lighting", whereas Denise used "sistema de iluminação PÚBLICA", which is masculine. The result of this was that Janisa's response options were "Muito boa", "Boa", "Razoável", "Ruim" and "Muito ruim", while Denise chose "Muito bom", "Bom", "Razoável", "Ruim" and "Muito ruim".

In view of the fact that all nine different versions of the translations, from all the stages of comments, alterations and scores, total well over a hundred pages, I shall summarize the process that led to the final translation, which is shown in Appendix E.

I shall begin with the alterations, suggestions and comments that Janisa made on Denise's translation. Janisa made no changes or comments on the title page. Her first alteration was to insert the word "aproximadamente" in the following introductory statement on the first page of the questionnaire proper, "Em primeiro lugar, algumas perguntas sobre o seu bairro. Nesta entrevista, bairro é definido como a área que fica a uma distância de *aproximadamente* 800 metros ou 10 minutos caminhando de sua casa." Janisa commented that she felt it was important to state that the measurement is approximate.

Since the English statement is simply "the area within one-half mile or a ten-minute walk from your home", and half a mile is 805 metres, it may seem superfluous to add this detail. However, the same comment appears again before item 3.15, where a "comunidade" is defined as the "área que fica a uma distância de *aproximadamente* 15 quilômetros ou 20 minutos de carro de sua casa." The English version has 10 miles, which is a little over 16 kilometres.

Notwithstanding whether the difference in measurement warrants the insertion, and the insertion was preserved in the final version, what is important is that a researcher using this method can see that there has been a debate over the item and decide whether to use more usual approximations, as in the IQOLA system, in which "standards of equivalence were set to convert one mile, several blocks, and one block into metric equivalents of one kilometer, several hundred meters, and one hundred meters" (Bullinger et al. 1998: 921).

The next intervention Janisa made was not until item 3.19, which is one of a series of subitems that, in the English version, follow the introductory question, "Please tell me if you yourself USE any of the following resources and facilities in your community." The subitem is a list of similar facilities, which in English is "Parks/playgrounds/sports fields?". Denise had translated this list as "Parques/praças/quadros esportivos" and Janisa suggested "Parques/praças/campos esportivos" as a better translation.

Note that she did not query "playground" translated as "praças", which may appear strange to English-speaking readers, since it refers to public squares. Here in Brazil it is common practice for municipal authorities to erect both basic exercise equipment such as isometric bars and also children's playground equipment such as slides and see-saws in public squares. In contrast with a back-translation test, which might simply identify "praças" as not being equivalent to "playgrounds", a competent translator embedded in the target culture is perfectly capable of understanding another translators' choice, even though she herself chose to retain "playground", which also exists in Brazilian Portuguese as a loanword.

The final suggestion that Janisa made to improve Denise's translation was to alter her translation of a response to item 3.22, which in English is "Do you use physical activity programs and facilities at a place of worship?" The English response is "Yes- R USES FACILITIES AT PLACE OF WORSHIP IN COMMUNITY"[29] and Denise had translated this as "Sim - UTILIZA O ESPAÇO DE IGREJAS DA COMUNIDADE". Janisa pointed out that this excluded several forms of worship that are common in Brazil, giving the example of an *umbanda* centre where people worship but which is not a church.

---

[29] R stands for respondent; I shall discuss the changing of first/second person for interviewer/respondent to first/third person (and back again) when I present Denise's comments.

Janisa's translation was "Sim – USA INSTALAÇÕES EM LOCAL DE CULTO RELIGIOSO NA COMUNIDADE" and when commenting on this Denise stated that "locais de culto religioso" was the best option for "place of worship". In the final version of the questionnaire a composite of both translations is used, as follows, "Sim – USA LOCAIS DE CULTO RELIGIOSO NA COMUNIDADE".

Denise made many more comments on Janisa's translation, although not all were suggestions for improvements or corrections. At several points in Janisa's translation, Denise expressed her opinion that Janisa had chosen better solutions or that Janisa's approach had inspired her to change some of her own decisions. Denise's comments, being more extensive, also reveal more of her translation methods.

I shall begin by presenting Denise's suggestions and alterations, before covering points where she commented on a difference, but concluded that it made no difference and then end the presentation of her interventions by discussing those where she considered Janisa had found better solutions.

Denise made extensive alterations to the title page as translated by Janisa. The first of these was to pluralize Janisa's translation of the subtitle of the questionnaire. In English this is "SOCIAL AND PHYSICAL ENVIRONMENT SURVEY", but Denise points out in her comments that at the top of every page except the title page, the running header is "Social and Physical Environments". Denise was therefore recommending standardising the titles.

Denise then substituted "Centro de Pesquisa sobre Prevenção" with "Centro de Pesquisa em Prevenção" and "Escola de Saúde Pública Norman J. Arnold" by "Faculdade de Saúde Pública Norman J. Arnold". She did not justify these changes, but, once more, hypothetical researchers who had requested the translation would be able to make their own decision, once the comment had alerted them to the need for a decision.

Denise makes what is a correction of Portuguese usage, adding an "a" that was missing from "Pesquisador*a* principal". Once more there is no comment, and one is unnecessary, assuming that Barbara Ainsworth is female, which she is.

Denise's final intervention on the title page is to insert a composite citation including the original title in English in parentheses. The citation is the same she used in her own translation (see Appendix C). She also inserts the English into Janisa's translation of the financial support statement, in parentheses.

She justified these interventions saying that she felt that some form of the citation was necessary and that she had re-inserted the words "SPECIAL INTEREST PROJECT #4-99", and "Cooperative Agreement #U48/CCU409664, Centers for Disease Control and Prevention" in parentheses on the basis that readers might wish to search electronically for details.

Since Janisa did not comment when she accepted these alterations at a later stage, it is hard to deduce why she removed the citation, but the most obvious answer would be to make room for an equivalent Portuguese citation for the culturally adapted version.

It is also conjecture, but one possible explanation for Denise's concern with authorial attributions is that she is a partner at Scientific Linguagem and deals with the customers, who are very often authors, but also include editorial boards of journals that Scientific translate *en bloc*. This may make her more sensitive to author acknowledgement.

The next change that Denise made is repeated at several points in the questionnaire because it is one of the response options in a series that recurs. In English the series is "Very physically active", "Somewhat physically active", "Not very physically active" and "Not at all physically active" and the gradations are repeated in other configurations such as "Very pleasant", "Somewhat pleasant", "Not very pleasant" and "Not at all pleasant". Janisa had used the series of the format "Muito agradável", "Um pouco agradável", "Não muito agradável" and "Nem um pouco agradável", and Denise altered "Um pouco agradável" to "Relativamente agradável" and did the same to all other response options using the same formula.

She justified this by saying that the concepts "um pouco" and "não muito" were too close and that the second option required a solution that was a little more positive to improve the distribution of the scale. She also offered "razoavelmente" on the basis that Janisa had used "razoável" in later items.

This was the only cause of disagreement between the two translators and proved to demonstrate that the scoring system I had devised for step four was flawed, as I shall discuss shortly.

There are points in favour of both sides and I shall jump forward to Janisa's explanation for not accepting Denise's change in order to deal with the issue all in one place. Janisa defended her original choice saying that she understood Denise's concern, but still felt that "um pouco" was positive in comparison with "não muito" and that she did not agree with the suggestions of "relativamente" or "razoavelmente" because they were more complex concepts and it would not be

understandable to everybody, since there was no educational criterion for administration of the questionnaire.

These are exactly the types of discussions that medical researchers have when designing original monolingual questionnaires and also when adapting them for other cultures. Once more, the ultimate decision on whether to prioritise intelligibility over scaling, on the basis that rank can be more important than the word chosen, as shown by Perneger et al., would fall to the researcher, but the cooperative and communicative aspects of this method mean that not only is the existence of an issue identified, but the translators offer solutions proactively.

It might be considered that Denise's point that Janisa uses "razoável" later on in the questionnaire counts against her argument that razoavelmente is too complex. However, it is used in the sense of acceptable, rather than as a degree of physical activity, as in "the standard of street lighting is reasonable" rather than "relatively active physically" or "reasonably active physically", which were the options Janisa rejected.

In the English version, item 3.6 on street lighting reads "For walking at night, would you describe the STREET lighting in your neighborhood as…" and Janisa had translated "the STREET lighting" as "a iluminação das RUAS". Denise swapped this for "a iluminação PÚBLICA", which is the normal way to refer to street lighting in Brazil. While Janisa's translation was perfectly intelligible and could be considered more specific to the context of walking at night, it was unidiomatic. Janise accepted this alteration.

In a similar vein, item 3.8 asks "For walking in your neighborhood, would you say that unattended dogs are…." and offers a series of responses ranging from "A big problem" to "Not a problem at all". Janisa had translated "unattended dogs" as "cachorros sem dono", which would back-translate literally as "ownerless dogs", but means something closer to stray dogs. Denise changed this for "cachorros soltos na rua", which would back-translate literally as "dogs loose on the street", but since "na rua" can mean "outside" in general, the idiomatic meaning is something close to "dogs loose in public". Janise also accepted this alteration.

Item 3.11 contains the words "Does your neighborhood have public recreation facilities" in the English version. Janisa had translated "public recreation facilities" as "instalações de recreação pública", which Denise changed to "instalações públicas de recreação", pointing out that the adjective public referred to the facilities not the recreation.

This correction recurred in several different items and Janise accepted it in all cases.

The next alteration that Denise made to Janisa's translation was a formatting change. The English version has an introductory question before item 3.15 that relates to item 3.15 and the five items that follow. The question is, "*Please tell me if you yourself USE any of the following resources and facilities in your community.*" and it is followed by a note to the interviewer to "**Emphasize you/yourself.**", both formatted as shown here. Janisa's translation had followed the English formatting, with "*Por favor, diga se você mesmo USA qualquer um desses recursos ou instalações em sua comunidade.*", but Denise altered the formatting so that the translation read as follows, "*Por favor, diga se VOCÊ MESMO USA qualquer um desses recursos ou instalações em sua comunidade.*" This is interesting in that it is a departure from the "original", but is actually more consistent with the context than the original formatting had been.

The question is directed at the interviewee, indeed all the introductory questions and statements are within inverted commas and are in italics to emphasize this, while the note is of course directed at the interviewer. Since this questionnaire is designed to be administered by telephone, emphasis would have to be purely a matter of the way the question is read to the respondent, and the capitalisation must be taken as an indication of that emphasis. If the intention is to emphasize you/yourself, then capitalization of "USE" is a misdirection to the reader. This is an example of the way that translation detects inconsistencies in source texts.

Denise made no further critical comments, alterations or suggestions on Janisa's translation, but the communicative interaction was by no means limited to these more negative reactions to Janisa's translation. Denise also made several comments acknowledging points where she felt Janisa's options were preferable to her own, identified certain points of difference that she felt were not significant to the questionnaire's function and even identified what she described as an error in her own translation on the basis of one of Janisa's choices.

I shall begin with the differences that Denise considered unimportant and then present the elements that she considered Janisa had translated more satisfactorily, including the point at which she identified an error in her own work.

The first solution that Denise identified as different, but considered not to have an impact was the response option "Don't know/ not sure", which recurs throughout the questionnaire and is invariably

followed by "Refused". Both Denise and Janisa retained the third person for their translation of "Refused", coinciding on "Recusou-se a responder". For "Don't know/ not sure", however, Denise retained the first person with "Não sei / não tenho certeza" while Janisa chose the third person with "Não sabe/não tem certeza". Denise noted this and commented that she didn't think it made any difference.

An expert in psychometric testing might respond that it does make a difference since the questionnaire is interviewer-administered, arguing that responses in the first person are offered to the respondent whereas a response in the third person prompts the interviewer to make a judgement, so if a respondent gave an answer that didn't make sense to the interviewer they might choose "Doesn't know/not sure" in a case in which the respondent might not choose "Don't know/ not sure". This is a valid argument and once more the defence of this method is that the discussion process makes the choice explicit meaning that a researcher can choose between the two options.

Furthermore, with relation to this particular questionnaire, it does not appear that the question of third or first person was considered critical by the developers, since there is no standardisation of this aspect.

Items 3.1 to 3.14 have a number of options with no person, such as "Yes", "No", "Excellent", "Good", "Quite safe" and "Slightly safe", followed by one option in the first and one in the third person ("Don't know/not sure" and "Refused"). In contrast, items 3.15 to 3.23 have a different format. The questions refer to different resources that the respondent may or may not use for physical activity, "waterways", for example. The first two responses are in the third person and are all of the form "Yes- R USES WATERWAYS ..." and No- R DOES NOT USE WATERWAYS ...", where "R" presumably stands for "Respondent". The third begins with the first person possessive pronoun and is of the form "My community does not have any waterways ..." and the fourth and fifth responses are the first and third person formulations "Don't know/ not sure" and "Refused".

This is very strange since it sets up a situation in which the interviewer makes a judgement on whether the respondent uses the resource in question, since there is no response for yes or no in the respondent's voice, but the respondent decides if there is no such resource or if they do not know or are unsure if they use such a resource, but the interviewer makes a judgement on whether the respondent refused or not. This raises the question of what would happen if they disagreed! However, with respect to back-translation, the important factor is that the back-translation-free method detects such

inconsistencies in the source text, which questionnaire developers would hopefully see as an opportunity rather than a challenge.

The second point at which Denise identified a difference, but considered it insignificant was in the conversion of one mile to 15 or 16 kilometres. In view of Perneger et al.'s finding that translating one mile as 1km or as 2km made no difference, I am inclined to agree with Denise, but once more what is important is that a researcher using this method would be alerted to the need to decide on how significant the degree of precision is to the function of the questionnaire.

The first of Janisa's solutions that Denise considered preferable to her own first attempt was the questionnaire's main title. The English title is "Environmental Supports For Physical Activity Questionnaire" and Denise had translated this as "Questionário Sobre o Papel do Ambiente na Realização de Atividade Física". In her comment on Janisa's translation, Denise explained that she had found the title difficult to translate and that Janisa's solution of "Questionário Sobre Suportes Ambientais Para Atividade Física" was more concise and therefore better. Denise's more cumbersome solution was therefore rejected by both translators and the final version used was Janisa's, followed by her subtitle as pluralized by Denise.

The next option that Denise preferred and adopted in subsequent drafts was the choice of "vizinhança" as a translation of "neighborhood", rather than "bairro" as she had chosen. She explained that she had decided that "vizinhança" was more appropriate to the definition of "the area within one-half mile or a ten-minute walk". I would add that since "neighborhood" has "neighbor" as its root, then "vizinhança" is also more appropriate because of its link to "vizinho(a)".

Denise identified what she described as an error in her own translation after reading Janisa's translation of one of the responses to item 3.9. The question in English was "How safe from crime do you consider your neighbourhood to be? Would you say…" and the response options were "Extremely safe", "Quite safe", "Slightly safe", "Not at all safe", "Don't know/not sure" and "Refused". Denise had translated the first four of these as follows, "Extremamente seguro", "Relativamente seguro", "Pouco seguro" and "Nem um pouco seguro", but after reading Janisa's options of "Extremamente segura", "Bastante segura" "Um pouco segura" and "Nem um pouco segura" realised that she had not noticed that the gradation of these responses was different to the pattern that had been hitherto established.

Up to this point in the questionnaire, the response scales of this type had followed the gradation of "Very pleasant", "Somewhat

pleasant", "Not very pleasant" and "Not at all pleasant", where the first two are positive and the last two negative with relation to pleasantness, level of physical activity or whatever the item is asking about. Denise's third item "Pouco seguro" follows this pattern, it is on the side of "unsafe", not of "safe". In contrast, the third option in English, "Slightly safe" is positive, if not very positive, and only "Not at all safe" is negative. In the final version, Janisa's translation was used for this entire item, i.e. the question and all six response options.

Item 3.15 in the English asks about use of walking trails and offers the following response options in addition to don't know and refused, "Yes – R USES WALKING TRAILS IN COMMUNITY", "No – R DOES NOT USE WALKING TRAILS IN COMMUNITY" and "My community does not have any walking trails", which Denise had translated rather formally as "Sim – UTILIZA TRILHAS DISPONÍVEIS NA COMUNIDADE", "Não – NÃO UTILIZA TRILHAS DISPONÍVEIS NA COMUNIDADE" and "Minha comunidade não dispõe de trilhas para caminhada".

Janisa's translation follows her explicitly stated objective of using accessible language, which she used to justify rejecting "relativamente" and "razoavelmente" in favour of "um pouco". She translated the same responses "USA TRILHAS PARA CAMINHADA NA COMUNIDADE", "NÃO USA TRILHAS PARA CAMINHADA NA COMUNIDADE" and "Minha comunidade não tem trilhas para caminhada", which Denise also adopted on the grounds that they were more concise than her solutions.

Item 3.23 asks about using waterways and includes the following proviso "DO NOT INCLUDE NON-PHYSICAL ACTIVITIES SUCH AS BOATING". Denise had translated this as "NÃO INCLUIR ATIVIDADES NÃO FÍSICAS, COMO PASSEAR DE BARCO", but rejected this solution in favour of Janisa's less cumbersome "NÃO INCLUA ATIVIDADES QUE NÃO ENVOLVAM PRÁTICA FÍSICA, COMO PASSEIO DE BARCO", which was used in the final version.

The same item (3.23) includes two out of three examples of the type of cultural mismatch that can only be solved in consultation with the (in this case hypothetical) researchers who will use the instrument and with reference to the underlying concepts they wish to tap.

Denise considered Janisa's choice of "cursos d'água" to be a better translation than "vias fluviais" for the English "waterways". She also commented that she had rejected Janisa's choice of "caiaquismo" for "kayaking" because it had too few occurrences, by which she means

it didn't return as many results on Brazilian Portuguese web pages when used as a search term.

Related to these items is a discussion that began with Janisa's comments on Denise's translation and is continued in Denise's comments. It relates to the list of facilities in Item 3.19 which, in English, are "Parks/playgrounds/sports fields". Denise rejects Janisa's use of the loanword "playground", replacing it with "praças", as in her translation, but rather than change Janisa's "campos esportivas" for "quadras esportivas" as she had used in her own translation and mirroring what Janisa did in her translation, Denise suggested using both terms. She argued that she had chosen to use sports courts rather than sports grounds on the basis that there are more likely to be courts than grounds in Brazil and, although she acknowledged that Janisa's option was more "faithful" (she used inverted commas), she suggested that using sports grounds would reduce the number of positive answers and so the best option would be to use both terms.

This discussion of which elements are appropriate to Brazil demonstrates that the translators are aware of the tension between adherence to the content of the source text and compatibility with the cultural reality of the target setting. It also demonstrates their awareness of the fact that non-compatibility could lead to distorted results.

As I have mentioned, this demonstration is by necessity an artificial situation and the lack of a researcher or research team who will administer the translated questionnaire means that these issues of cultural differences cannot be solved by the translators, although I maintain that they could be solved by the translators in conjunction with a target-setting researcher.

In order to gain an idea of how a questionnaire such as this one could be adapted to cater for different environmental supports in different cultures, I asked a professor of sports medicine at UFSC to comment on two lists of facilities from the questionnaire and suggest modifications for use in Brazil.

The two lists are part of questions. The first question was "Does your neighborhood have public recreation facilities (such as public swimming pools, parks, walking trails, bike paths, recreation centers, etc.)?". The item I sent to Professor Edio was from the final translation and was as follows, "A sua vizinhança tem espaços públicos de recreação (como piscinas públicas, parques, trilhas para caminhada, ciclovias, centros de recreação, etc.)?".

The second question was "Do you use nearby waterways such as creeks, rivers, and lakes for water-related physical activities such as

canoeing, kayaking, swimming, or skiing?" and the translation sent to Professor Edio, also from the final translation, was "Você usa cursos d'água, como riachos, rios e lagos, para praticar atividades físicas aquáticas, tais como canoagem, caiaquismo, natação ou esqui aquático?"

Professor Edio consulted colleagues and suggested adding "Praças públicas, academias ao ar livre, pistas de skate, praias, calçadão para caminhada/corrida, arenas," to the first list. These would translate something like "public squares, open air gyms, skate parks, beaches, pedestrianised area for walking/jogging, arenas". Professor Edio also suggested removing "centros de recreação" and replacing it with "quadras poliesportivas, quadras de tênis", which would translate as "multi-use sports courts, tennis courts".

For the second question, Professor Edio and colleagues suggested adding "praias e baías" (beaches and bays) to the list of waterways and "surf, mergulho, remo" (surfing, diving, rowing) to the list of sports.

As I claimed in section 4.3 on cultural issues, these types of elements that simply do not exist or are not used in the same way from one culture to another cannot be substituted by a translator alone without detailed explanations of the underlying concepts being probed. After a cooperative process such as the one I have demonstrated here, researchers and translators can work together to solve these issues with reference to the intended function of the instrument.

It will have been noted that I have only discussed the results of steps one to three in detail and mentioned some of the options from the final version. This is because the demonstration detected an inherent flaw in the process as I had designed it for this thesis. In stage four, during which I had envisaged the translators marking as many as a dozen items in order of preference, only one translation solution was chosen by each as being superior to the other's.

Unfortunately both translators chose the same element, which was the translation of "Somewhat", which recurs as "Somewhat physically active", "Somewhat pleasant", "Somewhat maintained", "Somewhat of a problem", "Somewhat important", "Somewhat safe" and "Somewhat unsafe". Janisa maintained that "Um pouco" was the better choice, whereas Denise remained convinced that "Relativamente" was more suitable.

This led to a discussion in which it was agreed that there was no way they could reach consensus since both accepted the other's arguments and considered the other's translation acceptable, but felt that on this point their own solution was better. The result was that the final version was a development of Denise's first draft since she had already

incorporated all of Janisa's suggestions (except "um pouco") and had also incorporated many of Janisa's solutions without Janisa having commented.

The final version is a good translation and, once the decisions to start with Denise's draft had been taken, was arrived at by both translators working together, but it cannot be considered a consensus version, as I has hoped, since consensus was not reached on the starting point.

This impasse was the reason I had incorporated the scoring stage, but it becomes evident that numbers are not a good way to determine a tie-break in this case, since both participants will play to the rules. By only selecting one item each, Janisa and Denise reduced my system to a scale with a total amplitude of zero to one, without fractional divisions. As I argued with relation to Brislin's performance criterion test, on which nobody made more than one error and the vast majority made zero errors, if this can happen then the assumptions on which the measurement is based are flawed and I fully accept that this is also the case here..

In Brislin's case the assumption was that the task was more difficult than it actually was and in my case it was the assumption that the translators would engage in some sort of bargaining, accepting one solution in return for rejecting another. This was naive of me since translations must be judged as complete texts and questionnaires are no exception. In the absence of any hierarchy between the translators there was no basis on which to break the impasse without abandoning the protocol and asking me what to do next, which is what happened.

In the next chapter I shall present a simplified version in which only one translation is produced and the second translator acts as a proof reader or editor and which abandons any attempt to try and rank translation solutions with numerical values.

However, before ending this chapter, in the next subsection I shall discuss what comparison of a back-translation of Denise's first draft with the English source text would have offered.

5.2.6   *Back-translation*

I commissioned a back-translation of Denise's first draft from a professional translator who has lived and worked in Florianopolis for over fifteen years. Despite my commitment to naming translators I cannot name this translator since I neglected to request permission when negotiating the contract and could not obtain it retrospectively.

Notwithstanding, this is a competent expert translator, but he had never done back-translation before and did not have a background in public health, physical education or any other subject related to the questionnaire, but a degree in journalism. He could therefore be considered "naive". As I mentioned in Chapter 2, Grunwald and Goldfarb provide the only information I have found in the literature on how much back-translation costs. To add to this scant data I can report that I paid a total of R$223 for the back-translation, which was around 115 US dollars at the time.

I chose Denise's first draft for the back-translation on the basis that it was altered extensively before becoming the basis for the final version and should therefore offer many opportunities for back-translation to demonstrate its merit. The back-translation is shown in Appendix F.

The differences between the back-translation and the English source text can basically be divided into five categories. The first three categories equate to Ozolins' "translation noise". These are (i) differences that are simply different ways to "say the same thing", not affecting the function in any "substantive" manner (to use Grunwald and Goldfarb's term), (ii) discrepancies that are the result of differences between English and Portuguese and (iii) differences that have been introduced during the back-translation phase and give a false impression of an error in the forward translation. These three categories account for the great majority of differences between the source text and the back-translation.

In addition to these are a series of differences that are related to issues that could indeed be problematic. The great majority of these had been discussed between Denise and Janisa and solved by the final draft.

However, there is also a fifth category, which could be described as "back-translation silence", to extend Ozolins' metaphor. These are cases in which there was a possible problem with Denise's first draft (whether identified by Denise herself, by Janisa or by me), but which was "silenced" by back-translation, since the back-translation coincides exactly with the source text, meaning that someone comparing the two without reference to the forward translation would have no way of knowing that there was an issue to be addressed.

I shall now deal with each of these categories in turn, starting with translation noise caused by different word choices with substantially the same meaning. Since there are a large number of such occurrences, I shall provide a few typical examples, rather than list the differences verbatim.

Table 5.3 below lists some typical examples of different wording in the source text and back-translation that would essentially make no difference in terms of administration of the instrument. For ease of visualisation I have italicised in the source text and underlined in the back-translation to indicate differences.

Differences such as these occurred constantly throughout the questionnaire. A minimum of two responses options for every question contained wording differences (in many cases all responses differed), and only two very simple question stems were identical: "3.16 Public pools?" and "3.19 Parks/playgrounds/sports fields?".

Table 5.3 – Back-translation noise (i) – non-substantive differences

Location	Source text	Back-translation
Title page	Principal *investigator*	Principal <u>researcher</u>
Introduction	I will *be asking you* some questions about the neighborhood *in which* you live	I will <u>ask</u> some questions about the neighborhood <u>where</u> you live
3.1	*Overall*, how would you *rate* your neighborhood	<u>In general</u>, how would you <u>classify</u> your neighborhood
3.5	*Skip* directly to question 3.7	<u>Go</u> directly to question 3.7
3.10	*Generally speaking*, would you say that most people in your neighborhood *can be trusted*?	<u>In general</u>, would you say that most people in your neighborhood are <u>trustworthy</u>?
3.21	My community does not have *a* shopping *mall*	My community does not have <u>any</u> shopping <u>centers</u>

In addition to the types of differences illustrated in Table 5.3 above, which would probably be considered unproblematic by all but the most literal-minded monolingual raters, there are also a series of differences that could appear to be errors to a monolingual rater, but which are reflections of asymmetrical language relations and would be rejected by someone with access to the intermediate forward translation.

This type of difference occurs because the target language and/or culture performs a given function in a different manner to the source language and/or culture and back-translation does not reverse the changes made during the forward translation. In terms of having confidence in the final translation this type of back-translation noise is not the methodological pitfall that back-translation "silence" represents, but, as the comments reproduced by Ozolins show, it contributes greatly to the translators' feelings of frustration.

A typical example would be the note to the interviewer that qualifies the response options for item 3.1. In the source text, this reads

as follows, "Less than one year is entered as months and more than twelve months is entered as whole years only." In the back-translation, this phrase is very different, reading as follows, "Less than one year should be informed with the number of months and more than 12 months should be informed only in the number of complete years.", which is sufficiently unidiomatic to cause concern in a monolingual researcher. However, this wording is a reflection of the fact that the back-translator is "naive" and unaccustomed to the specialised language of questionnaires. The forward translation reads, "Menos de um ano deve ser informado em número de meses, e mais de 12 meses deve ser informado apenas em número de anos inteiros." The back-translation has preserved the cognate of "informado", which is both idiomatic and appropriate to data entry in Portuguese, but is confusing in English, since it seems to suggest that the interviewer should tell the respondent their age, as the verb inform is not used with inanimate objects such as a questionnaire or form.

In a similar vein, the stem to item 3.7 reads as follows in the source text, "For walking at night, would you describe the STREET lighting system in your neighborhood as ...", but in the back-translation, street lighting has become "PUBLIC lighting". In Brazilian Portuguese, "iluminação pública" describes what in English is covered by "street lighting", and so Denise altered "iluminação das ruas" in Janisa's translation, as being unidiomatic. When this is back-translated into English as "public lighting", the more appropriate Portuguese translation appears to be at error, while one that had retained "iluminação das RUAS" would have gone unnoticed.

It could be argued, in favour of a back-translation protocol such as the IQOLA that does not specify using a naive translator, that this type of translation noise is the result of having contracted a naive translator, since one experienced in questionnaire adaptation might have translated "iluminação pública" as the more idiomatic "street lighting". However, many protocols specify that the back-translator produce a "more literal" translation, which would encourage even a non-naive translator not to do this.

Notwithstanding, even if a non-naive translator was used and was not requested to perform a more literal translation than normal, there are still pitfalls that back-translation cannot avoid. The third type of back-translation noise is when the back-translator has made a change, whether in order to conform to the demands of the language or culture of the translator or whether in error, but that change does not reflect the forward translation which does not contain the same change. Table 5.4

below lists some examples of this type. For ease of visualisation I have italicised in the source text and underlined in the back-translation to indicate differences and used bold typeface in the forward translation to indicate that the change occurred in the back-translation.

Table 5.4 – Back-translation noise (iii) – differences in back-translation only

Location	Source text	Forward translation	Back-translation
3.2	people in your *neighborhood* are	as pessoas no seu **bairro** são	people in your neighborhoods are
3.4	the *motorized* traffic	tráfego de veículos auto**motores**	the automobile traffic
3.13	*Thinking about how public money is spent*	**Pensando** no **dinheiro público** que **é gasto**	Considering government investments made

In the example from item 3.2, the back-translation suggests an erroneous pluralisation of "neighborhood", but the forward translation has actually used the singular, correctly. In the example from item 3.4, the back-translation has restricted the range of vehicles included in "traffic" to cars, but the forward translation has not, including all forms of motorised transport. Finally, in the example from item 3.13, the scope of funding sources has been restricted in the back-translation (public money does not have to be government money), but the scope of types of funding has been relaxed (since investment could include loans, i.e. the money is not spent but lent). Once more, the forward translation has maintained the scope as in the source text. The money is public, and it is spent.

Notwithstanding, if one is willing to put translators through the corrosive experience entailed, consultation with both forward and backward translators could elucidate all three types of back-translation noise and discard the resulting erroneous conclusions of problematic translation solutions in the forward translation. However, as I mentioned above, there are also points at which either Janisa, Denise or I identified issues in the forward translation that might need to be reworked and therefore required discussion, but which the back-translation did not reveal. There are only three such examples, but the potential for such an occurrence is a serious blow to the confidence that back-translation is supposed to afford. Table 5.5 below lists these examples. For ease of visualisation I have italicised in the source text and underlined in the

forward-translation to indicate differences and used bold typeface to indicate where the changes have been suppressed in the back-translation.

Table 5.5 – Back-translation silence – differences hidden in back-translation

Location	Source text	Forward translation	Back-translation
3.2	people in your *neighborhood* are	as pessoas no seu <u>bairro</u> são	people in your **neighborhoods** are
3.19	Parks/*playgrounds*/ sports fields?	Parques/<u>praças</u>/ quadras esportivas?	Parks/**playgrounds**/ sports fields?
3.23	*kayaking*	<u>caiaque</u>	**kayaking**

As mentioned on page 311 above, Denise had commented on the use of "bairro" for "neighborhood" and the final version used Janisa's option of "vizinhança", which is more appropriate because of the size of area specified and because of its link to "vizinho(a)", which parallels the relationship between "neighborhood" and "neighbor". The back-translation, however, has hidden this detail and, consequently, lost the opportunity for discussion of the issue.

The third example is fairly trivial, but it is not the nature of the difference ignored in this particular case that is important. The important factor here is the way that back-translation can ignore differences, which, in combination with monolingual raters, results in their being ignored. In this case, the forward translation had broken the sequence of names of sports with "caiaque", which is the name of the boat itself, not the name of the sport. In a later stage of the parallel method, Denise had identified this as a weakness in her own translation and improved it by adopting Janisa's solution. The back-translation, however, has rendered "caiaque" as "kayaking", re-standardising it with the remainder of the list of sports and hiding the existence of an inconsistency.

The second example, however, is the most serious. When I compared the back-translation of "playgrounds" with the forward translation of "praças", I could not understand how "praças" had led to "playground" without prior knowledge that it had been a translation of "playground" in the first place. I have to admit that my first reaction was that the back-translator must have "cheated" – i.e. despite promising not to, he had consulted the source questionnaire, which is freely available on the internet and for which the link is provided on the title page. This is not a pleasant reaction, but it is symptomatic of the suspicion introduced by back-translation. Furthermore, despite the fact that I commissioned the back-translation as a means to the end of discrediting

back-translation, I had subconsciously adopted the "policing" attitude that it engenders.

However, I do not believe that the translator consulted the questionnaire, not only because I have a professional relationship with him and trust him not to do so, having promised not to, but also on the basis of analysis of differences, such as "government investments" that are not only different to the source text, but are also departures from the forward translation and indicate a tendency towards lateral thinking in the back-translator's interpretations. In fact, there are at least two possible alternative explanations for this surprising observation that do not rely on assumptions of unethical behaviour.

The first is that the back-translator was attempting to find a suitable solution for one of the vocabulary items on the questionnaire and he employed the substitution technique using a search engine. I use this technique frequently myself. It works as follows. If I am attempting to find the correct term in a sentence that contains other terms for which I have already chosen translations, I will often copy the entire sentence, in the target language, into the Google search box, with the exception of the word or words that I have not yet translated. It is fairly common for the summaries of the first five or so results to contain the exact term I am looking for, purely because it collocates with the other words in the sentence. When this happens, I will not therefore need to even click on the links to the search results themselves, since I will already have found the answer in the summarised results.

If the back-translator had used this technique to try to find a term it is not only not possible, but actually quite probable that he would have been presented with "playgrounds" in the context of "parks", "sports fields" and "community", precisely because he was looking at results from Barbara Ainsworth's questionnaire. He would not necessarily guess that this was the source and if he never clicked on the link, he would never find out.

The second possible explanation for why the back-translator translated "praças" as "playgrounds" is related to the fact that although he is a United States national, he has lived in Brazil for at least fifteen years. It is possible that he has so internalised the relationship between playgrounds and public squares that "playgrounds" was simply the translation that occurred to him as most appropriate

Notwithstanding, while these hypotheses excuses the translator of any unethical behaviour, it is in fact irrelevant whether the back-translator chanced across the original, copied it in bad faith, was simply inspired or came to exactly reverse Denise's translation solution for

some other reason. In all cases the failure to indicate that the intermediate translation was not "playground", or another term describing a similar phenomenon, but "praças", reflects badly on back-translation.

This is because it once more highlights the fact that a monolingual researcher reading source and back-translation has no basis whatsoever on which to judge how they came to coincide or differ. The assumption when reading "playgrounds" in both source text and back-translation is that the translation contains a word or words in the target language that describes or describe the same phenomenon as "playground" describes in the source language. Since "praças" describes the phenomena that in English would be referred to by "squares" and since one word in Brazilian Portuguese for the phenomenon described by "playground" in English is (as shown by Janisa's translation) the loan word "playground", it is clear that any monolingual researcher who read the source text and the back-translation and, on that basis, came to the conclusion that the forward translation was semantically equivalent to both, would have been seriously misled by the back-translation technique.

It is not hard to envisage a scenario in which the difference between public parks and squares and playgrounds could be materially significant to the results of the questionnaire. There may be respondents, for example, who take their children to private playgrounds and engage in physical activity while there, but who would never frequent a public square for a similar purpose.

In addition to "translation noise" and "translation silence", there are also differences between the back-translation and the source text that do indicate some type of issue or difference in the forward translation. However, with a single exception which I shall cover shortly, all of these points of interest had been identified and commented on by Denise and Janisa during the course of the parallel translation process. For ease of visualisation I have italicised in the source text and underlined in the forward-translation to indicate points of interest and used bold typeface to indicate where these have been preserved in the back-translation.

Table 5.6 – Real translation issues identified by back-translation

Location	Source text	Forward translation	Back-translation
Introduction	neighborhood is defined as the area	bairro é definido como a área que fica a uma	neighborhood is defined as the area that is a

	within *one-half mile*	distância de até <u>800</u> metros	distance of up to **800 meters**
3.6	Not at all maintained	Nem um pouco <u>bem</u> mantidas	Not at all **well** maintained
3.7	Fair	<u>Razoável</u>	**Reasonable**
3.9	Quite safe	<u>Relativamente</u> segura	**Relatively** safe
3.15	community is defined as the area within ten miles	comunidade é definida como a área que fica a uma distância de até <u>15 quilômetros</u>	community is defined as the area that is a distance of up to **15 kilometers**
3.15	Please indicate if you <u>yourself</u> USE	Favor informar se VOCÊ UTILIZOU	Please indicate if YOU USE

All of the items in table 5.6 above were commented on by Janisa or Denise. The two statements defining distances were discussed and the final versions have the word "aproximadamente" inserted before the distance. Whether this is the most appropriate solution is not as important as the fact that the exchange of comments would alert a hypothetical researcher to the fact that there is a decision to be made. The error identified in item 3.6, where "not at all maintained" had become "not at all *well* maintained" was spotted by Denise and corrected in the final version, as was the translation of "quite safe" as "relativamente segura".

Indeed, the solution finally chosen - "bastante" - has a similar indeterminate quality to "quite" in that sometimes it can mean "almost entirely" "almost totally" "almost fully" etc. and sometimes it can mean "sufficiently so" (consider the different meanings of "quite" in the following: "the professor was quite dead when we arrived"; "he's had quite enough for today"; "that's quite a nice car").

In common with these two issues, Denise and Janisa also addressed the missing "yourself" in item 3.15, adding "mesmo" in the final version.

Finally, there is one apparent discrepancy that back-translation indicates, and which is present in both Denise's and Janisa's first drafts and remained in their final version. This is a translation of the question to item 3.9, which in the source text is "How safe from crime do you consider your neighborhood?", but which has been back-translated as "How safe do you consider your neighborhood?", with no mention of "crime".

This is the only difference between the back-translation and the source text that could be used to justify back-translation since it is the only difference that does appear to make a substantive difference, does reflect a missing semantic element in the forward translation and was

not changed during my cooperative process. It could, therefore, be considered as evidence that in a critical situation, back-translation might still be justified, despite all of the false positives and negatives and the corrosive effect that they have on the translator-client relationship, as long as there is a chance that it could prevent even one error.

In fact, the true situation is not nearly so simple. It is true that there is no semantic element in the Portuguese translation which could be considered the "equivalent" of "from crime", but this does not mean that the translation is at fault. Denise's translation of "How safe from crime do you consider your neighborhood?" was "Em sua opinião, qual o nível de segurança da sua vizinhança?" A word-for-word back-translation of this would be something like "In your opinion, what is the level of safety of your neighborhood", but the actual back-translation, "How safe do you consider your neighborhood?" has reversed the structural changes and reverted to the same part of speech as in the source (safe).

Had Denise used the word "seguro" (safe), the forward translation would indeed be lacking any reference to crime, as indicated by the back-translation. However, the unmarked meaning of "segurança" is actually "safety from crime", at least here in the South of Brazil where these translations were produced. In other words, if "segurança" is not qualified as employment safety, child safety or some other type of safety, its default meaning (context allowing) is closer to "security" than "safety". I do not expect non-Portuguese-speaking readers to take this claim on trust, so I shall demonstrate its validity.

The newspaper with the largest circulation here in the state of Santa Catarina is called the *Diário Catarinense*. The section of that newspaper that might be called something like "crime" or "law and order" in an English-language newspaper is simply entitled "Segurança" in the *Diário Catarinense*.

Table 5.7 below lists some recent headlines from *Diário Catarinense* with my translations of them into English. All were published under the title "Segurança".

Table 5.7 – Headlines of articles in the "*Segurança*" section of *Diário Catarinense*

Date	Portuguese headline	English translation
23 March 2013	A fuga, os esconderijos e a prisão da mulher de traficante e líder do PGC	The getaway, hideouts and arrest of the wife of a drug dealer and PGC* boss.
24 March 2013	Prisões na capital durante final de semana têm ligações com tráfico	Arrests in the capital over the weekend are linked to the drug trade

	Dois integrantes do PGC morrem após troca de tiros com a polícia no Sul de SC	Two members of the PGC shot dead in a shoot-out with police in the South of SC.
29 March 2013	"Atiravam até quando estavam deitados no chão", diz diretor do DEIC sobre tiroteio em Major Gercino	"They were still shooting even when they were on the floor" says the head of DEIC† after shootout in Major Gercino
30 March 2013	Bope prende ladrões fugitivos no assalto a banco no Vale do Rio Tijucas	Bope‡ catch escaped robbers during bank raid in the Vale do Rio Tijucas

*PGC: *Primeiro Grupo Catarinense* - an organised crime gang in Santa Catarina state, Brazil.
†DEIC: *Departamento de Investigações sobre Crime Organizado* – a special police unit dedicated to investigating organised crime.
‡BOPE: *Batalhão de Operações Policiais Especiais* – a police shock unit, equivalent to a SWAT team.

Hopefully it is clear from the above that, in the context that Janisa and Denise used it, "segurança" is a perfectly acceptable translation. Explicitation would be unidiomatic and it is not therefore surprising that neither translator objected to the other's omission of an explicit reference to "crime".

However, there is a more important point at stake here than whether Janisa and Denise correctly translated "safety from crime". As the authors I presented in subsection 3.4.2 above all agree, translation is semi-automatic (if the translator is competent!) and in fact most competent uses of language are also semi-automatic.

Stopping to think about the structural configuration of what one will say (or write) in formal terms is a sure sign that language competence has not yet been internalised sufficiently for true fluency. This is true of translators too, not just second language learners, but with translators, the stage of translation expertise can only begin after language competence has been internalised – before that point, novice language competence precludes true translation competence.

As Hurtado Albir and Alves have shown, the most experienced translators have a kind of "autopilot" that controls the bulk of the work, but surrenders control to more conscious processes when necessary. Notwithstanding, I maintain that even these more conscious processes rarely include consideration of formal parts of speech or of grammatical rules. It is highly likely that neither Janisa nor Denise would remember having translated "safe from crime" because the phrase presented no problems and they both translated it automatically.

This opens the door to what is possibly the most destructive effect wreaked by the antagonistic nature of back-translation. In such a situation, the very automatic nature of the translation process means that

a translator who was challenged by a client troubled by the missing two words in the back-translation would not have pre-prepared explanations or arguments with which to justify their choices and would quite probably accept the insertion of an explicitation, possibly even with an apology for having omitted "from crime" in their translation.

If the process of challenging the translator occurs in real-time, whether face-to-face during an expert committee meeting or virtually during a teleconference, the translator will also have to justify the original translation in conditions that are entirely unlike the ideal conditions for translation of written texts.

As this example shows, despite the fact that the translators' initial instincts were correct, back-translation challenges their "associations and linguistic reflexes" (Hönig 1997: 17) and the automatic nature of these associations and reflexes means that an unprepared translator would be defenceless in the face of such accusations.

Translators, who are empathetic people, working in a service industry, therefore accept fault where there is none. In turn, this "victory" for back-translation reinforces the initiator's belief in the efficacy of back-translation, while leaving a doubt in the mind of the translator, further increasing the corrosive effect of back-translation.

5.2.7   *Conclusions and discussion*

I believe that this demonstration, while not perfect because of the failure of the consensus-reaching mechanism, has succeeded in showing that expert translators working together are capable of producing high-quality translations of questionnaires. Furthermore, not only was back-translation incapable of detecting anything significant that they had not already discussed and/or corrected, it resulted in an inordinately large number of false positives and false negatives, all of which contribute to undermining translators, but not to improving the quality of translations.

It might be thought that Denise and Janisa exerted some type of favouritism because we have a business relationship. In Nord's terms it might be suggested that they have more "loyalty" to me than they would have to other clients. This may be true and I cannot rule it out.

However, the demonstration shows that they are capable of translating to this level of quality. If it is remembered that translation is an expert activity and that translation competence is built up over time as translators become more experienced, it seems inevitable that they must have acquired this expertise by producing translations (presumably

for their clients) and that at least some of those translations were of as high quality as these translations are.

Furthermore, as a translator myself, I see no reason why a translator would translate badly on purpose. Translators may fail to reread, edit or revise a translation as many times or with as great a degree of thoroughness as ideal, because of time pressures, low pay or just plain laziness, but, in the absence of the second-guessing imposed by back-translation, they have nothing to gain by using a weaker or less appropriate solution to a translation problem if a stronger or more appropriate one has already occurred to them spontaneously (and in my experience, backed up by Hurtado Albir and colleagues, the more experienced a translator becomes the more spontaneous and apparently automatically appropriate solutions occur to them).

Notwithstanding, the problems I encountered with reaching consensus and the anticipated criticism that I have privileged access to better-than-average translations encouraged me to attempt to develop a process that did not need a consensus-building stage and to test it with translators who were not known to me. The next chapter describes this process and its results.

6   STREAMLINED TRANSLATION PROTOCOL

6.1 Demonstration of the Streamlined Translation Method for Questionnaires

The primary objective of the streamlined method for questionnaire translation was to overcome the impasse caused by the fact that the consensus-building stage of the parallel method had been a failure. Up to the point at which Denise and Janisa had to construct a final version, the parallel method worked well, but at that point both translators accepted that the other's version was valid, but preferred their own version and so the choice to use Denise's translation as a basis for the final translation was arbitrary. I therefore considered the method to be only a partial success as a substitute for the translation → back-translation step, since it is not ready to "slot in" to the gap that is left. However, I believe the main objective of demonstrating that expert translators can identify problems, offer solutions and produce high quality translations, whereas back-translation produces false negatives and false positives and offers no solutions, was achieved.

It was suggested during my qualification hearing that perhaps the root of the problem had been assigning the same role to both translators and that possibly a solution would be to only produce one translation, to be prepared by a professional translator, ask a second professional translator to revise it, but not to prepare their own translation, and then return it to the first translator for preparation of the final version. This is what I therefore decided to do.

In addition to the problem with consensus, I also decided to attempt to address the issue of how a medical researcher, with no contacts in the translation industry, might identify expert translators with whom to work. For the demonstration of this method I therefore decided to source professional translators unknown to me, via the internet. Unfortunately this was not an entirely successful venture, as I shall explain shortly.

6.1.1   The instrument

The instrument translated was the same instrument described in subsection 5.2.1 above, the "Social and Physical Environment Survey" from the "Environmental Supports for Physical Activity Questionnaire" (SIP 4-99 Research Group: 2002).

6.1.2   The translators, translation brief and translation protocol

In addition to attempting to solve the consensus problem and to contracting translators with whom I had no pre-existing professional relationship, it was also suggested that I should source translators from a different region of Brazil (Brazil is conceptually divided into five regions, North, North East, Mid-West, South East and South).

I therefore ran an internet search for translation agencies specialising in technical and medical translation and selected the first three that were not in the South of Brazil. These agencies were Easy Translation Services in Rio de Janeiro (www.easyts.com), Nativo Traduções in São Paulo (www.nativotraducoes.com.br) and All Traduções, also in São Paulo (http://www.alltraducoes.com).

I wrote, in English, to all three agencies via email explaining that I needed two experienced translators to conduct a translation for me in a specific manner and that I wished to read their CVs and know their names. The letter I sent to the agencies is shown in Appendix G. This is the point at which the problems with the streamlined method began to emerge, since none of the agencies were willing to disclose the names of their translators. It could be that they do not wish their clients to "cut out the middle man" and go directly to their translators in future projects, but this is speculation, since they simply stated it was against their policy.

Notwithstanding, after negotiating with the agencies I contracted two translators via Easy Translation Services on the basis that of the three contacts their project manager seemed to have the best grasp of what I wanted the translators to do.

To add to the information provided by Grunwald and Goldfarb, I can report that I paid a total of R$620 for the forward translation and two revisions, which was around 310 US dollars at the time.

After agreeing the terms and the delivery date, I sent a translation brief and a translation protocol (shown in Appendix G) to the agency. The brief was similar to that for the first method, but the protocol required just one translation which would be revised by a second translator, who had not translated the questionnaire, and then returned to the first translator who would have carte-blanche to prepare the final version incorporating or not the second translator's alterations and to make any further alterations deemed necessary. Since the agency did not inform me of the translators' names, I shall refer to them as Translator 1 and Translator 2.

6.1.3   First draft

The first draft produced by Translator 1 is shown in Appendix H. Translator 1 is evidently both competent and talented and some of his or her solutions are very creative. For example, options including the word "somewhat" were the only point at which Denise and Janisa did not consider the other's translation acceptable. An example is the sequence,

"Very physically active", "Somewhat physically active", "Not very physically active" and "Not at all physically active" in the source text. Denise had originally translated this sequence as "Bastante fisicamente ativas", "Relativamente fisicamente ativas", "Não muito fisicamente ativas" and "Nem um pouco fisicamente ativas", whereas Janisa had chosen "Muito ativas fisicamente", "Um pouco ativas fisicamente", "Não muito ativas fisicamente" and "Nem um pouco ativas fisicamente" on the basis that "relativamente" was too complex a concept.

Translator 1 chose a very simple and, in my opinion, preferable solution for all of these response options. He or she simply translated somewhat as zero, as follows, "Muito ativas fisicamente", "Ativas fisicamente", "Não muito ativas fisicamente" and "Inativas fisicamente", bypassing the problem Denise identified with Janisa's option that it was not positive enough, but also sidestepping the problem Janisa had identified with Denise's translation.

This pattern is repeated at all points where "somewhat" is the qualifier for second rank in a scale of four and it is appropriate in each case. For example, "Muito bem conservadas", "Bem conservadas", "Não muito bem conservadas" and "Não conservadas", or "Muito seguras", "Seguras", "Relativamente inseguras" and "Totalmente inseguras".

Another example of a fitting solution to a response option is the translation of "Fair", in the sequence "Excellent", "Good", "Fair" and "Poor". Translator 1 chose "Regulares" in the sequence "Excelentes", "Bons", "Regulares" and "Ruins".

Translator 1 also anticipated Janisa and Denise's solution to whether "quadras esportivas" or "campos esportivos" would be a better translation for "sports grounds", using a composite of both options right from the first draft: "campos ou quadras esportivas".

Other interesting options chosen by Translator 1 include "auxílios ambientais" for "environmental supports", the use of "no raio" to define the area within a neighbourhood ("a vizinhança é definida como a área no raio de uma milha e meia") and the choice of "navegação" for

"boating" as a non-physical activity (Denise had used "passear de barco" and Janisa "passeio de barco"). Furthermore, Translator 1 used "segura" for "safe" in his or her translation of "How safe from crime do you consider your neighborhood to be?", making it necessary to qualify the type of safe as safe from crime, as follows, "O quão segura você considera sua vizinhança, em relação à criminalidade?"[30]

In other cases, the choices made by Translator 1 coincided with those made by Denise, Janisa or both. All three used the same translations for "almost always", "often", "seldom" and "never", choosing "sempre", "frequentemente", "raramente" and "nunca".

In the first draft, Translator 1 coincided with Denise on choices such as "Pesquisa" for "Study", "Escola de Saúde Pública" for "School of Public Health", "Investigador(a)" for "Investigator", "iluminação pública" for "street lighting", "shoppings" for "malls" and "você" for "you yourself" (Janisa had used "você mesmo"). Translator 1 also retained the suggested citation, as did Denise, but without translating any part of it.

Translator 1 coincided with Janisa on choices such as "vizinhança" for "neighborhood", "mora" for "live", "Nem um pouco segura" for "not at all safe", "playgrounds" for "playgrounds", "pule" for "skip" and the choice to put the translation of "Don't know/not sure" in the third person.

In general, Translator 1's first draft is a good basis for a final translation, but it contains some solutions that are questionable and would benefit from discussion, in addition to a small number of errors.

Examples of phrases that could possibly be improved with discussion are "Caso o tipo de recurso ou comodidade mencionado não esteja disponível em sua comunidade, favor notificar." for "If the type resource or facility I mention is not available in your community, please let me know.", since it is more formal than the English. This is alos true of both "Quão seguras são as instalações de recreação pública em sua comunidade? Você as considera:" for "How safe are the public recreation facilities in your community? Would you say…" and "Preocupações a respeito da segurança em instalações de recreação pública influenciam seu uso das mesmas?" for "Do concerns about safety at the public recreation facilities in your community influence

---

[30] This in no way invalidates my argument that, here in the South of Brazil, "safety from crime" is the unmarked meaning of "segurança", since (i) the same is not true of "seguro" and (ii) Translator 1 is from Rio de Janeiro, over 1100 kilometres away from Florianopolis and 1800 kilometres from Porto Alegre.

your using them?" because in both cases the register is more formal and the constructions are complex.

In addition to entire phrases that could be improved, there are certain specific translation choices that should at least be discussed. These include "prazerosa" for "pleasant" in the translation of "Overall, how would you rate your neighborhood as a place to walk? Would you say… Very pleasant, Somewhat pleasant, Not very pleasant, Not at all pleasant", the use of "cães abandonados" for "unattended dogs", "dinheiro público é investido" for "public money is spent", "canais" for "waterways", "comodidade" for "facility" and, probably most importantly, the decision not to convert miles to kilometres and metres, but to translate "miles" as "milhas" instead.

Finally, there are also a very small number of what could be considered errors. There is an error of agreement on the title page, where Barbara Ainsworth, who is female, is described as Investigador Principal, without the final "a" needed for agreement. Janisa used the term "Pesquisador", but she also omitted the final "a" in her first draft, which was picked up by Denise. Along the same lines, there is also a typing error by which "tais como" has become "tis como".

Translator 1 has also made an omission and a misinterpretation of the type that the proponents of back-translation like to use as evidence of its worth. The first is a misreading of the English "within one-half mile" (< ½ mile). Translator 1 translated this as "no raio de uma milha e meia" and although the use of "no raio" is succinct and creative, he or she has clearly misread the English as "within one *and a* half mile*s*" (< 1½ miles).

Additionally, there is no part of Translator 1's first draft that corresponds to the information offered by the words in italics in the following, "do you use ANY private *or membership only* recreation facilities?" since this item is translated as follows, "você utiliza QUALQUER local privado para recreação?".

This omission could be justified on the basis that private could include members-only facilities, but the change in the distance defining neighbourhood is of fundamental importance to the responses to the first 14 of the 27 items. This is the type of error that one would hope the translator undertaking the revision would identify and correct.

6.1.4   *Revision*

Unfortunately, it was exactly at this point that the streamlined process failed. Translator 2 made exactly one change to Translator 1's draft,

which was to remove "etc" from the Question "Sua vizinhança possui locais para recreação pública (tais como piscinas públicas, parques, trilhas de caminhada, ciclovias, centros recreativos, etc)?". Considering that the "etc" was present in the English source text (albeit with a full stop at the end – "etc."), it is hard to understand why this change was made. Translator 1 replaced "etc" in the final translation returned to me, although with the addition of the missing full stop.

I have re-read the translation protocol sent to both Translator 1 and Translator 2 and I can only conclude that the cause of Translator 2's behaviour was the following sentence "Please do not correct or improve just for the sake of it. If the translation you are revising does not have any errors and you cannot see any way of improving it, then please simply send it back to me unaltered." However, the instructions also request Translator 2 to "... review the translation and correct anything they consider to be an error, whether because they believe there is a problem with the Portuguese or because they believe the translator has misunderstood the original." Therefore, even if Translator 2 did not consider any of the more nuanced issues I have discussed in the last subsection warranted improvement, the misinterpretation, omissions and grammatical errors should still have been identified.

The best light I can put on this occurrence is that Translator 2 was not in fact qualified to revise Translator 1's translation, whether for simple errors of Portuguese usage such as the failure of agreement and the typing mistake, or for omissions and misrepresentations in the translation of information offered by the source text, as in the one-half versus one-and-a-half confusion.

The worst interpretation is that Translator 2 saw that there was no obligation to do any work to earn the money offered and simply returned the questionnaire after deleting "etc". The reason is fairly immaterial to the method's failure.

6.1.5   Third draft

Translator 1 received the (unaltered) questionnaire back and, despite the opportunity provided in the protocol to return to the first draft and do no work whatsoever, Translator 1 actually made a number of changes to his or her first draft.

On the title page, Translator 1 reverted to the English titles for the "Norman J. Arnold School of Public Health" and the "University of South Carolina", he or she changed "auxílios ambientais" to "auxílios do ambiente" and "acordo cooperativo" to "acordo de cooperação" and

changed from "investigador" to "pesquisador" for "investigator". He or she also extended the recurrent response option "Recusou" to "Recusou-se a responder".

Unfortunately, Translator 1 did not correct the agreement when changing "investigador" to "pesquisador" (referring to Barbara, who is female) and did not change all of the "Recusou" responses, leaving the first occurrence as it had been in the first draft, thereby losing standardization.

Translator 1 altered "canais" for "waterways" to "canais aquáticas" and "em locais de culto" to "em locais de atividade religiosa", and while any improvement is only marginal, this would alert a researcher to the existence of an issue to be addressed in each case.

Translator 1 did identify the omission of "membership only", adding the words "apenas para membros", and also noticed and corrected the typing error.

However, although Translator 1 did make an alteration to the most important error in the first draft, he or she did not correct the error and in fact compounded it. In the first draft, Translator 1 had left both the definition of the size of a neighbourhood and the definition of the size of a community in miles, translating the name of the measurement rather than converting to the measurements used here in Brazil.

In the revision, Translator 1 converted the first of these distances to kilometres, but evidently did not check back with the source text, since the distance converted is the incorrect 1½ miles rather than the correct ½ mile. Translator 1 therefore "corrected" "no raio de uma milha e meia" to the equally incorrect (but metric) "no raio de aproximadamente 2,4 km".

Furthermore, the second distance had been correct in the first draft, albeit culturally incongruous, at "dez milhas". During the revision, Translator 1 changed this to "aproximadamente dez quilômetros", which is no longer the correct distance, although now the units are those used in Brazil. It is not hard to imagine how this happened. When Translator 1 decided to convert the first distance, he or she probably remembered that there was another distance measured in miles and decided to convert that too. However, what he or she actually did was to change "milhas" to "quilômetros" and insert "aproximadamente", but never come back and convert the number, resulting in a distance of just over six miles rather than ten.

6.1.6   Conclusions and discussion

Even though there were only two of them, I must accept that the errors relating to the size of a neighbourhood and a community, by which the neighbourhood is defined at three times the radius intended and a community at 62% of the radius intended, are exactly the type of problem that the proponents of back-translation trust it to solve.

While I do not believe that back-translation is a trustworthy solution, I cannot claim therefore that this second method is foolproof either. Translator 1 is evidently reasonably competent and had he or she received the kind of feedback that Denise provided, the method would almost certainly have succeeded. The mistake was probably to introduce too many new variables into the new method and it might have been more advisable to use Denise and Janisa once more, creating a protocol in which Janisa was the translator and Denise revised, on the basis that Denise gave more feedback the first time. However, to retain the same translators would have demanded a different questionnaire, reducing the possibilities for comparing one method with another.

With hindsight, once the decision to use different translators had been taken, it might have been better to source the translators via online profiles on a special-interest website such as Translator's Cafe or ProZ, since such profiles are for individuals rather than agencies and many are accompanied by CVs. I had wished to avoid using "insider" knowledge in this demonstration, but it may be that one of the things I have inadvertently demonstrated is that insider knowledge is necessary and so researchers should be engaging the services of linguistic consultants.

Notwithstanding the reasons for selecting one professional who in the event added nothing to the final product (but who also did not antagonise the other professional as back-translation would have done), what the failure of this second method has demonstrated is that it is necessary to start with professionals who are known to be competent in this type of work. If that means conducting a number of preliminary exercises for the purposes of identifying competent translators, then this is still preferable to back-translation, since a pool of competent translators would build up in a relatively short time.
.

7 FINAL COMMENTS

In Chapter 2 I presented the back-translation literature. I began with work by Richard Brislin conducted within cross-cultural psychology (Brislin 1970; Brislin et al. 1973; Brislin 1980; Brislin 1986) in which he formalised the technique and advocated its use for ensuring equivalency in translation.

I then presented the most important contemporary literature on back-translation for cross-cultural adaptation of health-related questionnaires (Guillemin et al. 1992; Sperber et al. 1994; Bullinger et al. 1998; Beaton et al. 2000; Sperber 2004), showing how they differed methodologically from Brislin's methods and from each other, primarily in the type and number of additional steps, since the back-translation step itself has remained unchanged.

Finally, I presented three articles containing criticism of back-translation, two that reached positive conclusions (Grunwald & Goldfarb 2004; Ozolins 1999) and one that concluded, with surprise, that back-translation was unproven (Perneger et al. 1999).

I hope that in so doing I achieved my first objective, which was to bring the existence of the back-translation practice in healthcare to the attention of scholars studying translation and show that it is a widespread and distinctive application of translation that merits further study.

In Chapter 3 I presented theories and concepts developed within translation studies and in Chapter 4 I utilised these elements to argue, and I believe demonstrate, the following:

- The original experiments conducted by Brislin employed amateur translators with suboptimal English competence and back-translation provided an indication of the degree of that language deficiency, but that is not the same as evaluating translation quality.

- The value added by back-translation as a step in cross-cultural adaptation was never proven.

- Back-translation can lead to dysfunctional translations if followed blindly.

- Back-translation is founded on a simplistic concept of language and of translation and on an *a priori* expectation of translation failure.

- Back-translation methods perpetuate the illusion of symmetry in translation.

- Use of back-translation anchors the translation to the source text and the source culture.

- Back-translation enforces hegemonic power relations, enabling monolingual control of a multilingual process and reinforcing cultural insecurity in "peripheral" cultures.

- Use of back-translation is further perpetuated by past success as defined within the publication-dominated academic patronage system.

- Back-translation is dismissive of translators' expertise and denies translators the necessary conditions in which to exercise their expertise.

- Expert translators should not need back-translation since they have adequate problem identification and solution skills.

- In cultures such as Brazil, in which translators and translation are not low status, the imported back-translation paradigm is corrosive since it instils and reinforces mistrust of translations and translators.

- Back-translation is not an acceptable means for giving translators a voice, since communication between translator and client should be proactive and ongoing, not reactive and defensive, in response to an accusatory list of "discrepancies".

I hope that by demonstrating the above, I have achieved my second objective, which was to show that not only is back-translation not necessarily the best means of achieving the objectives of cross-cultural adaptation, but that it can actually lead to several undesirable consequences.

My final objective was to demonstrate that expert professional translators are capable of translating health-related questionnaires to a high standard without back-translation and that a translations process involving extensive communication between those involved can not only produce acceptable translations, but can also offer opportunities for solving problems that back-translation cannot.

While I fully admit that neither the full parallel process described in Chapter 6, nor the streamlined process described in Chapter 7, were ideal, particularly not the second of the two, I do not believe that invalidates my claim to have shown that expert professional translators are capable of translating health-related questionnaires to a high

standard without back-translation. The fact that my process did not lead Denise and Janisa to arrive at a consensus does not invalidate the high-quality translations they created, nor, more importantly, the useful feedback that their interaction provided. Furthermore, the fact that the back-translation did not detect any errors that Denise and Janisa had not addressed, but did hide certain issues that they had discussed, is further confirmation of the dangers of trusting to back-translation.

Along the same lines, my failure to contract two competent professionals for the second process does not disprove my arguments in any way. Rather, it reinforces them. Having contracted a translator who was either unqualified or unwilling to make the effort to revise the first draft of the translation, I had broken my own stricture, which was to work with expert professional translators. While Translator 2 may well be expert and professional at other tasks, bearing in mind that Translator 1 revised his or her own first draft extensively, the failure to notice a single defect demonstrates that either Translator 2 was not expert, if he or she truly read through the entire translation and found nothing, or that Translator 2 was not professional, if he or she accepted the fee without doing the work.

In short, what the partial failure of the parallel method and the total failure of the streamlined method do show is that I failed in the secondary objective that I set myself at the end of Chapter 4, which was to provide a substitute translation stage that would produce a translated questionnaire that was ready for the committee stage. I maintain, however, that the level of expertise demonstrated by Denise, Janisa and Translator 1 is sufficient to adequately translate health-related questionnaires, if the correct communicative framework is used, which was not the case for Translator 1, who was left to work alone with no feedback in a situation that, in the final analysis, was not much different from the forward translation step of a back-translation protocol.

I have constantly stressed the importance of collaboration and communication between those that commission translations and their translators and the greatest limitation of my demonstrations was that I did not achieve this.

The underlying reason, I believe, is that I commissioned these translations, but I am not a medical researcher and as such I am not in a position to provide the necessary feedback.

I had provided the translators with an email address for Barbara Ainsworth, lead author of the questionnaire, and they had her permission to request any clarification or additional information they might need. However, none of the translators or agencies contacted

Barbara. I hypothesise that this was because Barbara was not the commissioner. Furthermore, had they done so, her lack of experience of the target setting would have meant that, despite being the original author, she would not have been the ideal informant.

My suggestion for future research into alternatives to back-translation is therefore that a truly interdisciplinary effort be mounted in which scholars of translation work together with target-culture healthcare researchers to develop a framework for the communicative collaboration needed to ensure the success of translation.

As Professor Edio's suggestions for widening the amplitude of responses to the environmental supports questionnaire showed, and as I have argued, certain translation decisions can only be made with knowledge of the ultimate function of the translation. My demonstrations did not have a real life application and were not commissioned by a healthcare researcher.

In an ideal process, researchers and translators would work together to define the function of the translated questionnaire, analyzing each item for potentially problematic elements and identifying solutions on the basis of the underlying health concepts being measured.

The failure to achieve the level of collaboration and communication that I had wished for affects both demonstrations, although it was only a relative failure in the first case, but an absolute failure in the second case. The second demonstration, however, also suffered from a limitation that would make it worthless for the purposes of building up a pool of recognizedly experienced questionnaire translators. This is the fact that I was unable to convince the translation agency to provide the translators' names or CVs.

The result of this is that Translator 1, who I judged reasonably expert, remains anonymous and cannot be sought out by name. If this proves to be the rule with translation agencies, then there are two alternatives. The first is to engage in "supplier education" in addition to client education and attempt to convince agencies that although they may have concerns relating to their customers and their translators bypassing them for future projects if they allow contact, they should be able to trust their translators to adhere to professional ethics and request any clients who were to attempt to contract them directly, having first worked with them through the agency, to contract future work through the agency too.

If an agency gained a reputation as the contact point for expert translators, who were known and cited, then they too would gain from lifting the cloak of anonymity. If, however, such educative efforts prove

ineffective and agencies categorically refuse to divulge their translators' names, the second option is to source freelance translators directly, thereby eliminating anonymity.

It is my firm opinion that naming translators is the most direct route to achieving several desirable objectives. Naming translators would allow researchers control over the most important element in any translation project – the choice of translator. Current anonymous methods are based on an assumption that all translators are alike, but healthcare professionals know better than anyone that all human beings are different. Naming translators would also allow competent and talented translators to build reputations on the strength of past work and this would not only raise the status of questionnaire translators among healthcare professionals, it would also raise the status of questionnaire translating as a task among translators. Finally, naming translators would also be a disincentive to translators accepting commissions for which they lacked the necessary expertise.

I believe that with this thesis I have opened a hitherto unexplored area of the translation industry to academic scrutiny. Despite the length of this thesis, for which I apologise wholeheartedly, a single researcher can only hope to scratch the surface of a phenomenon as complex and widespread as back-translation and I hope that others will be inspired by my first steps to continue to explore this avenue. I would like to end by listing some areas that I believe of interest, but which were beyond the scope of this project:

- investigating series of translations produced using back-translation, such as the IQOLA and EORTC questionnaires;

- investigating whether there is a relationship between economic factors, such as GDP and trade surpluses/deficits and societies' propensity to import questionnaires via translation;

- investigating the extent to which rank and context affect people's reactions to different response options;

- investigating whether a "documentary" approach to back-translation might be more useful to researchers;

- investigating whether, rather than translating the questionnaires, it might be preferable to translate the supporting literature, which would allow the use of controlled medical vocabularies, and have target-setting researchers create de novo questionnaires.

REFERENCES

AAOS. *AAOS On-Line Service January 1997 Bulletin*. Online publication, available from: http://www2.aaos.org/aaos/archives/bulletin/jan97/modems.htm, accessed Feb. 2011d

AAOS. *About the AAOS*. Online publication, available from: http://www.aaos.org/about/about.asp, accessed Feb. 2011a.

AAOS. *Outcomes Information*. Online publication, available from: http://www.aaos.org/research/outcomes/outcomesmain.asp, accessed Feb. 2011b.

AAOS. *Outcomes Instruments and Information*. Online publication, available from: http://www.aaos.org/research/outcomes/outcomes.asp, accessed Feb. 2011c

Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez J, Filiberti A, Flechtner H, Fleishman SB, de Haes JCJM, Kaasa S, Klee M, Osoba D, Razavi D, Rofe PB, Schraub S, Sneeuw K, Sullivan M, Takeda F. *The European Organization for Research and Treatment of Cancer QLQ-C30: A Quality-of-Life Instrument for Use in International Clinical Trials in Oncology* JNCI J Natl Cancer Inst 85, 5, 1993, 365-376. Abstract available from: http://jnci.oxfordjournals.org/content/85/5/365.abstract

Ajzen I, Fishbein M. Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin*, 84, 1977, 888-918

Ajzen I, Fishbein M. *Understanding attitudes and predicting social* behavior. Englewood Cliffs, NJ: Prentice-Hall. 1981

Arend D. *Scientific translators in Brazil: a preliminary profile description* Masters dissertation, University of Aston. 2008

Arrojo R, Rajagopalan K. A noção de literalidade: metáfora primordial. In: Arrojo R. (Org.) *O signo desconstruído: Implicações para a tradução, a leitura e o ensino* (2nd ed.) Campinas:Pontes. 2003 (1992a), 47-57

Arrojo R, Rajagopalan K. O ensino da leitura e a escamoteação da ideologia. In: Arrojo R. (Org.) *O signo desconstruído: Implicações para a tradução, a leitura e o ensino* (2nd ed.) Campinas:Pontes. 2003 (1992b), 87-93

Arrojo R. (Org.) A desconstrução do logocentrismo e a origem do significado. In: *O signo desconstruído: Implicações para a tradução, a leitura e o ensino* (2nd ed.) Campinas:Pontes. 2003 (1992b), 35-41

Arrojo R. (Org.) *O signo desconstruído: Implicações para a tradução, a leitura e o ensino* (2nd ed.) Campinas:Pontes. 2003 (1992)

Arrojo R. A pesquisa em teoria da tradução ou o que pode haver de novo no front. In: Arrojo R. (Org.) *O signo desconstruído: Implicações para a tradução, a leitura e o ensino* (2nd ed.) Campinas:Pontes. 2003 (1992f), 99-107

Arrojo R. Arrojo R. (Org.) A noção do inconsciente e a desconstrução do sujeito cartesiano. In: *O signo desconstruído: Implicações para a tradução, a leitura e o ensino* (2nd ed.) Campinas:Pontes. 2003 (1992a), 13-19

Arrojo R. As questões teóricas da tradução e a desconstrução do logocentrismo: algumas reflexões. In: Arrojo R. (Org.) *O signo desconstruído: Implicações para a tradução, a leitura e o ensino* (2nd ed.) Campinas:Pontes. 2003 (1992d), 67-71

Arrojo R. Compreender x interpretar e a questão da tradução. In: Arrojo R. (Org.) *O signo desconstruído: Implicações para a tradução, a leitura e o ensino* (2nd ed.) Campinas:Pontes. 2003 (1992c), 67-71

Arrojo R. O ensino da tradução e seus limites: por uma abordagem menos ilusória. In: Arrojo R. (Org.) *O signo desconstruído: Implicações para a tradução, a leitura e o ensino* (2nd ed.) Campinas:Pontes. 2003 (1992e), 93-99

Arrojo R. *Oficina de tradução: a teoria na pratica* (5ª ed.) São Paulo: Atica. 2007 (1986)

Arrojo R. Os estudos da tradução na pós-modernidade, o reconhecimento da diferença e a perda da inocência. *Cadernos de Tradução* (UFSC) 1, 1996, 53-69 Available from:
http://www.periodicos.ufsc.br/index.php/traducao/article/view/5064/4567

Arrojo R. *Tradução, desconstrução e psicanálise* Rio de Janeiro: Imago. 1993

ASA *Bio – Uldis Ozolins*. Online publication, disponível em http://www.asatours.com.au/heritage-cities-baltic/82-leaders/243-bio-uldis-ozolins accessed Mar. 2013

Associação Médica Brasileira, Conselho Federal de Medicina. *Texto introdutório: Projeto diretrizes* Associação Médica Brasileira e Conselho Federal de Medicina. Online publication, 2008. Available from: http://www.projetodiretrizes.org.br/projeto_diretrizes/texto_introdutorio.pdf, accessed Dec. 2010.

ASTM International. *F 2575-06 Standard Guide for Quality Assurance in Translation*. 2006.

Baker M. Linguistic Perspectives on Translation. In France P (Ed) *The Oxford Guide to Literature in English Translation* Oxford: Oxford University Press. 2000: 20 – 26

Bartman BA *Health Status Measurement, Floor and Ceiling Effects*. The Health Science. Online publication, available from http://thehealthscience.com/wiki/Health-Status-Measurement-Floor-and-Ceiling-Effects, last updated: 06 May. 2012.

Bassnet S, Lefevere A. General editors' preface. In Lefevere A. *Translation, rewriting and the manipulation of literary fame* London/NewYork: Routledge 1992

Bassnet S, Lefevere A. General editors' preface. In Venuti L. *The translator's invisibility: A history of translation*. Taylor & Francis e-Library. 2004 (1995), vii-viii

BBC News. *Micronesia country profile*. Online publication, last updated: Feb. 9, 2011, available from: http://news.bbc.co.uk/2/hi/asia-pacific/country_profiles/1300849.stm, accessed Feb. 2011a

BBC News. *Palau country profile.* Online publication, last updated: Feb. 9, 2011, available from: http://news.bbc.co.uk/2/hi/asia-pacific/country_profiles/2693697.stm, accessed Feb. 2011b

BBC News. *Regions and territories: Guam*. Online publication, last updated: Feb. 10, 2011, available from: http://news.bbc.co.uk/2/hi/asia-pacific/country_profiles/4102043.stm, accessed Feb. 2011c

Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-reported measures. *Spine* 25, 2000, 3186-91.

Behling O, Law KS. *Translating questionnaires and other research instruments: problems and solutions.* Thousand Oaks: Sage; 2000.

Berman A. *A tradução e a letra ou o albergue longínquo*. Tradução de Torres M, Furlan M, Guerini A. Rio de Janeiro: 7Letras. 2007

Berry JW. On cross-cultural comparability *International Journal of Psychology,* 4, 1969, 119-128

Bland JM, Altman, DG. A note on the use of the intraclass correlation coefficient in the evaluation of agreement between two methods of measurement. *Comput. Biol. Med.* 20, 5, 1990, 337-340

Bland JM, Altman, DG. Statistics notes: Cronbach's alpha *BMJ* 314, 1997, 572

Brislin RW, Freimanis, C. Back-translation: a tool for cross-cultural research. In: Sin-Wai Chan, Pollard DE (eds.) *An encyclopaedia of translation: Chinese-English, English-Chinese*. Chinese University Press. 2001

Brislin RW, Lonner WJ, Thorndike RM, *Cross-cultural research methods*. New York: Wiley. 1973

Brislin RW. Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology*, Vol. 1, No. 3, 1970, 185-216

Brislin RW. The wording and translation of research instruments. In: Lonner WJ, Berry JW, (eds.), *Field methods in cross-cultural research*. Newbury Park, CA: Sage. 1986, 137-164.

Brislin RW. Translation and content analysis of oral and written material. In: Triandis H, Berry JW, (eds.) *Handbook of cross-cultural psychology vol 2*. Boston: Allyn and Bacon. 1980, 389-444

Britto PH. Desconstruir para quê? *Cadernos de Tradução* (UFSC), 2, 8, 2001, 41-50. Available from:
http://www.periodicos.ufsc.br/index.php/traducao/article/view/5883/5563

Britto PH. Fidelidade em tradução poética: o caso Donne. *Terceira Margem.* X, 15, Jul.-Dec. 2006. Available from:
http://www.ciencialit.letras.ufrj.br/terceiramargemonline/numero15/terceiramar gem15.pdf

Britto PH. Tradução e criação, *Cadernos de Tradução* (UFSC), 1, 4, 1999, 239-262. Available from:
http://www.periodicos.ufsc.br/index.php/traducao/article/view/5534/4992

Britto PH. Why deconstruct? Online publication, last updated 21 May 2007, available from:
https://docs.google.com/file/d/0B1HLT669NYMFY2I1OGMzZWMtODczNC0 0MjRhLWJiZDYtOGUyOTZjYzkwNDIy/edit?hl=en_US&pli=1, accessed Aug. 2012

Bueno de Paula, M. *Borges e As mil e uma noites: leitura, tradução e criação.* Doctoral thesis, Florianópolis: UFSC, 2011. Available from:
http://www.pget.ufsc.br/curso/teses/Marcelo_Bueno_de_Paula_-_Tese.pdf

Burisch M. Approaches to personality inventory construction: A comparison of merits. *American Psychologist* 39, 1984, 214-227

Bushnell DM, Martin ML. Quality of life and Parkinson's disease: translation and validation of the US Parkinson's Disease Questionnaire (PDQ-39). *Qual Life Res* 8 1999, 345-50

Butcher JN, Clark LA Recent trends in cross-cultural MMPI research and application. In Butcher JN (Ed.) *New developments in the use of the MMPI.* Minneapolis: University of Minnesota Press. 1979

Butcher JN, Garcia R. Cross-national application of psychological tests. *Personnel and Guidance*, 56, 1978, 472-475

Campbell D, Brislin R, Stewart V, Werner O. Back-translation and other translation techniques in cross-cultural research. Paper to be submitted to the *International Journal of Psychology*. 1970

Campbell DT. A cooperative multinational opinion sample exchange. *Journal of Social Issues*, 24, 1968, 245-258

Cardozo M. Heidermann W, Weininger M. *A escola tradutológica de Leipzig.* Frankfurt am Main: Peter Lang. 2009

Catford J. *A linguistic theory of translation: An essay in applied linguistics.* London: Oxford University Press. 1965

Catford J. *Uma teoria lingüística da tradução: um ensaio de lingüística aplicada.* Translated by the Centro de Especialização de Tradutores de Inglês do Instituto de Letras da Pontifícia Universidade Católica de Campinas; proofreader Maria da Gloria Novak; consultant Antonio Paulo Carrozo. São Paulo: Cultrix; Campinas: Pontifícia Universidade Católica de Campinas. 1980

Chapanis A *Man-machine engineering*. Belmont, CA: Wadsworth. 1965

Chesterman A, Wagner E. *Can theory help translators?: A dialogue between the ivory tower and the wordface*. Manchester: St. Jerome Publishing. 2002

Chesterman A. Description, explanation, prediction: A response to Gideon Toury and Theo Hermans. In Schäffner C (Ed.) *Translation and norms*. Beijing: Foreign Language Teaching and Research Press. 2007 [1999], 90-97

Cohen J. *Statistical power analysis for the behavioural sciences* 2nd ed. Hillsdale: Lawrence Erlbaum Associates, Inc., Publishers 1977

Coracini MJ. O cientista e a noção de sujeito na lingüística: expressão de liberdade ou submissão? In: Arrojo R. (Org.) *O signo desconstruído: Implicações para a tradução, a leitura e o ensino* (2nd ed.) Campinas:Pontes. 2003 (1992), 13-19

Crowne D, Marlowe D. *The approval motive.* New York: John Wiley. 1964.

Crowne D, Marlowe D. The Crowne-Marlow social desirability scale. (1964) In appendix to Tatman AW, Swogger MT, Love K, Cook MD. Psychometric properties of the Marlowe-Crowne social desirability scale with adult male sexual offenders *Sex Abuse* Mar. 2009, 21, 1, 21-34. Available from: http://sax.sagepub.com/content/21/1/21.full.pdf+html

Crowne DP *From response style to motive* ISI Citation Classics, 30, 1991. Available from: http://garfield.library.upenn.edu/classics1991/A1991FW06500001.pdf

Cull A, Sprangers M, Bjordal K, Aaronson N, West K, Bottomley A. *EORTC quality of life group translation procedure* 2nd ed. 2002. Online publication, available from: http://www.ipenproject.org/pdf_file/EORTC_translation.pdf accessed Nov. 2012.

da Silva F. *Análise comparativa dos resultados de mecanismos de tradução automática baseados em regras e estatística*. Masters dissertation, Florianopolis: UFSC. 2010.

Darlington RB. Factor analysis. Cornell University. 1997 Online publication, available from: http://www.psych.cornell.edu/darlington/factor.htm accessed Feb. 2011

David VM, Bousounis G, Kapakoulakis T, Champion R, Masman K, McCullough K. Correlation of MAPT scores with clinical and radiographic assessment of patients awaiting THR/TKR *ANZ Journal of Surgery* 2011 Jul.-Aug.; 81 (7-8): 543-546.

Derrida J. White mythology: Metaphor in the text of philosophy. In *Margins of Philosophy* (translation by Alan Bass). Chicago: The University of Chicago Press. 1982, 207-272.

Dewolf L, Koller M, Velikova G, Johnson C, Scott N, Bottomley A (on behalf of the EORTC Quality of Life Group). *EORTC Quality of life group translation procedure*. EORTC, Brussels, 2009

Díario Catarinense *Segurança*. 2013. Available online from: http://diariocatarinense.clicrbs.com.br/sc/policia/

Dworkin SF, LeResche L. Research diagnostic criteria for temporomandibular disorders: review, criteria, examinations and specifications, critique. *Journal of craniomandibular disorders: facial & oral pain*. 1992, 6, 4, 301-55.

Dworkin SF, LeResche L. *Research diagnostic criteria for temporomandibular disorders.* No date, available from: http://www.rdc-tmdinternational.org/, accessed: Dec. 2012.

EORTC. *Activities: EORTC QLQ-C30 Development*. Online publication, available from: http://groups.eortc.be/qol/qolg_activities.htm accessed Feb. 2011.

EORTC. *Quality of Life Group*. Online publication, available from: http://groups.eortc.be/qol/quality-life-group accessed Jun. 2013.

Even-Zohar, I. The position of translated literature within the literary polysystem. In: Hrushovski B, Even-Zohar I. (Eds.) *Papers on Poetics and Semiotics* 8, 1978, 21-7.

Even-Zohar, I. The position of translated literature within the literary polysystem. In: Even-Zohar, I. (Ed.) *Poetics Today* 11, 1, 1990, 45-51

Ferraz MB. Cross cultural adaptation of questionnaires: what is it and when should it be performed [editorial; comment]? *J Rheumatol* 24, 1997 2066-68

First Clinical *Guidelines for authors* Available from: http://firstclinical.com/journal/guidelines.html Accessed Nov. 2012

Fleck MP, Leal OF, Louzada S, Xavier M, Chachamovich E, Viera G, dos Santos L, Pinzon V. Desenvolvimento da versão em português do instrumento de avaliação de qualidade de vida da OMS (WHOQOL-100) *Rev Bras Psiquiatr*, 21 (1), 1999

Fleuri, LJ. *O perfil ideacional dos itens lexicais translator/tradutor em "Translator Through History" e em "Os Tradutores na História"*. Masters dissertation, Florianópolis: UFSC, 2006. Available from: http://www.pget.ufsc.br/curso/dissertacoes/Lilian_Fleuri_-_Dissertacao.pdf

González YM, Díaz W, Maupome G, Monserratte RS, Rendón E, Zazueta A, Fischman S, Beneduce C, Ohrbach R. *Criterios Diagnósticos para la Investigación de los Trastornos Temporomandibulares*. 2003. Publicação electrónica, available from: http://www.rdc-tmdinternational.org/TMDAssessmentDiagnosis/RDC-TMD/Translations/Spanish.aspx, accessed Dec. 2012.

Google Scholar. *Articles citing Beaton: Guidelines for the process of cross-cultural adaptation of self-report measures*. Google, available from: http://scholar.google.com/scholar?hl=en&lr=&cites=463242340466986131&um=1&ie=UTF-8&ei=YH1uTdiaAoe5twf38LX8Dg&sa=X&oi=science_links&ct=sl-citedby&resnum=6&ved=0CEQQzgIwBQ, accessed Mar. 2011d

Google Scholar. *Articles citing Brislin: Back-translation for cross-cultural research* Google, available from: http://scholar.google.com/scholar?hl=en&lr=&cites=463242340466986131&um=1&ie=UTF-8&ei=YH1uTdiaAoe5twf38LX8Dg&sa=X&oi=science_links&ct=sl-citedby&resnum=6&ved=0CEQQzgIwBQ, accessed Mar. 2011a

Google Scholar. *Articles citing Brislin: Questionnaire wording and translation.* Google, available from: http://scholar.google.com/scholar?cites=10131994531441996907&as_sdt=2005&sciodt=0,5&hl=en, accessed Aug. 2012a

Google Scholar. *Articles citing Brislin: The wording and translation of research instruments.* Google, available from: http://scholar.google.com.br/scholar?hl=en&lr=&cites=10131994531441996907&um=1&ie=UTF-8&sa=X&ei=v_osULObLuXo0QH6o4DgDg&ved=0CEUQzgIwAA, accessed Aug. 2012c

Google Scholar. *Articles citing Brislin: Translation and content analysis of oral and written material.* Google, available from: http://scholar.google.com/scholar?cites=11620388172381542323&as_sdt=2005&sciodt=0,5&hl=en, accessed Aug. 2012b

Google Scholar. *Articles citing Guillemin: Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines* Google, available from: http://scholar.google.com/scholar?cites=16901126221661871113&as_sdt=2005&sciodt=0,5&hl=en, accessed Mar. 2011c

Google Scholar. *Articles citing Perneger: Cross-cultural adaptation of a psychometric instrument: two methods compared*. Google, available from: http://scholar.google.com/scholar?cites=11408110993940574067&as_sdt=2005&sciodt=0,5&hl=en, accessed Apr. 2013

Google Scholar. *Articles citing Sperber: Translation and validation of study instruments for cross-cultural research*. Google, available from: http://scholar.google.com/scholar?hl=en&lr=&cites=1988023338497126542&um=1&ie=UTF-8&ei=vXxuTb34K8KqlAfKiOmTAQ&sa=X&oi=science_links&ct=sl-citedby&resnum=1&ved=0CCIQzgIwAA, accessed Mar. 2011b

Gorsuch R. Measurement: The boon and bane of investigating religion. *American Psycholigist*. 39, 1984, 228-236

Gough HG. *California Psychological Inventory*. Palo Alto, CA: Consulting Psychologists Press. 1969

Grigoletto, M. A constituição do sentido em teorias de leitura e a perspectiva desconstrutivista. In: Arrojo R. (Org.) *O signo desconstruído: Implicações para a tradução, a leitura e o ensino* (2nd ed.) Campinas: Pontes. 2003 (1992), 93-99

Grunwald D, Goldfarb NM, Back Translation for Quality Control of Informed Consent Forms *Journal of Clinical Research Best Practices* 2, 2, 2006

Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J Clin Epidemiol* 46, 1993, 1417-32

Hall E *The silent language* Garden City, NY: Doubleday. 1959

Hermans J. Simoens P, Jansen P. (Eds.) *Taal, vertaling, management: verkenningen in een economisch niemandsland* (Publications of the CERA Chair for Translation, Communication and Cultures 1). Leuven: CERA Chair. 1994.

Hermans T, *Translation in systems: Descriptive and systematic approaches explained*. Manchester: St. Jerome Publishing. 1999

Hermans T. Translation and Normativity. In Schäffner C (Ed.) *Translation and norms*. Beijing: Foreign Language Teaching and Research Press, 2007 [1999], 50-71.

Hönig HG. Translating: The constructive way. *Ilha do Desterro* 33, 1997, 11-23

HRAF *Human Relations Area Files* Available from: http://www.yale.edu/hraf/ accessed Nov. 2012.

Hrushovski B, Even-Zohar I. (Eds.) *Papers on Poetics and Semiotics* 8, 1978, 21-7.

Hurtado Albir A. A aquisição da competência tradutória: aspectos teóricos e didáticos. Translated by Alves F. In Pagano A, Magalhães C, Alves F (eds). *Competência em tradução: cognição e discurso*. Belo Horizonte: Editora UFMG, 2005.

Igoa RL. *Transblanco de Octavio Paz/Haroldo de Campos: análisis paratextual del acuerdo autor/traductor*. Masters dissertation, Florianópolis: UFSC, 2011. Available from:
http://www.pget.ufsc.br/curso/dissertacoes/Rosario_Lazaro_Igoa_-_Dissertacao.doc

IQOLA *SF-36 Health Survey Update*. Online publication, available from: http://www.sf-36.org/tools/SF-36.shtml, accessed Nov. 2012

IQOLA *The IQOLA Project*. Online publication, available from: http://www.iqola.org/project.aspx, accessed Feb. 2011a.

IQOLA *The SF Instruments*. Online publication, available from: http://www.iqola.org/instruments.aspx, accessed Feb. 2011b.

Jakobson, R. Aspectos lingüisticos da tradução. In: Jakobson, R. *Linguistica e comunicação*. Translated by Bilkstein I, Paes JP. São Paulo: Cultrix, 1971, 63-72

Kaiser-Cooke M. Translatorial expertise – a cross-cultural phenomenon from an inter-disciplinary perspective. In Snell-Hornby M, Pöchhacker F, Kaindl K (Eds.) *Translation studies – an interdiscipline: selected papers from the Translation Studies Congress, Vienna 9-12 September 1992* Amsterdam: John Benjamins. 1994, 135-139

Kirtland KA, Porter DE, Addy CL, Neet MJ, Williams JE, Sharpe PA, Neff LJ, Kimsey CD, Ainsworth BE. Environmental measure of physical activity supports: perception versus reality. *American Journal of Preventive Medicine* 2003; 24 (4):323-331

Koller M, Aaronson NK, Blazeby J, Bottomley A, Dewolf L, Fayers P, Johnson C, Ramage J, Scott N, West K (on behalf of the EORTC Quality of Life Group). *Translation procedures for standardised quality of life questionnaires: The European Organisation for Research and Treatment of Cancer (EORTC) approach.* Eur J Cancer. 43(12):1810-1820, 2007. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0959-8049(07)00462-5

Lambert J, Hermans J. From translation markets to language management: the implications of translation services. In Delabastita D, D'hulst L and Meylaerts R (eds) *Functional approaches to culture and translation: Selected papers by José Lambert* Amsterdam: John Benjamins. 2006 (1995), 131-146

Lambert J. and Van Gorp H On describing translations. In Delabastita D, D'hulst L and Meylaerts R (eds) *Functional approaches to culture and translation: Selected papers by José Lambert* Amsterdam: John Benjamins. 2006 (1985), 37-47

Lambert J. In quest of literary world maps. In Delabastita D, D'hulst L and Meylaerts R (eds) *Functional approaches to culture and translation: Selected papers by José Lambert* Amsterdam: John Benjamins. 2006 (1991), 63-74

Lambert J. Language and translation as management problems: A new task for education, in Dollerup C and Appel V (eds) Teaching translation and interpreting 3 Amsterdam: John Benjamins. 1996

Lambert J. Literatures, translation and (de)colonization. In Delabastita D, D'hulst L and Meylaerts R (eds) *Functional approaches to culture and translation: Selected papers by José Lambert* Amsterdam: John Benjamins. 2006 (1995), 87-103

Lambert J. Twenty years of research on literary translation at the Katholieke Universiteit Leuven. In Delabastita D, D'hulst L and Meylaerts R (eds) *Functional approaches to culture and translation: Selected papers by José Lambert* Amsterdam: John Benjamins. 2006 (1988), 49-62

Lefevere A. *Translation, rewriting and the manipulation of literary* fame London/NewYork: Routledge, 1992

List T, Huggins K, Dworkin S. *Diagnostiska Kriterier för Temporomandibulär Dysfunktion*. 2003, Online publication, available from: http://www.rdc-tmdinternational.org/TMDAssessmentDiagnosis/RDC-TMD/Translations/Swedish.aspx, accessed Dec. 2012

Lobbezoo F, van Selms MKA, van Os-Ravesloot AL. *Onderzoekscriteria voor de Diagnostiek van Temporomandibulaire Dysfunctie*. 2004. Online publication, available at: http://www.rdc-tmdinternational.org/TMDAssessmentDiagnosis/RDC-TMD/Translations/Dutch.aspx, accessed Dec. 2012

Lörscher W (org.) *Translation studies in Germany: Ilha do Desterro* 33 1997, 39-53

Lyssenko, V. Jonsson, A. Almgren: Pulizzi, N. Isomaa, B. Tuomi, T. Berglund, G. Altshuler, D. Nilsson: and Groop, L. Clinical Risk Factors, DNA Variants, and the Development of Type 2 Diabetes. *N Engl J Med* 2008; 359:2220-2232, Nov. 20 2008. DOI: 10.1056/NEJMoa0801869

Macri LA, Deli V, Deli R. Development of a reliable and clinically useful Italian version of the Axis II of the Research Diagnostic Criteria for Temporomandibular Disorders (RDC-TMD). *Italian Journal of Public Health* 2008, 6, 5, 2: 165-167

McConnell S, Beaton DE, Bombardier C. *The DASH outcome measure: a user's manual*. Toronto, Ontario: Institute for Work & Health, 1999.

Michelotti A, Beneduce C, Ferretti M, Ferretti L, Ohrbach R. Criteri Diagnostici di Ricerca per i Disordini Temporomandibolari. 2002. Online publication, available from: http://www.rdc-tmdinternational.org/TMDAssessmentDiagnosis/RDC-TMD/Translations/Italian.aspx, accessed Dec. 2012

Miller G. What is information measurement? *American Psychologist*, 9, 1953, 3-11

Minaglia S, Ozel B, Nguyen JN, Mishell jr DR. Validation of Spanish version of pelvic pain and urgency/frequency (PUF) patient symptom scale. *Urology*, 65, 4, 2005, 664-9

Montalt V, Gonzalez Davies M. *Medical translation step by step: learning by drafting*. Manchester: St. Jerome Publishing. 2007.

National Health Service (UK). *Different levels of evidence* National Health Service, publicação eletrônica, last updated: Sep. 1, 2010, available from: http://www.patient.co.uk/doctor/Different-Levels-of-Evidence-(Critical-Reading).htm, accessed Nov. 2010

Nida E. *The theory and practice of translation*. Leiden: E J Brill. 1969

Nida E. Theories of translation. *TTR: traduction, terminologie, rédaction*, 4, 1, 1991: 19-32.

Nida E. *Towards a science of translating: with special reference to principles and procedures involved in Bible translating*. Leiden: E J Brill. 1964

NIDCR. *Prevalence of TMJD and its signs and symptoms.* Online publication, available from: http://www.nidcr.nih.gov/datastatistics/finddatabytopic/facialpain/prevalencetmj d.htm, accessed Jun. 2013.

Nilsson P, Baigi A, Marklund B, Månsson J. Cross-cultural adaptation and determination of the reliability and validity of PRTEE-S (Patientskattad Utvärdering av Tennisarmbåge), a questionnaire for patients with lateral epicondylalgia, in a Swedish population. *BMC Musculoskeletal Disorders* 79, 9, 2008

Nord C. Defining translation functions: the translation brief as a guideline for the trainee translator, *Ilha do Desterro* 33, 1997b, 39-53

Nord C. *Text analysis in Translation: theory, methodology, and didactic application of a model for translation-oriented text analysis*, (2nd ed.) translated by Nord C, Sparrow P. Amsterdam: Editions Rodopi BV. 2005 [1991]

Nord C. Training functional translators. *Cadernos de Tradução* (UFSC) 1, 5, 2000. Available from: http://www.periodicos.ufsc.br/index.php/traducao/article/view/5606/5083

Nord C. *Translating as a purposeful activity: functionalist approaches explained*. Manchester: St. Jerome Publishing. 1997a

Ohrbach R. *Introduction to Translation and Cultural Equivalency*. Online publication, last updated: Aug., 2007, available from http://www.rdc-tmdinternational.org/Portals/18/Guidelines/Introduction%20to%20Translation%20and%20Cultural%20Equivalency.pdf

Ozolins U. Back translation as a means of giving translators a voice. *Interpreting and Translation* 1, 2, 2009, 1-13

Pagano A, Magalhães C, Alves F (eds). *Competência em tradução* Belo Horizonte: Editora UFMG. 2005

Parra FC, Amado RC, Lambertucci JR, Rocha J, Antunes CM, Pena SDJ. Color and genomic ancestry in Brazilians. *Proc Natl Acad Sci U S A.* 100, 1, 2003, 177-182

Pena, SDJ, Bastos-Rodrigues L, Pimenta JR, Bydlowski SP. DNA tests probe the genomic ancestry of Brazilians. *Braz J Med Biol Res*, 42, 10, 2009, available from http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-879X2009001000001&lng=en&nrm=iso accessed Feb. 2011

Pereira FJ, Huggins KH, Dworkin SF, Ohrbach R, Favilla E. *Critérios de Diagnóstico para Pesquisa das Desordens Temporomandibulares*. 2009 [2004], Online publication, available from: http://www.rdc-tmdinternational.org/TMDAssessmentDiagnosis/RDC-TMD/Translations/Portuguese(Brazil).aspx, accessed Dec. 2012

Pereira Júnior FJ, Favilla EE, Dworkin S, et al. Critérios de diagnóstico para pesquisa das disfunções temporomandibulares (RDC-TMD). Tradução oficial para a língua portuguesa. J Bras Clin Odontol Integr. 2004; 8(47):384-95.

Perneger TV, Etter JF, Rougemont A. Prospective versus retrospective measurement of change in health status: a community based study in Geneva, Switzerland *Journal of Epidemiology and Community Health* 51, 3 1997, 320-325

Perneger TV, Etter JF, Rougemont A. Switching Swiss enrollees from indemnity health insurance to managed care: the effect on health status and satisfaction with care. *Am J Public Health*. 1996 March; 86(3): 388–393.

Perneger TV, Leplège A, Etter JF. Cross-cultural adaptation of a psychometric instrument: two methods compared. *J Clin Epidemiol*. 52, 11, 1999, 1037-46

Petterson, B. The Postcolonial Turn in Literary Translation Studies: Theoretical Frameworks Reviewed. *Canadian Aesthetics Journal* 4, summer 1999. Available from http://www.uqtr.uquebec.ca/AE/vol_4/petter.htm

Polchlopek, SA. *O Mundo pós "11 de setembro": Tecendo fios/textos entre tradução e a narratividade jornalística*. Doctoral thesis, Florianópolis: UFSC, 2011. Available from: http://www.pget.ufsc.br/curso/teses/Silvana_Ayub_Polchlopek_-_Tese.pdf

Przeworski A, Teune, H. Equivalence in cross-national research *Public Opinion Quarterly*, 30, 1966, 33-43

Przeworski A, Teune, H. *The logic of social comparative enquiry* New York: Wiley 1970

PubMed *Search results: ((back-translation) OR back translation) OR "back translation"* PubMed. Online, at:
http://www.ncbi.nlm.nih.gov/pubmed?term=((back-translation) OR back translation) OR "back translation"&cmd=correctspelling Accessed June 2013.

Pym A. *Doubts about deconstruction as a general theory of translation*. TradTerm (São Paulo) 2 (1995), 11-18. Available from:
http://usuaris.tinet.cat/apym/on-line/research_methods/decon.html

Pym A. *Epistemiological problems in translation and its teaching* Calaceit: Edicions Caminade. 1993

Pym A. Okay, so how are translation norms negotiated? A question for Gideon Toury and Theo Hermans. In Schäffner C (Ed.) *Translation and norms*. Beijing: Foreign Language Teaching and Research Press, 2007 [1999], 106-112

Rajagopalan K, Arrojo R. A crise da metalinguagem: uma perspectiva interdisciplinar. In: Arrojo R. (ed) *O signo desconstruído: Implicações para a tradução, a leitura e o ensino* (2nd ed.) Campinas:Pontes. 2003 (1992), 57-63

Rajagopalan K. A trama do signo: O significado e sua gênese: algumas anotações avulsas. In: Arrojo R. (ed) *O signo desconstruído: Implicações para a tradução, a leitura e o ensino* (2nd ed.) Campinas:Pontes. 2003 (1992a), 41-47

Rajagopalan K. O conceito de interpretação na lingüística: seus alicerces e seus desafios. In: Arrojo R. (ed) *O signo desconstruído: Implicações para a tradução, a leitura e o ensino* (2nd ed.) Campinas:Pontes. 2003 (1992b), 63-67

Rand Health. *36-Item Short Form Survey from the Medical Outcomes Study*. Rand Health, available from:
http://www.rand.org/health/surveys_tools/mos/mos_core_36item.html, accessed, Jan. 2012

Rand Health. *Medical Outcomes Study: 36-Item Short Form Survey Instrument*. Rand Health, available from:
http://www.rand.org/health/surveys_tools/mos/mos_core_36item_survey.html, accessed Mar. 2010

RDC-TMD *Spanish*. Online publication, available from: http://www.rdc-tmdinternational.org/TMDAssessmentDiagnosis/RDC-TMD/Translations/Spanish.aspx, accessed Dec. 2012.

Robinson E. The racket with standardized test scores. *The Washington Post*. 1 Apr. 2013. Available at: http://www.washingtonpost.com/opinions/eugene-robinson-the-racket-with-standardized-test-scores/2013/04/01/1bbd61ea-9b01-11e2-a941-a19bce7af755_story.html

Ruuskanen DDK. The translation of Finnish medical texts: who is the expert? In: Snell-Hornby M, Pöchhacker F, Kaindl K (eds) *Translation studies – an interdiscipline: selected papers from the Translation Studies Congress, Vienna 9-12 September 1992*. Amsterdam: John Benjamins. 1994, 291-300

SAGE Journals Online. *Articles citing this article: Back-translation for cross-cultural research.* SAGE Publications, available from: http://jcc.sagepub.com/content/1/3/185.short?rss=1&ssource=mfc, accessed Feb. 2011

Schäffner C (Ed.) *Translation and norms*. Beijing: Foreign Language Teaching and Research Press, 2007 [1999]

Schäffner C The concept of norms in translation studies. In Schäffner C (Ed.) *Translation and norms*. Beijing: Foreign Language Teaching and Research Press, 2007 [1999], 1-8.

Schleiermacher F Sobre os diferentes métodos de tradução. Translation by Margarete von Mühlen Poll. In Heidermann W (Ed) *Clássicos da teoria da tradução: Antologia bilíngüe* Florianópolis: UFSC/NUT, 2001

Scientific Linguagem *Empresa: Scientific* available from: http://www.scientific.com.br/empresa/, accessed Feb. 2011

Sheppard N. *Alitjinya ngura tjukurmankuntjala* [Alitji in Dreamland]. Berkely: Ten Speed Press, 1992

SIP 4-99 Research Group. *Environmental Supports for Physical Activity Questionnaire*. Prevention Research Center, Norman J. Arnold School of Public Health, University of South Carolina 2002, available from: http://prevention.sph.sc.edu/tools/Env_Supports_for_PA.rtf, accessed Feb. 2011.

Sperber A, DeVellis R, Boehlecke B. Cross-cultural translation: methodology and validation. *J Cross-Cult Psychol* 25, 4, 1994, 501-524

Steiner P. *Russian Formalism*. Ithaca and London: Cornell University Press, 1984

Swift J. *Gulliver's travels into several remote nations of the world*. Project Gutenberg eBook, available at www.gutenberg.org, first published 1726-7, this edition 1892.

Tatman AW, Swogger MT, Love K, Cook MD. Psychometric properties of the Marlowe-Crowne social desirability scale with adult male sexual offenders *Sex Abuse* Mar. 2009, 21, 1, 21-34

available from: http://sax.sagepub.com/content/21/1/21.full.pdf+html

Toury G. A handful of paragraphs on 'translation' and 'norms'. In Schäffner C (Ed.) *Translation and norms*. Beijing: Foreign Language Teaching and Research Press, 2007 (1999), 9-31

Toury G. A Rationale for Descriptive Translation Studies. In Hermans T. (Ed.) *The Manipulation of Literature: Studies in Literary Translation*, London & Sydney: Croom Helm. 1985, 16-41.

Toury G. The nature and role of norms in translation. In: Toury G *Descriptive translation studies and beyond* Amsterdam/Philadelphia: John Benjamins Publishing Company. 1995a, 53-71

Toury G. The notion of 'Assumed Translation' - An invitation to a new discussion. In Bloemen H, Hertog E, Segers W. (Eds.) *Letterlijkheid, Woordelijheid / Literality, Verbality*. Antwerpen/Harmelen: Fantom. 1995b, 135-147. Available from: http://www.tau.ac.il/~toury/works/assumed.htm

Travassos C, Williams, DR, The concept and measurement of race and their relationship to public health: a review focused on Brazil and the United States. Rio de Janeiro: *Cad. Saúde Pública* 20, 3, 2004, 660-678

Trochim WM. Convergent & discriminant validity *Research Methods Database*. Online publication, last updated: Oct. 20, 2006, available from: http://www.socialresearchmethods.net/kb/convdisc.php, accessed Nov. 2012

U.S. Preventive Services Task Force. *Appendix A: taskforce ratings in guide to clinical preventive services: report of the U.S. Preventive Services Task Force*, 2nd ed. Lippincott Williams & Wilkins 1996, 860-865

Available from: http://odphp. osophs.dhhs.gov/pubs/guidecps/PDF/APPA.PDF

Venuti L. *The scandals of translation: Towards an ethics of difference*. London: Routledge. 1998.

Venuti L. *The translator's invisibility: A history of translation* (1st ed.) New York: Routledge. 1995

Venuti L. *The translator's invisibility: A history of translation*. Taylor & Francis e-Library. 2004 (1995).

Venuti L. *The translator's invisibility: A history of translation* (2nd ed.) New York: Routledge. 2008 (1995)

Vermeer HJ. *Skopos und Translationsauftrag – Aufsätze*. Heidleberg: Universiẗst (thw – translatorishes handeln wissenschaft 2). 2nd ed. 1989

Vermeer HJ. Translation today: old and new problems. In Snell-Hornby M, Pöchhacker F, Kaindl K (eds) *Translation studies – an interdiscipline: selected papers from the Translation Studies Congress, Vienna 9-12 September 1992* Amsterdam: John Benjamins. 1994, 3-16

Viaggio S. The limitations of the strictly socio-historical description of norms: A response to Theo Hermans and Gideon Toury. In Schäffner C (ed) *Translation and norms*. Beijing: Foreign Language Teaching and Research Press. 2007 [1999], 122-128

Vitnay K. The New ASTM Translation Guide *The Translation Directory* (2007) Online publication, available from: http://www.translationdirectory.com/articles/article1315.php Last updated 2007, accessed Nov. 2012

von Wright GH. *Norm and Action. A logical Enquiry*. London: Routledge & Kegan Paul. 1963.

Ware JE. *SF-36® Health Survey Update* SF-36.org. Online publication, available from: http://www.sf-36.org/tools/SF-36.shtml, accessed Feb. 2011

WebMD *Temporomandibular Disorders* WebMD. Online publication, available from: http://www.webmd.com/oral-health/temporomandibular-disorders, accessed Feb. 2013

Werner O, Campbell DT. Translating, working through interpreters, and the problem of decentering. In Naroll R, Cohen R (eds.) *A handbook of method in cultural anthropology*. New York: American Museum of Natural History. 1970, 398-420

Wesley F, Karr C. Problems in establishing norms for cross-cultural comparisons. *International Journal of Psychology*, 1, 1966, 257-262.

WHO – World Health Organization *Management of substance abuse: Process of translation and adaptation of instruments*. Online publication, available from: www.who.int/substance_abuse/research_tools/translation/en/, accessed Feb. 2012

WHOQOL Group. Division of Mental Health, World Health Organization. Study protocol for the World Health Organization project to develop a Quality of Life assessment instrument (WHOQOL). *Qual Life Res* 1993; 2: 153–159.

Wikipedia. *Brazil.* Wikipedia. Online publication, available from: http://en.wikipedia.org/wiki/Brazil, accessed Feb. 2013a.

Wikipedia. *Byte*. Wikipedia. Online publication, available from: http://en.wikipedia.org/wiki/Byte, accessed Feb. 2013b.

Wikipedia. *Chamorro*. Wikipedia. Online publication, available from: http://en.wikipedia.org/wiki/Chamorro_language accessed May. 2013f.

Wikipedia. *Glasgow Coma Scale*. Wikipedia. Online publication, available from: http://en.wikipedia.org/wiki/Glasgow_Coma_Scale, accessed Mar. 2013d.

Wikipedia. *Guam*. Wikipedia. Online publication, available from: http://en.wikipedia.org/wiki/Guam accessed May. 2013e.

Wikipedia. *Immigration to Italy*. Online publication, available from: http://en.wikipedia.org/wiki/Immigration_to_Italy, accessed Dec. 2012b.

Wikipedia. *List of countries by GDP (nominal)*. Wikipedia. Online publication, available from: http://en.wikipedia.org/wiki/List_of_countries_by_GDP_(nominal)#Graph_of_l argest_economies, accessed Feb. 2013c.

Wikipedia. *Race in Brazil*. Wikipedia. Online publication, available from: http://en.wikipedia.org/wiki/Race_in_Brazil#cite_ref-54, accessed Feb. 2011.

Wikipedia. *The Aleut people*. Online publication, available from:
http://en.wikipedia.org/wiki/Aleut_people, accessed Dec. 2012a.

APPENDIX A – ENVIRONMENTAL SUPPORTS FOR PHYSICAL ACTIVITY QUESTIONNAIRE

ENVIRONMENTAL SUPPORTS FOR PHYSICAL ACTIVITY QUESTIONNAIRE

(SOCIAL AND PHYSICAL ENVIRONMENT SURVEY)

Prevention Research Center
Norman J. Arnold School of Public Health
University of South Carolina

Principal Investigator:
Barbara E. Ainsworth, PhD, MPH

Section 3: Social and Physical Environments

"I will be asking you some questions about the neighborhood in which you live, followed by some questions about the community in which you live."

"First, some questions about the neighborhood in which you live. For the purpose of this interview, neighborhood is defined as the area within one-half mile or a ten-minute walk from your home.

3.1   How long have you lived at your current address?

Number of months (twelve months or less)

___ ___

Number of years (one year or more)...................................... ___ ___

Don't know/ not sure.. 77

Refused.. 99

<Note to interviewer: Less than one year is entered as months and more than twelve months is entered as whole years only. E.g. 5 years, not 5 years and 4 months.>

3.2   In general, would you say that the people in your neighborhood are….

a.   Very physically active... 1

b.   Somewhat physically active 2

c.   Not very physically active.. 3

d.   Not at all physically active 4

Don't know/ not sure.. 7

Refused.. 9

3.3   Overall, how would you rate your neighborhood as a place to walk?

Would you say…

a.   Very pleasant 1

b.   Somewhat pleasant... 2

c.   Not very pleasant 3

d.   Not at all pleasant.. 4

Don't know/ not sure.. 7

Refused.. 9

Section 3: Social and Physical Environments

3.4   In general, would you say the motorized traffic in your neighborhood is…

   a.   Heavy,.. 1

   b.   Moderate, OR.. 2

   c.   Light?.. 3

   Don't know/not sure.. 7

   Refused 9

3.5   Does your neighborhood have any sidewalks?

   a.   Yes... 1

   b.   No *(Skip to question 3.7)*.. 2

   Don't know/not sure.. 7

   Refused 9

3.6   For walking in your neighborhood, would you say your sidewalks are…

   a.   Very well maintained.. 1

   b.   Somewhat maintained... 2

   c.   Not very well maintained.. 3

   d.   Not at all maintained.. 4

   Don't know/ not sure.. 7

   Refused 9

3.7   For walking at night, would you describe the STREET lighting in your
      neighborhood as…

   a.   Very good 1

   b.   Good 2

   c.   Fair... 3

   d.   Poor.. 4

   e.   Very poor 5

   Don't know/ not sure.. 7

   Refused 9

Section 3: Social and Physical Environments

3.8 For walking in your neighborhood, would you say that unattended dogs are….

    a.   A big problem 1

    b.   Somewhat of a problem 2

    c.   Not very much of a problem 3

    d.   Not a problem at all.. 4

    Don't know/ not sure 7

    Refused.. 9

3.9 How safe from crime do you consider your neighborhood to be?
Would you say…

    a.   Extremely safe 1

    b.   Quite safe 2

    c.   Slightly safe 3

    d.   Not at all safe 4

    Don't know/not sure 7

    Refused.. 9

3.10 Generally speaking, would you say most people in your neighborhood can be
trusted?

    a.   Yes 1

    b.   No 2

    Don't know/not sure 7

    Refused.. 9

3.11 Does your neighborhood have public recreation facilities (such as public swimming
pools, parks, walking trails, bike paths, recreation centers, etc.)?

    a.   Yes 1

    b.   No *(Skip to question 3.13)* 2

    Don't know/not sure 7

    Refused.. 9

Section 3: Social and Physical Environments

3.12 In general, how would you rate the condition of these public recreation facilities?
Would you say…

    a.   Excellent 1

    b.   Good 2

    c.   Fair... 3

    d.   Poor.. 4

    Don't Know/Not Sure 7

    Refused 9

3.13 Thinking about how public money is spent on recreation facilities, which of the
following statements is most accurate…

    a.   My neighborhood almost always gets its fair share........ 1

    b.   My neighborhood often gets its fair share...................... 2

    c.   My neighborhood seldom gets its fair share................... 3

    d.   My neighborhood never gets its fair share 4

    Don't Know/Not Sure 7

    Refused 9

3.14 For physical activity, do you use ANY private or membership only recreation
facilities?

(… including those outside of your neighborhood)

    a.   Yes... 1

    b.   No 2

    Don't know/not sure... 7

    Refused 9

Section 3: Social and Physical Environments

"For the next several questions, think about the community in which you live. For the purposes of this interview, community is defined as the area within ten miles or a twenty-minute drive from your home.

"Please tell me if you yourself USE any of the following resources and facilities in your community. If the type resource or facility I mention is not available in your community, please let me know."

<Note to interviewer: Emphasize you/yourself. The question is asking about personal use, not their family or community's use.>

3.15 Walking trails?

    a.   Yes – R USES WALKING TRAILS IN COMMUNITY 1

    b.   No – R <u>DOES</u> <u>NOT</u> USE WALKING TRAILS IN COMMUNITY 2

    c.   My community does not have any walking trails 3

    Don't know/ not sure 7

    Refused... 9

PROBE: BY WALKING TRAILS WE MEAN PUBLIC TRAILS THAT ARE DESIGNATED FOR WALKING.

3.16 Public swimming pools?

    a.   Yes - R USES POOLS IN COMMUNITY.. 1

    b.   No- R <u>DOES</u> <u>NOT</u> USE POOLS IN COMMUNITY 2

    c.   My community does not have any public swimming pools...................... 3

    Don't know/ not sure 7

    Refused... 9

3.17 Public Recreation Centers?

    a.   Yes - R USES PUBLIC RECREATION CENTERS IN COMMUNITY 1

    b.   No- R <u>DOES</u> <u>NOT</u> USE PUBLIC RECREATION CENTERS IN
        COMMUNITY 2

    c.   My community does not have any public recreation centers.................... 3

    Don't know/ not sure 7

Refused   9

Section 3: Social and Physical Environments

3.18 Bicycle paths or bike trails?

    a.   Yes - R USES BIKE TRAILS IN COMMUNITY..1

    b.   No- R <u>DOES</u> <u>NOT</u> USE BIKE TRAILS IN COMMUNITY........................2

    c.   My community does not have any bike paths or bike trails......................3

Don't know/ not sure...7

Refused 9

3.19 Parks/playgrounds/sports fields?

    a.   Yes - R USES PARKS IN COMMUNITY 1

    b.   No - R <u>DOES</u> <u>NOT</u> USE PARKS IN COMMUNITY..................................2

    c.   My community does not have any parks/playgrounds/sports fields 3

Don't know/ not sure...7

Refused 9

3.20 Schools that are open for public recreation activities?

    a.   Yes- R USES SCHOOLS FOR REC IN COMMUNITY 1

    b.   No - R <u>DOES</u> <u>NOT</u> USE SCHOOLS FOR REC IN COMMUNITY 2

    c.   Schools in my community are not open for the public to use...................3

Don't know/ not sure...7

Refused 9

3.21 Do you use a shopping mall for physical activity/walking programs?

    a.   Yes- R USES MALLS FOR PA IN COMMUNITY 1

    b.   No- R <u>DOES</u> <u>NOT</u> USE MALLS FOR PA IN COMMUNITY 2

    c.   My community does not have a shopping mall..3

Don't know/ not sure...7

Refused 9

Section 3: Social and Physical Environments

3.22 Do you use physical activity programs and facilities at a place of worship?

    a.   Yes- R USES FACILITIES AT PLACE OF WORSHIP IN COMMUNITY . 1

    b.   No- R <u>DOES</u> <u>NOT</u> USE FACILITIES AT PLACE OF WORSHIP IN

        COMMUNITY 2

    c.   My community does not have any places of worship with physical

        activity programs.. 3

    Don't know/ not sure 7

    Refused... 9

3.23 Do you use nearby waterways such as creeks, rivers, and lakes for water-related physical activities such as canoeing, kayaking, swimming, or skiing? (DO NOT INCLUDE NON-PHYSICAL ACTIVITIES SUCH AS BOATING)

    a.   Yes- R USES WATERWAYS FOR PA IN COMMUNITY 1

    b.   No- R <u>DOES</u> <u>NOT</u> USE WATERWAYS FOR PA IN COMMUNITY......... 2

    c.   My community does not have any waterways to use for physical

        activity 3

    Don't know/ not sure 7

    Refused... 9

"The next questions concern your opinion about physical activity facilities in your community."

3.24 For your own physical activity, how important are recreational/ physical activity clubs, programs, or organized recreational events in your community…

    a.   Very important 1

    b.   Somewhat important 2

    c.   Not very important... 3

    d.   Not at all important 4

    e.   My community does not have any physical activity clubs or programs ... 5

    Don't know/ not sure 7

    Refused... 9

Section 3: Social and Physical Environments

3.25 In your community, would you say that all people have equal access to public recreation facilities?

    a.   Yes..1

    b.   No 2

    My community does not have any public recreation facilities..........................3

    (Skip to question 4.1)

    Don't Know/Not Sure 7

    Refused 9

3.26 How safe are the public recreation facilities in your community? Would you say…

    a.   Very safe..1

    b.   Somewhat safe 2

    c.   Somewhat unsafe 3

    d.   Not at all safe 4

    Don't Know/Not Sure 7

    Refused 9

3.27 Do concerns about safety at the public recreation facilities in your community influence your using them?

    a.   Yes..1

    b.   No 2

    c.   My community does not have any public recreation facilities 3

    Don't Know/Not Sure 7

    Refused 9

APPENDIX B – TRANSLATION BRIEF – PARALLEL, COOPERATIVE, CONSENSUS-BASED METHOD

Parallel, cooperative, consensus-based translation of the "Environmental Supports for Physical Activity Questionnaire: (Social and Physical Environment Survey)" from English into Brazilian Portuguese

The source text

Target text addressees

The questionnaire is designed to be administered by an interviewer, who will complete the answers in the spaces provided. Therefore, the questions and all answers except "refused" are addressed to the interviewee. The interviewer, however, must interpret the response and assign it to one of the options provided. In addition to the questions and possible responses, there are notes to the interviewer to explain how certain items should be interpreted and instructions to the interviewer on the sequence of questions.

The questionnaire was designed for adults (the validation sample was aged 18-96). There are no other exclusion criteria, so the target population for this translation is the adult Brazilian-Portuguese-speaking population.

Text reception

The questionnaire is designed to be administered by an interviewer by telephone.

Motive for translation

This questionnaire is being translated as part of a doctoral research project that aims to demonstrate that competent translators working in the areas in which they have experience are capable of producing very high quality translations of medical research instruments without the

need for back-translation. The motivation for this is to raise awareness of the true level of professional translation competence available and to contribute to improving the professional standing of competent translators in the eyes of health professionals and health sciences researchers.

Text function

The original questionnaire was developed in order to investigate the relationship between respondents' perceptions about their neighbourhoods and communities and their levels of physical activity. A brief description of the questionnaire can be found here: http://www.activelivingresearch.org/node/10645 and the validation study can be found here: http://www.ajpm-online.net/article/S0749-3797(03)00021-7/fulltext.

The translation has the same intended function as the original, with the only differences being the target population, their native language and the country of administration (Brazil).

The translation protocol

Since the objective of the project is to demonstrate translator competence, there are no restrictions whatsoever on the methods each translator may use to produce her initial translation: reference materials, translation memory, internet searches, consultation with experts, etc. are all acceptable. You also have permission to e-mail the original author to ask for clarification of any ambiguities, but all decisions on translation solutions should be your own. The purpose of these instructions is not to police (*fiscalizar*) the translators, but to provide a framework for the parallel, cooperative, consensus-based translation method and to ensure documentation of the process for the purposes of the research project.

Stage 1: First translations

Each translator will produce her own first translation working on her own. As stated above – you may use whatever resources you see fit, with the single exception of consulting with the other translator (at this point). The main priorities are clarity and consistency of terminology, particularly for the Likert scales. If software other than Microsoft Word is used for the translation, please export the translation to Word and save it as a .doc file (Word 97 – 2003 format). Each translator should then email her translation to the other translator and send a copy to me.

Stage 2: Revision of first translations

Before starting, please ensure that "track changes" is enabled (*controlar alterações* in Portuguese Word).

Each translator should review the other's translation and correct anything they consider to be an error, whether because they believe there is a problem with the Portuguese or because they believe the translator has misunderstood the original. Please make these corrections using blue type. Each translator should also indicate any suggestions they have for improving the other's translation. These improvements

may be to increase the clarity or readability of the Portuguese or in order to achieve a better translation. Please make these improvements in green type.

If you wish to explain why you have made any or all of the alterations, please do so in a comment. There is not, however, any obligation whatsoever to explain any of your alterations, but the other translator will later have the opportunity to reject or accept them and explanations could improve the chances of her retaining your suggestions.

Please do not correct or improve just for the sake of it. If the translation you are revising does not have any errors and you cannot see any way of improving it, then please simply send it back to me unaltered.

Stage 3: Rejecting or accepting alterations.

I will send each translator the revised version of her translation. Each translator is now at liberty to accept or reject the alterations made by the other translator to the extent she sees fit; up to and including rejecting all of the alterations and returning to her original version. Please send me the new version if you have rejected any but not all of the alterations. If you have accepted or rejected all of the alterations, simply inform me of this fact by email.

Stage 4: Scoring the revised (or not) translations

For the purposes of this exercise, I shall refer to "items". What I mean by an "item" is a section of text bounded by two carriage returns. In the example below, Question 3.23 is therefore made up of 6 items:

3.23 Do you use nearby waterways such as creeks, rivers, and lakes for water-related physical activities such as canoeing, kayaking, swimming, or skiing? (DO NOT INCLUDE NON-PHYSICAL ACTIVITIES SUCH AS BOATING)

    a.   Yes- R USES WATERWAYS FOR PA IN COMMUNITY 1

    b.   No- R <u>DOES</u> <u>NOT</u> USE WATERWAYS FOR PA IN COMMUNITY......... 2

    c.   My community does not have any waterways to use for physical activity 3

Don't know/ not sure..7

Refused 9

I will send each translator the other translator's final version and each translator will then compare the two versions. Each translator should first indicate any items where she feels strongly that her version is superior to the other translator's version. She can do this using highlighting. Once she has identified all such items she should give each of these items a number, from 1 to the total number of items she has selected, starting with the item about which she feels most strongly and ending with the item about which she feels least strongly.

Once more, please do not indicate items simply to avoid not selecting anything. If you do not consider that any of your items are sufficiently better than the other translator's to warrant substitution, then please simply inform me of this by email. I shall use these scores for the purposes of analysis and in the event that consensus cannot be reached in stage 5.

Stage 5: Producing the final version

Working together, the translators will produce a final version that they both agree is the best that can be achieved.

There is no obligation to restrict yourselves to solutions that have been suggested in the previous stages if by working together you produce a new version that is different from any of the previous versions and which both of you agree is better than previous versions.

If you are truly unable to reach consensus on any of the items, you should indicate this, including each translator's preferred solution, and explanations or not as you see fit, and send me the near-consensus that you have arrived at. If you are able to reach a consensus, you should send me the final version. In either case, your work is complete.

Thank you both.

APPENDIX C – FIRST TRANSLATION - DENISE

QUESTIONÁRIO SOBRE O PAPEL DO AMBIENTE NA REALIZAÇÃO DE ATIVIDADE FÍSICA

(PESQUISA SOBRE AMBIENTES SOCIAIS E FÍSICOS)

Centro de Pesquisa em Prevenção
Faculdade de Saúde Pública Norman J. Arnold
University of South Carolina

Investigadora principal:
Barbara E. Ainsworth, PhD, MPH

Seção 3: Ambientes Sociais e Físicos

"Farei algumas perguntas sobre o bairro onde você mora, seguidas de perguntas sobre a comunidade onde você vive."

"Em primeiro lugar, algumas perguntas sobre o seu bairro. Nesta entrevista, bairro é definido como a área que fica a uma distância de até 800 metros ou 10 minutos caminhando de sua casa.

3.1 Há quanto tempo você vive no seu endereço atual?

Número de meses (doze meses ou menos) .... ___ ___

Número de anos (um ano ou mais) .... ___ ___

Não sei / não tenho certeza 77

Recusou-se a responder

     99

<Nota ao entrevistador: Menos de um ano deve ser informado em número de meses, e mais de 12 meses deve ser informado apenas em número de anos inteiros. Exemplo: 5 anos em vez de 5 anos e 4 meses.>

3.2 Em geral, você diria que as pessoas no seu bairro são….

    a.   Bastante fisicamente ativas 1

    b.   Relativamente fisicamente ativas 2

    c.   Não muito fisicamente ativas 3

    d.   Nem um pouco fisicamente ativas.. 4

Não sei / não tenho certeza 7

Recusou-se a responder 9

3.3 Em geral, como você classifica seu bairro enquanto local para se fazer uma caminhada? Você diria que ele é…

    a.   Muito agradável 1

    b.   Relativamente agradável 2

    c.   Não muito agradável 3

    d.   Nem um pouco agradável.. 4

Não sei / não tenho certeza 7

Recusou-se a responder 9

Seção 3: Ambientes Sociais e Físicos

3.4   Em geral, você diria que o tráfego de veículos automotores em seu bairro é…

    a.   Pesado 1

    b.   Moderado, OU 2

    c.   Leve? 3

    Não sei / não tenho certeza 7

    Recusou-se a responder 9

3.5 Seu bairro tem calçadas?

    a.   Sim 1

    b.   Não *(Passe diretamente para a questão 3.7)* 2

    Não sei / não tenho certeza 7

    Recusou-se a responder 9

3.6   Para caminhar, você diria que as calçadas em seu bairro são…

    a.   Muito bem mantidas 1

    b.   Relativamente bem mantidas 2

    c.   Não muito bem mantidas 3

    d.   Nem um pouco bem mantidas 4

    Não sei / não tenho certeza 7

    Recusou-se a responder 9

3.7 Para caminhar à noite, você descreveria o sistema de iluminação PÚBLICA do seu bairro como…

    a.   Muito bom 1

    b.   Bom 2

    c.   Razoável 3

    d.   Ruim 4

    e.   Muito ruim 5

    Não sei / não tenho certeza 7

    Recusou-se a responder 9

Seção 3: Ambientes Sociais e Físicos

3.8   Para caminhar em seu bairro, você diria que a presença de cães soltos na rua causa….

    a.   Um grande problema.. 1

    b.   Certo problema.. 2

    c.   Não muito problema... 3

    d.   Nenhum problema 4

    Não sei / não tenho certeza 7

    Recusou-se a responder 9

3.9 Quão seguro você considera seu bairro? Você diria que ele é…

    a.   Extremamente seguro.. 1

    b.   Relativamente seguro 2

    c.   Pouco seguro... 3

    d.   Nem um pouco seguro 4

    Não sei / não tenho certeza 7

    Recusou-se a responder 9

3.10 De forma geral, você diria que a maioria das pessoas no seu bairro é confiável?

    a.   Sim.. 1

    b.   Não.. 2

    Não sei / não tenho certeza 7

    Recusou-se a responder 9

3.11 Seu bairro dispõe de espaços públicos de recreação (por exemplo, piscinas públicas, parques, trilhas para caminhada, ciclovias, centros de recreação, etc.)?

    a.   Sim.. 1

    b.   Não *(Passe diretamente para a questão 3.13)* 2

    Não sei / não tenho certeza 7

    Recusou-se a responder 9

Seção 3: Ambientes Sociais e Físicos

3.12 Em geral, como você classifica as condições desses espaços públicos de recreação? Você diria que as condições são…

    a.   Excelentes 1

    b.   Boas 2

    c.   Razoáveis 3

    d.   Ruins 4

    Não sei / não tenho certeza 7

    Recusou-se a responder 9

3.13 Pensando no dinheiro público que é gasto em espaços públicos de recreação, qual das afirmações a seguir você considera mais adequada?

    a.   Meu bairro quase sempre recebe a fatia de recursos que lhe cabe 1

    b.   Meu bairro muitas vezes recebe a fatia de recursos que lhe cabe 2

    c.   Meu bairro raramente recebe a fatia de recursos que lhe cabe 3

    d.   Meu bairro nunca recebe a fatia de recursos que lhe cabe 4

    Não sei / não tenho certeza 7

    Recusou-se a responder 9

3.14 Para atividade física, você utiliza ALGUM tipo de espaço de recreação particular ou de associações? (… inclusive fora do seu bairro)

    a.   Sim 1

    b.   Não 2

    Não sei / não tenho certeza 7

    Recusou-se a responder 9

Seção 3: Ambientes Sociais e Físicos

"Para as perguntas a seguir, pense na comunidade onde você vive. Nesta entrevista,

comunidade é definida como a área que fica a uma distância de até 15 quilômetros ou 20

minutos de carro de sua casa.

"Favor informar se VOCÊ UTILIZOU qualquer um dos recursos e instalações listados a seguir em sua comunidade. Se algum dos recursos ou instalações mencionados não estiver disponível em sua comunidade, favor indicar."
<Nota ao entrevistador: Dê ênfase ao "VOCÊ". A pergunta se refere a uso pessoal/próprio do entrevistado, e não uso pela família ou pela comunidade.>

3.15 Trilhas para caminhada?

    a.   Sim – UTILIZA TRILHAS DISPONÍVEIS NA COMUNIDADE 1

    b.   Não – NÃO UTILIZA TRILHAS DISPONÍVEIS NA COMUNIDADE 2

    c.   Minha comunidade não dispõe de trilhas para caminhada 3

Não sei / não tenho certeza 7

Recusou-se a responder 9

OBS.: ENTENDEMOS POR TRILHAS PARA CAMINHADA TRILHAS PÚBLICAS DESIGNADAS ESPECIFICAMENTE PARA A REALIZAÇÃO DE CAMINHADAS.

3.16 Piscinas públicas?

    a.   Sim - UTILIZA PISCINAS DISPONÍVEIS NA COMUNIDADE 1

    b.   Não - NÃO UTILIZA PISCINAS DISPONÍVEIS NA COMUNIDADE 2

    c.   Minha comunidade não dispõe de piscinas públicas 3

Não sei / não tenho certeza 7

Recusou-se a responder 9

3.17 Centros públicos de recreação?

    a.   Sim - UTILIZA CENTROS PÚBLICOS DISPONÍVEIS NA COMUNIDADE

    1

    b.   Não - NÃO UTILIZA CENTROS PÚBLICOS DISPONÍVEIS NA

    COMUNIDADE.. 2

    c.   Minha comunidade não dispõe de centros públicos de recreação.................... 3

Não sei / não tenho certeza 7

Recusou-se a responder 9

Seção 3: Ambientes Sociais e Físicos

3.18 Ciclovias ou trilhas para bicicletas?

    a.   Sim - UTILIZA CICLOVIAS/TRILHAS DISPONÍVEIS NA COMUNIDADE 1

    b.   Não - NÃO UTILIZA CICLOVIAS/TRILHAS DISPONÍVEIS NA COMUNIDADE 2

    c.   Minha comunidade não dispõe de ciclovias ou trilhas para bicicletas............ 3

    Não sei / não tenho certeza.. 7

    Recusou-se a responder 9

3.19 Parques/praças/quadras esportivas?

    a.   Sim - UTILIZA PARQUES DISPONÍVEIS NA COMUNIDADE 1

    b.   Não - NÃO UTILIZA PARQUES DISPONÍVEIS NA COMUNIDADE 2

    c.   Minha comunidade não dispõe de parques/praças/quadras esportivas 3

    Não sei / não tenho certeza.. 7

    Recusou-se a responder 9

3.20 Escolas abertas a atividades públicas de recreação?

    a.   Sim - UTILIZA ESCOLAS DA COMUNIDADE PARA RECREAÇÃO...... 1

    b.   Não - NÃO UTILIZA ESCOLAS DA COMUNIDADE PARA RECREAÇÃO 2

    c.   As escolas na minha comunidade não são abertas para uso público................ 3

    Não sei / não tenho certeza.. 7

    Recusou-se a responder 9

3.21 Você utiliza algum shopping para praticar atividade física/caminhada?

    a.   Sim - UTILIZA SHOPPINGS DA COMUNIDADE PARA REALIZAR ATIV. FÍS. 1

    b.   Não - NÃO UTILIZA SHOPPINGS DA COMUNIDADE PARA ATIV. FÍS. 2

    c.   Minha comunidade não dispõe de shoppings 3

    Não sei / não tenho certeza.. 7

    Recusou-se a responder 9

Seção 3: Ambientes Sociais e Físicos

3.22 Você participa de programas ou utiliza o espaço de igrejas/locais de prática religiosa para realizar atividade física?

    a.   Sim - UTILIZA O ESPAÇO DE IGREJAS DA COMUNIDADE.................. 1

    b.   Não - NÃO UTILIZA O ESPAÇO DE IGREJAS DA COMUNIDADE 2

    c.   Minha comunidade não dispõe de igrejas/locais de prática religiosa que ofereçam programas de atividade física... 3

    Não sei / não tenho certeza 7

    Recusou-se a responder 9

3.23 Você utiliza as vias fluviais locais, como córregos, rios e lagos, para a prática de atividades aquáticas, tais como canoagem, caiaque, natação ou esqui aquático? (NÃO INCLUIR ATIVIDADES NÃO FÍSICAS, COMO PASSEAR DE BARCO)

    a.   Sim - UTILIZA VIAS FLUVIAIS DISPONÍVEIS NA COMUNIDADE 1

    b.   Não - NÃO UTILIZA VIAS FLUVIAIS DISPONÍVEIS NA COMUNIDADE 2

    c.   Minha comunidade não dispõe de vias fluviais para a prática de atividade física 3

    Não sei / não tenho certeza 7

    Recusou-se a responder 9

"As próximas perguntas dizem respeito à sua opinião sobre espaços para a realização de atividade física na sua comunidade."

3.24 Qual a importância de clubes e programas de atividades físicas ou de recreação, ou ainda de eventos organizados de recreação, para que você realize atividade física?

    a.   Muito importantes 1

    b.   Relativamente importantes 2

    c.   Não muito importantes 3

    c.   Não muito importantes 3

    e.   Minha comunidade não dispõe de clubes ou programas de atividade física 5

    Não sei / não tenho certeza 7

    Recusou-se a responder 9

Seção 3: Ambientes Sociais e Físicos

3.25 Na sua comunidade, você diria que todas as pessoas têm acesso igual a espaços públicos de recreação?

    a.   Sim..1

    b.   Não 2

    c.   Minha comunidade não dispõe de espaços públicos de recreação

    (Passe diretamente para a questão 4.1)..3

    Não sei / não tenho certeza..7

    Recusou-se a responder 9

3.26 Quão seguros são os espaços públicos de recreação na sua comunidade? Você diria que eles são…

    a.   Muito seguros 1

    b.   Relativamente seguros 2

    c.   Relativamente inseguros 3

    d.   Nem um pouco seguros..4

    Não sei / não tenho certeza..7

    Recusou-se a responder 9

3.27 Sua decisão de utilizar ou não os espaços públicos de recreação disponíveis na sua comunidade é influenciada por preocupações com segurança?

    a.   Sim..1

    b.   Não 2

    c.   Minha comunidade não dispõe de espaços públicos de recreação..................3

    Não sei / não tenho certeza..7

    Recusou-se a responder 9

APPENDIX D – FIRST TRANSLATION – JANISA

QUESTIONÁRIO SOBRE SUPORTES AMBIENTAIS PARA ATIVIDADE FÍSICA

(INQUÉRITO DE AMBIENTE FÍSICO E SOCIAL)

Centro de Pesquisa sobre Prevenção
Escola de Saúde Pública Norman J. Arnold
Universidade da Carolina do Sul

Pesquisador principal:
Barbara E. Ainsworth, Doutora, Mestre em Saúde Pública

Seção 3: Ambientes Sociais e Físicos

"Primeiro, vou fazer algumas perguntas sobre a sua vizinhança e, depois, sobre a comunidade em que você mora".

"Vamos começar com algumas perguntas sobre a sua vizinhança. Para os objetivos desta entrevista, vizinhança é uma área de aproximadamente 800 metros de distância ou uma caminhada de dez minutos da sua casa.

3.1 Há quanto tempo você mora no seu endereço atual?

Número de meses (doze meses ou menos) .... ___ ___

Número de anos (um ano ou mais) .... ___ ___

Não sabe/não tem certeza 77

Recusou-se a responder 99

<Nota para o entrevistador: Menos de um ano deve ser registrado em meses e mais de doze meses deve ser registrado somente como anos inteiros. P. ex.: 5 anos, em vez de 5 anos e 4 meses.>

3.2 Em geral, você diria que as pessoas da sua vizinhança são...

a. Muito ativas fisicamente 1

b. Um pouco ativas fisicamente 2

c. Não muito ativas fisicamente 3

d. Nem um pouco ativas fisicamente.. 4

Não sabe/não tem certeza 7

Recusou-se a responder 9

3.3 Em geral, como você classificaria a sua vizinhança como um lugar para caminhar? Você diria que é...

a. Muito agradável 1

b. Um pouco agradável 2

c. Não muito agradável 3

c. Nem um pouco agradável.. 4

Não sabe/não tem certeza 7

Recusou-se a responder 9

Seção 3: Ambientes Sociais e Físicos

3.4   Em geral, você diria que o tráfego de veículos em sua vizinhança é...

    a.   Pesado,... 1

    b.   Moderado, OU 2

    c.   Leve? 3

    Não sabe/não tem certeza 7

    Recusou-se a responder 9

3.5 Existem calçadas na sua vizinhança?

    a.   Sim... 1

    b.   Não **(Pule para a pergunta 3.7)** 2

    Não sabe/não tem certeza 7

    Recusou-se a responder 9

3.6   Para caminhar na sua vizinhança, você diria que as calçadas têm...

    a.   Manutenção muito boa... 1

    b.   Manutenção razoável 2

    c.   Manutenção não muito boa 3

    d.   Nenhuma manutenção... 4

    Não sabe/não tem certeza 7

    Recusou-se a responder 9

3.7   Para caminhar à noite, você descreveria a iluminação das RUAS na sua vizinhança como...

    a.   Muito boa.. 1

    b.   Boa... 2

    c.   Razoável 3

    d.   Ruim 4

    e.   Muito ruim... 5

    Não sabe/não tem certeza 7

    Recusou-se a responder 9

Seção 3: Ambientes Sociais e Físicos

3.8 Para caminhar na sua vizinhança, você diria que os cachorros sem dono são um problema...

    a.   Muito importante...1

    b.   Um pouco importante...2

    c.   Não muito importante 3

    d.   Nem um pouco importante..4

    Não sabe/não tem certeza 7

    Recusou-se a responder 9

3.9 Em sua opinião, qual o nível de segurança da sua vizinhança?
Você diria que é...

    a.   Extremamente segura 1

    b.   Bastante segura 2

    c.   Um pouco segura 3

    d.   Nem um pouco segura...4

    Não sabe/não tem certeza 7

    Recusou-se a responder 9

3.10 De um modo geral, você diria que a maioria das pessoas da sua vizinhança é de confiança?

    a.   Sim...1

    b.   Não...2

    Não sabe/não tem certeza 7

    Recusou-se a responder 9

3.11 A sua vizinhança tem instalações de recreação pública (como piscinas públicas, parques, trilhas para caminhada, ciclovias, centros de recreação, etc.)?

    a.   Sim...1

    b.   Não **(Pule para a pergunta 3.13)**...2

    Não sabe/não tem certeza 7

    Recusou-se a responder 9

Seção 3: Ambientes Sociais e Físicos

3.12 Em geral, como você classificaria a condição dessas instalações de recreação pública? Você diria que é...

    a.   Excelente... 1

    b.   Boa... 2

    c.   Razoável 3

    d.   Ruim 4

    Não sabe/não tem certeza 7

    Recusou-se a responder 9

3.13 Analisando como o dinheiro público é gasto em instalações de recreação, qual das afirmações a seguir é a mais verdadeira...

    a   A minha vizinhança quase sempre recebe uma parcela justa de investimentos 1

    b.   A minha vizinhança muitas vezes recebe uma parcela justa de investimentos 2

    c.   A minha vizinhança raramente recebe uma parcela justa de investimentos 3

    d.   A minha vizinhança nunca recebe uma parcela justa de investimentos........... 4

    Não sabe/não tem certeza 7

    Recusou-se a responder 9

3.14 Para praticar atividades físicas, você usa ALGUMA instalação de recreação particular ou somente para sócios? (... incluindo aquelas localizadas fora da sua vizinhança)

    a.   Sim.. 1

    b.   Não 2

    Não sabe/não tem certeza 7

    Recusou-se a responder 9

Seção 3: Ambientes Sociais e Físicos

"Para as próximas perguntas, pense sobre a comunidade em que você mora. Para os objetivos desta entrevista, comunidade é a área de aproximadamente 16 quilômetros de distância ou vinte minutos de carro da sua casa".

"Por favor, diga se você mesmo USA qualquer um desses recursos ou instalações em sua comunidade. Por favor, informe se o tipo de recurso ou instalação que eu citar não estiver disponível em *sua comunidade".*

<Nota para o entrevistador: Enfatize "você mesmo". A pergunta é sobre uso pessoal, não se refere ao uso familiar ou comunitário.>

3.15 Trilhas para caminhada?

    a.   Sim – USA TRILHAS PARA CAMINHADA NA COMUNIDADE 1

    b.   Não – NÃO USA TRILHAS PARA CAMINHADA NA COMUNIDADE ... 2

    c.   Minha comunidade não tem trilhas para caminhada 3

    Não sabe/não tem certeza 7

    Recusou-se a responder 9

DICA: TRILHAS PARA CAMINHADA SIGNIFICAM TRILHAS PÚBLICAS QUE SÃO DESTINADAS A CAMINHADAS.

3.16 Piscinas públicas?

    a.   Sim – USA PISCINAS NA COMUNIDADE 1

    b.   Não – NÃO USA PISCINAS NA COMUNIDADE 2

c.   Minha comunidade não tem piscinas públicas 3

    Não sabe/não tem certeza 7

    Recusou-se a responder 9

3.17 Centros públicos de recreação?

    a.   Sim – USA CENTROS PÚBLICOS DE RECREAÇÃO NA COMUNIDADE
    1

    b.   Não – NÃO USA CENTROS PÚBLICOS DE RECREAÇÃO NA
    COMUNIDADE 2

    c.   Minha comunidade não tem centros públicos de recreação 3

    Não sabe/não tem certeza 7

    Recusou-se a responder 9

Seção 3: Ambientes Sociais e Físicos

3.18 Ciclovias ou trilhas para bicicletas?

    a.   Sim – USA TRILHAS PARA BICICLETAS NA COMUNIDADE.............. 1

    b.   Não – NÃO USA TRILHAS PARA BICICLETAS NA COMUNIDADE..... 2

    c.   Minha comunidade não tem trilhas para bicicletas.. 3

    Não sabe/não tem certeza 7

    Recusou-se a responder 9

3.19 Parques/playgrounds/campos esportivos?

    a.   Sim – USA PARQUES NA COMUNIDADE 1

    b.   Não – NÃO USA PARQUES NA COMUNIDADE...................................... 2

    c.   Minha comunidade não tem parques/playgrounds/campos esportivos............ 3

    Não sabe/não tem certeza 7

    Recusou-se a responder 9

3.20 Escolas abertas a atividades públicas de recreação?

    a.   Sim – USA ESCOLAS PARA RECREAÇÃO NA COMUNIDADE............. 1

    b.   Não – NÃO USA ESCOLAS PARA RECREAÇÃO NA COMUNIDADE... 2

    c.   As escolas da minha comunidade não estão abertas ao público 3

    Não sabe/não tem certeza 7

    Recusou-se a responder 9

3.21 Você usa shopping centers para praticar atividade física/realizar programas de caminhada?

    a.   Sim – USA SHOPPING CENTERS PARA ATIVIDADE FÍSICA NA COMUNIDADE 1

    b.   Não – NÃO USA SHOPPING CENTERS PARA ATIVIDADE FÍSICA NA COMUNIDADE 2

    c.   Minha comunidade não tem um shopping center 3

    Não sabe/não tem certeza 7

    Recusou-se a responder 9

Seção 3: Ambientes Sociais e Físicos

3.22 Você usa programas e instalações para atividade física em locais de culto religioso?

    a.   Sim – USA INSTALAÇÕES EM LOCAL DE CULTO RELIGIOSO NA COMUNIDADE.. 1

    b.   Não – NÃO USA INSTALAÇÕES EM LOCAL DE CULTO RELIGIOSO NA COMUNIDADE 2

    c.   Minha comunidade não tem locais de culto religioso que ofereça programa de atividade física.. 3

    Não sabe/não tem certeza 7

    Recusou-se a responder 9

3.23 Você usa cursos d'água, como riachos, rios e lagos, para praticar atividades físicas aquáticas, tais como canoagem, caiaquismo, natação ou esqui aquático? (NÃO INCLUA ATIVIDADES QUE NÃO ENVOLVAM PRÁTICA FÍSICA, COMO PASSEIO DE BARCO)

    a.   Sim – USA CURSOS D'ÁGUA PARA PRATICAR ATIVIDADE FÍSICA NA COMUNIDADE 1

    b.   Não – NÃO USA CURSOS D'ÁGUA PARA PRATICAR ATIVIDADE FÍSICA NA COMUNIDADE 2

    c.   Minha comunidade não tem cursos d'água para praticar atividade física........ 3

    Não sabe/não tem certeza 7

    Recusou-se a responder 9

"As próximas perguntas estão relacionadas à sua opinião sobre as instalações para atividade física em sua comunidade".

3.24 Para a sua prática de atividade física, qual a importância de clubes, programas ou eventos organizados de atividade física/recreação na sua comunidade...

    a.   Muito importante.. 1

    b.   Um pouco importante.. 2

    c.   Não muito importante 3

    d.   Nem um pouco importante 4

    e.   Minha comunidade não tem clubes ou programas de atividade física 5

    Não sabe/não tem certeza 7

Recusou-se a responder 9

Seção 3: Ambientes Sociais e Físicos

3.25 Na sua comunidade, você diria que todas as pessoas têm a mesma facilidade de acesso a instalações de recreação pública?

    a.   Sim 1

    b.   Não 2

    c.   Minha comunidade não tem instalações de recreação pública, **(Pule para a pergunta 4.1)** 3

    Não sabe/não tem certeza 7

    Recusou-se a responder 9

3.26 Qual o nível de segurança das instalações de recreação pública da sua comunidade? Você diria que é...

    a.   Muito seguro 1

    b.   Um pouco seguro 2

    c.   Um pouco perigoso 3

    d.   Nem um pouco seguro 4

    Não sabe/não tem certeza 7

    Recusou-se a responder 9

3.27 As preocupações com segurança nas instalações de recreação pública da sua comunidade influenciam o seu uso dessas instalações?

    a.   Sim 1

    b.   Não 2

    c.   Minha comunidade não tem instalações de recreação pública 3

    Não sabe/não tem certeza 7

    Recusou-se a responder 9

QUESTIONÁRIO SOBRE SUPORTES AMBIENTAIS PARA ATIVIDADE FÍSICA

(INQUÉRITO DE AMBIENTES FÍSICOS E SOCIAIS)

Centro de Pesquisa em Prevenção
Faculdade de Saúde Pública Norman J. Arnold
Universidade da Carolina do Sul

Pesquisadora principal:
Barbara E. Ainsworth, Doutora, Mestre em Saúde Pública

Seção 3: Ambientes Sociais e Físicos

"Primeiro, vou fazer algumas perguntas sobre a sua vizinhança e, depois, sobre a comunidade onde você mora".

"Vamos começar com algumas perguntas sobre a sua vizinhança. Nesta entrevista, vizinhança é uma área de aproximadamente 800 metros de distância ou uma caminhada de 10 minutos da sua casa.

3.1   Há quanto tempo você mora no seu endereço atual?

Número de meses (doze meses ou menos) .... ___ ___

Número de anos (um ano ou mais) .... ___ ___

Não sabe/não tem certeza 77

Recusou-se a responder 99

<Nota para o entrevistador: Menos de um ano deve ser registrado em número de meses e mais de 12 meses deve ser registrado somente em número de anos inteiros. Exemplo: 5 anos em vez de 5 anos e 4 meses.>

3.2   Em geral, você diria que as pessoas da sua vizinhança são...

a.   Muito ativas fisicamente 1

b.   Um pouco ativas fisicamente 2

c.   Não muito ativas fisicamente 3

d.   Nem um pouco ativas fisicamente... 4

Não sabe/não tem certeza 7

Recusou-se a responder 9

3.3 Em geral, como você classificaria a sua vizinhança como um lugar para fazer uma caminhada? Você diria que é...

a.   Muito agradável 1

b.   Um pouco agradável 2

c.   Não muito agradável 3

c.   Nem um pouco agradável... 4

Não sabe/não tem certeza 7

Recusou-se a responder 9

Seção 3: Ambientes Sociais e Físicos

3.4 Em geral, você diria que o tráfego de veículos em sua vizinhança é...

    a.   Pesado,.. 1

    b.   Moderado, OU 2

    c.   Leve? 3

    Não sabe/não tem certeza 7

    Recusou-se a responder 9

3.5 Sua vizinhança tem calçadas?

    a.   Sim... 1

    b.   Não *(Passe para a pergunta 3.7)*.. 2

    Não sabe/não tem certeza 7

    Recusou-se a responder 9

3.6 Para caminhar, você diria que as calçadas da sua vizinhança têm...

    a.   Manutenção muito boa... 1

    b.   Manutenção razoável 2

    c.   Manutenção não muito boa 3

    d.   Nenhuma manutenção... 4

    Não sabe/não tem certeza 7

    Recusou-se a responder 9

3.7 Para caminhar à noite, você descreveria a iluminação PÚBLICA na sua vizinhança como...

    a.   Muito boa... 1

    b.   Boa... 2

    c.   Razoável 3

    d.   Ruim 4

    e.   Muito ruim... 5

    Não sabe/não tem certeza 7

    Recusou-se a responder 9

Seção 3: Ambientes Sociais e Físicos

3.8   Para caminhar na sua vizinhança, você diria que os cachorros soltos na rua são um problema...

 a. Muito importante...1

 b. Um pouco importante...2

 c. Não muito importante 3

 d. Nem um pouco importante...4

 Não sabe/não tem certeza 7

 Recusou-se a responder 9

3.9 Em sua opinião, qual o nível de segurança da sua vizinhança?

Você diria que é...

 a. Extremamente segura 1

 b. Bastante segura 2

 c. Um pouco segura 3

 d. Nem um pouco segura..4

 Não sabe/não tem certeza 7

 Recusou-se a responder 9

3.10 De um modo geral, você diria que a maioria das pessoas da sua vizinhança é de confiança?

 a. Sim...1

 b. Não...2

 Não sabe/não tem certeza 7

 Recusou-se a responder 9

3.11 A sua vizinhança tem espaços públicos de recreação (como piscinas públicas, parques, trilhas para caminhada, ciclovias, centros de recreação, etc.)?

 a. Sim...1

 b. Não *(Passe para a pergunta 3.13)* 2

 Não sabe/não tem certeza 7

 Recusou-se a responder 9

Seção 3: Ambientes Sociais e Físicos

Em geral, como você classificaria a condição desses espaços públicos de recreação? Você diria que é...

  a. Excelente... 1

  b. Boa.. 2

  c. Razoável 3

  d. Ruim 4

  Não sabe/não tem certeza 7

  Recusou-se a responder 9

3.13 Pensando no dinheiro público que é gasto em espaços públicos de recreação, qual das afirmações a seguir é a mais verdadeira?

  a A minha vizinhança quase sempre recebe uma parcela justa de investimentos.......................

  b. A minha vizinhança muitas vezes recebe uma parcela justa de investimentos 2

  c. A minha vizinhança raramente recebe uma parcela justa de investimentos 3

  d. A minha vizinhança nunca recebe uma parcela justa de investimentos........... 4

  Não sabe/não tem certeza 7

  Recusou-se a responder 9

3.14 Para praticar atividades físicas, você usa ALGUM tipo de espaço de recreação particular ou de associações? (... inclusive fora da sua vizinhança)

  a. Sim... 1

  b. Não 2

  Não sabe/não tem certeza 7

  Recusou-se a responder 9

Seção 3: Ambientes Sociais e Físicos

"Para as próximas perguntas, pense sobre a comunidade em que você mora. Nesta entrevista, comunidade é a área de aproximadamente 16 quilômetros de distância ou 20 minutos de carro da sua casa".

"Por favor, diga se VOCÊ MESMO USA qualquer um desses recursos e instalações em sua comunidade. Por favor, informe se o tipo de recurso ou instalação mencionado não estiver disponível em *sua comunidade".*

<Nota para o entrevistador: Enfatize "você mesmo". A pergunta é sobre uso pessoal, não se refere ao uso familiar ou comunitário.>

3.15 Trilhas para caminhada?

    a.   Sim – USA TRILHAS PARA CAMINHADA NA COMUNIDADE 1

    b.   Não – <u>NÃO</u> USA TRILHAS PARA CAMINHADA NA COMUNIDADE ... 2

    c.   Minha comunidade não tem trilhas para caminhada.. 3

    Não sabe/não tem certeza 7

    Recusou-se a responder 9

OBS.: ENTENDEMOS POR TRILHAS PARA CAMINHADA TRILHAS PÚBLICAS DESIGNADAS ESPECIFICAMENTE PARA A REALIZAÇÃO DE CAMINHADAS.

3.16 Piscinas públicas?

    a.   Sim – USA PISCINAS NA COMUNIDADE 1

    b.   Não – NÃO USA PISCINAS NA COMUNIDADE 2

c.   Minha comunidade não tem piscinas públicas........................ 3

    Não sabe/não tem certeza 7

    Recusou-se a responder 9

3.17 Centros públicos de recreação?

    a.   Sim – USA CENTROS PÚBLICOS DE RECREAÇÃO NA COMUNIDADE ....

    b.   Não – <u>NÃO</u> USA CENTROS PÚBLICOS DE RECREAÇÃO NA COMUNIDADE... 2

    c.   Minha comunidade não tem centros públicos de recreação............................ 3

    Não sabe/não tem certeza 7

    Recusou-se a responder 9

Seção 3: Ambientes Sociais e Físicos

3.18 Ciclovias ou trilhas para bicicletas?

    a.   Sim – USA TRILHAS PARA BICICLETAS NA COMUNIDADE................ 1

    b.   Não – <u>NÃO</u> USA TRILHAS PARA BICICLETAS NA COMUNIDADE..... 2

    c.   Minha comunidade não tem trilhas para bicicletas... 3

Não sabe/não tem certeza 7

Recusou-se a responder 9

3.19 Parques/praças/quadras ou campos esportivos?

    a.   Sim – USA PARQUES NA COMUNIDADE 1

    b.   Não – NÃO USA PARQUES NA COMUNIDADE...................................... 2

    c.   Minha comunidade não tem parques/praças/quadras ou campos esportivos ... 3

Não sabe/não tem certeza 7

Recusou-se a responder 9

3.20 Escolas abertas a atividades públicas de recreação?

    a.   Sim – USA ESCOLAS DA COMUNIDADE PARA RECREAÇÃO............. 1

    b.   Não – <u>NÃO</u> USA ESCOLAS DA COMUNIDADE PARA RECREAÇÃO ... 2

    c.   As escolas da minha comunidade não são abertas ao público 3

Não sabe/não tem certeza 7

Recusou-se a responder 9

3.21 Você usa *shopping* para praticar atividade física/caminhada?

    a.   Sim – USA *SHOPPING* DA COMUNIDADE PARA ATIVIDADE FÍSICA 1

    b.   Não – <u>NÃO</u> USA *SHOPPING* DA COMUNIDADE PARA ATIVIDADE

FÍSICA... 2

    c.   Minha comunidade não tem um *shopping*.. 3

Não sabe/não tem certeza 7

Recusou-se a responder 9

Seção 3: Ambientes Sociais e Físicos

3.22 Você participa de programas e usa o espaço de locais de culto religioso para realizar atividade física?

    a.   Sim – USA LOCAIS DE CULTO RELIGIOSO NA COMUNIDADE 1

    b.   Não – <u>NÃO</u> USA LOCAIS DE CULTO RELIGIOSO NA COMUNIDADE . 2

    c.   Minha comunidade não tem locais de culto religioso que ofereçam programas de atividade física 3

    Não sabe/não tem certeza 7

    Recusou-se a responder 9

3.23 Você usa cursos d'água, como riachos, rios e lagos, para praticar atividades físicas aquáticas, tais como canoagem, caiaquismo, natação ou esqui aquático? (NÃO INCLUA ATIVIDADES QUE NÃO ENVOLVAM PRÁTICA FÍSICA, COMO PASSEIO DE BARCO)

    a.   Sim – USA CURSOS D'ÁGUA DA COMUNIDADE 1

    b.   Não – <u>NÃO</u> USA CURSOS D'ÁGUA DA COMUNIDADE.......................... 2

    c.   Minha comunidade não tem cursos d'água para a prática de atividade física .. 3

    Não sabe/não tem certeza 7

    Recusou-se a responder 9

"As próximas perguntas estão relacionadas à sua opinião sobre os espaços para a realização de atividade física em sua comunidade".

3.24 Para a sua prática de atividade física, qual a importância de clubes e programas de atividades físicas ou de recreação, ou ainda de eventos organizados de recreação...

    a.   Muito importante.. 1

    b.   Um pouco importante... 2

    c.   Não muito importante 3

    d.   Nem um pouco importante 4

    e.   Minha comunidade não tem clubes ou programas de atividade física 5

    Não sabe/não tem certeza 7

    Recusou-se a responder 9

Seção 3: Ambientes Sociais e Físicos

3.25 Na sua comunidade, você diria que todas as pessoas têm acesso igual a espaços públicos de recreação?

    a.   Sim 1

    b.   Não 2

    c.   Minha comunidade não tem espaços públicos de recreação *(Passe para a pergunta 4.1)* 3

    Não sabe/não tem certeza 7

    Recusou-se a responder 9

3.26 Qual o nível de segurança dos espaços públicos de recreação da sua comunidade? Você diria que é...

    a.   Muito seguro 1

    b.   Um pouco seguro 2

    c.   Um pouco perigoso 3

    d.   Nem um pouco seguro 4

    Não sabe/não tem certeza 7

    Recusou-se a responder 9

3.27 As preocupações com segurança nos espaços públicos de recreação da sua comunidade influenciam o seu uso dessas instalações?

    a.   Sim 1

    b.   Não 2

    c.   Minha comunidade não tem espaços públicos de recreação 3

    Não sabe/não tem certeza 7

    Recusou-se a responder 9

APPENDIX F – BACK-TRANSLATION

ENVIRONMENTAL SUPPORTS FOR PHYSICAL ACTIVITY QUESTIONNAIRE

(STUDY OF SOCIAL AND PHYSICAL ENVIRONMENTS)

Prevention Research Center
Norman J. Arnold School of Public Health
University of South Carolina

Principal researcher:
Barbara E. Ainsworth, PhD, MPH

Section 3: Social and Physical Environments

"I will ask some questions about the neighborhood where you live, followed by questions about the community where you live."

"First, some questions about your neighborhood. In this interview, neighborhood is defined as the area that is a distance of up to 800 meters or a 10-minute walk from your house.

3.1 For how long have you lived at your current address?

Number of months (twelve months or less) ....___ ___

Number of years (one year or more)..___ ___

I don't know / I'm not certain.. 77

Refused to answer 99

<Note to interviewer: Less than one year should be informed with the number of months and more than 12 months should be informed only in the number of complete years. Example: 5 years instead of 5 years and 4 months.>

3.2 In general, would you say that the people in your neighborhoods are...

a. Very physically active...1

b. Relatively physically active 2

c. Not very physically active...3

d. Not at all physically active 4

I don't know / I'm not certain.. 7

Refused to answer ....

3.3 In general, how would you classify your neighborhood as a place to walk? Would you say that it is…

a. Very pleasant 1

b. Relatively pleasant 2

c. Not very pleasant 3

d. Not at all pleasant...4

I don't know / I'm not certain.. 7

Refused to answer 9

Section 3: Social and Physical Environments

3.4   In general, would you say that the automobile traffic in your neighborhood is...

    a.   Heavy... 1

    b.   Moderate, OR.. 2

    c.   Light?... 3

    I don't know / I'm not certain... 7

    Refused to answer 9

3.5 Does your neighborhood have sidewalks?

    a.   Yes... 1

    b.   No *(Go directly to question 3.7)* 2

    I don't know / I'm not certain... 7

    Refused to answer 9

3.6   For walking, would you say that the sidewalks in your neighborhood are...

    a.   Very well maintained.. 1

    b.   Relatively well maintained.. 2

    c.   Not very well maintained.. 3

    d.   Not at all well maintained 4

    I don't know / I'm not certain... 7

    Refused to answer 9

3.7 For walking at night, would you describe the PUBLIC lighting system in your neighborhood as...

    a.   Very good 1

    b.   Good 2

    c.   Reasonable... 3

    d.   Bad... 4

    e.   Very bad 5

    I don't know / I'm not certain... 7

    Refused to answer 9

Section 3: Social and Physical Environments

3.8   For walking in your neighborhood, would you say that the presence of dogs loose in the street causes...

    a.   A big problem 1

    b.   A limited problem 2

    c.   Not much of a problem 3

    d.   No problem 4

    I don't know / I'm not certain.. 7

    Refused to answer 9

3.9 How safe do you consider your neighborhood? Would you say that it is...

    a.   Extremely safe 1

    b.   Relatively safe.. 2

    c.   Not very safe.. 3

    d.   Not at all safe 4

    I don't know / I'm not certain.. 7

    Refused to answer 9

3.10 In general, would you say that most people in your neighborhood are trustworthy?

    a.   Yes 1

    b.   No 2

    I don't know / I'm not certain.. 7

    Refused to answer 9

3.11 Does your neighborhood have public recreation spaces (for example, public pools, parks, walking paths, bicycle paths, recreation centers, etc.)?

    a.   Yes 1

    b.   No *(Go directly to question 3.13)* 2

    I don't know / I'm not certain.. 7

    Refused to answer 9

Section 3: Social and Physical Environments

In general, how would you classify the conditions of these public recreation spaces? Would you say that the conditions are...

    a.   Excellent 1

    b. Good.. 2

    c. Reasonable... 3

    d. Poor 4

    I don't know / I'm not certain.. 7

    Refused to answer 9

3.13 Considering government investments made in public recreation spaces, which of the statements below do you consider most suitable?

    a.   My neighborhood almost always receives its proper share of investments 1

    b.   My neighborhood often receives its proper share of investments 2

    c.   My neighborhood rarely receives its proper share of investments 3

    d.   My neighborhood never receives its proper share of investments................... 4

    I don't know / I'm not certain.. 7

    Refused to answer 9

3.14 To practice physical activities, do you use ANY type of private recreation space or clubs (...even if it is outside of your neighborhood)

    a.   Yes... 1

    b.   No 2

    I don't know / I'm not certain.. 7

    Refused to answer 9

Section 3: Social and Physical Environments

"For the questions below, think of the community where you live. In this interview,

community is defined as the area that is a distance of up to 15 kilometers or 20 minutes

by car from your house.

"Please indicate if YOU USE any one of the resources and facilities listed below in your

community. If any of the resources or facilities mentioned is not available in your

community, please indicate this."

<Note to the interviewer:
Emphasize "YOU." The question refers to personal use by the person interviewed,
and not use by the family or community.>

3.15 Walking paths?

    a.   Yes – USES WALKING PATHS AVAILABLE IN THE COMMUNITY 1

    b.   No – DOES <u>NOT</u> USE WALKING PATHS AVAILABLE IN THE

    COMMUNITY 2

    c.   My community does not have walking paths 3

    I don't know / I'm not certain 7

    Refused to answer 9

Note: WE UNDERSTAND WALKING PATHS TO BE PUBLIC PATHS
SPECIFICALLY DESIGNATED FOR TAKING WALKS.

3.16 Public pools?

    a.   Yes – USES POOLS AVAILABLE IN THE COMMUNITY 1

    b.   No – DOES <u>NOT</u> USE POOLS AVAILABLE IN THE COMMUNITY 2

    c.   My community does not have public pools 3

    I don't know / I'm not certain 7

    Refused to answer 9

3.17 Public recreation centers?

    a.   Yes – USES PUBLIC CENTERS AVAILABLE IN THE COMMUNITY 1

    b.   No - DOES <u>NOT</u> USE PUBLIC CENTERS AVAILABLE IN THE

    COMMUNITY 2

    c.   My community does not have public recreation centers 3

    I don't know / I'm not certain 7

    Refused to answer 9

Section 3: Social and Physical Environments

3.18 Bicycle paths or trails?

    a.   Yes – USES BICYCLE LANES OR PATHS AVAILABLE IN THE COMMUNITY 1

    b.   No – DOES <u>NOT</u> USE BICYCLE LANES OR PATHS AVAILABLE IN THE COMMUNITY 2

    c.   My community does not have bicycle lanes or paths 3

    I don't know / I'm not certain 7

    Refused to answer 9

3.19 Parks/playgrounds/sports fields?

    a.   Yes – USES PARKS AVAILABLE IN THE COMMUNITY 1

    b.   No – DOES <u>NOT</u> USE PARKS AVAILABLE IN THE COMMUNITY 2

    c.   My community does not have parks /playgrounds/sports fields 3

    I don't know / I'm not certain 7

    Refused to answer 9

3.20 Schools open to public recreation activities?

    a.   Yes - USES SCHOOLS IN THE COMMUNITY FOR RECREATION 1

    b.   No – DOES <u>NOT</u> USE SCHOOLS IN THE COMMUNITY FOR RECREATION 2

    c.   The schools in my community are not open for public use 3

    I don't know / I'm not certain 7

    Refused to answer 9

3.21 Do you use any shopping center to practice physical activity/walking?

    a.   Yes – USES SHOPPING CENTERS IN THE COMMUNITY FOR PHYSICAL ACTIVITY 1

    b.   No –DOES <u>NOT</u> USE SHOPPING CENTERS IN THE COMMUNITY FOR PHYSICAL ACTIVITY 2

    c.   My community does not have any shopping centers 3

    I don't know / I'm not certain 7

    Refused to answer 9

Section 3: Social and Physical Environments

3.22 Do you participate in programs or use space in churches or places of religious practice to conduct physical activity ?

    a.   Yes – USES CHURCH SPACE IN THE COMMUNITY 1

    b.   No – DOES NOT USE CHURCH SPACE IN THE COMMUNITY.............. 2

    c.   My community does not have any churches or places of religious practice that offer physical activity programs 3

    I don't know / I'm not certain... 7

    Refused to answer 9

3.23 Do you use local waterways, such as streams, rivers and lakes to practice aquatic activities such as canoeing, kayaking, swimming or water skiing? (DO NOT INCLUDE NON-PHYSICAL ACTIVITIES SUCH AS BOATING)

    a.   Yes – USES WATERWAYS AVAILABLE IN THE COMMUNITY 1

    b.   No – DOES NOT USE WATERWAYS AVAILABLE IN THE COMMUNITY... 2

    c.   My community does not have waterways for practicing physical activities.... 3

    I don't know / I'm not certain... 7

    Refused to answer 9

"The following questions concern your opinion about the spaces for conducting physical activity in your community."

3.24 What is the importance of clubs and programs for physical or recreational activities or even of organized recreational events, for you to engage in physical activity?

    a.   Very important 1

    b.   Relatively important... 2

    c.   Not very important 3

    d.   Not at all important 4

    e.   My community does not have clubs or programs for physical activity 5

    I don't know / I'm not certain... 7

    Refused to answer 9

Section 3: Social and Physical Environments

3.25 In your community, would you say that all people have equal access to public recreation spaces?

    a.   Yes.. 1

    b.   No 2

    c.   My community does not have public recreation spaces *(**Go directly to question 4.1)*** 3

    I don't know / I'm not certain.. 7

    Refused to answer 9

3.26 How safe are the public recreation spaces in your community? Would you say they are…

    Very safe 1

    Relatively safe.. 2

    Relatively unsafe.. 3

    Not at all safe 4

    I don't know / I'm not certain.. 7

    Refused to answer 9

3.27 Is your decision to use or not the public recreation spaces available in your community influenced by safety concerns?

    a.   Yes.. 1

    b.   No 2

    c.   My community does not have public recreation spaces 3

    I don't know / I'm not certain.. 7

    Refused to answer 9

APPENDIX G – LETTER TO TRANSLATION AGENCIES AND TRANSLATION BRIEF – STREAMLINED METHOD

Dear sir/madam,

I am currently engaged in a doctoral research project for which I need to translate a questionnaire about physical activity. I need the translation to be done in a specific way by a translator with specific characteristics. I do not want a back-translation and will not have the resulting translation back-translated, but I do need to contract a bilingual proof-reader/editor in addition to the translator.

The translator must have a minimum of seven years' experience or have translated a minimum of 1 million words of English into Portuguese, preferably of medical translation and obligatorily of technical translation.

The editor/proof reader must have seven years' experience or a relevant postgraduate qualification.

The translation procedure would be as follows:

I need the translator to translate my questionnaire and I need a copy of their translation before it is sent to the editor.

I then need the editor to read and (if necessary) correct the translation, with reference to the original. In addition to correcting anything they feel is incorrect, whether because it is incorrectly translated or because it is not written in good Portuguese. I also need the editor to suggest any changes they feel would improve the translation, using "track changes" or a similar method to indicate their revisions.

If the editor feels the translation is correct and cannot suggest improvements they should say so and would be paid the same amount as if they had corrected the translation. I need a copy of the document at this point too (even if it is identical to the first translation).

The questionnaire would then be sent back to the translator (even if the editor has not made any changes or suggestions) who would have full autonomy over whether to accept the editor's changes (if there are any) and would also be authorized to make any changes that occurred to him/her at this point. I would need a copy of the translation after this stage, which would be the final version of the translation

I would now like to ask the following questions:

1: Can you provide these services?

2: Would you be willing to provide the CVs of the professionals involved (I need to be able to state that I have checked their credentials)?

3: How much would such a service cost (in Brazilian Reais) for a 1,500 word questionnaire written in American English (target language: Brazilian Portuguese)?

NB: I have already obtained permission from the original author and the copyright holders.

Thank you for your time,

Robert Coulthard

Translation brief for a cooperative translation of the "Environmental Supports for Physical Activity Questionnaire: (Social and Physical Environment Survey)" from English into Brazilian Portuguese

The source text

Target text addressees

The questionnaire is designed to be administered by an interviewer, who will complete the answers in the spaces provided. Therefore, the questions and all answers except "refused" are addressed to the interviewee. The interviewer, however, must interpret the response and assign it to one of the options provided. In addition to the questions and possible responses, there are notes to the interviewer to explain how certain items should be interpreted and instructions to the interviewer on the sequence of questions.

The questionnaire was designed for adults (the validation sample was aged 18-96). There are no other exclusion criteria, so the target population for this translation is the adult Brazilian-Portuguese-speaking population.

Text reception

The questionnaire is designed to be administered by an interviewer by telephone.

Motive for translation

This questionnaire is being translated as part of a doctoral research project that aims to demonstrate that competent translators working in the areas in which they have experience are capable of producing very high quality translations of medical research instruments without the need for back-translation. The motivation for this is to raise awareness of the true level of professional translation competence available and to contribute to improving the professional standing of competent translators in the eyes of health professionals and health sciences researchers.

Text function

The original questionnaire was developed in order to investigate the relationship between respondents' perceptions about their neighbourhoods and communities and their levels of physical activity. A brief description of the questionnaire can be found here: http://www.activelivingresearch.org/node/10645 and the validation study can be found here: http://www.ajpm-online.net/article/S0749-3797(03)00021-7/fulltext.

The translation has the same intended function as the original, with the only differences being the target population, their native language and the country of administration (Brazil).

<div align="center">The translation protocol</div>

Stage 1: First translation

Since the objective is to demonstrate translator competence, there are no restrictions whatsoever on the methods the translator may use to produce the initial translation: reference materials, translation memory, internet searches, consultation with experts, etc. are all acceptable. You also have permission to e-mail the original author to ask for clarification of any ambiguities (Barbara.Ainsworth@asu.edu), but all decisions on translation solutions should be your own. The purpose of these instructions is not to police (*fiscalizar*) translators, but to grant them freedom to work. The main priorities are clarity and consistency of terminology, particularly for the Likert scales. If software other than Microsoft Word is used for the translation, please export the translation to Word and save it as a .doc file (Word 97 – 2003 format).

Stage 2: Revision of first translation

Before starting, please ensure that "track changes" is enabled (*controlar alterações* in Portuguese Word). The second translator should review the translation and correct anything they consider to be an error, whether because they believe there is a problem with the Portuguese or because they believe the translator has misunderstood the original. The second translator may also indicate any suggestions they have for improving the other's translation. If you wish to explain why you have made any or all of the alterations, please do so in a comment. There is not, however, any obligation whatsoever to explain any of your alterations, but the first translator will later have the opportunity to reject or accept them and explanations could improve the chances of her retaining your suggestions.

Please do not correct or improve just for the sake of it. If the translation you are revising does not have any errors and you cannot see any way of improving it, then please simply send it back to me unaltered.

Stage 3: Rejecting or accepting alterations and producing the final version.

The revised version is sent back to the first translator who is now at liberty to accept or reject the alterations made by the other translator; up to and including rejecting all of the alterations and returning to the original version. The first translator does not need to explain anything, but may provide comments as he or she sees fit. There is no obligation to restrict yourself to solutions that have been used in the previous stages if you think of a new version that is better.

Thank you both.

<u>**APPENDIX H – FIRST DRAFT - ETS**</u>

QUESTIONÁRIO SOBRE AUXÍLIOS AMBIENTAIS PARA A PRÁTICA DE EXERCÍCIOS FÍSICOS

(PESQUISA DE AMBIENTE SOCIAL E FÍSICO)

Centro de Pesquisa de Prevenção
Escola de Saúde Pública Norman J. Arnold
Universidade da Carolina do Sul

Investigador Principal:
Barbara E. Ainsworth, PhD, MPH

Seção 3: Ambientes Físicos e Sociais

"Farei algumas perguntas sobre a vizinhança na qual você vive, e em seguida algumas perguntas sobre a comunidade na qual você vive."

"Primeiramente, aqui vão algumas perguntas sobre a vizinhança na qual você vive. Para esta entrevista, a vizinhança é definida como a área no raio de uma milha e meia, ou uma caminhada de dez minutos, a partir de sua residência.

3.1 Há quanto tempo você mora em seu endereço atual?

Número de meses (doze meses ou menos) .... ___ ___

Número de anos (um ano ou mais) .... ___ ___

Não sabe / Indeciso.. 77

Recusou 99

< Nota ao entrevistador: menos de um ano deverá ser inserido como número de meses, e mais de doze meses deverão ser inserido como anos completos. Por exemplo: 5 anos, e não 5 anos e 4 meses.>

3.2 Em geral, você diria que as pessoas da sua vizinhança são:

a. Muito ativas fisicamente 1

b. Ativas fisicamente... 2

c. Não muito ativas fisicamente 3

d. Inativas fisicamente... 4

Não sabe / Indeciso... 7

Recusou 9

3.3 De forma geral, como você classificaria sua vizinhança como um local para caminhadas? Você a considera:

a. Muito prazerosa 1

b. Prazerosa.. 2

c. Não muito prazerosa 3

d. Nem um pouco prazerosa.. 4

Não sabe / Indeciso... 7

Recusou 9

Seção 3: Ambientes Físicos e Sociais

3.4   Em geral, você considera o tráfego de veículos motorizados em sua vizinhança como:

    a.   Pesado 1

    b.   Moderado, OU 2

    c.   Leve? 3

    Não sabe / Indeciso 7

    Recusou... 9

3.5 Sua vizinhança possui calçada?

    a.   Sim... 1

    b.   Não *(pule para a pergunta 3.7)* 2

    Não sabe / Indeciso 7

    Recusou... 9

3.6   Para caminhar em sua vizinhança, você diria que suas calçadas são:

    a.   Muito bem conservadas 1

    b.   Bem conservadas 2

    c.   Não muito bem conservadas 3

    d.   Não conservadas 4

    Não sabe / Indeciso 7

    Recusou... 9

3.7 Para caminhar à noite, você descreveria a ILUMINAÇÃO pública em sua vizinhança como...

    a.   Muito boa... 1

    b.   Boa... 2

    c.   Regular... 3

    d.   Ruim 4

    e.   Muito ruim 5

    Não sabe / Indeciso 7

    Recusou... 9

Seção 3: Ambientes Físicos e Sociais

3.8   Para caminhar em sua vizinhança, você diria que a questão dos cães abandonados é:

a.   Um grande problema... 1

b.   Um problema considerável.. 2

c.   Um pequeno problema 3

d.   Não é um problema 4

Não sabe / Indeciso.. 7

Recusou 9

3.9 O quão segura você considera sua vizinhança, em relação à criminalidade?
Você a considera:

a.   Extremamente segura 1

b.   Muito segura 2

c.   Razoavelmente segura... 3

d.   Nem um pouco segura... 4

Não sabe / Indeciso.. 7

Recusou 9

3.10 Em geral, você acha que a maioria das pessoas em sua vizinhança é confiável?

a.   Sim.. 1

b.   Não.. 2

Não sabe / Indeciso.. 7

Recusou 9

3.11 Sua vizinhança possui locais para recreação pública (tais como piscinas públicas, parques, trilhas de caminhada, ciclovias, centros recreativos, etc)?

a.   Sim.. 1

b.   Não *(pule para a pergunta 3.13)* 2

Não sabe / Indeciso.. 7

Recusou 9

Seção 3: Ambientes Físicos e Sociais

Em geral, como você classifica as condições destes locais de recreação pública? Você os considera:

    a.   Excelentes 1

    b.   Bons 2

    c.   Regulares 3

    d.   Ruins 4

    Não sabe / Indeciso 7

    Recusou 9

3.13 Levando em consideração a forma como o dinheiro público é investido em locais de recreação, aponte qual das afirmações abaixo é mais condizente com a realidade da sua vizinhança:

    Minha vizinhança sempre recebe sua parte devida 1

    Minha vizinhança frequentemente recebe sua parte devida 2

    Minha vizinhança raramente recebe sua parte devida 3

    Minha vizinhança nunca recebe sua parte devida 4

    Não sabe / Indeciso 7

    Recusou 9

3.14 Para atividades físicas, você utiliza QUALQUER local privado para recreação?
(...incluindo aqueles fora de sua vizinhança)

    a.   Sim 1

    b.   Não 2

    Não sabe / Indeciso 7

    Recusou 9

Seção 3: Ambientes Físicos e Sociais

"Para as próximas perguntas, pense sobre a comunidade na qual você vive. Para esta entrevista, uma comunidade é definida como a área em um raio de dez milhas, ou vinte minutos de automóvel, a partir de sua residência.

"Por favor, me diga se você UTILIZA quaisquer dos seguintes recursos e comodidades em sua comunidade. Caso o tipo de recurso ou comodidade mencionado não esteja disponível em sua comunidade, favor notificar."

< Nota ao Entrevistador: Enfatize a palavra "você". A pergunta é feita em nível pessoal, não uso familiar ou comunitário.>

3.15 Trilhas para caminhada?

    a.   Sim – R UTILIZA TRILHAS DE CAMINHADA NA COMUNIDADE 1

    b.   Não – R <u>NÃO UTILIZA</u> TRILHAS DE CAMINHADA NA COMUNIDADE ....

    c.   Minha comunidade não possui trilhas para caminhada 3

    Não sabe / Indeciso 7

    Recusou 9

NOTA: SÃO CONSIDERADAS TRILHAS DE CAMINHADA QUAISQUER TRILHAS OU CAMINHOS PÚBLICOS PROJETADOS PARA SEREM PERCORRIDOS A PÉ.

3.16 Piscinas públicas?

    a.   Sim – R UTILIZA PISCINAS PÚBLICAS NA COMUNIDADE 1

    b.   Não – R <u>NÃO UTILIZA</u> PISCINAS PÚBLICAS NA COMUNIDADE 2

    c.   Minha comunidade não possui piscinas públicas 3

    Não sabe / Indeciso 7

    Recusou 9

3.17 Centros de Recreação Pública?

    a.   Sim – R UTILIZA CENTROS DE RECREAÇÃO PÚBLICA NA COMUNIDADE 1

    b.   Não – R <u>NÃO UTILIZA</u> CENTROS DE RECREAÇÃO PÚBLICA NA COMUNIDADE 2

    c.   Minha comunidade não possui centros de recreação pública 3

    Não sabe / Indeciso 7

    Recusou 9

Seção 3: Ambientes Físicos e Sociais

3.18 Ciclovias ou trilhas para prática de ciclismo?

    a.   Sim – R UTILIZA CICLOVIAS NA COMUNIDADE 1

    b.   Não – R NÃO UTILIZA CICLOVIAS NA COMUNIDADE 2

    c.   Minha comunidade não possui ciclovias ou trilhas para ciclismo 3

    Não sabe / Indeciso 7

    Recusou 9

3.19 Parques / playgrounds / campos ou quadras esportivas?

    a.   Sim – R UTILIZA PARQUES NA COMUNIDADE 1

    b.   Não – R NÃO UTILIZA PARQUES NA COMUNIDADE 2

    c.   Minha comunidade não possui parques / playgrounds / campos ou quadras

    esportivas 3

    Não sabe / Indeciso 7

    Recusou 9

3.20 Escolas que sejam abertas a atividades de recreação pública?

    a.   Sim – R UTILIZA ESCOLAS PARA RECREAÇÃO NA COMUNIDADE . 1

    b.   Não – R NÃO UTILIZA ESCOLAS PARA RECREAÇÃO NA

    COMUNIDADE 2

    c.   Escolas na minha comunidade não são abertas para uso público 3

    Não sabe / Indeciso 7

    Recusou 9

3.21 Você utiliza um shopping center para atividades físicas / caminhadas?

    a.   Sim – R UTILIZA SHOPPINGS PARA ATIVIDADES FÍSICAS NA

    COMUNIDADE 1

    b.   Não – R NÃO UTILIZA SHOPPINGS PARA ATIVIDADES FÍSICAS NA

    COMUNIDADE 2

    c.   Minha comunidade não possui um shopping center 3

    Não sabe / Indeciso 7

    Recusou 9

Seção 3: Ambientes Físicos e Sociais

3.22 Você utiliza programas e recursos de atividade física em locais de culto?

    a.   Sim – R UTILIZA INSTALAÇÕES EM LOCAIS DE CULTO NA COMUNIDADE..1

    b.   Não – R <u>NÃO UTILIZA</u> INSTALAÇÕES EM LOCAIS DE CULTO NA COMUNIDADE..2

    c.   Minha comunidade não possui locais de culto com programas de atividade física 3

    Não sabe / Indeciso...

    Recusou..9

3.23 Você utiliza canais próximos tais como riachos, rios e lagos para atividades físicas aquáticas tis como canoagem, caiaque, natação ou esqui aquático? (NÃO INCLUA ATIVIDADES NÃO-FÍSICAS, COMO NAVEGAÇÃO)

    a.   Sim – R UTILIZA CANAIS PARA ATIVIDADES FÍSICAS AQUÁTICAS NA COMUNIDADE..1

    b.   Não – R <u>NÃO UTILIZA</u> CANAIS PARA ATIVIDADES FÍSICAS AQUÁTICAS NA COMUNIDADE 2

    c.   Minha comunidade não possui canais para uso em atividades físicas 3

    Não sabe / Indeciso..7

    Recusou..9

"As próximas perguntas abordam sua opinião a respeito de instalações para atividades físicas em sua comunidade."

3.24 Para sua própria atividade física, você considera clubes, programas ou outras atividades recreativas / físicas organizadas em sua comunidade como:

    a.   Muito importantes..1

    b.   Importantes 2

    c.   Não muito importantes...3

    d.   Não importantes 4

    e.   Minha comunidade não possui clubes ou programas de atividade física.........5

    Não sabe / Indeciso..7

    Recusou..9

Seção 3: Ambientes Físicos e Sociais

3.25 Em sua comunidade, você diria que todas as pessoas possuem acesso igualitário a instalações de recreação pública?

    a.   Sim... 1

    b.   Não 2

    c.   Minha comunidade não possui quaisquer instalações de recreação pública

    (Pule para a pergunta 4.1) 3

    Não sabe / Indeciso 7

    Recusou.. 9

3.26 Quão seguras são as instalações de recreação pública em sua comunidade? Você as considera:

    Muito seguras.. 1

    Seguras.. 2

    Relativamente inseguras.. 3

    Totalmente inseguras... 4

    Não sabe / Indeciso 7

    Recusou.. 9

3.27 Preocupações a respeito da segurança em instalações de recreação pública influenciam seu uso das mesmas?

    a. Sim 1

    b.   Não 2

    c.   Minha comunidade não possui quaisquer instalações de recreação pública.... 3

    Não sabe / Indeciso 7

    Recusou.. 9

<u>**APPENDIX I – THIRD DRAFT - ETS**</u>

QUESTIONÁRIO SOBRE AUXÍLIOS DO AMBIENTE PARA A PRÁTICA DE EXERCÍCIOS FÍSICOS

(PESQUISA DE AMBIENTE SOCIAL E FÍSICO)

Centro de Pesquisa em Prevenção
Norman J. Arnold School of Public Health
University of South Carolina

Pesquisador Principal:
Barbara E. Ainsworth, PhD, MPH

Seção 3: Ambientes Físicos e Sociais

"Farei algumas perguntas sobre a vizinhança na qual você vive, e em seguida algumas perguntas sobre a comunidade na qual você vive."

"Primeiramente, aqui vão algumas perguntas sobre a vizinhança na qual você vive. Para esta entrevista, a vizinhança é definida como a área no raio de aproximadamente 2,4 km, ou uma caminhada de dez minutos a partir de sua residência.

3.1   Há quanto tempo você mora em seu endereço atual?

Número de meses (doze meses ou menos) .... ___ ___

Número de anos (um ano ou mais) .... ___ ___

Não sabe / Indeciso.. 77

Recusou.. 99

< Nota ao entrevistador: menos de um ano deverá ser inserido como número de meses, e mais de doze meses deverão ser inserido como anos completos. Por exemplo: 5 anos, e não 5 anos e 4 meses.>

3.2   Em geral, você diria que as pessoas da sua vizinhança são:

a.   Muito ativas fisicamente 1

b.   Ativas fisicamente... 2

c.   Não muito ativas fisicamente 3

d.   Inativas fisicamente.. 4

Não sabe / Indeciso.. 7

Recusou-se a responder 9

3.3 De forma geral, como você classificaria sua vizinhança como um local para caminhadas? Você a considera:

a.   Muito prazerosa 1

b.   Prazerosa... 2

c.   Não muito prazerosa 3

d.   Nem um pouco prazerosa.. 4

Não sabe / Indeciso.. 7

Recusou-se a responder 9

Seção 3: Ambientes Físicos e Sociais

3.4   Em geral, você considera o tráfego de veículos motorizados em sua vizinhança como:

    a.   Pesado 1

    b.   Moderado, OU 2

    c.   Leve? 3

    Não sabe / Indeciso 7

    Recusou-se a responder 9

3.5 Sua vizinhança possui calçada?

    a.   Sim 1

    b.   Não *(pule para a pergunta 3.7)* 2

    Não sabe / Indeciso 7

    Recusou-se a responder 9

3.6   Para caminhar em sua vizinhança, você diria que suas calçadas são:

    a.   Muito bem conservadas 1

    b.   Bem conservadas 2

    c.   Não muito bem conservadas 3

    d.   Não conservadas 4

    Não sabe / Indeciso 7

    Recusou-se a responder 9

3.7 Para caminhar à noite, você descreveria a ILUMINAÇÃO pública em sua vizinhança como...

    a.   Muito boa 1

    b.   Boa 2

    c.   Regular 3

    d.   Ruim 4

    e.   Muito ruim 5

    Não sabe / Indeciso 7

    Recusou-se a responder 9

Seção 3: Ambientes Físicos e Sociais

3.8 Para caminhar em sua vizinhança, você diria que a questão dos cães abandonados é:

    a.   Um grande problema..1

    b.   Um problema considerável..2

    c.   Um pequeno problema 3

    d.   Não é um problema 4

    Não sabe / Indeciso...7

    Recusou-se a responder 9

3.9 O quão segura você considera sua vizinhança, em relação à criminalidade?

Você a considera:

    a.   Extremamente segura 1

    b.   Muito segura 2

    c.   Razoavelmente segura...3

    d.   Nem um pouco segura...4

    Não sabe / Indeciso...7

    Recusou-se a responder 9

3.10 Em geral, você acha que a maioria das pessoas em sua vizinhança é confiável?

    a.   Sim...1

    b.   Não...2

    Não sabe / Indeciso...7

    Recusou-se a responder 9

3.11 Sua vizinhança possui locais para recreação pública (tais como piscinas públicas, parques, trilhas de caminhada, ciclovias, centros recreativos, etc.)?

    a.   Sim...1

    b.   Não *(pule para a pergunta 3.13)* 2

    Não sabe / Indeciso...7

    Recusou-se a responder 9

Seção 3: Ambientes Físicos e Sociais

Em geral, como você classifica as condições destes locais de recreação pública? Você os considera:

    a.   Excelentes 1

    b.   Bons 2

    c.   Regulares 3

    d.   Ruins 4

    Não sabe / Indeciso 7

    Recusou-se a responder 9

3.13 Levando em consideração a forma como o dinheiro público é investido em locais de recreação, aponte qual das afirmações abaixo é mais condizente com a realidade da sua vizinhança:

    a.   Minha vizinhança sempre recebe sua parte devida.. 1

    b.   Minha vizinhança frequentemente recebe sua parte devida 2

    c.   Minha vizinhança raramente recebe sua parte devida 3

    d.   Minha vizinhança nunca recebe sua parte devida... 4

    Não sabe / Indeciso 7

    Recusou-se a responder 9

3.14 Para atividades físicas, você utiliza QUAISQUER locais de recreação privados ou instalações de recreação apenas para membros? (...incluindo aqueles fora de sua vizinhança)

    a.   Sim 1

    b.   Não 2

    Não sabe / Indeciso 7

    Recusou-se a responder 9

Seção 3: Ambientes Físicos e Sociais

"Para as próximas perguntas, pense sobre a comunidade na qual você vive. Para esta entrevista, uma comunidade é definida como a área em um raio de aproximadamente dez quilômetros, ou a vinte minutos de automóvel a partir de sua residência.

"Por favor, me diga se você UTILIZA quaisquer dos seguintes recursos e comodidades em sua comunidade. Caso o tipo de recurso ou comodidade mencionado não esteja disponível em sua comunidade, favor notificar."

< Nota ao Entrevistador: Enfatize a palavra "você". A pergunta é feita em nível pessoal, não uso familiar ou comunitário.>

3.15 Trilhas para caminhada?

    a.   Sim – R UTILIZA TRILHAS DE CAMINHADA NA COMUNIDADE 1

    b.   Não – R <u>NÃO UTILIZA</u> TRILHAS DE CAMINHADA NA COMUNIDADE ....

    c.   Minha comunidade não possui trilhas para caminhada 3

    Não sabe / Indeciso 7

    Recusou-se a responder 9

NOTA: SÃO CONSIDERADAS TRILHAS DE CAMINHADA QUAISQUER TRILHAS OU CAMINHOS PÚBLICOS PROJETADOS PARA SEREM PERCORRIDOS A PÉ.

3.16 Piscinas públicas?

    a.   Sim – R UTILIZA PISCINAS PÚBLICAS NA COMUNIDADE 1

    b.   Não – R <u>NÃO UTILIZA</u> PISCINAS PÚBLICAS NA COMUNIDADE 2

    c.   Minha comunidade não possui piscinas públicas 3

    Não sabe / Indeciso 7

    Recusou-se a responder 9

3.17 Centros de Recreação Pública?

    a.   Sim – R UTILIZA CENTROS DE RECREAÇÃO PÚBLICA NA COMUNIDADE 1

    b.   Não – R <u>NÃO UTILIZA</u> CENTROS DE RECREAÇÃO PÚBLICA NA COMUNIDADE 2

    c.   Minha comunidade não possui centros de recreação pública 3

    Não sabe / Indeciso 7

    Recusou-se a responder 9

Seção 3: Ambientes Físicos e Sociais

3.18 Ciclovias ou trilhas para prática de ciclismo?

    a.   Sim – R UTILIZA CICLOVIAS NA COMUNIDADE 1

    b.   Não – R NÃO UTILIZA CICLOVIAS NA COMUNIDADE 2

    c.   Minha comunidade não possui ciclovias ou trilhas para ciclismo 3

    Não sabe / Indeciso 7

    Recusou-se a responder 9

3.19 Parques / playgrounds / campos ou quadras esportivas?

    a.   Sim – R UTILIZA PARQUES NA COMUNIDADE 1

    b.   Não – R NÃO UTILIZA PARQUES NA COMUNIDADE 2

    c.   Minha comunidade não possui parques / playgrounds / campos ou quadras

    esportivas 3

    Não sabe / Indeciso 7

    Recusou-se a responder 9

3.20 Escolas que sejam abertas a atividades de recreação pública?

    a.   Sim – R UTILIZA ESCOLAS PARA RECREAÇÃO NA COMUNIDADE . 1

    b.   Não – R NÃO UTILIZA ESCOLAS PARA RECREAÇÃO NA

    COMUNIDADE 2

    c.   Escolas na minha comunidade não são abertas para uso público 3

    Não sabe / Indeciso 7

    Recusou-se a responder 9

3.21 Você utiliza um shopping center para atividades físicas / caminhadas?

    a.   Sim – R UTILIZA SHOPPINGS PARA ATIVIDADES FÍSICAS NA

    COMUNIDADE 1

    b.   Não – R NÃO UTILIZA SHOPPINGS PARA ATIVIDADES FÍSICAS NA

    COMUNIDADE 2

    c.   Minha comunidade não possui um shopping center 3

    Não sabe / Indeciso 7

    Recusou-se a responder 9

Seção 3: Ambientes Físicos e Sociais

3.22 Você utiliza programas e recursos de atividade física em locais de atividade religiosa?

a.   Sim – R UTILIZA INSTALAÇÕES EM LOCAIS DE ATIVIDADE RELIGIOSA NA COMUNIDADE 1

b.   Não – R <u>NÃO UTILIZA</u> INSTALAÇÕES EM LOCAIS DE ATIVIDADE RELIGIOSA NA COMUNIDADE 2

c.   Minha comunidade não possui locais de atividade religiosa com programas de atividade física 3

Não sabe / Indeciso 7

Recusou-se a responder 9

3.23 Você utiliza canais próximos tais como riachos, rios e lagos para atividades físicas aquáticas como canoagem, caiaque, natação ou esqui aquático? (NÃO INCLUA ATIVIDADES NÃO-FÍSICAS, COMO NAVEGAÇÃO)

a.   Sim – R UTILIZA CANAIS PARA ATIVIDADES FÍSICAS AQUÁTICAS NA COMUNIDADE 1

b.   Não – R <u>NÃO UTILIZA</u> CANAIS PARA ATIVIDADES FÍSICAS AQUÁTICAS NA COMUNIDADE 2

c.   Minha comunidade não possui canais aquáticas para uso em atividades físicas 3

Não sabe / Indeciso 7

Recusou-se a responder 9

"As próximas perguntas abordam sua opinião a respeito de instalações para atividades físicas em sua comunidade."

3.24 Para sua própria atividade física, você considera clubes, programas ou outras atividades recreativas / físicas organizadas em sua comunidade como:

a.   Muito importantes 1

b.   Importantes 2

c.   Não muito importantes 3

d.   Não importantes 4

e.   Minha comunidade não possui clubes ou programas de atividade física 5

Não sabe / Indeciso 7

Recusou 9

Seção 3: Ambientes Físicos e Sociais

3.25 Em sua comunidade, você diria que todas as pessoas possuem acesso igualitário a instalações de recreação pública?

    a.   Sim 1

    b.   Não 2

    c.   Minha comunidade não possui quaisquer instalações de recreação pública

    (Pule para a pergunta 4.1) 3

    Não sabe / Indeciso 7

    Recusou-se a responder 9

3.26 Quão seguras são as instalações de recreação pública em sua comunidade? Você as considera:

    a.   Muito seguras 1

    b.   Seguras 2

    c.   Relativamente inseguras 3

    d.   Totalmente inseguras 4

    Não sabe / Indeciso 7

    Recusou-se a responder 9

3.27 As preocupações a respeito da segurança em instalações de recreação pública influenciam seu uso das mesmas?

    a. Sim 1

    b.   Não 2

    c.   Minha comunidade não possui quaisquer instalações de recreação pública 3

    Não sabe / Indeciso 7

    Recusou-se a responder 9