

**UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO DE COMUNICAÇÃO E EXPRESSÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM LINGÜÍSTICA**

FLÁVIA AZEVEDO

**Desambiguação do item lexical *correto* através de
etiquetadores semânticos: uma abordagem baseada em
corpus.**

Florianópolis, 09 de julho de 2007.

**Desambiguação do item lexical *correto* através de
etiquetadores semânticos: uma abordagem baseada em
corpus.**

Dissertação apresentada ao Programa
de Pós-graduação em Lingüística da
Universidade Federal de Santa
Catarina como requisito parcial para a
obtenção do título de Mestre em
Lingüística.

Orientador: Prof. Dr. MARCO ROCHA

Florianópolis, 09 de julho de 2007.

AGRADECIMENTOS

Cabe aqui um agradecimento a todos que contribuíram decisivamente para o desenvolvimento desse trabalho.

Ao professor Marco Rocha, que me incentivou a dar o pontapé inicial do caminho que me fez chegar até esse “final”.

À minha querida amiga Sabrina pelas tardes de estudo nos primeiros semestres do curso.

Às minhas queridas amigas Gabriela e Fernanda, que apesar de não terem acompanhado esse processo de perto, sempre estiveram ao meu lado me apoiando nos momentos mais difíceis.

Aos meus queridos pais pelo apoio e incentivo incondicional em todas as etapas da minha vida.

Muito obrigada.

SUMÁRIO

INTRODUÇÃO.....	1
------------------------	----------

1 – A teoria do princípio colocacional e sua aplicação na desambiguação automática de sentidos.....	5
--	----------

- 1.1 Introdução.....5
- 1.2 O princípio colocacional.....7
- 1.3 Anotação de corpus.....10
- 1.4 O problema da polissemia.....14
- 1.5 A base de dados lexicais WordNet.....19
- 1.6 A desambiguação automática de sentidos.....23
- 1.7 Concluindo o capítulo.....27

2 – Metodologia.....	31
-----------------------------	-----------

- 2.1 Introdução.....31
- 2.2 A observação dos padrões de colocações.....33
- 2.3 Concluindo o capítulo.....40

3 – Resultados.....	49
----------------------------	-----------

- 3.1 Introdução.....49
- 3.2 A definição do conjunto de etiquetas.....50
 - 3.2.1 Sentido 1 (exato).....50
 - 3.2.2 Sentido 2 (adequado)52
 - 3.2.3 Sentido 3 (isento de erros).....53
 - 3.2.4 Sentido 4 (válido).....55
 - 3.2.5 Sentido 5 (justo).....56
 - 3.2.6 Sentido 6 (aceitável).....58

3.2.7	Sentido 7 (com a razão).....	59
3.2.8	Sentido 8 (verdade).....	59
3.2.9	“Sentido” 9 (marcador de discurso).....	60
3.2.10	Sentido 10 (honesto).....	61
3.2.11	Sentido 11 (ideal).....	62
3.2.12	Sentido 12 (bem-feito).....	63
• 3.3	Concluindo o capítulo.....	64
4	–Modelo de processamento baseado em n-gramas.....	65
CONCLUSÃO.....		132
REFERÊNCIAS BIBLIOGRÁFICAS.....		135
APÊNDICE.....		137

RESUMO

Esta pesquisa tem como objetivo investigar os possíveis sentidos do item lexical *correto* em um corpus de português escrito. Espera-se utilizar os resultados como subsídio para fundamentar lingüisticamente um sistema capacitado a realizar a desambiguação deste item lexical. A análise do contexto imediato das ocorrências do item lexical no corpus é vista como a fonte fundamental de informações lingüísticas tendo em vista a mencionada desambiguação. O material utilizado no presente trabalho foi coletado do corpus NILC (Núcleo Institucional de Lingüística Computacional), e a análise foi feita através do programa WordSmith (Scott, 1995), um software para a análise de corpora. O estudo deve ser visto como uma contribuição inicial para a definição de uma metodologia de análise de itens lexicais de polissemia complexa que venha a servir de base para investigações semelhantes, uma vez que os diferentes sentidos encontrados podem ser transformados em um conjunto de etiquetas para uso em um etiquetador semântico dentro de um sistema que utiliza tecnologia de análise lingüística.

Palavras-Chave: lingüística de corpus; polissemia; desambiguação de sentido de palavras; etiquetador semântico.

INTRODUÇÃO

No âmbito das áreas de pesquisa, pode-se dizer que estamos vivendo um momento de revolução no pensamento lingüístico. Dentro do meio acadêmico constantemente têm surgido questões sobre o que é a língua, como esta se organiza e como a mesma deve ser estudada e ensinada atualmente. Grande parte dessa revolução veio como consequência da inserção da tecnologia dentro da área de estudos das ciências humanas, mais especificamente com o advento do computador digital que trouxe novas perspectivas no âmbito da investigação científica.

Segundo Berber Sardinha (2004), o computador, com sua grande capacidade de armazenamento e processamento de dados, “começa a desempenhar nas ciências humanas, o papel transformador que o telescópio teve na física e nas ciências exatas”. É pertinente afirmar que essa revolução está relacionada ao desenvolvimento da Lingüística de Corpus. O termo **Lingüística de Corpus** apareceu pela primeira vez em um livro denominado *Corpus Linguistics*, organizado por Aarts e Meijs (1984), onde se encontra associado à utilização de computadores como ferramenta de análise na área de pesquisa lingüística.

Ainda não há um consenso no meio acadêmico sobre a definição do status da área: afinal, a Lingüística de Corpus é uma disciplina ou uma metodologia? Não entrarei no mérito dessa discussão (para uma discussão mais detalhada ler Berber Sardinha 2004), mas deixarei expressa a minha simpatia pela definição defendida por Hoey: “Lingüística de Corpus não é um ramo da lingüística, mas a rota para a lingüística” (Hoey 1997 *apud* Berber Sardinha 2004). Partindo dessa premissa, a Lingüística de Corpus pode ser entendida como uma *abordagem*, e por esse motivo gostaria de definir esse trabalho como uma *abordagem baseada em corpus*.

Tradicionalmente, o termo corpus tem sido usado por lingüistas para designar um conjunto de dados de certa língua natural que pode ser utilizado como base para pesquisas lingüísticas. Levando-se em consideração todas as características importantes atribuídas a essa abordagem, a definição abaixo é bem adequada:

Um conjunto de dados lingüísticos (pertencentes ao uso oral ou escrito da língua, ou a ambos), sistematizados segundo determinados critérios, suficientemente extensos em amplitude e profundidade, de maneira que sejam representativos da totalidade do uso lingüístico ou de algum de seus âmbitos, dispostos de tal modo que possam ser processados por computador, com a finalidade de propiciar resultados vários e úteis para a descrição e análise. (Sánchez 1996 *apud* Berber Sardinha 2004)

Partindo dessa definição, pode-se deduzir que o uso de corpora em uma pesquisa implica em uma observação detalhada de fenômenos lingüísticos em amostras autênticas de uso da língua. Apesar do esforço de alguns lingüistas do passado, que corajosamente utilizaram corpora que eram coletados, mantidos e analisados manualmente, fica evidente que, antes da tecnologia dos computadores, esse tipo de estudo era praticamente inviável, não só devido à dificuldade de se fazer uma análise detalhada de palavras dentro de um contexto, mas também devido a pouca confiabilidade dos resultados, uma vez que o ser humano não é talhado para trabalhos dessa natureza.

Além disso, fatores como o trabalho e a grande quantidade de pessoas que estudos com base nesse tipo de abordagem envolveriam também contribuíram para a inviabilidade desse tipo de pesquisa durante muitos anos. Mesmo assim, investigações de base em corpus em lingüística e lexicografia foram produzidas desde o século XVIII, até a década de cinquenta com os estruturalistas norte-americanos (para uma resenha

das abordagens de base em corpus antes do aparecimento de recursos computacionais, ver Francis 1992). Felizmente, com o advento do computador e a possibilidade de processar e armazenar um grande número de informações, hoje esse tipo de pesquisa se tornou mais fácil e viável, fazendo parte de uma área de estudo que se encontra em plena expansão.

A partir da metade da década de oitenta, um grande número de estudos estruturados a partir da metodologia com base em corpus passou a ser mais freqüente na área de investigações lingüísticas contemporâneas. Com o barateamento dos custos das máquinas, tornou-se possível armazenar e processar um grande número de dados para fins de pesquisa sem grandes dificuldades.

Para se ter uma breve idéia, o primeiro corpus lingüístico eletrônico foi lançado em 1964, o Brown University Standard Corpus of Present-day American English. Esse corpus continha 1 milhão de palavras, uma quantidade de dados considerada impressionante para a época, mas pouco significativa se comparada a um corpus atual como o BNC (British National Corpus), que conta hoje com aproximadamente 100 milhões de palavras.

Portanto, pode-se perceber que a utilização intensiva das máquinas viabilizou o processamento automático de uma grande quantidade de dados de línguas naturais, fazendo com que a utilização de corpora se tornasse uma rotina entre diversos pesquisadores da área.

Além disso, é importante destacar que esse tipo de abordagem serve como uma alternativa para aqueles pesquisadores que resistem ao distanciamento da teoria lingüística em relação à língua utilizada no cotidiano. A Lingüística de Corpus permite que não apenas lingüistas, mas também professores, tradutores, lexicógrafos e outros

profissionais da área possam questionar paradigmas da teoria lingüística que foram estabelecidos sem uma abordagem empírica.

O objetivo desse trabalho é fazer uma análise das colocações do item lexical *correto* e definir quais significados essa palavra polissêmica pode assumir dentro do corpus. A partir dessas informações espera-se descobrir significados novos, diferenças estilísticas e suas colocações típicas dentro de vários contextos diferentes.

Por fim, pode-se afirmar que a principal justificativa desse trabalho é a sua possível contribuição para a definição de uma metodologia de análise de itens lexicais de polissemia complexa que possa servir de base para possíveis investigações semelhantes com base em corpus. Em termos de aplicação prática, os resultados podem ser transformados em um conjunto de etiquetas para uso em um etiquetador semântico dentro de um sistema que utiliza tecnologia de análise lingüística.

No primeiro capítulo será apresentada uma resenha da literatura, na qual os trabalhos dos autores mais importantes para esta pesquisa serão discutidos, bem como a sua inserção nesta mesma pesquisa; no segundo capítulo será apresentada a metodologia que orientou as atividades da pesquisa realizadas; no terceiro capítulo uma discussão detalhada dos resultados encontrados; e no quarto capítulo o modelo de n-gramas que pode ser utilizado para desambiguar o item lexical polissêmico em questão.

CAPÍTULO 1

REFERENCIAL TÉORICO:

A TEORIA DO PRINCÍPIO COLOCACIONAL E SUA APLICAÇÃO NA DESAMBIGUAÇÃO AUTOMÁTICA DE SENTIDOS

Para uma ampla classe de casos – embora não para todos – no qual empregamos a palavra sentido, ela pode ser definida da seguinte maneira: o sentido de uma palavra é seu uso na língua.

(Wittgenstein 1968: 43)

1.1 INTRODUÇÃO

O objetivo dessa resenha da literatura não é ser exaustiva, mas sim fazer uma breve discussão sobre aspectos das abordagens de base em corpus em lingüística, tratando de assuntos relevantes para esta pesquisa, tais como anotação de corpus, programas de desambiguação de sentidos, o fenômeno da polissemia e as dificuldades que esta apresenta para dicionários e sistemas computacionais, o princípio da desambiguação via colocação e a descrição do banco de dados lexicais eletrônico, o WordNet.

De acordo com Berber Sardinha (2004) “a Lingüística de Corpus ocupa-se da coleta e da exploração de corpora, ou conjuntos de dados lingüísticos textuais coletados criteriosamente, com o propósito de servirem para pesquisa de uma língua ou variedade lingüística”. Esse tipo de abordagem tem como objetivo a exploração da linguagem por meio de evidências empíricas extraídas com o auxílio do computador.

Um dos aspectos que atrai muito a atenção dos lingüistas de corpus é a busca por evidências que mostrem certa padronização do léxico, isto é, “uma regularidade nos

tipos de associação a que se submetem as palavras de uma língua” (Berber Sardinha, 2004). A descrição desse tipo de padrão já foi estudada por vários autores, como afirma Berber Sardinha (2004):

Os padrões de uma palavra podem ser definidos como todas as palavras e estruturas com as quais são regularmente associadas e que contribuam para seu significado. Um padrão pode ser identificado se uma combinação de palavras ocorre com relativa freqüência, se é dependente de uma palavra específica, e se há um significado claro associado.

A área que se ocupa da descrição de padrões se preocupa com questões como descobrir quais os padrões lexicais dos quais a palavra faz parte; investigar se a palavra se associa regularmente com outros sentidos específicos; determinar em quais estruturas ela aparece e estabelecer se há uma correlação entre o uso/sentido da palavra e as estruturas das quais ela participa e verificar se a mesma está associada com uma certa posição organizacional no texto (Berber Sardinha, 2004).

Todas as questões citadas acima se referem à padronização como regularidade expressa na recorrência sistemática de unidades co-ocorrentes de várias ordens (lexical, gramatical, sintática etc.). Segundo Berber Sardinha (2004), esses padrões podem ser formalizados em três conceitos principais:

1) Colocação: associação entre itens lexicais, ou entre o léxico e campos semânticos. Por exemplo, em termos lexicais, *stark* associa-se a *contrast*; *sheer*, a *scale*, *number* e *force*. Em termos de campos semânticos, *jam* relaciona-se com itens no campo dos alimentos: *tarts*, *butty* e *doughnuts*.

2) Coligação: associação entre itens lexicais e gramaticais. Por exemplo, *start* é mais comum com sintagmas nominais e orações -ing, enquanto *begin* é mais usado com um complemento *to*.

3) Prosódia semântica: associação entre itens lexicais e conotação (negativa, positiva ou neutra) ou instância avaliativa. O nome deve-se ao fato de certas palavras prepararem o ouvinte, ou o leitor, para o conteúdo semântico que está por vir, da mesma maneira que a prosódia na fala indica para o interlocutor que tipos de sons estão por vir a seguir. Por exemplo, *cause* tem uma prosódia semântica negativa, pois se associa a palavras desfavoráveis como *problem(s)*, *damage*, *death(s)*, *disease*, *concern* e *cancer*. Já *provide* possui uma prosódia semântica positiva ou neutra, uma vez que se associa a palavras desse tipo, tais como *assistance*, *care*, *jobs*, *opportunities* e *training*.

Na área de lingüística de corpus, o fenômeno da colocação é o mais tradicionalmente focado, e este será o recurso utilizado para fazer a análise do item lexical *correto* neste trabalho. Para esse fim, parece pertinente falar um pouco mais detalhadamente sobre o termo colocação.

1.2 O PRINCÍPIO COLOCACIONAL

De acordo com Partington (1998), o principal tipo de pesquisa com base em corpus dentro do léxico investiga a frequência das palavras e seus sentidos em tipos diferentes de textos ou variedades da língua e suas colocações, isto é, seus padrões de combinações com outras palavras. Esse tipo de estudo era muito difícil de realizar sem a utilização de grandes quantidades de dados que os corpora oferecem.

Partington (1998) também afirma que a maior parte das análises lingüísticas feitas utilizando corpora computadorizados vem de uma filosofia metodológica estatística, ou seja, uma procura por padrões recorrentes. A análise é baseada nos

conceitos de frequência e probabilidade. Em outras palavras, se algum evento acontece frequentemente em uma língua, pode ser considerado relevante. É relevante precisamente porque essa ocorrência freqüente, ou regularidade, pode ser usada como base para deduzir como outros aspectos da língua ainda não analisados vão se comportar, e por fim fazer hipóteses de uma descrição de como o universo inteiro do discurso que está sendo analisado é construído.

O termo colocação, propriamente dito, foi introduzido por Firth juntamente com o famoso slogan que diz: “you shall judge a word by the company it keeps”, ou seja, “diga-me com quem a palavra anda e eu te direi quem ela é”. Para Firth, “sentido por colocação é uma abstração no nível sintagmático e não está diretamente relacionado à abordagem conceitual ou ideal do sentido das palavras. Um dos sentidos de *noite* é a sua colocação com *escura*, e, de *escura*, é claro, sua colocação com *noite*”. (Firth 1957, *apud* Partington 1998).¹

Outros escritores posteriormente se apropriaram de diferentes aspectos das idéias de Firth, e outras definições para o termo “colocação” acabaram surgindo. Partington (1998) diz que há três definições de colocação principais na literatura. A primeira seria a definição textual (textual definition), especificada inicialmente por Sinclair: “colocação é a ocorrência de duas ou mais palavras no âmbito de um pequeno espaço de texto”.² A segunda seria a definição psicológica ou associativa, atribuída a Leech: “o sentido colocacional consiste das associações que uma palavra realiza por conta dos sentidos das outras palavras que tendem a ocorrer no seu ambiente”.³ E por fim a definição estatística de Hoey: “colocação tem sido o nome dado à relação que um

¹ Tradução da autora. Meaning by collocation is an abstraction at the syntagmatic level and is not directly concerned with the conceptual or idea approach to the meaning of words. One of the meanings of *night* is its collocability with *dark*, and, of *dark*, of course, collocation with *night*. (Firth 1957: 196)

² Tradução da autora. Collocation is the occurrence of two or more words within a short space of each other in a text.

³ Tradução da autora. Collocative meaning consists of the associations a Word acquires on account of the meanings of words which tend to occur in its environment.

item lexical tem com itens que aparecem com probabilidade maior do que a que poderia ser atribuída ao acaso no seu contexto textual”.⁴

Segundo Partington (1998), a definição estatística é muito boa para aqueles que estudam na área de lingüística de corpus, “onde grandes quantidades de texto podem estar disponíveis para análise através de computador. A co-ocorrência de dois itens se torna interessante se parece acontecer por uma razão, e especialmente se é repetida, se há “padrões de colocação”. As associações habituais de uma palavra com outros itens podem ser estudadas tanto através das concordâncias da palavra em questão quanto através de uma lista de suas principais colocações”.⁵

A partir do fenômeno da colocação, um fato marcante da língua observado por Firth, Sinclair (que foi aluno de Firth na Universidade de Londres) formulou e introduziu o princípio colocacional como termo em lingüística de corpus. O princípio colocacional é, portanto, uma noção teórica com base na qual se postula a atuação de um segundo elemento, juntamente com um princípio sintático, na produção e compreensão das línguas humanas.

É importante ressaltar que os termos “colocação” e “princípio colocacional” são conceitos distintos. As colocações são um fato da língua e o princípio colocacional é uma noção teórica que não pode ser utilizada sem adaptação. Portanto, para os fins desta pesquisa, o princípio colocacional será adaptado, ou seja, o objetivo inicial será detectar com quais palavras o item lexical *correto* aparece regularmente associado formando bigramas que podem ser utilizados como base para desambiguá-lo nos diversos contextos em que este se encontra dentro do corpus.

⁴ Tradução da autora. Collocation has long been the name given to the relationship a lexical item has with items that appear with greater than random probability in its (textual) context.

⁵ Tradução da autora. (...) where large quantities of text can be made available for computer analysis. The co-occurrence of two items becomes interesting if it seems to happen for a purpose, and especially if it is repeated, if there are “patterns of collocation”. The habitual associations of a word with other items can thus be studied both by calling up concordances of that word and by obtaining lists of its most frequent collocates.

1.3 ANOTAÇÃO DE CORPUS

Além da observação das principais colocações de um determinado item lexical dentro de um contexto, existem outras maneiras de se obter conhecimento lingüístico a partir de um corpus. A prática de anotação de corpus é um bom exemplo e merece ser destacado. A anotação de corpus pode ser definida como “a prática de adicionar informação lingüística interpretativa em um corpus eletrônico de língua escrita ou falada” (Garside et al, 1997) ⁶. Segundo esses autores, o termo “anotação” também pode fazer referência ao produto final desse processo, ou seja, os símbolos lingüísticos que são colocados junto à representação eletrônica do material lingüístico propriamente dito”. ⁷

Um exemplo típico de anotação de corpus é a etiquetagem gramatical. Nesse caso, uma etiqueta é associada a uma palavra para indicar sua classe gramatical. Por exemplo, em ‘*correto_ADJ*’, a etiqueta *_ADJ* mostra que a palavra *correto* é um adjetivo. Mas esse processo de anotação não é tão simples como parece. Nesse caso mencionado acima, é pouco esperado que exista alguma resistência em relação a considerar essa palavra como um provável adjetivo, mas outras palavras podem trazer problemas. No caso de uma palavra como *fala*, por exemplo, teria que se ter o cuidado de determinar em quais contextos ela seria um substantivo ou uma forma flexionada do verbo falar.

Garside et al. (1997) afirma que a anotação de corpus é, em certo nível, “interpretativa”, uma vez que é produto do entendimento da mente humana sobre o

⁶ Tradução da autora. (...) the practice of adding interpretative, linguistic information to an electronic corpus of spoken and/or written data.

⁷ Tradução da autora. ‘Annotation’ can also refer to the end-product of this process: the linguistic symbols which are attached to, linked with, or interspersed with the electronic representation of the language material itself.

texto. Portanto, não há um objetivo puro, uma maneira mecanicista de decidir quais etiquetas devem ser atribuídas a um determinado fenômeno.⁸ “A anotação de um texto é metalingüística: em vez de nos dizer o que o próprio texto abrange, ela nos dá informações sobre a língua do texto”.⁹

Mas afinal, qual seria a importância de se anotar um corpus? Garside et al. (1997) apresenta três motivos para se anotar um corpus. O primeiro seria a extração de informações, uma vez que “os corpora só tem utilidade se nós pudermos extrair conhecimento ou informação deles. O fato é que, para extrair informações de um corpus, normalmente é necessário começar adicionando informações ao mesmo, ou seja, anotando o corpus”.¹⁰

“O corpus original não contém nenhum tipo de informação em sua forma ortográfica sobre gramática, por exemplo. Isso pode impedir que muitas das aplicações que o corpus poderia ter sejam de fato executadas. Se um corpus é anotado de maneira satisfatória, todas as palavras virão acompanhadas de uma etiqueta que indica sua classe de palavra. Esse é um pré-requisito para qualquer profissional que esteja utilizando o corpus para fazer ou melhorar dicionários”, por exemplo.¹¹

O segundo motivo para se anotar um corpus é o princípio da reutilização. “A identificação de classes de palavras (ou qualquer outro fenômeno lingüístico) não pode ser tratada como um problema isolado. Uma vez que a anotação tenha sido feita, o

⁸ Tradução da autora. (...) annotation is, at least in some degree, the product of human mind's understanding of the text. There is no purely objective, mechanistic way of deciding what label or labels should be applied to a given linguistic phenomenon.

⁹ Tradução da autora. (...), the annotation of a text is *metalinguistic*: instead of telling us what the text itself comprises, it gives information *about* the language of the text.

¹⁰ Tradução da autora. Corpora are useful only if we can extract knowledge or information of them. The fact is that to extract information from a corpus, we often have to begin by building information in – that is, by adding annotations.

¹¹ Tradução da autora. The ‘raw corpus’ in its orthographic form contains no direct information, for example, about grammar – and this can hinder many of the applications to which a corpus can be put. (...) if a corpus is successfully grammatically tagged, each occurrence of *left* will be accompanied by a label indicating its word-class. This is a pre-requisite for anyone using a corpus for making or improving dictionaries.

corpus resultante desse processo pode ser considerado uma fonte muito mais valiosa de informações do que o corpus original, podendo ser manipulado por outros usuários”.¹² Segundo os autores, “esse argumento de reutilização é muito poderoso, já que a anotação de corpus tende a ser um trabalho muito caro que consome muito tempo”.¹³

O terceiro motivo apresentado é a multifuncionalidade (multi-functionality), já que a anotação de corpus tem muitos propósitos e diversas aplicações. Garside et al. (1997) afirma que “o ponto principal trazido pelos autores é que a anotação dá um ‘valor adicional’ ao corpus de uma maneira geral: ela traz informações lingüísticas claras que podem ser utilizadas para propósitos variados, múltiplos. Além disso, a etiquetagem gramatical, um tipo básico de anotação, pode ser o primeiro passo para níveis de anotação mais difíceis e complexos como os sintáticos e semânticos”.¹⁴

Em relação às abordagens utilizadas no processo de etiquetagem, Sampson (1996) observa que “técnicas baseadas em regras lógicas rígidas serão sempre frágeis para dar conta dos dados disponibilizados por um corpus”. Ele salienta que conseguiu um sucesso notável ao usar um método probabilístico para realizar a tarefa de etiquetagem automática de palavras no âmbito da identificação das funções gramaticais de palavras em um texto.¹⁵

Logo após a compilação do LOB (Lancaster Oslo-Bergen) Corpus, a equipe de Lancaster já começou a produzir sua versão etiquetada. Com a árdua tarefa de lidar

¹² Tradução da autora. (...), the identification of word-classes (or any other linguistic phenomena) cannot be treated as an isolated problem. (...): once the annotation has been added to the corpus, the resulting annotated corpus can now be handed on to other users.

¹³ Tradução da autora. This argument of ‘re-usability’ is a powerful one, since corpus annotation tends to be an expensive and time consuming business.

¹⁴ Tradução da autora. (...) the general point to make is that annotation gives ‘added value’ to a corpus in the general sense: it adds overt linguistic information, which can then be used for a multitude of purposes. Thus grammatical tagging is often considered a kind of ‘base camp’ annotation which can be the first step towards more difficult levels of annotation such as those of syntax and semantics.

¹⁵ Tradução da autora. Techniques based on rigid logical rules would always be too fragile to cope with our material. In fact the Lancaster team (before I became involved with it) had scored a notable success using a probabilistic approach to one particular natural language analysis task: the task of automatic *word-tagging*, i.e. identifying the grammatical roles of the words of a text.

com um milhão de palavras, era desejável que, à medida do possível, esse trabalho pudesse ser feito automaticamente, e a equipe conseguiu com sucesso criar um etiquetador automático, conhecido como ‘CLAWS’, cujo desempenho foi muito satisfatório. Segundo Sampson (1996), o programa etiquetou corretamente entre 96% e 97% das palavras (dependendo do tipo de texto), sendo que apenas uma pequena quantidade de pós-edição considerada aceitável foi necessária.¹⁶

“O programa ‘CLAWS’ utilizou um algoritmo baseado em probabilidades de transição das etiquetas: isto é, o tipo de conhecimento sobre a língua inglesa incorporada ao sistema foi o conhecimento das frequências relativas com os quais diferentes pares de etiquetas são encontrados adjacentes a outro em textos em inglês. Dada uma série de palavras, cada uma associada a um grupo de uma ou mais etiquetas, o programa escolhe qual caminho pelo qual as etiquetas “candidatas” para palavras sucessivas maximiza a probabilidade de acerto, medida como o produto das probabilidades para transições sucessivas entre pares de palavras etiquetadas adjacentes no caminho (Sampson,1996)”¹⁷.

Concluindo, pode-se dizer que o sucesso da etiquetação de um corpus está diretamente relacionado ao uso de um modelo probabilístico com todas as informações que possam ser obtidas de um corpus, apesar das exigências possivelmente excessivas para os recursos atuais. No entanto, o etiquetador ‘CLAWS’ parece ser uma evidência

¹⁶ Tradução da autora. (...), so once the LOB Corpus was complete the Lancaster group undertook to produce a tagged version. With a million word-tokens to deal with, it was desirable for as much as possible of this work to be achieved automatically; and they succeeded in developing an automatic word-tagger, ‘CLAWS’, that worked extremely well, correctly tagging between 96% and 97% of word-tokens (depending on text type), so that only an acceptably small level of postediting was needed.

¹⁷ Tradução da autora. CLAWS used an algorithm based on tag-transition probabilities: that is, the kind of knowledge about English incorporated in the system – apart from knowledge about what tags were candidates for a word out of context – was knowledge of the relative frequencies with which different pairs of tags (say, preposition and past participle, or definite article and adjective) are found adjacent to one another in English texts. Given a string of words, each associated with a set of one or more candidate tags, CLAWS chose that path through the candidates for successive words which maximized probability, measured as the product of the probabilities for successive transitions between pairs of adjacent wordtags on the path.

de que modelos probabilísticos têm importância fundamental e papel definitivo na eficácia de qualquer programa de etiquetagem automática.

1.4 O PROBLEMA DA POLISSEMIA

Além dos temas até então abordados, é de fundamental importância para o presente trabalho abordar a questão da polissemia e as dificuldades que esta apresenta para dicionários e sistemas computacionais. Pustejovsky levanta em sua obra intitulada *Lexical Semantics* (1996) a seguinte questão: “qual é a representação de um item lexical que o leva a assumir diferentes sentidos em diversos contextos na composição da semântica? Isto é, o que existe na representação de um item lexical que dá origem a extensões de sentidos e ao fenômeno da polissemia lógica”?¹⁸

Essa questão certamente é difícil de ser respondida, uma vez que “a ambigüidade lexical é um dos problemas mais difíceis em estudos de processamento computacional de línguas, e não é surpreendente que esteja no âmbito de pesquisas em semântica lexical. É certamente verdade que a maioria das palavras em uma língua tem mais de um sentido, mas as maneiras como as palavras podem carregar sentidos múltiplos podem variar” (Pustejovsky, 1996).¹⁹

De acordo com Pustejovsky (1996), a ambigüidade lexical é um fenômeno heterogêneo, com pelo menos três fatores distintos que contribuem para o aparecimento contextual de sentidos para um determinado item lexical: a ambigüidade contrastiva, que normalmente é resolvida pelo contexto e pelo conhecimento do discurso; a

¹⁸ Tradução da autora. (...) what is the representation of a lexical item such that it may assume different senses in diverse contexts in composition in the semantics? That is, what is it about the representation of a lexical item that gives rise to sense extensions and to the phenomenon of logical polysemy?

¹⁹ Tradução da autora. Lexical ambiguity is one of the most difficult problems in language processing studies and, not surprisingly, is at the core of lexical semantics research. It is certainly true that most words in a language have more than one meaning, but the ways in which words carry multiple meanings can vary.

ambigüidade complementar (ou polissemia lógica) que seria resolvida pela composição no contexto sintático da sentença; e extensão de sentidos, mediados por regras lexicais e condições específicas relacionadas ao falante e ao contexto.

“A ambigüidade contrastiva, tradicionalmente conhecida como homonímia, é a situação onde um item lexical é associado com pelo menos dois sentidos distintos e não relacionados”, como nos exemplos abaixo citados por Pustejovsky (1996):²⁰

(1) a. The judge asked the defendant to approach the bar.

b. The defendant was in the pub at the bar.

“Foi convencionalmente assumido que homonímias como o par acima são distribuídas em contextos diferentes e não representariam um grande problema de desambiguação em um texto. Essa é a posição defendida por autores como Boguraev (1979), Waltz & Pollack (1985) e Hirst (1987), onde a interpretação heurística do contexto lexical ajuda a estreitar a tarefa de selecionar os sentidos de palavras ambíguas”.²¹

Pustejovsky (1996) afirma que “apesar dessa estratégia poder ser útil se aplicada à terminologia em domínios específicos, para palavras de frequência geral com sentidos comuns de usos e, possivelmente, sentidos especializados também, o problema é mais difícil. Por exemplo, a palavra *bar* tem pelo menos vinte e cinco sentidos diferentes na maior parte dos dicionários de língua inglesa”.²²

²⁰ Tradução da autora. Contrastive ambiguity, traditionally known as homonymy, is the situation where a lexical item is associated with at least two distinct and unrelated meanings.

²¹ Tradução da autora. (...) it has been conventionally assumed that homonyms such as the (a)-(b) pairs above are distributed in different contexts and would therefore not present a real challenge to disambiguation in a text. This is the position taken by *priming-based* disambiguation strategies, such as Boguraev (1979), Waltz & Pollack (1985) e Hirst (1987), where the heuristic interpretation of a lexical context helps narrow the sense selection task for ambiguous words.

²² Tradução da autora. Although the strategy can be usefully applied to terminology in specialized domains, for general frequency words with common use senses and, possibly, specialized ones too, the problem is more difficult. For example, the word *bar* has at least twenty-five distinct senses in most unabridged English dictionaries.

Os sentidos contrastivos utilizados podem parecer facilmente distinguíveis, mas como Asher e Lascarides (*apud* Pustejovsky 1996) demonstram, “o domínio primário não é suficiente para desambiguar tais sentidos lexicais no discurso natural do dia-a-dia. O que seria preciso, de acordo com eles, é uma abordagem de base semântica para a seleção de sentidos, onde a lógica desenvolvida do discurso e as relações retóricas que existem entre as sentenças é que ajudam a determinar a desambiguação do item lexical no discurso”.²³

Na ambigüidade complementar, diferentemente da homonímia, os sentidos a seguir “exibem uma polissemia complementar, onde as leituras alternativas são manifestações do mesmo grupo de sentidos à medida que ocorrem em diferentes contextos”:²⁴

- (2) a. The bank raised its interest rates yesterday. (the institution)
 - b. The store is next to the bank. (the building)
- (3) a. John crawled through the window.
 - b. The window is closed.

Pustejovsky (1996) afirma que nesses casos “a semântica tem, de alguma maneira, que explicar como um banco pode ser tanto uma instituição como um prédio, e como uma janela pode ser tanto uma abertura como um objeto físico. A conexão lógica entre os sentidos lexicais é o que motivou uma representação semântica mais rica para substantivos e adjetivos, conhecida como *qualia structure* (cf. Pustejovsky 1996). O

²³ Tradução da autora. (...), domain-priming is not sufficient to disambiguate such lexical senses in very natural and everyday discourses. What is needed, they argue, is not a priming-based strategy, but a *semantics-based* approach to sense selection, where it is the unfolding logic of a discourse and the rhetorical relations that exist between sentences that help determine the disambiguation of a lexical item in the discourse.

²⁴ Tradução da autora. (...) the senses exhibit a complementary polysemy, where the alternative readings are manifestations of the same core sense as it occurs in different contexts.

termo *qualia* se refere aos modos de explicação para o objeto em questão.”²⁵ As frases abaixo são exemplos do fenômeno da extensão de sentidos:

(4) a. I am parked out back.

b. Ringo squeezed himself into the parking space.

“Essas frases ilustram dois tipos de transferência referencial: um tipo de combinação entre sujeito e predicado em (a), onde é o carro que está parado, não o indivíduo, e uma combinação entre verbo e objeto, junto com uma não-identidade entre o antecedente e a anáfora na relação em (4b). Na contribuição atribuída a Nunberg para esse assunto (*apud* Pustejovsky 1996), tais extensões de sentido são conhecidas como transferências de predicado (*predicate transfers*). Em particular, ele se diz contra uma análise metonímica, onde o sujeito *I* em (4a) e o objeto *himself* em (4b) são interpretados como *my car* e *his car* respectivamente. Sua posição é que existem condições pragmáticas licenciadas que permitem que o predicado extenda seu sentido, onde é reescrito para selecionar os sujeitos que estão presentes na sintaxe”.²⁶

A polissemia não é, portanto, um fenômeno que possa ser considerado isolado. “É antes o resultado de operações composicionais na semântica, como co-composição, e de efeitos contextuais, como a estrutura de relações retóricas no discurso

²⁵ Tradução da autora. The semantics must somehow account for how a bank can be both an institution and a building, and how a window or door can be both an aperture and a physical object. This logical connection between lexical senses is what motivated a richer semantic representation for nouns and adjectives, known as *qualia-structure* (cf. Pustejovsky 1991). The *qualia* refer to modes of explanation for the object.

²⁶ Tradução da autora. These sentences illustrate two types of referential transfer: a type-mismatch between Boguraev (1979), Waltz & Pollack (1985) e Hirst (1987), en subject and predicate in (a), where it is the car that is parked, not the individual; and a mismatch between verb and object, together with a non-identity between antecedent and anaphor in the binding relation in (4b). In Nunberg’s contribution in this issue, such extensions of meaning are referred to as *predicate transfers*. In particular, he argues against a metonymic analysis, where the subject *I* in (4a) and the object *himself* in (4b) are interpreted as *my car* and *his car* respectively. Rather, his position is that there are pragmatically licensed conditions which allow the predicate to extend its sense, where it is retyped to select for the subjects that are present in the syntax.

e limitações pragmáticas em co-referência” (Pustejovsky 1996).²⁷ Por ser um fenômeno tão difícil de ser abordado, a grande questão é desvendar como outros componentes no processo de interpretação da língua natural interagem com o léxico. A partir desse tipo de informação é que se torna possível desambiguar e determinar completamente a semântica de palavras em um contexto.

²⁷ Tradução da autora. (...), it is the result of both compositional operations in the semantics, such as co-composition, and of contextual effects, such as the structure of rhetorical relations in discourse and pragmatic constraints on co-reference.

1.5 A BASE DE DADOS LEXICAIS WORDNET

Nessa seção parece pertinente mencionar uma base de dados lexicais (WordNet), que apesar de não ter sido elaborada com base em corpus, teve um grande impacto na área de lexicografia e acabou sendo utilizada em diversas pesquisas na área de lingüística de corpus. Essa base de dados acabou surgindo como uma proposta de combinação de informações lexicográficas tradicionais com os recursos modernos e velozes da computação e será brevemente comentada nesse trabalho.

Nas últimas décadas, “lingüistas passaram a se interessar sobre as informações que o léxico deve conter para que os componentes lexicais, sintáticos e fonológicos trabalhem juntos na compreensão e produção diária de mensagens lingüísticas, e estas propostas passaram a ser incluídas no trabalho dos psicolingüistas. Começando com estudos de associações de palavras no início do século e continuando com sofisticadas tarefas experimentais dos últimos vinte anos, psicolingüistas descobriram muitas propriedades sincrônicas do léxico mental que podem ser exploradas em lexicografia (Miller, 1993)”.²⁸

Como consequência desse súbito interesse pelas propriedades e organização do léxico mental, um grupo de psicólogos e lingüistas desenvolveu na Universidade de Princeton uma base de dados lexicais feita com base nesse tipo de investigação no ano de 1985. O WordNet é o resultado desse trabalho. “O WordNet é um sistema de referência lexical on-line cuja concepção é inspirada por teorias psicolingüísticas da

²⁸ Tradução da autora. (...), linguists became increasingly explicit about the information a lexicon must contain in order for the phonological, syntactic, and lexical components to work together in the everyday production and comprehension of linguistic messages, and those proposals have been incorporated into the work of psycholinguistics. Beginning with word association studies at the turn of the century and continuing down to the sophisticated experimental tasks of the past twenty years, psycholinguistics have discovered many synchronic properties of the mental lexicon that can be exploited in lexicography.

memória lexical humana. Nomes, verbos e adjetivos são organizados em conjuntos de sinônimos, cada um representando um conceito lexical subjacente. Diferentes relações unem o conjunto de sinônimos” (Miller 1993).²⁹ Esses conjuntos de sinônimos são formalmente denominados synsets.

O conceito de sinonímia é considerado a mais importante relação do WordNet. Miller (1993) diz que “de acordo com a definição (normalmente atribuída a Leibniz) duas expressões são sinônimas se a substituição de uma pela outra nunca altera o valor de verdade de uma sentença na qual a substituição seja feita”.³⁰ O autor ainda aponta que é a própria definição de sinonímia em termos de substituição que faz necessária a partição do WordNet em substantivos, verbos, adjetivos e advérbios.

O WordNet pode ser considerado, portanto, um dicionário baseado em princípios psicolinguísticos, e o que o difere basicamente de um dicionário padrão é a divisão do seu léxico nas cinco categorias mencionadas acima. “A característica mais ambiciosa desse projeto, de acordo com Miller (1993), foi sua tentativa de organizar informações lexicais em termos de significados de palavras, em vez de forma de palavras”.³¹

Marrafa (2001) afirma que, diferentemente do que ocorre usualmente nos dicionários, que têm subjacente uma concepção estática do léxico, no WordNet os sentidos das unidades lexicais não são listados e definidos através de perífrases. Os sentidos são sempre inferidos a partir das relações que estruturam a rede. “Em conseqüência, embora a cada conceito esteja associada uma glosa, com função idêntica

²⁹ Tradução da autora. WordNet is an on-line lexical reference system whose design is inspired by current psycholinguistic theories of human lexical memory. English nouns, verbs, and adjectives are organized into synonym sets, each representing one underlying lexical concept. Different relations link the synonym sets.

³⁰ Tradução da autora. According to one definition (usually attributed to Leibniz) two expressions are synonymous if the substitution of one for the other never changes the truth value of a sentence in which the substitution is made.

³¹ Tradução da autora. The most ambitious feature of WordNet, however, is its attempt to organize lexical information in terms of word meanings, rather than word forms.

à das definições clássicas dos dicionários convencionais, as glosas (não se pretende que sejam definições) constituem apenas informação complementar, não tendo qualquer papel na determinação do sentido” (Marrafa 2001).

Como consequência da construção de WordNets de outras línguas pelo mundo afora, foi criada uma associação com objetivos mais amplos, conhecida como Global WordNet Association (www.hum.uva.nl/~ewn/gwa.htm). Essa associação tem a proposta de criar uma rede mundial de interligação de todas as WordNets existentes. Uma proposta muito interessante que, segundo Marrafa (2001) “contribuirá de forma significativa para conferir estatuto de igualdade - a nível científico, técnico e sócio-político - das línguas representadas em tão importante projeto”.

Todas as WordNets existentes têm como base os pressupostos da pioneira *WordNet* de Princeton, que foi desenvolvida para o inglês americano a partir de uma experiência sobre a organização do léxico mental dirigida por George Miller (Miller 1986). É possível afirmar que o projeto WordNet tornou-se um dos recursos de maior impacto no domínio do Processamento das Línguas Naturais da atualidade. “As unidades lexicais são organizadas em conjuntos de sinônimos que têm subjacente um conceito, sendo cada um desses conjuntos representados por um nó na rede. As ligações entre os diferentes nós exprimem relações semânticas diversas, como por exemplo, as de generalização e de especialização. O sentido emerge, assim, do complexo de relações que a rede permite exprimir (Marrafa 2001)”.

A grande inovação do WordNet está no fato da rede não apenas diferenciar os conceitos entre si, como acontece nos dicionários tradicionais, mas de codificar toda a informação necessária para sua construção, organizando-a numa rede de relações que, ao serem investigadas, é capaz de fazer uma correspondência com a organização do léxico mental. Portanto, como afirma Marrafa (2001), “a construção de uma WordNet

assenta num pressuposto fundamental: o significado das unidades lexicais é basicamente derivado das suas relações com o sentido de outras unidades lexicais. A existência de uma estreita relação semântica entre as unidades lexicais em que este tipo de relação se registra justifica igualmente a sua especificação, por ser de grande utilidade em diversos domínios da Engenharia da Linguagem”.

Os synsets da base TeP (<http://www.nilc.icmc.usp.br/>) foram o ponto de partida utilizado neste trabalho. Os synsets são “conjuntos de sinônimos cognitivos, sendo que cada sinônimo expressa um conceito distinto e são interligados através da semântica conceitual e relações lexicais”.³² (<http://wordnet.princeton.edu/>).

A tentativa de classificar os sentidos encontrados no corpus com base nesse conjunto de sinônimos gerou alguns problemas como a omissão de sentidos muito freqüentes, por exemplo. Esse fato poderia abrir caminhos para a proposta de um WordNet de base em corpus, onde os synsets fossem constituídos a partir da análise dos dados.

³² Tradução da autora. Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets) each expressing a distinct concept. Synsets are interlinked by means of conceptual and lexical relations.

1.6 A DESAMBIGUAÇÃO AUTOMÁTICA DE SENTIDOS

A desambiguação automática de sentido (*word sense disambiguation*, WSD) consiste no processo de selecionar o sentido mais apropriado para uma palavra baseado no contexto em que ela ocorre. Para esse propósito, é pressuposto que um conjunto de sentidos possíveis seja previamente determinado.

Por exemplo, ao se pressupor que o adjetivo *pesado* tem o seguinte conjunto de sentidos possíveis: algo que tem muito peso; árduo; e indigesto, quando utilizado em um contexto “*Coube a ele o serviço pesado todos os dias, em jornadas de até 14 horas*”, o leitor humano entende imediatamente que *pesado* está sendo usado com o sentido de *árduo*. Entretanto, um programa de computador tentando realizar a mesma tarefa enfrenta um grande problema, uma vez que ele não pode contar com um conhecimento lingüístico prévio, como acontece com seres humanos. Por esse motivo, a criação de programas de desambiguação de sentidos acaba se tornando uma tarefa árdua e complexa, que exige um estudo detalhado dos contextos em que a palavra polissêmica está inserida.

Dentro desse âmbito de desambiguação de sentidos, é pertinente comentar sobre o SENSEVAL-2, o segundo workshop internacional sobre avaliação de sistemas de desambiguação de sentidos que aconteceu em julho de 2001 em Toulouse. Esse evento reuniu pesquisadores com o intuito de discutir e solucionar problemas pertinentes ligados a sistemas de desambiguação automática de sentidos. Seu objetivo foi avaliar os pontos fortes e os problemas encontrados nos algoritmos de desambiguação automática de sentidos de palavras diferentes em contextos lingüísticos diversos.

Um dos trabalhos apresentados nesse evento fazia referência à desambiguação automática de sentidos de adjetivos e substantivos da língua inglesa

(English Lexical Sample Task). O corpus escolhido para esse trabalho foi o British National Corpus. O objetivo do trabalho, em princípio, era utilizar dados do inglês americano em conjunto com o BNC, porém, devido à quantidade limitada de dados disponíveis do inglês americano (em função das restrições ocasionadas por direitos autorais), os dados do corpus de inglês britânico acabaram sendo predominantes.

Para fazer a escolha dos sentidos, o WordNet acabou sendo utilizado. O maior problema em relação a essa escolha foi o fato do WordNet ter sido organizado em torno de grupos de palavras de sentidos semelhantes (os synsets), não em torno das palavras e seus sentidos variados. A desambiguação de sentido de palavras, por outro lado, é uma tarefa que exige uma distinção de sentidos muito bem definida e clara. Portanto, para resolver o problema dos casos em que o WordNet fazia uma distinção entre dois sentidos que não era totalmente clara, todas as palavras na amostra lexical e seus respectivos verbetes foram revisados por um lexicógrafo, com uma preocupação especial pelos sentidos não suficientemente distintos. Posteriormente, as revisões propostas foram incluídas na versão do WordNet (1.7).

O desempenho dos melhores sistemas de desambiguação nesse trabalho acabou ficando em torno de 14%. Os pesquisadores acreditam que esses resultados são justificados pela escolha do léxico, nesse caso o WordNet. Eles ressaltam que a alta exatidão da desambiguação de sentido das palavras só é possível quando o léxico faz uma distinção clara e bem definida dos sentidos e oferece informações suficientes sobre as distinções para a construção do algoritmo.

Em julho de 2004, na cidade de Barcelona, foi sediado o SENSEVAL-3, o terceiro workshop internacional sobre avaliação de sistemas de desambiguação de sentidos. Esse evento foi realizado em conjunto com a Associação para Linguística Computacional (Association for Computational Linguistics) e teve grande repercussão.

Foram apresentados 14 trabalhos diferentes e 55 times participaram do evento com mais de 160 sistemas. Dentro desses trabalhos estavam incluídos trabalhos de desambiguação automática de sentidos para diversas línguas, como basco, catalão, chinês, inglês, italiano, romeno e espanhol.

Um dos trabalhos apresentados nesse evento foi o “Word-Sense Disambiguation of WordNet Glosses”, ou seja, Desambiguação Automática das Glosas do WordNet. Segundo Litkowski (2004) esse trabalho “foi criado para encorajar o desenvolvimento de tecnologia que fizesse uso de padrões de recursos lexicais. O trabalho foi baseado na disponibilidade de glosas de desambiguação de sentidos feitas à mão no projeto de extensão do WordNet”.³³

De acordo com o autor, “o processo de anotação seguido do projeto XWN, com as etiquetas usadas nesse trabalho, mais uma vez indicaram dificuldades com o conjunto de sentidos do WordNet. Esse fato permanece devido ao WordNet não ter tido o benefício de recursos lexicográficos suficientes na construção de suas glosas e na aquisição de outras informações lexicográficas em seus verbetes. O projeto WordNet continua com seu esforço para adicionar informação, mas com recursos limitados.”³⁴

Aproveitando os corpora utilizados no SENSEVAL, Pedersen (2001) desenvolveu uma abordagem de desambiguação através do princípio colocacional. Ele apresentou uma abordagem de base em corpus onde uma árvore de decisões atribui um sentido para uma palavra ambígua baseado nos n-gramas de que faz parte. Nessa abordagem, a árvore de decisões é determinada a partir de um número de sentenças

³³ Tradução da autora. The SENSEVAL-3 task to perform word-sense disambiguation of WordNet glosses was designed to encourage development of technology to make use of standard lexical resources. The task was based on the availability of sense disambiguated hand-tagged glosses created in the eXtended WordNet project.

³⁴ Tradução da autora. The annotation process followed by the XWN project, with the taggings used in this task, has again indicated difficulties with the WordNet sense inventory. The fact remains that WordNet has not had the benefit of sufficient lexicographic resources in the construction of its glosses and in the acquisition of other lexicographic information in its entries. The WordNet project continues its efforts to add information, but with limited resources.

onde cada instância da palavra ambígua foi anotada manualmente com uma etiqueta de sentido que denota o sentido mais apropriado para aquele contexto.

Os bigramas são seqüências de duas palavras que ocorrem em um texto. “O contexto em que a palavra ocorre é representado por um número de traços binários que indicam se um determinado bigrama ocorreu em aproximadamente 50 palavras à direita ou à esquerda da palavra a ser desambiguada” (Pedersen, 2001).³⁵ Pedersen (2001) justifica essa abordagem argumentando que “traços lexicais como bigramas, colocações e co-ocorrências geralmente contribuem muito para um bom desempenho dos programas de desambiguação”.³⁶

Depois de todas as questões abordadas, parece evidente que a Lingüística de Corpus acabou exercendo uma maior influência na área de lexicografia, tanto que praticamente todos os grandes dicionários da língua inglesa são feitos com base nesse tipo de abordagem hoje em dia. Esse fenômeno reflete também um interesse no âmbito empresarial de investir em estudos baseados em corpora. Um bom exemplo disso foi o Cobuild, uma parceria entre a Universidade de Birmingham (Grã-Bretanha) e a editora Collins. No âmbito do Cobuild foram produzidos não apenas dicionários com base em corpora, mas também gramáticas e livros didáticos para o ensino do inglês (Berber Sardinha, 2004).

Outro exemplo de projeto que, nesse caso, envolve o uso de corpora da língua portuguesa é o “Dicionário UNESP do Português Contemporâneo” (Borba, 2004), publicado pela Editora UNESP. Esse dicionário, segundo seus organizadores, foi desenvolvido com base em um corpus de 90 milhões de palavras.

³⁵ Tradução da autora. The context in which an ambiguous word occurs is represented by some number of binary features that indicate whether or not a particular bigram has occurred within approximately 50 words to the left or right of the word being disambiguated.

³⁶ Tradução da autora. (...) surface lexical features like bigrams, collocations, and co-occurrences often contribute a great deal to disambiguation accuracy.

Borba (2004) afirma que o conjunto das entradas foi estabelecido pelo critério de ocorrência do corpus que faz parte do banco de dados do Laboratório de Lexicografia da Faculdade de Ciências de Letras de Araraquara. No entanto, ressalta que, como em todo dicionário de língua, a principal informação é de natureza semântica, sendo que o sistema definitório do dicionário não é diferente do adotado pela maioria dos dicionários. O autor diz que “para patentear ou esclarecer melhor as acepções, utilizou-se um sistema de contextualização constituído por frases e expressões extraídas de textos reais do corpus, com adaptações e condensações para melhor cumprir seus objetivos”.

O autor não esclarece se os verbetes do dicionário foram estabelecidos com base nos sentidos encontrados no corpus. Ao que parece, o corpus só foi utilizado para definir o conjunto das entradas do dicionário e prover exemplos. Tradicionalmente, na área de lexicografia que faz utilização de corpus, os verbetes são definidos a partir das ocorrências encontradas no mesmo, ou seja, “os dados são o ponto de partida para a construção do dicionário”. (Rocha 2001)

1.7 CONCLUINDO O CAPÍTULO

O presente capítulo apresentou em quadro de alguns trabalhos de investigação que se situam dentro da Linguística de Corpus, umas das áreas voltadas ao estudo da linguagem que se encontra em plena expansão. A definição de Sardinha que caracteriza essa área como uma busca por uma padronização do léxico, ou seja, “uma regularidade nos tipos de associação a que se submetem as palavras de uma língua”, será de grande utilidade para este trabalho, uma vez que os sentidos do item lexical *correto* serão definidos a partir das associações que esta palavra assumir com outras palavras dentro dos contextos estudados. Portanto, será investigado se há algum tipo de

relação entre o uso e o sentido do item lexical em questão e as estruturas em que o mesmo aparece, procurando verificar se os seus respectivos sentidos estão associados com uma certa posição organizacional nos textos do corpus.

As definições de princípio colocacional e co-ocorrência de Sinclair (1991) são de fundamental importância, uma vez que serão as ferramentas mais utilizadas para fazer a análise do item lexical *correto* neste trabalho. A desambiguação via colocação de Pedersen e o modelo de bigrama apresentado também será fundamental na análise dos dados e na definição das etiquetas que em trabalhos futuros poderão ser utilizadas em um programa de desambiguação deste item lexical de polissemia complexa.

A parte teórica sobre anotação de corpus também será útil na orientação da pesquisa, já que a análise dos sentidos será interpretativa e os resultados poderão ser utilizados para fazer a anotação semântica do item lexical *correto* em um corpus. A versão do corpus utilizado não é etiquetada, porém é possível que a observação do adjetivo na versão do corpus NILC etiquetada traga contribuições interessantes para o estudo do mesmo ao longo da pesquisa.

A discussão de Pustejovsky sobre a representação de um item lexical polissêmico e os seus sentidos assumidos em diversos contextos na composição da semântica também é muito pertinente a este trabalho, já que um dos seus objetivos é a tentativa de descobrir as extensões de sentidos do item lexical *correto* dentro de um corpus.

Finalizando, os trabalhos apresentados no evento SENSEVAL são os que contêm mais similaridades com a pesquisa descrita neste projeto e foram a inspiração que me levou a fazer um trabalho semelhante com um item lexical polissêmico da língua portuguesa. A finalidade desse projeto é, portanto, fazer um levantamento de todos os sentidos possíveis do item lexical *correto* e, a partir da definição precisa de seu

quadro de polissemia, definir um conjunto de etiquetas que possa ser utilizado como base teórica para um sistema de desambiguação automática dos sentidos do mesmo.

O corpus da Língua Portuguesa escolhido para o presente trabalho foi o NILC (Núcleo Interinstitucional de Lingüística Computacional), um corpus de aproximadamente 35 milhões de palavras composto de textos escritos. Apesar de conter textos de natureza variada, a maior parte do corpus é composta por textos jornalísticos em sua íntegra. O NILC pode ser considerado um corpus cuja autoria é de língua nativa, ou seja, a maior parte dos autores são falantes nativos da língua portuguesa.

Acredito que, para os fins dessa pesquisa, o corpus NILC possa ser considerado representativo, estando consciente de que a definição precisa de representatividade de um corpus seja seguramente uma grande ilusão. Rocha (2001) coloca que “é fácil contestar a representatividade de um *corpus*, embora seja praticamente impossível assegurá-la, o que deixa em posição privilegiada os críticos que exploram a dificuldade de demonstrar cabalmente a representatividade de um *corpus* para invalidar iniciativas de investigação lingüística a partir de *corpus*”.

A despeito da afirmação apresentada acima, posso afirmar que a extensão do corpus NILC foi capaz de oferecer uma amostragem suficiente e satisfatória de exemplos e posso afirmar com segurança que as conclusões alcançadas a partir da análise do mesmo podem ser vistas, não obstante, como válidas, principalmente devido ao seu caráter empírico. Empírico no sentido de que os sentidos do item lexical analisado foram definidos a partir dos fatos da língua em uso no âmbito de situações da vida real.

Após ter justificado a escolha do corpus e ter descrito as teorias lingüísticas que permeiam esse trabalho, parece pertinente levantar uma pequena discussão sobre a metodologia utilizada no mesmo e as suas vantagens. O primeiro ponto seria que a

grande quantidade de informação viabilizada pelo uso do corpus nos permite fazer afirmações sobre significados, padrões, e uso de palavras com muito mais confiança e exatidão de detalhes. A observação de determinada palavra dentro de um contexto nos fornece informações sobre padrões e fenômenos da língua que passam despercebidos, já que as pessoas fazem uso da língua de uma maneira inconsciente. Elas não conseguem examiná-la em detalhe, mas simplesmente a usam com o intuito de comunicar-se.

As línguas humanas envolvem mecanismos complexos, e as palavras possuem maneiras diversas de se manifestar dentro de um contexto. As gramáticas e dicionários “tradicionais” não conseguem explicar integralmente tais fenômenos e abordar a grande quantidade de significados e usos que as palavras podem ter dentro de um contexto. Nesse sentido, o uso de computadores facilita a identificação de fenômenos sobre a língua que não podem ser descobertos apenas através da reflexão, leitura e consulta de dicionários e gramáticas tradicionais.

Portanto, os estudos realizados através da manipulação de corpus podem ser considerados uma ferramenta poderosa que nos permite desvendar esses mecanismos complexos existentes nas línguas humanas que as gramáticas e dicionários “tradicionais” (elaborados sem base em corpus) não conseguem explicar integralmente.

A metodologia consiste em retirar uma amostra das ocorrências, que são um pequeno fragmento do texto onde se encontra a palavra (neste caso o item lexical *correto*) dentro do corpus NILC. O programa utilizado para se obter estas amostragens do item lexical dentro de um contexto chama-se WordSmith. Esse programa, escrito por Mike Scott e publicado pela Oxford University Press, viabiliza a geração de concordâncias e coloca à disposição do pesquisador uma série de recursos que podem ser considerados muito úteis na análise de vários aspectos da linguagem.

CAPÍTULO 2

METODOLOGIA

In the beginning was the word.
(Philip King).

2.1 INTRODUÇÃO

De acordo com o que já foi explicitado na resenha da literatura, o corpus escolhido para o presente trabalho foi o NILC, Núcleo Interinstitucional de Linguística Computacional (<http://www.nilc.icmc.usp.br/>). O Corpus NILC possui aproximadamente 35 milhões de palavras, e apesar de ser composto de textos escritos de natureza variada, a sua maior parte é constituída por textos jornalísticos em sua íntegra.

Para que pudesse ser feito um bom aproveitamento dos dados disponibilizados pelo corpus, foi fundamental adotar um programa flexível e fácil de usar para manipular os dados. O programa que foi utilizado nessa análise é o WordSmith Tools, escrito por Mike Scott e publicado pela Oxford University Press. Esse programa disponibiliza para o analista uma série de recursos que são muito úteis na análise de vários aspectos da linguagem, como a composição lexical, a temática de textos selecionados e a organização retórica e composicional de gêneros discursivos (Berber Sardinha, 2004).

O programa computacional WordSmith funciona da seguinte maneira: o palavra a ser analisada (keyword) é selecionada; após esta operação o programa “varre” todos os textos selecionados do corpus a procura de todas as frases em que a palavra em questão aparece, separando-as em ocorrências que mostram parte do contexto em que a palavra está inserida, ou seja, gerando uma concordância.

“Uma concordância, ou simplesmente uma palavra chave dentro de um contexto, é uma lista de linhas desconectadas de um texto, que foram coletadas pelo programa de concordância a partir de um corpus de computador, ou seja, uma coleção de textos em formato que é legível por máquina. No centro de cada linha está o item que está sendo analisado (keyword). O resto da linha contém o co-texto imediato à direita e à esquerda da palavra. Tal lista permite ao analista olhar para os padrões no co-texto (surrounding co-text) adjacente à palavra, que fornece evidências (clues) sobre o uso do item analisado. Essa procura por evidências é a base da metodologia no estudo dos casos. Se a keyword for, por exemplo, um substantivo, e o usuário deseja saber quais são os adjetivos que são colocados com ele, é só pedir ao sistema para organizar (sort) as linhas em ordem alfabética da palavra à direita da keyword.” (Partington 1998).³⁷

Além de localizar estas ocorrências, o programa fornece uma série de ferramentas que nos permite observar quais palavras aparecem mais freqüentemente ao lado do adjetivo e outros recursos que viabilizam uma visão detalhada das ocorrências. Uma das ferramentas de análise do programa é o re-sort, que permite que as palavras que ocorrem à direita e à esquerda da palavra-chave *correto* sejam destacadas e organizadas alfabeticamente, o que possibilita a observação dos padrões de colocações.

³⁷ Tradução da autora. A concordance, or rather KWIC (Key Word in Context) concordance, is a list of unconnected lines of text, which have been summoned by the concordance program from a computer corpus, that is a collection of texts held in a form which is accessible to the computer. At the centre of each line is the item being studied (keyword or node). The rest of each line contains the immediate co-text to the left and right of the keyword. Such a list enables the analyst to look for eventual patterns in the surrounding co-text, which proffer clues to the use of the key word item. The search for such clues, is the basis of the methodology in the case studies (...). (...) if the keyword is, for example, a noun, and the user wishes to know what adjectives it collocates with, then they will ask the system to sort the lines in alphabetical order of the word to the left of the keyword (to “left sort”).

2.2 A OBSERVAÇÃO DOS PADRÕES DE COLOCAÇÃO

A observação dos padrões de colocações foi feita da seguinte forma: primeiramente os sentidos do item lexical *correto* foram definidos a partir dos exemplos que esta palavra assumiu dentro dos contextos estudados. Em seguida, foi investigado se havia algum tipo de relação entre o uso e o sentido do item lexical em questão e as estruturas em que o mesmo aparece, procurando verificar se os seus respectivos sentidos estavam associados com alguma posição organizacional dentro dos textos.

Era esperado que a partir dos padrões de co-ocorrência desse item lexical com outras palavras fosse possível encontrar soluções para a desambiguação do mesmo. Foi observado, por exemplo, que sempre que o item lexical *correto* aparecia precedido da palavra *posicionamento*, assumia um determinado sentido. Portanto, a partir da observação e análise de padrões de co-ocorrência dessa palavra com outras, pôde-se estabelecer os parâmetros iniciais para a fundamentação das etiquetas que possivelmente pudessem desambiguar esse item lexical polissêmico.

O ponto de partida para se fazer uma análise inicial do material coletado foi um conjunto de sete significados retirados da Base de Dados TeP (thesaurus do português, elaborada pelo NILC) para classificar as ocorrências do adjetivo *correto* encontradas no corpus. Esses significados encontram-se listados abaixo juntamente com seus antônimos:

Sentido 1:

³⁸S={certo; correto; exato;}

A={errado; incorreto; inexato;}

Sentido 2:

S={acertado; adequado; apropriado; certo; correto;}

³⁸ S = significado / A = antônimo

Sentido 3:

S={airoso; capaz; correto; decente; decoroso; digno; direito; honesto; honrado;
íntegro; probo; pundonoroso; reto; sério;}

A={amazelado; desairoso; deselegante; desonesto; desonrado; ímprobo; indecente;
indecoroso;}

Sentido 4:

S={castiço; correto; puro; vernacular; vernáculo;}

A={errado; espúrio; impuro; invernáculo;}

Sentido 5:

S={acertado; consertado; correto; corrigido; emendado; refeito; retificado; revisto;}

Sentido 6:

S={correto; incensurável; irrepreensível;}

A={censurável; condenável; criticável; culpável; estranhável; repreensível;
reprovável;}

Sentido 7:

S={certo; correto; válido; }

Assim que o grupo de significados iniciais foi definido, foi gerada uma concordância que apresentou 956 casos do item lexical *correto* encontrados no corpus NILC. A partir dos sentidos sugeridos pelo TeP, iniciei o processo de etiquetagem, ou seja, analisei cada ocorrência individualmente e coloquei uma etiqueta ao lado de cada uma delas identificando o seu sentido naquele contexto. À medida que o trabalho prosseguia, começaram a surgir ocorrências que, no meu ponto de vista, não se enquadravam em nenhum dos sete sentidos disponíveis.

No exemplo (1) abaixo, percebi que o adjetivo não assumia nenhum dos sete sentidos disponíveis, mas sim o caráter de *'justo'*. Portanto, adicionei esse sentido aos demais e passei a considerá-lo ao fazer a análise das ocorrências subseqüentes.

- (1) A Constituição, em seu artigo 5, inciso XXXIII, estabelece um princípio geral bastante correto: “Todos têm direito a receber dos órgãos públicos informações de seu interesse particular, ou de interesse coletivo ou geral, que serão prestadas no prazo da lei, sob pena de responsabilidade (...)”.

A expressão *politicamente correto* teve uma freqüência enorme, foram 204 ocorrências que foram incluídas nesse mesmo grupo de sentido, ou seja, *justo*. Não que a expressão em si tenha o sentido de *justo*, mas o significado isolado do item lexical *correto* nesse padrão de co-ocorrência aproxima-se mais deste sentido, como pode ser observado no exemplo abaixo:

- (2) Na ótica americana do “politicamente correto”, as duas “mulheres-não brancas” seriam um prato cheio. Mas elas também fazem parte de uma outra minoria: um grupo restrito da literatura americana contemporânea que não produz “thrillers”.

Além dos exemplos acima, outras ocorrências exigiram a criação de novos significados ao longo do processo de análise. O exemplo (3) abaixo é um desses casos, que levou à construção de mais um sentido. Nesse contexto, o item lexical assume o sentido de *'aceitável'*:

- (3) Mas foram comprometidos por um risoto duro de cavaquinhas; pelo acompanhamento do bom magret de pato (repolho salgadíssimo com

presunto), e por um confit de pato com lentilhas que, da mesma forma que as pappardelles com ragu de coelho era apenas correto. Mesmo as sobremesas (...) ficaram longe do que prometia o prestígio do Palladin.

Em outras ocorrências encontradas no corpus, o item lexical assumiu os sentidos de ‘*com a razão*’ e ‘*verdade*’, como nos respectivos exemplos (4) e (5) que ilustram os casos que culminaram na criação desses novos grupos de sentidos:

- (4) Os recorrentes aumentos acima da inflação de mercadorias produzidas por oligopólios demonstram que o governo está correto em reduzir as tarifas de importação.
- (5) Há muitas opiniões sobre a mídia, principalmente em sua sessão eletrônica, mas é importante ter opiniões próprias para saber o que é correto, assim, como não se deve ser controlado por opiniões em que haja um interesse por trás delas.

Após anotar sistematicamente todas as ocorrências do item lexical *correto* seguindo essa metodologia, concluí que havia a necessidade de se fazer o caminho inverso para que o trabalho pudesse realmente ser considerado científico. Afinal, se a pesquisa proposta seria baseada em corpus, tradicionalmente os sentidos deveriam ser definidos a partir das ocorrências encontradas no mesmo. Portanto resolvi abandonar o conjunto de sete significados retirados da Base de Dados TeP e adotar outra metodologia que fosse mais coerente com a proposta desse trabalho.

A estratégia utilizada consistiu em olhar para os dados e, a partir deles, inferir os diferentes sentidos e usos do item lexical polissêmico em questão. A partir da observação de cada ocorrência, os sentidos eram determinados utilizando o conceito de

sinonímia, ou seja, procurando um outro item lexical que tivesse o mesmo sentido do item lexical correto em cada contexto das ocorrências encontradas no corpus.

Depois de fazer uma análise sistemática do material da amostra coletada no corpus selecionado, contendo 956 casos do item lexical *correto*, as ocorrências acabaram sendo classificadas em doze diferentes significados, determinados a partir do conceito de sinonímia. Analisei ocorrência por ocorrência tentando definir qual o sentido assumido pelo item lexical nos diferentes contextos.

No exemplo (6), por exemplo, concluí que o item lexical *correto* assumia o sentido de *honesto*. A fim de confirmar essa informação, substituí o item lexical pelo sentido em questão para confirmar se sintaticamente e cognitivamente essa informação era verdadeira e se a substituição não alteraria o sentido da sentença.

(6) “O ágio é uma excrescência. Esse dinheiro não é **correto**. O **dinheiro** **correto** vem do trabalho”, afirmou Amato, que participou ontem no Rio do almoço de lançamento do livro “ABC da Propriedade Industrial - Patentes e Marcas”, publicado pela CNI.

(6.1) “O ágio é uma excrescência. Esse dinheiro não é **correto**. O **dinheiro** **honesto** vem do trabalho”, afirmou Amato, que participou ontem no Rio do almoço de lançamento do livro “ABC da Propriedade Industrial - Patentes e Marcas”, publicado pela CNI.

O método utilizado consistiu, portanto, em identificar o possível sentido e, em seguida, fazer um teste sintático de substituição com o sinônimo escolhido para testar a validade do procedimento. Uma vez que o novo sentido era definido e testado, ele era incluído em uma lista de sentidos para que fosse utilizado para classificar outras ocorrências em que o sentido do item lexical fosse semelhante.

À medida que as demais ocorrências eram analisadas, pude perceber que havia outros casos em que o item lexical *correto* assumia o sentido de *honesto*, como no exemplo (7). Nesses casos, a ocorrência marcada com uma etiqueta que definisse que naquele caso o item lexical também assumia o sentido de *honesto*.

- (7) O marido morreu logo e a jovem viúva casou-se com um português comerciante, muito **correto**. Acontece que ela começou a sonhar que o primeiro marido aparecia e queria dormir com ela.

O conceito de sinonímia é de extrema relevância para este trabalho e deve ser considerado válido para os fins desta pesquisa. Mas é necessário lembrar que sinônimos representam nuances de uma palavra e não há garantias de que a substituição de uma palavra por outra garanta a integridade semântica em todos os casos. Lyons (1979) afirma que “a sinonímia é uma questão de grau e que todo conjunto de unidades lexicais pode ser organizado numa escala de semelhança e de diferença de sentido, de modo que, por exemplo, *a* e *b* possam ser considerados como de sentido idêntico (sinônimos no sentido estrito) *a* e *c* relativamente semelhantes quanto ao sentido (sinônimos no sentido lato), *a* e *d* menos semelhantes quanto ao sentido, e assim por diante”.

No caso do exemplo (7), estou considerando que a substituição do item lexical *correto* pelo adjetivo *honesto* garante a integridade semântica da sentença. Uma vez que esta integridade é assegurada, o conceito de sinonímia é adotado e pode ser utilizado para desambiguar o item lexical *correto* neste contexto sem grandes problemas.

O próximo passo da pesquisa foi a criação de um padrão inicial de etiquetas que resolvessem o problema da ambigüidade dos casos analisados. A partir da

identificação dos significados assumidos pelo item lexical dentro do corpus, a próxima meta do projeto foi propor uma solução baseada em módulos, ou seja, criar critérios separados para identificar os diversos sentidos da palavra em questão. Como a etiquetagem está diretamente associada aos elementos que precedem e antecedem a palavra, neste caso, foi necessário fazer uma análise detalhada da pontuação e a posição em que a palavra se encontra na sentença.

Nos casos em que o número de ocorrências foi muito pequeno e, portanto, insuficiente para se fazer afirmações sobre padrões de colocação, foi feita uma busca na WWW via WebCorp (<http://www.webcorp.org.uk/>) para aperfeiçoar o perfil das colocações. Esse site de busca foi escolhido porque ele gera automaticamente os resultados sob a forma de concordância, o que facilita muito a observação de possíveis padrões de ocorrência.

2.3 CONCLUINDO O CAPÍTULO

Todas as ocorrências do item lexical *correto* encontradas no corpus desde o início da pesquisa foram classificadas e deram origem a uma destas doze formas de estratégia de processamento. A classificação descrita abaixo, deste modo, procura sistematizar as informações coletadas inicialmente no corpus, e foram o ponto de partida para uma investigação sistemática que permitiu definir como estas relações textuais se materializaram na língua portuguesa de um modo geral, a partir da amostra analisada na pesquisa. Abaixo, encontram-se listados os resultados classificatórios do item lexical em questão:

Sentido 1: exato

Ao primeiro grupo pertencem os casos em que o item lexical *correto* assume o sentido de *exato*.

- (8) Esse resultado não é contraditório com o preâmbulo do argumento EPR, que aceitava o **caráter correto** da teoria, mas recusava sua completeza: é com efeito num nível ulterior de refutabilidade que o teorema de Bell se coloca e, se ele fosse verificado, os resultados adquiridos da mecânica quântica permaneceriam, mas na condição de uma aproximação, mostrando-se o domínio de validade da teoria doravante limitado.

- (9) Auditoria: é uma outra aplicação natural para hipertextos, porque é baseada na geração e análise de relatórios de informação originários de várias fontes. A tarefa de uma auditoria inclui reunir e produzir um grande número de documentos, e interconectá-los para comprovar o **conteúdo correto** das informações.

No exemplo (10), o item lexical *correto* faz referência a um cálculo, que por sua natureza numérica só pode ter o sentido *exato*. Portanto, neste caso, fica claro que a intenção do item lexical *correto* é sinalizar que calculadoras simples não podem fazer o cálculo exato.

(10) Calculadoras simples não fazem o **cálculo correto**. Deve-se usar uma HP. No Roteiro de Crédito deste caderno você encontra uma tabela prática para calcular o juro aproximado. O Procon-SP também fornece tabela semelhante.

Já nos exemplo (11), (12), (13), (14) e (15), o item lexical *correto* faz referência a um determinado tipo de localização de natureza precisa. No exemplo (11), fica claro que para que o time do Inter tivesse um bom aproveitamento seria necessário que os passes chegassem ao destino exato para obter sucesso na partida. Em (12), mesmo com a confirmação de que os bloqueios estivessem no local exato a CET não sobe para verificar se houve invasão.

(11) Quanto ao aproveitamento, o Inter é o que possui a pior marca: 30,12% dos passes do time não chegam ao **destino correto**.

(12) A CET informou que os funcionários só fazem checagem dos bloqueios (feitos com correntes e barris vazios). Mesmo que eles não estejam no **local correto**, a CET não sobe para verificar se houve invasão.

No exemplo (13), é evidente que caso o jogador não encontre o lugar exato para o maior número de cartas não será possível ganhar o jogo. Os exemplos (13) e (14) também foram incluídos nesse grupo de sentido, já que o mapa tem que ser exato para

poder auxiliar os pilotos e o momento também tem que ser exato para possibilitar a observação do fenômeno a ser observado.

(13) Cada jogador recebe as cartas com as capitais dos Estados, bandeiras ou pontos turísticos. Os jogadores colocam as cartas no tabuleiro. Ganha quem acertar o **lugar correto** para o maior número de cartas.

(14) O sistema também pode usar sensores acoplados à espaçonave para varrer o solo abaixo do veículo, comparar os dados encontrados com os dados armazenados no computador e então dar automaticamente aos pilotos o **mapa correto**, baseado no ponto em que os sensores indicam que o avião está localizado.

(15) “O impacto produzido por ele será um show espetacular para um observador que esteja no lugar certo e no **momento correto**”, diz Zdenek Sekanina. Esse é o problema. Como a explosão se dará na face oculta de Júpiter, pouco se poderá ver da Terra, mesmo com telescópios potentes, como o Hubble.

Sentido 2: adequado

A característica deste grupo é que o adjetivo *correto* assume o sentido de *adequado*, como no exemplo (16). Os exemplos (17) e (18) também se encaixam nesta categoria.

(16) **Armazenamento correto** também contribui com a conservação dos alimentos e figura entre as orientações anti-desperdício.

(17) Produzir um bem ou serviço com qualidade, com a entrega no prazo determinado, com o **atendimento correto** (inclusive na assistência pós-

venda) ao menor custo (logo, menor preço), são as novas regras de mercado.

(18) - O sr. acha um **comportamento correto** sair da sala para não votar?

As ocorrências da expressão *ecologicamente correto* foram incluídas nesse grupo, já que nesse caso, o adjetivo denota o sentido de adequado, como pode ser observado nos exemplos abaixo:

(19) Até a criação de animais silvestres com fins comerciais é capaz de ser um negócio **ecologicamente correto**. Os criadouros, autorizados por lei e fiscalizados pelo Ibama, são um fator de preservação de espécies, como o jacaré do Pantanal e da capivara, que antes estavam no mercado exclusivamente pelas mãos de caçadores clandestinos.

(20) **Ecologicamente correto**, o barco, na falta de vento, será impulsionado por energia solar. O motor a diesel só será acionado quando a bateria solar estiver descarregada, em noites de calmaria.

Sentido 3: isento de erro, acertado

Sob esta denominação pouco esclarecedora, por falta de melhor, foram agrupados casos em que o item lexical *correto* assume o sentido de *acertado*, *isento de falha*, *erro*, ou *defeito*, como mostra o exemplo (21). Também estão nesse grupo os casos em que o item lexical assume o sentido de ‘que corresponde a certo padrão ou norma’, como no exemplo (22) abaixo.

(21) Ruy Martins Altenfelder Silva disse que o diagnóstico da Folha, em que baseia a sua sugestão para assembléia revisora exclusiva, está “**absolutamente correto**”.

(22) Não fosse pelo **acento correto** sobre o “e” da última palavra, o leitor desavisado poderia imaginar que o redator decidira abolir esses sinais de duas outras, “tem” e “pé”.

Sentido 4: válido

Fazem parte deste grupo as ocorrências em que o item lexical *correto* assume o sentido de *válido*. No exemplo (23), o laudo não será refeito por ter sido considerado válido. Já em (24), o princípio em questão é válido se o advogado estiver tentando evitar o arbítrio contra um cliente:

(23) Segundo Cunha, o laudo é **correto** e não será refeito.

(24) Se o advogado estiver tentando evitar o arbítrio contra um cliente, o princípio é **correto**; mas a formulação é vaga demais e pode ser invocada por qualquer advogado a qualquer momento.

Sentido 5: justo

Fazem parte deste grupo os casos em que o item lexical *correto* assume o sentido de *justo*. No exemplo (25) abaixo, o item lexical está qualificando o processo em termos de justiça. As ocorrências da expressão *politicamente correto* também fazem parte deste grupo de sentidos.

(25) A CPI do Congresso não resultou em nada. O processo hoje posto em prática é **correto**.

(26) O pensamento **politicamente correto** está passando dos limites.

Sentido 6: aceitável

Fazem parte deste grupo as ocorrências em que o item lexical *correto* é utilizado sem dar muita ênfase ou credibilidade àquilo que está caracterizando. No exemplo (27), o adjetivo vem acompanhado da palavra *apenas*, que reforça a idéia de que o prato era considerado *aceitável*, mas deixava a desejar.

(27) O linguado com ervas é também delicado. Por outro lado, a cavaquinha grelhada na manteiga passou do ponto; e o pato assado com cèpes (cogumelos porcini) em conserva, servido com molho do assado, é **apenas correto** - a não ser pelo fraco risoto que o acompanha.

Sentido 7: com a razão.

A este grupo pertencem os casos em que o adjetivo *correto* assume o sentido da expressão '*com a razão*', como no exemplo (28):

(28) Note que para o professor Pastore os encargos são entendidos segundo seus impactos sobre o custo do empregador. Mas se a discussão refere-se aos efeitos sobre o mercado de trabalho está **correto** o professor Amadeo ao afirmar que o importante é a diferença entre o custo percebido pelo empregador e o salário recebido pelo trabalhador.

Sentido 8: verdade

Pertencem a este grupo os casos em que o item lexical *correto* assume o sentido de *verdade*. No exemplo (29), o autor usa o item lexical *correto* para informar

ao leitor sobre a confiabilidade da informação, neste caso o desaparecimento de uma pequena parte das empresas.

(29) É **correto**, entretanto, que uma pequena parte das empresas venha a desaparecer, fruto de um ajuste competitivo. É aceitável ainda que o governo se utilize das tarifas de importação para impor a empresários irresponsáveis um ambiente competitivo e, dessa forma, controle o crescimento dos preços.

Sentido 9:

Esse grupo não deve ser considerado como um grupo de sentido propriamente dito. Ele foi criado para caracterizar as ocorrências em que o item lexical *correto* tem a função de marcador discursivo no corpus. São os casos em que ele aparece, por exemplo, no início de uma sentença confirmando uma informação prévia. Esse fenômeno é bastante comum em entrevistas, quando um dos falantes fornece uma informação esperando pela confirmação do outro, como no exemplo (30) abaixo:

(30) Ayala - **Correto**, e eu começo minha resposta lembrando a primeira explosão atômica, o teste em Alamogordo. O diretor do Projeto Manhattan, Robert Oppenheimer, disse então que a ciência tinha conhecido o pecado.

Sentido 10: honesto

Pertencem a este grupo as ocorrências em que o item lexical *correto* assume o sentido de *honesto*, caracterizando principalmente pessoas, como nos exemplos (31) e (32) citados abaixo:

(31) Os italianos reclamaram do que acharam ter sido pênalti sobre Dino Baggio e os mexicanos não gostaram de um impedimento marcado ao final do primeiro tempo, o que impediu a validação de um gol feito na seqüência. Mas Lamolina foi um **juiz correto**.

(32) Dessa forma, poder-se-á aliviar a carga tributária individual, beneficiando o **contribuinte correto**.

Sentido 11: ideal

A característica deste grupo é que o item lexical *correto* assume o sentido de *ideal*. Os exemplos (33) e (34) ilustram as ocorrências em que o item lexical tem como sinônimo esta palavra:

(33) Isto provoca um êxodo prematuro dos pilotos e assim não acontece um **aprendizado correto** de todos os fundamentos.

(34) Mas é, por um outro ponto de vista, o **desfecho correto** e lógico para todos esses jogos inúteis que marcaram e foram a tônica desse torneio que deveria ser demonstrativo do panorama atual do futebol tetracampeão.

Sentido 12: bem feito

Fazem parte deste grupo os casos em que o item lexical *correto* assume o significado da expressão '*bem feito*', como nos exemplos (35) e (36):

(35) A perua tem **acabamento correto** e bons bancos, com apoio central de braços. O revestimento das portas é sofrível.

(36) De outro lado, fazem um disco, como mencionado acima, **correto**. Sem mais apodrecimentos. Regravado, o hit "It's Rainin' Men" é a última faixa do disco e deve ser a primeira a ser ouvida, só pra inspirar.

CAPÍTULO 3

RESULTADOS

Computer tools don't do the thinking.
(WordSmith)

3.1 INTRODUÇÃO

Abaixo se encontra uma tabela com o número e a percentagem de ocorrência de cada um dos sentidos descritos na seção anterior. Em seguida, há uma breve discussão sobre os padrões de co-ocorrência encontrados em cada um dos grupos de sentidos que representam um esboço das etiquetas que foram desenvolvidas neste trabalho. Portanto, a classificação descrita em seguida é uma tentativa de definir estratégias de processamento adequadas para a identificação dos diferentes usos do item lexical *correto* encontrados no corpus analisado.

Tabela 1 - Frequência dos sentidos de correto

Sentido 1: exato	71 ocorrências	7,42 % dos casos
Sentido 2: adequado	231 ocorrências	24,16% dos casos
Sentido 3: isento de erro	168 ocorrências	17,57% dos casos
Sentido 4: válido	20 ocorrências	2,09% dos casos
Sentido 5: justo	284 ocorrências	29,70% dos casos
Sentido 6: aceitável	21 ocorrências	2,19% dos casos

Sentido 7: com a razão	8 ocorrências	0,83% dos casos
Sentido 8: verdade	15 ocorrências	1,56% dos casos
Sentido 9: m. discursivo	7 ocorrências	0,73% dos casos
Sentido 10: honesto	46 ocorrências	4,81% dos casos
Sentido 11: ideal	69 ocorrências	7,21% dos casos
Sentido 12: bem feito	16 ocorrências	1,67% dos casos

Para determinar as etiquetas de cada conjunto de sentidos foi utilizada a noção teórica do princípio colocacional devidamente adaptada para os fins desta pesquisa. Ou seja, foi feita uma observação das palavras com as quais o item lexical *correto* se combinava formando bigramas que pudessem ser utilizados como base para desambiguá-lo nos diversos contextos em que este foi encontrado no corpus.

3.2 A DEFINIÇÃO DO CONJUNTO DE ETIQUETAS

O conjunto das etiquetas encontra-se listado abaixo e tem relação direta com os diversos sentidos do item lexical em questão:

3.2.1 Sentido 1: exato

Ao assumir esse sentido, o item lexical *correto* apresenta várias colocações regulares com outras palavras formando bigramas que podem ser suficientes para diferenciar esse sentido dos demais. Ele assumiu colocações regulares ao ser precedido imediatamente pelas seguintes palavras: os substantivos *caráter, cálculo, conteúdo, destino, horário, local, lugar mapa, momento, nível, número, perfil, ponto, posicionamento, resultado, tamanho, termo, traçado, trajeto e valor.*

(37) A assinatura de carteira de trabalho é difícil, e raras vezes pelo **valor** **correto**, a rotina de trabalho é estafante e indigna, e o roteiro de folgas, estúpido.

Na tentativa de buscar outros bigramas que pudessem desambiguar o item lexical *correto*, foi observado que em algumas ocorrências ele veio precedido pelo verbo *estar* (ex. está correto). Infelizmente foi detectado que esse padrão isolado não serviria para desambiguar esses casos, já que em praticamente todos os outros grupos de sentidos o adjetivo também apareceu precedido por esse verbo. Mas foi observado que nas ocorrências desse sentido em especial, o verbo *estar* sempre veio precedido em alguma posição colocacional das palavras *cálculo*, *número* e *valor*. Portanto, o padrão do verbo precedido por essas palavras específicas também pode ser usado como ferramenta para desambiguar esse sentido, como pode ser observado no exemplo abaixo:

(38) Esse **cálculo** está **correto**?

Outro padrão que se repetiu nesse grupo de sentido foi aquele em que o item lexical *correto* era precedido pelo artigo “o”. Esse padrão isolado não é suficiente para diferenciar esse sentido do grupo de sentido denominado *ideal* (onde esse padrão também se repete), mas foi observado que em todos os casos que o item lexical era precedido pelo artigo “o” e simultaneamente em alguma posição à esquerda ou à direita havia um valor numérico ele assumia o sentido de *exato*, e não de *ideal*.

Conclui-se que nesses casos a posição do adjetivo pode ser utilizada para desambiguá-lo, já que os padrões de co-ocorrência com essas palavras identificados são regulares e não se repetem nos outros grupos de sentidos.

Outra possibilidade, devido ao fato do item lexical em questão ser usado como substantivo na maioria destes casos, o sentido poderia ser identificado a partir da saída do processamento de um etiquetador de classes de palavras, anterior ao processo de desambiguação. Este último poderia se encontrar apenas na distinção entre os dois sentidos possíveis para a estrutura ART *correto*, um ótimo exemplo de léxico-gramática. Foram catalogados 71 casos como pertencentes a esta categoria, ou seja, 7,42% dos casos.

Sentido 2: adequado

Ao assumir o sentido de adequado, o adjetivo *correto* também apresenta várias colocações regulares com palavras que parecem ser suficientes para desambiguar o seu sentido. Neste caso, o adjetivo apresentou padrões de co-ocorrência regulares ao ser precedido imediatamente pelas seguintes palavras: os substantivos *armazenamento, atendimento, comportamento, conceito, discurso, dispositivo, dúvida, equipamento, exercício, filme, funcionamento, instrumento, jeito, manejo, manuseio, média, método, modelo, modo, paliativo, procedimento, teste, título, tratamento, tom e uso.*

(39) Outro fator que contribui para que o banco dure é o **uso correto**.

O item lexical *correto* também assumiu esse sentido ao ser precedido pelos seguintes advérbios: *conceitualmente, culturalmente, ergonomicamente, estatisticamente, formalmente, historicamente, inteiramente, plenamente, sexualmente, sheerazadianamente, tecnicamente, totalmente e virtualmente.* Vale destacar a expressão *ecologicamente correto* (40), que foi incluída nesse grupo, já que em todos os contextos dessas ocorrências o item lexical *correto* assumiu o sentido de *adequado*. Esse padrão de co-ocorrência do item lexical com o advérbio *ecologicamente* pode

perfeitamente ser utilizado para desambiguar esse sentido. Foram encontradas 19 ocorrências dessa expressão.

- (40) **Ecologicamente correto**, o barco, na falta de vento, será impulsionado por energia solar.

Outro padrão de colocação importante detectado nesse sentido foi o a co-ocorrência do adjetivo com o advérbio *mais* (ex. *mais correto*), que representa 13,85 % dos casos desse subgrupo, ou seja, 32 ocorrências.

- (41) Nesse sentido, talvez fosse **mais correto** dizer que a prática do zapping não faz senão potencializar, ou mesmo tirar conseqüências daquilo que é o funcionamento normal da televisão, (...).

Pode-se concluir, portanto, que a colocação do item lexical *correto* com os substantivos e advérbios mencionados acima podem resolver a desambiguação da maioria dos casos. Dentre os casos analisados na amostra, foram encontrados 231 casos do tipo descrito nesta subseção, ou seja, 24,16 % dos casos.

Sentido 3: isento de erros

Fazem parte desse grupo os casos em que o item lexical *correto* assume o sentido de isento de falha ou de erros, e também os casos em que ele equivale a algo que ‘corresponde a certo padrão, que não apresenta falha, ou seja, algo perfeito’. Optei por denominá-lo assim devido à dificuldade de encontrar um termo mais genérico, ou seja, uma única palavra sinônima para definir esse sentido (como aconteceu com parte dos demais).

Esse sentido, assim como os demais, também apresenta padrões de ocorrência que podem ser considerados suficientes para desambiguar o item lexical *correto*. Em todos os casos que o item lexical em questão apareceu precedido pelo advérbio *absolutamente* ele assumiu o sentido de ‘isento de erros’, como pode ser observado no exemplo (42). O advérbio nesse caso dá ênfase à informação, mostrando que não há dúvidas quanto à precisão do fundamento:

(42) E a jurisprudência se tem revoltado contra ela (RJTJESP 93/237), inclusive sob o fundamento, **absolutamente correto**, de que o cumprimento da precatória sai mais barato que a publicação de editais.

Outro padrão que chamou a atenção foi a colocação desse item lexical precedido pela palavra *caminho*, um padrão que se repetiu em 15 ocorrências, como pode ser observado nos exemplos (43). Além de co-ocorrer com essa palavra, nos casos em que o item lexical aparecia com os substantivos *andamento*, *conhecimento*, *diagnóstico*, *encaminhamento*, *endereço*, *ensino e entendimento*, também assumiu esse sentido.

(43) O governo do Estado considera que as atribuições federais “ditadas pela Constituição Federal” são o **caminho correto** “para superar as decepções” resultantes dos esforços das polícias estaduais, que travam combate com os traficantes só na venda final dos tóxicos.

Outra característica marcante desse uso é a colocação de *correto* com as palavras *acento*, *português*, *inglês*, *gramaticalmente* e *sintaticamente*, sinalizando o sentido ‘isento de erros’ no que se refere ao uso da língua. Em função disso, a expressão

falar correto também foi inserida nesse grupo, onde grande parte das ocorrências podem ser desambiguadas a partir dessas colocações.

Em alguns casos, o item lexical *correto* faz referência a um conjunto de informações que deve ser transmitido seguindo certo padrão, sem apresentar falha ou erros, como nos exemplos abaixo. Nessas ocorrências o item lexical apareceu precedido pelos substantivos *informe, nome, planejamento, preenchimento, processo, projeto, raciocínio, relato, rumo, sentido e título*.

Os casos em que as palavras *correto* e *incorreto* apareciam juntas formando um contraste também foram incluídos neste grupo e foram considerados um par de sinônimos para as expressões “*isento de erro*” e “*com erro*” ou *certo* e *errado*. É o caso do exemplo (44):

- (44) (...) 2) também a fronteira entre o **correto** e **incorreto** fica na dependência das forças em luta e da conjuntura.

Também foi detectado que, sempre que o item lexical *correto* era precedido pela palavra *tudo*, assumia o sentido de *isento de falha* ou *erro*: A amostra contém 168 casos que pertencem a este grupo, o que significa 17,57% do total.

Sentido 4: válido

Nesse grupo o principal padrão de ocorrência observado foi a colocação do item lexical *correto* precedido pelo verbo *estar* e seguido imediatamente por uma conjunção integrante (dizer que), como no exemplo (45) abaixo:

- (45) Acho que **é correto dizer que** a dança, embora tomando diferentes formas, caminha em ciclos de dança pura e dança expressiva.

Em apenas uma ocorrência o item lexical *correto* foi precedido pela palavra *como* (ex.46) e assumiu o sentido de válido. Na ausência de mais ocorrências para confirmar se esse bigrama poderia ser considerado para desambiguar o item lexical, foi feita uma busca no WebCorp. Foram geradas 172 ocorrências onde o item lexical *correto* era precedido pelo pronome demonstrativo *como*. Como esse padrão não se repetiu em nenhum dos outros grupos de sentido, esse padrão de co-ocorrência pode ser utilizado para diferenciar esse caso dos demais.

- (46) Dentro do que aceita **como correto** a terapia de vidas passadas, uma pessoa que tem medo de altura ou de fogo pode ter adquirido a fobia em função de acidentes anteriores, como, por exemplo, ter sido empurrada de um precipício ou ter morrido queimada.

Foram encontradas 20 ocorrências que pertencem a este grupo, o que significa 2,09% do total.

Sentido 5: justo

Esse grupo acabou se tornando um dos mais representativos em função da expressão *politicamente correto*, que apareceu 204 vezes na concordância. Retomando o que foi colocado na discussão dos sentidos na seção da metodologia, é necessário ressaltar que a expressão em si não tem o sentido de *justo*, mas o significado isolado do item lexical *correto* nesse padrão de co-ocorrência aproxima-se mais deste sentido. Portanto, a desambiguação deste caso é fácil de ser determinada, uma vez que, em todas as colocações do adjetivo com o advérbio *politicamente*, ele assumiu o sentido de *justo*. Portanto, o padrão de colocação desse adjetivo com o advérbio em questão pode ser considerada suficiente para desambiguar 76,11% dos casos desse sentido.

(47) Os dois autores, num tom alegre e **politicamente correto**, dizem que ir bem no QI não é o mais importante.

As demais ocorrências desse grupo de sentidos também podem ser facilmente desambiguadas através da posição do item lexical, pois apresenta muitas outras colocações regulares com outras palavras. Por exemplo, foi observado que em todas as ocorrências em que *correto* aparece precedido dos verbos *achar* ou *considerar*, ele assume o sentido de ‘justo’, como pode ser observado nos exemplos (48), (49), abaixo:

(48) O senhor **acha correto** que a Universidade ofereça uma série de serviços gratuitamente, por ser uma universidade pública? Isso não poderia ser explorado comercialmente, para gerar recursos para a própria Universidade?

(49) Ele disse não **considerar correto** remover ou sacrificar o golfinho que está atacando turistas em Caraguatatuba. “Ele está no seu habitat natural. Os banhistas não. Além do mais, os golfinhos só atacam quando incomodados”, afirmou.

Ao ser precedido pelos advérbios *administrativamente*, *constitucionalmente*, *impoliticamente* e *moralmente* o item lexical em questão também assumiu o sentido de *justo*. Os padrões de co-ocorrência do item lexical *correto* com as palavras *feminismo*, *preço* e *princípio* também fazem parte deste grupo de sentidos.

Já foi mencionado que a co-ocorrência do item lexical *correto* com formas do verbo *ser* não pode ser utilizada para diferenciá-lo dos demais sentidos. Mas um fenômeno interessante foi observado nesse grupo de sentido. Em aproximadamente 51% dos casos em que o item lexical assumia o sentido de *justo*, as sentenças eram negativas, seguindo a estrutura sintática do exemplo abaixo:

(50) “É um cinismo. **Não é correto** vender uma idéia de que o saneamento está uma maravilha, além de gastar recursos públicos em publicidade na mesma semana que a Sabesd cancela uma concorrência que já durava 14 meses”, diz o empresário Antônio Carlos Germano Gomes, 36.

Pode-se dizer que esse sentido pode ser inteiramente desambiguado a partir da contigüidade simples do princípio colocacional, ou seja, através do posicionamento do item lexical com as palavras especificadas acima. Este grupo abrange 284 casos, somando 29,70% do total.

Sentido 6: aceitável

Nesse grupo de sentidos, o único padrão de ocorrência que se repetiu foi a colocação do item lexical *correto* precedido da palavra *apenas*, em quatro ocorrências, como no exemplo (51):

(51) O tenor Peter Dvorsky faz um Renato Des Grieux **apenas correto**. Não é um Plácido Domingo.

Além desse padrão detectado acima, não foi detectado nenhum outro padrão colocacional recorrente que pudesse desambiguar as 17 demais ocorrências deste grupo, o que leva à necessidade de procurar outras estratégias para determinar a

desambiguação das mesmas. Foram encontradas 21 ocorrências deste sentido, ou seja, 2,19% dos casos.

Sentido 7: com a razão

Em todas as ocorrências que o item lexical *correto* assumiu o sentido de ‘*com a razão*’, ele veio precedido do verbo *estar* conjugado no presente, na terceira pessoa do singular (*está* *correto*). Como o verbo *estar* não serve de parâmetro para a esperada desambiguação, uma vez que *está* presente em praticamente todos os sentidos, a solução parece estar em observar o padrão que se repetiu nas palavras localizadas à direita do item lexical *correto*. Nesse grupo de sentidos, o item lexical *correto* sempre apareceu seguido da preposição *em* e da palavra *ao*.

(52) Os recorrentes aumentos acima da inflação de mercadorias produzidas por oligopólios demonstram que o governo **está correto em** reduzir as tarifas de importação.

(53) É um técnico completo. **Está correto ao** pensar na figura do técnico como uma espécie de chefe de Estado de um clube.

Portanto a partir da colocação desses itens lexicais parece possível desambiguar esse sentido dos demais. Este grupo abrange apenas 8 casos, somando apenas 0,83 % do total.

Sentido 8: verdade

Das quinze ocorrências classificadas com o sentido de ‘*verdade*’, seis apresentam o item lexical *correto* precedido da forma verbal *seria* formando a

expressão *seria correto*. Como esse padrão não aconteceu nos demais sentidos, essa pode ser uma das estratégias para diferenciar esse sentido dos demais.

(54) Mas não **seria correto** dizer que Freud “gostava” de Shakespeare, como gostava de Goethe ou Milton.

(55) Folha - **Seria correto** dizer que a sra. saiu da Prefeitura mais pragmática e menos ideológica?

Nas nove ocorrências restantes este item lexical não apresentou padrões de co-ocorrência regulares que pudessem servir de subsídio para criação de etiquetas. Esse grupo representa 0,83% dos casos.

Sentido 9

Em princípio, os casos incluídos neste grupo haviam sido incluídos no primeiro grupo de sentidos, ou seja, *exato*. Isso aconteceu porque a substituição do item lexical *correto* pela palavra *exato* nessas ocorrências não alterava o sentido das sentenças.

Mas ficou claro, nesses casos, que o item lexical era usado com um propósito que merece ser destacado. Os casos de *correto* em posição inicial, ou posição final (confirmando uma informação previamente dada) seguidos de vírgula indicavam que ele está sendo utilizado como marcador conversacional.

Segundo Urbano (1995), os marcadores conversacionais são “elementos lingüísticos que estruturam o texto, considerado não só como uma construção verbal cognitiva, mas também como uma organização interacional interpessoal. Ou seja, são

recursos que sinalizam orientação ou alinhamento recíproco dos interlocutores em relação ao discurso”.

Urbano (1995) diz que esses marcadores “posicionam-se, de modo geral, antes ou depois das unidades conversacionais”. Em função disso, as ocorrências em que o item lexical *correto* situava-se no início da frase foram tratados como marcador conversacional, semelhante a outras formas afirmativas como *sim* ou *isso*, como pode ser observado nos exemplos abaixo:

(56) A genética não é um monte de pequenos projetos Manhattan prontos a explodir nas mãos da sociedade? Ayala - **Correto**, e eu começo minha resposta lembrando a primeira explosão atômica, o teste em Alamogordo.

(57) Ao mesmo tempo, em suas coreografias você mistura técnicas clássicas, dança moderna e movimentos cotidianos ... Cunningham - **Correto**, eu não rejeito o balé clássico.

Concluindo, nesse grupo a pontuação pode ser utilizada para desambiguar estes casos inseridos nesse grupo. Coincidentemente a maioria das ocorrências fazia parte de entrevistas, o que confirma o uso do item lexical como marcador conversacional. Foram encontradas 7 ocorrências do item lexical *correto* como marcador conversacional, ou seja, 0,73% dos casos.

Sentido 10: honesto

Nesse grupo o primeiro padrão de ocorrência observado foi a colocação típica de bigramas formados pelo item lexical *correto* com os substantivos *contribuinte*,

dinheiro, funcionário, homem, jornalismo, juiz, julgamento, pagamento, político, profissional e sujeito.

(58) Em realidade, alcançou, como sempre, o **contribuinte correto**, que já suportava uma pressão tributária elevada.

Nos casos em que o item lexical *correto* era precedido pelo advérbio de intensidade *muito* ele também assumia o sentido de *honesto*, como pode ser observado no exemplo (186) abaixo:

(59) Sabes que eu sempre fui **muito correto** nos meus amores.

Também foi observado que nos casos em que um nome próprio situava-se em qualquer posição à esquerda do item lexical ele também assumia esse sentido:

(60) Quem tentou opor-se a Covas se deu mal. **Covas é correto**, não há nada contra ele, está do lado do bem.

Foram encontradas 46 ocorrências que assumiam esse sentido na amostra, representando 4,81% do total.

Sentido 11: ideal

Pertencem a este grupo os casos em que o item lexical *correto* tem o sentido de *ideal*. Ao assumir esse sentido, o item lexical tem a tendência de ocorrer precedido do artigo definido *o* (padrão considerado significativo, ocorrendo em 71,01% dos casos) e de *do*. Os exemplos (61) e (62) ilustram esse tipo de uso do item lexical em questão:

(61) Provavelmente é isto que está acontecendo, e **o correto** seria reduzir a dose. Quando a tristeza normal se transforma em tristeza patológica, muitas vezes existe uma depressão.

(62) (...): pelotões de pesquisadores pensavam que se aproximavam mais e mais **do correto** ao elaborarem montanhas de papéis; ou, mais ainda, acreditavam que, apresentando números e tabelas, assemelhavam-se à ciência.

O item lexical em questão também assumiu o sentido de *ideal* ao ser precedido pelos substantivos *aprendizado*, *desfecho*, *papel* e ao ser seguido imediatamente pela palavra dimensionamento. Foram detectadas 69 ocorrências em que o item lexical assumia o sentido de *ideal*, ou seja, 7,21% dos casos.

Sentido 12: bem feito

Fazem parte desse grupo os casos em que o item lexical *correto* assume o sentido da expressão *bem feito*. Assim como no caso do grupo de sentido definido como *isento de erro*, houve certa dificuldade de encontrar uma única palavra para definir esse sentido.

Ao assumir o sentido de *bem feito*, o item lexical *correto* apresentou várias colocações regulares com os substantivos *acabamento*, *buffet*, *cronograma*, *desenho*, *desfile*, *disco*, *seriado*, *show*, *trabalho* e do advérbio *melodicamente*.

(63) Um **buffet correto**, bebidas idem, noivos embalados a “Unforgettable”; com Nat King Cole.

(64) Robert Cray volta com um **disco correto**, onde acaba se redimindo, talvez sem saber, das tolices que pode ter cometido.

A amostra contém 16 casos que pertencem a este grupo, o que significa 1,67% do total.

3.3 CONCLUINDO O CAPÍTULO

Após essa análise dos dados, pude constatar que grande parte dos casos de ocorrência deste item lexical de polissemia complexa podem ser desambiguadas através dos padrões regulares de colocação desse item lexical com outras palavras, ou seja, através de um conjunto de bigramas que serão descritos na subseção seguinte.

CAPÍTULO 4

MODELO DE PROCESSAMENTO BASEADO EM N-GRAMAS

O objetivo desse capítulo é fazer uma pequena amostra de modelo de processamento baseado em n-gramas e sistematizar as informações apresentadas no capítulo de resultados. O modelo terá como base as colocações que só foram encontradas em associação com um único sentido no corpus NILC.

O modelo de processamento baseado em n-gramas foi desenvolvido por Pedersen (2001). Neste modelo, uma árvore de decisões atribui um sentido para uma palavra ambígua baseado nos bigramas de que faz parte, ou seja, as seqüências de palavras que ocorrem em um texto. “O contexto em que a palavra ocorre é representado por um número de traços binários que indicam se um determinado bigrama ocorreu em aproximadamente 50 palavras à direita ou à esquerda da palavra a ser desambiguada” (Pedersen, 2001).

Para se determinar os possíveis n-gramas que pudessem ser utilizados como base de um sistema que fosse capaz de desambiguar automaticamente o item lexical *correto* dentro de um contexto, foram escolhidas as colocações que só foram encontradas em um determinado grupo de sentidos.

Por exemplo, todas as vezes que o item lexical *correto* apareceu precedido da palavra *valor*, o sentido atribuído nesses casos foi *exato*. Como o n-grama *valor correto* não se repetiu nos demais grupos de sentidos, esse modelo foi utilizado para determinar que, sempre que o sistema encontrasse o item lexical em questão precedido

pela palavra *valor*, esse assumiria o sentido de *exato*, como pode ser observado no esquema abaixo:

N-grama: [valor correto], correto = exato

Como em alguns casos o número de ocorrências dos n-gramas encontrados no corpus NILC era muito pequeno, foram feitas buscas no site do WebCorp (<http://www.webcorp.org.uk/>) a fim de verificar se os mesmos padrões se repetiam na WWW e se, conseqüentemente, poderiam ser considerados apropriados e significativos.

O WebCorp é uma interface de busca que oferece ferramentas que permitem acesso à World Wide Web como um corpus. Basta digitar o item que se deseja pesquisar e, assim como o WordSmith, o mecanismo de busca gera uma concordância que contém as ocorrências encontradas na WWW. A interface oferece três opções para o número de concordâncias geradas: 5, 10, 50 ou ilimitado (unlimited). Na opção “ilimitada” a própria interface define o número de concordâncias, sendo que o programa não especifica quais são os critérios que definem esse procedimento.

Devido à inviabilidade de tempo e recursos, o procedimento adotado foi observar apenas as cinco primeiras ocorrências de cada um dos n-gramas encontrados na WWW que foram listados na concordância. A opção de pedir ao WebCorp um número ilimitado de concordâncias pareceu mais interessante (apesar de existir a possibilidade de se gerar só as cinco concordâncias que seriam analisadas) para verificar se o número de ocorrências dos n-gramas encontradas na WWW poderia ser considerado significativo em relação ao número de ocorrências encontradas no corpus NILC.

A fim de exemplificar, foram encontradas 9 ocorrências do n-grama *valor correto* no corpus NILC. Na busca feita através do WebCorp foi gerada uma concordância contendo 283 ocorrências desse mesmo n-grama. Diante da impossibilidade já mencionada de observar todas as ocorrências, a metodologia adotada foi analisar somente as primeiras cinco ocorrências do n-grama *valor correto* listadas na concordância a fim de verificar se, nesses casos, o item lexical *correto* também assumia o sentido de *exato*.

O mesmo procedimento foi adotado sistematicamente para verificar todos os n-gramas dos demais grupos de sentido. As tabelas abaixo mostram os n-gramas determinados a partir da análise das ocorrências do corpus NILC. Na coluna do meio pode ser observado o número de ocorrências encontradas no corpus NILC. Ao lado dessa coluna foram listados os números de ocorrências dos mesmos n-gramas encontrados na WWW. Para cada n-grama será feita uma breve discussão a partir da análise dos exemplos retirados da WWW via WebCorp.

Sentido 1: exato

N-gramas	Número de ocorrências encontradas no corpus NILC	Número de ocorrências encontradas na Web
Destino correto	2 ocorrências	169 ocorrências
Local correto	3 ocorrências	232 ocorrências
Lugar correto	3 ocorrências	246 ocorrências
Mapa correto	4 ocorrências	59 ocorrências
Momento correto	6 ocorrências	173 ocorrências
Nível correto	2 ocorrências	157 ocorrências
Número correto	3 ocorrências	167 ocorrências
Perfil correto	2 ocorrências	135 ocorrências
Posicionamento correto	7 ocorrências	177 ocorrências
Tamanho correto	2 ocorrências	179 ocorrências
Traçado correto	2 ocorrências	63 ocorrências
Valor correto	9 ocorrências	283 ocorrências

Destino correto:

O WebCorp selecionou 169 ocorrências de *destino correto* encontradas na WWW. As cinco primeiras ocorrências foram analisadas e encontram-se listadas abaixo. Como pode ser observado, nas três primeiras as ocorrências pode-se afirmar que

o item lexical *correto* assume o sentido de *exato*. Parece seguro afirmar, portanto, que, pelo menos nesses casos, a co-ocorrência com a palavra *destino* permite prever que o sentido adequado para realizar a desambiguação é aquele já detectado na análise de corpus.

1. ANDEF, 1993 A Lei n° 7.802 exige que no rótulo de cada defensivo conste para a embalagem em questão, qual o [destino correto](#) que deve ser dado à mesma.
2. Quando o usuário e endereço de destino não coincidem com a identidade daquela máquina a mensagem é redirecionada para o [destino correto](#).
3. Se você está vendo que algo não está se encaminhando para o [destino correto](#), não espere o problema se instalar, resolva-o antes.
4. José Siqueira, responsável pela usina, diz que os habitantes estão "muito satisfeitos, pois estamos dando um [destino correto](#) para os resíduos urbanos da cidade".
5. A psicóloga carioca Sheila Fonseca avalia como enriquecedora a experiência de dar um [destino correto](#) à pena pecuniária.

No entanto, nos exemplos 4 e 5, uma interpretação do item lexical com o sentido de *adequado*, parece mais apropriada. O contexto do exemplo 4 sugere que um destino *adequado* está sendo dado para os resíduos da cidade. Já no exemplo 5, a substituição do item lexical *correto* por *exato* parece alterar o sentido da sentença. Portanto, nesse caso, a interpretação do item lexical com o sentido de *adequado* também parece ser mais apropriada.

Seria necessário observar as 169 ocorrências para verificar se há uma variação de sentidos muito grande e definir com segurança, se o n-grama *destino correto* pode ser utilizado para desambiguar o item lexical em questão.

Local correto:

O WebCorp selecionou 232 ocorrências de *local correto* encontradas na Web. Destas, também foram selecionadas as cinco primeiras ocorrências, que estão listadas abaixo.

1. Note que em alguns casos a peça pode já estar no [local correto](#), mas orientada incorretamente.
2. Então se o TAR foi gerado na pasta /tmp do servidor, é preciso gravá-lo após todas as tarefas concluídas, no [local correto](#).
3. A trepadeira vai crescer e envolver-se na estrutura. Para conduzi-la ao [local correto](#), basta colocar alguns arames.
4. * Se seu Windows não estiver instalado na pasta c:windows edite todas as entradas que tiverem essa indicação para o [local correto](#) do diretório do windows.*
5. Daí você aplica U [para colocar a posição UL (a de destino) no [local correto](#)] B U' B'.

Como pode ser observado, em todas as ocorrências o item lexical *correto* assume o sentido de *exato*. Parece seguro afirmar que a co-ocorrência com a palavra *local* permite prever que o sentido adequado para realizar a desambiguação é o mesmo já estipulado na análise de corpus.

Lugar correto:

O WebCorp selecionou 246 ocorrências de *lugar correto* encontradas na WWW. As cinco primeiras ocorrências listadas na concordância estão listadas abaixo:

1. Blackout Buddy foi projetado para permanecer conectado continuamente, estando assim no [lugar correto](#) na hora que ocorra falta de energia.

2. Quando essa luxação ocorre, a estrutura não volta ao [lugar correto](#) até que o músculo fique novamente relaxado e se estire novamente, aumentando assim seu comprimento.

3. Minhas Legendas - 1 Semana Atrás Pessoal, por favor, vamos postar no [lugar correto](#) e não num tópico que nada tem a ver com o assunto em questão.

4. O estator daquela seção teve que ser retirado, depois situei-o no [lugar correto](#) por meio de pedaços de folha de cobre de 0,1mm colocados entre as chapas e depois soldei novamente.

Qual é o recurso que começou a ser utilizado pelos juízes no campeonato paulista de futebol, objetivando que a barreira fique posicionada no [lugar correto](#)? Um spray de espuma começou a ser utilizado para fazer marcação do [lugar correto](#) da barreira na grama.

Em todas as ocorrências o item lexical *correto* apresenta o sentido de *exato*, sendo que esse padrão ocorreu como uma constante tanto nas ocorrências encontradas no corpus NILC, quanto na WWW . Parece seguro afirmar, portanto, que, pelo menos com uma frequência superior àquela que seria de se esperar com base no acaso, a co-ocorrência com a palavra *lugar* permite prever que o sentido *adequado* para realizar a desambiguação é o mesmo detectado na análise de corpus.

Mapa correto:

Foram encontradas 59 ocorrências do N-grama *mapa correto* na WWW. Ao analisar esses casos, pareceu ser possível uma interpretação do item lexical *correto* como *isento de erros* (sentido 3 apresentado neste trabalho). Mas como o conceito de mapa sugere uma informação de natureza precisa, esses casos foram mantidos nesse

grupo de sentidos que parece ser mais apropriado para a desambiguação do item lexical em questão.

1. Usando o sistema de localização de mapas, você pode simplesmente clicar em "create game" e o [mapa correto](#) será selecionado.
2. A primeira cena do filme parte do mapa de Cape Carrera ou Cyber Town 6.6 (ainda vou analisar qual o [mapa correto](#)), mas para se ter uma idéia do ambiente: uma estrada numa montanha.
3. Se as pessoas investigando um assunto tão importante não conseguem nem mesmo obter um [mapa correto](#) da estação para as testemunhas apontarem onde elas estavam, então qual a chance que o resto do caso tem", criticou Sue Thomason.
4. Esse seria o [mapa correto](#), o antigo Município de Santo Amaro até 1935, mais antigo do que a própria capital: <http://i127.photobucket.com/album/p156/thgteixeira/mapas/mapa2-1.jpg>
5. Para obter outro mapa desenhado para as condições da sua máquina (escapamento, filtros, velas, comando de válvulas alteradas, etc.) personalizado ou não, basta fazer o download do [mapa correto](#).

Portanto, em todas as ocorrências listadas acima o item lexical *correto* apresenta o sentido de *exato*, sendo que esse padrão ocorreu como uma constante tanto nas ocorrências encontradas no corpus NILC, quanto na WWW.

Momento correto:

Foram encontradas 173 ocorrências do n-grama *momento correto* na busca realizada na WWW via WebCorp. Ao analisar esses casos, não foi observada nenhuma variação de sentidos do item lexical *correto* no n-grama em questão, como pode ser observado nos exemplos abaixo:

1. A médium Silvia Browne diz que a data 2012 para o término deste ciclo de sofrimento está errada e o [momento correto](#) seria no término de 2003.
2. Carlos Franciscato - Elaboramos um cronograma interno de trabalho para 2007 e percebemos que todas as tarefas estão sendo encaminhadas no [momento correto](#).
3. Esquece o [momento correto](#) para substituição do produto e chega a ficar até três anos com a mesma palheta, o que representa grande risco, principalmente em épocas de chuvas constantes", afirma o diretor-geral da Cibié Service, Carlos-Henrique Rotella.
4. Alguns tipos de interatividade estão fortemente atrelados ao conteúdo que está sendo exibido e perdem o sentido se não forem exibidas no [momento correto](#).
5. Leão, por sua vez, não nega que Lulinha terá sua oportunidade na equipe, mas alega que ainda não chegou o [momento correto](#) para integrar o atleta ao elenco principal e também garantiu que não tem qualquer problema em trabalhar com garotos.

Concluindo, em todas as ocorrências listadas acima o item lexical *correto* apresenta o sentido de *exato*, sendo que esse padrão ocorreu como uma constante tanto nas ocorrências encontradas no corpus NILC, quanto na WWW.

Nível correto:

Foram encontradas 157 ocorrências de *nível correto* na busca realizada na WWW via WebCorp. Em todas as ocorrências listadas abaixo o item lexical *correto* apresenta o sentido de *exato*.

1. O que posso dizer é que o controlador tinha elementos que confirmavam que a aeronave não estava no [nível correto](#)", disse.
2. Verifique se o fluido de freio está no [nível correto](#) e se não há vazamentos de óleo.
3. Observar sempre o [nível correto](#) do óleo do motor e cumprir suas trocas periódicas é vital para que sua Honda esteja sempre em ordem.

4. O transformador — que adequa a tensão alternada da rede ao [nível correto](#) de tensão alternada que se deseja.
5. O [nível correto](#) do tanque é quando o combustível atinge o bico da bomba.

Portanto, esse modelo de n-grama (*nível correto*) pode ser considerado apropriado para desambiguar o item lexical em questão.

Número correto:

O WebCorp selecionou 167 ocorrências de *número correto* encontradas na WWW. As cinco primeiras ocorrências listadas na concordância estão listadas abaixo e em todas, sem exceção, o item lexical *correto* apresenta o sentido de *exato*, como pode ser observado:

1. Preencher novas solicitações com o [número correto](#) da matrícula e/ ou sem ele quando da impossibilidade de obtê-lo.
2. A não ser que a promiscuidade feminina se torne uma virtude, nenhuma mulher irá exagerar quantos parceiros sexuais já teve, ou mesmo informar o [número correto](#).
3. A tabela abaixo apresenta o [número correto](#) de dígitos que devem ter os números de protocolo e o número de registro.
4. Seja qual for o [número correto](#), todos concordam em que a população indígena no Brasil vem crescendo.
5. "Arrecadação do tributo da gasolina pode financiar obras" (pág. B3) que, neste ano, os programas sociais (de transferência de renda aos pobres) deverão custar R\$ 4,3 bi. Na retranca sobre os mesmos programas, no abre da pág. A9 (Brasil), o registro é de um orçamento de R\$ 4,8 bilhões para isso. Qual é o [número correto](#)?

A repetição deste padrão era esperada, já que o substantivo *número* sugere uma informação de natureza exata. Parece seguro afirmar, portanto, que, pelo menos com uma frequência superior àquela que seria de se esperar com base no acaso, a co-ocorrência com a palavra *lugar* permite prever que o sentido adequado para realizar a desambiguação é o mesmo detectado na análise do corpus NILC.

Perfil correto:

Foram encontradas 135 ocorrências de *perfil correto* na busca realizada na WWW via WebCorp. As cinco primeiras ocorrências listadas na concordância estão listadas abaixo e em todas, sem exceção, o item lexical *correto* apresenta o sentido de *exato*.

1. Eu sinto que está acontecendo uma inquietante proliferação indiscriminada, porque desconhecemos o [perfil correto](#) dessas rádios.
2. Isto nos permite captar esta projeção e traçar um [perfil correto](#) desses candidatos.
3. EVENTOS - O grande gargalo de elaborar bem a viagem de incentivo é exatamente ter o [perfil correto](#) do cliente para que se possa dimensionar os componentes da viagem?
4. Checam datas, nomes, endereços, escolas, cursos, conversam com ex-empregadores, ex-companheiros de trabalho e tentam se aprofundar ao máximo nos detalhes para traçar um [perfil correto](#) do candidato.
5. A evolução constante do método tem nos permitido alcançar cada vez mais resultados melhores na seleção do [perfil correto](#) para a área de desenvolvimento de software, bem como o refinamento dos critérios técnicos e pessoais para essa escolha.

Parece seguro afirmar, portanto, que, a co-ocorrência do item lexical *correto* com a palavra *perfil* permite prever que o sentido adequado para realizar a desambiguação é o mesmo detectado na análise do corpus NILC.

Posicionamento correto:

O WebCorp selecionou 177 ocorrências de *posicionamento correto* encontradas na WWW. As cinco primeiras ocorrências listadas na concordância estão listadas abaixo e em todas, sem exceção, o item lexical *correto* assume o sentido de *exato*.

1. Ao contrário, fica difícil a automatização do [posicionamento correto](#) da língua em ausência de movimento ou durante a fonação.
2. A versatilidade dos aparelhos, o [posicionamento correto](#) do corpo em cada movimento, a multiplicidade dos exercícios, as suas diversas modificações segundo as características e limitações de cada praticante, tudo isto demanda muita demanda muita observação, estudo, vivência intensa de execução e, sobretudo a prática do ensino.
3. Considerar-se o [posicionamento correto](#) da faca e a oscilação do corte é fundamental para um excelente resultado.
4. Pode-se constatar que o desenho é anatomicamente curvo, otimizando a adaptação tubo/dente, e a endentação bucal compatível com a superfície do molar, facilitando o [posicionamento correto](#) do tubo bucal.
5. O [posicionamento correto](#) do túnel femoral seria às 11 horas no joelho direito e a 1 hora no joelho esquerdo, deixando 1 a 2 mm de cortical posterior (Figura 2).

Pode-se afirmar, portanto, que a busca realizada na WWW confirmou que o padrão colocação *posicionamento correto* encontrado no corpus NILC pode ser considerado significativo e eficaz na desambiguação do item lexical em questão.

Tamanho correto:

Foram encontradas 179 ocorrências do n-grama *tamanho correto* na WWW. Mais uma vez, nas cinco primeiras ocorrências analisadas o item lexical *correto* assumiu o sentido de *exato*, como pode ser observado nos exemplos retirados da WWW e que estão listados abaixo:

1. PondPack no sentido de estabelecer os requisitos máximos de armazenamento - de modo que você possa inserir o [tamanho correto](#) de bacia de detenção no primeiro projeto de terraplenagem.
2. Escolha o [tamanho correto](#) da luva ao tamanho da mão do profissional que a veste.
3. Então, este teste inclui uma barra de calibragem que pode ser medida para determinar que as letras são do [tamanho correto](#).
4. Eu tenho uma tabela onde coloquei os tamanhos usados por cada um dos laboratórios e mando a foto no [tamanho correto](#).
5. Quando imprimo uma folha por vez sem ser no modo dupla-face sai no [tamanho correto](#) 28,2 cm.

Pode-se concluir, portanto, que esse padrão ocorreu como uma constante tanto nas ocorrências encontradas no corpus NILC, quanto na WWW e pode ser utilizado para desambiguar o item lexical em questão.

Traçado correto:

O WebCorp selecionou 63 ocorrências de *traçado correto* encontradas na WWW. As cinco primeiras ocorrências listadas na concordância estão listadas abaixo e em todas o item lexical *correto* assume o sentido de *exato*.

1. É fundamental que tenhamos um [traçado correto](#) de como tratar a empresa nacional, para fazê-la utilizar sua capacidade ociosa, criativa, produtiva, tornando-a participante de várias atividades nesse mercado.
2. O vídeo vai ensinar como fazer o [traçado correto](#) e a shadow vai te dar uma base de comparação para treinos.
3. A régua será um instrumento imprescindível para o [traçado correto](#) das linhas inclinadas.
4. Piquet justifica, com razão, que se quem escapou tentar voltar pelo [traçado correto](#) seria perigoso, pois alguém pode vir de frente e também perder o controle e termos um acidente grave.
5. Você pode acelerar até o último momento soltar o acelerador e começar a efetuar a curva, quando notar que o veículo está fazendo o [traçado correto](#) pode-se acelerar fundo rumo a vitória.

Pode-se afirmar, portanto, que a busca realizada na WWW confirmou que o padrão colocação *traçado correto* encontrado no corpus NILC pode ser considerado significativo e eficaz na desambiguação do item lexical em questão.

Valor correto:

O WebCorp selecionou 283 ocorrências de *valor correto* encontradas na WWW. Destas, como também dos demais casos, foram selecionadas as cinco primeiras ocorrências apresentadas na concordância. Estas ocorrências estão listadas abaixo.

1. Assim, quando recebermos seu pedido, calculamos o [valor correto](#) dos correios e informamos, seja através do e-mail (em caso de depósito ou transferência), seja no corpo do pedido que enviamos com seu boleto bancário.
2. Os campos 31 e 32 da GFIP estão sendo impressos com o dobro do [valor correto](#).

3. Conseqüentemente, este deve ser o [valor correto](#) da constante de equilíbrio para o sistema apresentado.

4. O [valor correto](#) é de 149,597870691 milhões de km.

5. Assim, negocie com [valor correto](#) da peça, levando em conta as possibilidades e os custos de uma restauração.

Como pode ser observado, todas as ocorrências se referem a um valor numérico, seja este valor dinheiro, distância ou ainda um outro valor matemático, como uma constante. Parece seguro afirmar, portanto, que, pelo menos com uma frequência superior àquela que seria de se esperar com base no acaso, a co-ocorrência com a palavra *valor* permite prever que o sentido adequado para realizar a desambiguação é aquele já detectado na análise de corpus.

Sentido 2: adequado

N-gramas	Número de ocorrências encontradas no corpus NILC	Número de ocorrências encontradas na Web
Atendimento correto	2 ocorrências	123 ocorrências
Comportamento correto	5 ocorrências	162 ocorrências
Discurso correto	3 ocorrências	92 ocorrências
Funcionamento correto	3 ocorrências	199 ocorrências
Método correto	4 ocorrências	148 ocorrências
Modelo correto	7 ocorrências	166 ocorrências
Modo correto	7 ocorrências	219 ocorrências

Procedimento correto	7 ocorrências	265 ocorrências
Uso correto	16 ocorrências	218 ocorrências
Ecologicamente correto	19 ocorrências	205 ocorrências
Mais correto	32 ocorrências	290 ocorrências
Tecnicamente correto	5 ocorrências	189 ocorrências

Atendimento correto:

Foram encontradas 123 ocorrências de *atendimento correto* na WWW. Destas, como também ocorreu nos demais casos, foram selecionadas as cinco primeiras ocorrências apresentadas na concordância. Estas ocorrências estão listadas abaixo.

1. É essencial passar a maior quantidade de informações para que a viatura que vai atender a ocorrência tenha as melhores condições possíveis de efetuar um [atendimento correto](#) procure manter a calma;
2. Muitas mortes maternas poderiam ter sido evitadas, caso os profissionais utilizasse o [atendimento correto](#).
3. Você não sabe se é a Santa Casa de Misericórdia de Catanduva que foi falha, que não deu [atendimento correto](#).
4. A segunda providência é mobilizar alguém para chamar o resgate e pedir que pessoas qualificadas sejam enviadas ao local para prestar [atendimento correto](#) e transportar a vítima com segurança para alguma instituição de saúde.

5. O projeto oferece assistência através de obstetras, infectologistas, pediatras, psicólogos, assistentes sociais e enfermeiras, para conscientizar a gestante HIV positiva e promover o [atendimento correto](#).

Como pode ser observado nas ocorrências acima, em todos os casos o item lexical *correto* assumiu o sentido de *adequado*. Parece seguro afirmar, portanto, que a co-ocorrência do item lexical *correto* com a palavra *atendimento* permite prever que o sentido adequado para realizar a desambiguação é o mesmo já detectado na análise do corpus NILC.

Comportamento correto:

Foram encontradas 162 ocorrências de *comportamento correto* na WWW. Nas cinco primeiras ocorrências analisadas o item lexical *correto* assumiu o sentido de *adequado*, como pode ser observado nos exemplos retirados da WWW que estão listados abaixo:

1. Aqueles homens tentaram ganhar a salvação por sua boa conduta, por seu [comportamento correto](#), porque não faziam isto, ou aquilo.
2. Assumimos que existe um [comportamento correto](#) ou conjunto de regras quando, de fato, são convenções que desenvolvemos coletivamente e que são usadas para camuflar relações de poder e autoridade não examinadas.
3. Normalmente, são pessoas que têm conhecimento da legislação e do [comportamento correto](#), mas querem sempre levar vantagem sobre os demais usuários, sem se importar se a sua atitude vai perturbar o fluxo de tráfego.
4. O guia deverá assumir um [comportamento correto](#) e gentil junto ao cliente.

5. Nas escolas, onde o jovem deve desde já ter um [comportamento correto](#) no trânsito (como pedestre) e estar alerta para o comportamento adequado do motorista, a lei não está sendo cumprida.

Pode-se concluir, portanto, que esse padrão ocorreu como uma constante tanto nas ocorrências encontradas no corpus NILC, quanto na WWW e pode ser utilizado para desambiguar o item lexical em questão.

Discurso correto:

Foram encontradas 92 ocorrências de *discurso correto* na WWW. Nas quatro primeiras ocorrências analisadas o item lexical *correto* também assumiu o sentido de *adequado*, como pode ser observado nos exemplos abaixo:

1. "Não sou bom de dar nota para ninguém não, só sei jogar futebol", respondeu rápido, retomando de uma vez o [discurso correto](#) que pode o levar a titular do Brasil logo em sua primeira participação em Copas do Mundo.
2. Existe um [discurso correto](#) até para pedir dinheiro emprestado a um amigo?
3. Outro ponto positivo é que a prefeita adotou um [discurso correto](#) de que não será uma salvadora da pátria, mas que irá se esforçar para fazer o possível para melhorar a cidade.
4. O [discurso correto](#) releva e justifica a sua causa, ate dá-lhe a certeza da escolha certa, mas a inquietação é latente.
5. A sala de imprensa do Vaticano divulgará nesta quarta-feira (8) o [discurso correto](#), que foi pronunciado pelo Pontífice na terça-feira (7).

Vale ressaltar que a ocorrência 5 parece definitivamente duvidosa. Aqui o sentido de *exato*, ou seja, exatamente como pronunciado pelo pontífice, é certamente possível. Apesar dessa exceção, ainda é possível afirmar que esse é mais um padrão que

ocorreu como uma constante tanto nas ocorrências encontradas no corpus NILC, quanto na WWW e pode ser utilizado para desambiguar o item lexical em questão.

Funcionamento correto:

O WebCorp selecionou 199 ocorrências de *funcionamento correto* encontradas na WWW. As cinco primeiras ocorrências listadas na concordância estão listadas abaixo e em todas, sem exceção, o item lexical *correto* assume o sentido de *adequado*.

1. Antes de começar o exercício, vou colocar em observação que todas as palavras entre aspas "" que aparecerem neste tutorial devem ser colocadas sem as aspas para ocorrer o [funcionamento correto](#) do exercício.
2. Se o detentor dos direitos autorais e/ou o LICENCIADO que distribui o TRABALHO não expressam nenhuma garantia dada por conta própria pelo [funcionamento correto](#) do TRABALHO, o detentor dos direitos autorais e/ou o LICENCIADO disponibilizam o TRABALHO como ele é sem garantia de nenhum tipo, expressas ou implícitas, incluindo, mas não se limitando a, garantias comerciais e de compatibilidade para um determinado fim específico.
3. O microcontrolador PIC gera todos os sinais de controle e dados para o LM629, programando-o para o [funcionamento correto](#).
4. O imóvel objeto deste contrato será entregue nas condições descritas no auto de vistoria, ou seja, com instalações elétricas e hidráulicas em perfeito funcionamento, com instalações elétricas e hidráulicas em perfeito funcionamento, com todos os cômodos e paredes pintados, sendo que portas, portões e acessórios se encontram também em [funcionamento correto](#), devendo o LOCATÁRIO mantê-lo desta forma.
5. A função do Apolo é, justamente, agregar inteligência ao [funcionamento correto](#) de uma empresa.

Pode-se afirmar, portanto, que a busca realizada na WWW confirmou que o padrão colocação *funcionamento correto* encontrado no corpus NILC pode ser considerado significativo e eficaz na desambiguação do item lexical em questão.

Método correto:

Foram encontradas 148 ocorrências de *método correto* na WWW. Nas cinco primeiras ocorrências listadas na concordância o item lexical *correto*, assim como foi observado no corpus NILC, assume o sentido de *adequado*. Os exemplos estão citados abaixo:

1. O [método correto](#) de coleta, assim como, um transporte seguro e a conservação da amostra são as principais exigências para o sucesso dos testes.
2. Daí a idéia da prática do seiza, que passou a ser considerado o [método correto](#) para a obtenção da tranqüilidade mental.
3. O [método correto](#) seria acrescentar itens nos menus da barra de operação.
4. Em cada envelope, além de uma camisinha, é incluído um folheto contendo todas as informações sobre o [método correto](#) de uso do preservativo.
5. O problema, entretanto, é que, em princípio, as comparações múltiplas de médias de tratamentos não são o [método correto](#) para detectar diferenças em bioeficácia.

Pode-se afirmar, portanto, que a busca realizada na WWW confirmou que o padrão de colocação *método correto* encontrado no corpus NILC pode ser considerado significativo e eficaz na desambiguação do item lexical em questão.

Modelo correto:

O WebCorp selecionou 166 ocorrências de *modelo correto* encontradas na WWW. Nas cinco primeiras ocorrências apresentadas na concordância, o item lexical *correto* também assumiu o sentido de *adequado*. As ocorrências em questão estão listadas abaixo.

1. Se tiver dúvidas quanto ao [modelo correto](#) para seu caso, entre em contato antes de solicitar.
- 2.. Durante a instalação do software, indique o [modelo correto](#) correspondente à placa para evitar erros.
3. É incrível que a agricultura, que tem gerado empregos e riqueza através do "agrobusiness", tenha ainda que discutir se este é o [modelo correto](#) de explorar a terra.
4. Você volta lá e escolhe o [modelo correto](#) e faz de novo... e pronto!!!
5. Solicitem [modelo correto](#) de instalação e aterramento que enviaremos via correio ou e-mail

Parece seguro afirmar, portanto, que a co-ocorrência do item lexical *correto* com a palavra *modelo* permite prever que o sentido adequado para realizar a desambiguação é aquele já detectado na análise do corpus NILC.

Modo correto:

O WebCorp selecionou 219 ocorrências de *modo correto* encontradas na WWW. Nas cinco primeiras ocorrências apresentadas na concordância, o item lexical *correto* assumiu o sentido de *adequado*. As ocorrências em questão estão listadas abaixo.

1. Antes de realizar a transferência de qualquer arquivo verifique se você está usando o [modo correto](#), isto é, no caso de arquivos-texto, o modo é ASCII, e no caso de arquivos binários (.exe, .com, .zip, .wav, etc.).
2. O programa utiliza textos para ensinar o [modo correto](#) de digitação e se auto-ajusta em diferentes níveis dependendo da qualidade de digitação do usuário.
3. Os alunos recebiam os visitantes no estande, onde forneciam informações aos pacientes sobre o [modo correto](#) de utilização dos medicamentos e os riscos da auto-medicação.
4. Seus próprios conceitos sobre o [modo correto](#) de viver, sobre felicidade, sobre o modo de pensar adequado, sobre respeito, sobre disciplina, sobre a verdade.
5. O manual da impressora ou do scanner pode indicar o [modo correto](#) a ser usado.

Para esses casos, portanto, é possível afirmar que a co-ocorrência do item lexical *correto* com a palavra *modo* permite prever que o sentido adequado para realizar a desambiguação é o mesmo detectado na análise do corpus NILC.

Procedimento correto:

Foram encontradas 265 ocorrências de procedimento *correto* na WWW. Nas cinco primeiras ocorrências apresentadas na concordância, o item lexical *correto* assumiu o sentido de *adequado* ao ser precedido da palavra *procedimento*. As ocorrências em questão estão listadas abaixo.

1. Caso seu cadastro conosco esteja desatualizado, o [procedimento correto](#) para a atualização será o envio de um fax para o número (11) 3849-1166, em papel timbrado de sua empresa.

2. Eu cobri este assunto (Pig Farming agosto e outubro de 1999) e desde então eu tenho coletado dados do custo efetivo de um [procedimento correto](#) de biossegurança - o correto programa está ilustrado na Tabela 1.

3. Para evitar esse tipo de erro explicamos abaixo o [procedimento correto](#) que deve ser efetuado na troca do equipamento que será substituído.

4. A empresa emitente das notas fiscais buscou orientação quanto ao [procedimento correto](#) a ser adotado no preenchimento das notas fiscais, onde foi orientado de que não havia necessidade de se emitir a nota fiscal de remessa.

5. Qual seria o [procedimento correto](#) neste caso, que aconteceu comigo...

Nesses casos, portanto, é possível afirmar que a co-ocorrência do item lexical *correto* com a palavra *procedimento* permite prever que o sentido adequado para realizar a desambiguação é o mesmo detectado previamente na análise do corpus NILC.

Uso correto:

Foram encontradas 218 ocorrências de *uso correto* na WWW. Nas quatro primeiras ocorrências analisadas o item lexical *correto* assumiu o sentido de *adequado*, como pode ser observado nos exemplos abaixo:

1. Projeto de Lei nº2151, de 2003 (Do Senhor Coronel Alves) Institui o "Dia do Alerta sobre o [uso correto](#) da cadeira e do cinto de segurança para crianças".

2. É necessário que campanhas de esclarecimento informem o [uso correto](#) destes dispositivos exigidos pelo Código de Trânsito Brasileiro, para que as famílias possam proteger de modo eficaz as crianças quando viajem nos automóveis.

3. O objetivo principal de nosso estande é orientar a população quanto aos riscos da má utilização de medicamentos, orientar sobre o [uso correto](#), as formas corretas de descartar o

medicamento ao final do uso e demonstrar ao paciente possíveis interações medicamentosas, orientando sobre o que deve ser feito para correção do problema.

4. Prescrição ou Receita Médica A receita médica é que vai fornecer todas as informações sobre o [uso correto](#) do medicamento.

5. Devido ao fato de que português tem na 2ª pessoa (você) o mesmo tratamento gramatical dado à 3ª pessoa (ele ou ela), o aluno frequentemente encontra dificuldade no [uso correto](#) dos pronomes possessivos em inglês.

Já no exemplo 5, o sentido ideal para o item lexical *correto* deveria ser *isento de erros* (sentido 3), já que neste caso em especial, *correto* está sinalizando o uso apropriado da língua. Mas por tratar-se de um caso bem específico, esse modelo de n-grama será mantido nesse grupo de sentido por poder ser utilizado para desambiguar a maior parte dos casos do item lexical em questão.

Ecologicamente correto:

Foram encontradas 205 ocorrências de *ecologicamente correto* na WWW. Nas cinco primeiras ocorrências analisadas o item lexical *correto* assumiu sistematicamente o sentido de *adequado*, como pode ser observado nos exemplos abaixo:

1. Vestido feito em couro de tilápia [ecologicamente correto](#), por Patrícia Viera FOTOS DA EXPOSIÇÃO O São Paulo Fashion Week chega ao fim de sua 22ª edição.
2. Produto [ecologicamente correto](#), extremamente seguro e de fácil aplicação.
3. Construção civil aposta no "[ecologicamente correto](#)" em SP.
4. A obra está sendo totalmente desenvolvida dentro do conceito do [ecologicamente correto](#), desde a fundação ao revestimento.

5. Para o diretor da Andrade Mendonça, Antônio Andrade Júnior, não existiu demagogia com a questão do [ecologicamente correto](#).

Nesses casos, portanto, é possível afirmar que a co-ocorrência do item lexical *correto* com a palavra *ecologicamente* permite prever que o sentido adequado para realizar a desambiguação é o mesmo detectado previamente na análise do corpus NILC.

Mais correto:

Foram encontradas 290 ocorrências de *mais correto* na WWW. Nas cinco primeiras ocorrências analisadas o item lexical *correto* assumiu sistematicamente o sentido de *adequado*, como pode ser observado nos exemplos abaixo:

1. O caminho [mais correto](#) para descobrir onde encontrá-la é perguntar a si mesmo: Onde você estaria agora?
2. Consulte um advogado pra descobrir o caminho legal [mais correto](#) pra fazer isso.
3. Com isso acho que você consegue passar um ano tranquilo, porem o [mais correto](#), aconselhável e que vai lhe possibilitar você curtir o seu som de qualquer forma é colocar uma bateria maior e melhor, não é muito caro.
4. Concordo plenamente, o [mais correto](#) na realidade seria ter o ID como obrigatório desde o início, talvez para a versão 2.0 isso fique obrigatório de vez , vou falar com o resto da equipe ok.
5. Os grupos de defesa dos homossexuais não gostam da palavra “entendido” e preferem apenas o termo direto: gay. “Eu também prefiro o termo direto. O [mais correto](#) é mesmo homossexual ou gay”.

Nesses casos, portanto, é possível afirmar que a co-ocorrência do item lexical *correto* com a palavra *mais* permite prever que o sentido adequado para realizar a desambiguação é o mesmo detectado previamente na análise do corpus NILC.

Tecnicamente correto:

O WebCorp selecionou 189 ocorrências de *tecnicamente correto* encontradas na WWW. Nas cinco primeiras ocorrências apresentadas na concordância, o item lexical *correto* assumiu o sentido de *adequado*, como pode ser observado nas ocorrências em questão listadas abaixo.

1. Um manual que registrasse o assunto [tecnicamente correto](#) e completo, mostrando projetos.
2. Isto acontece quando o fator humano é negligenciado. O trabalho, embora [tecnicamente correto](#), não envolve as pessoas adequadas nas fases anteriores e a implantação é iniciada sem o comprometimento de quem efetivamente faz parte da mudança.
3. Porque tudo o que você me disse está [tecnicamente correto](#), porém esta informação me é totalmente inútil, pois continuo perdido".
4. Giambiagi - Nós temos de combinar o que é [tecnicamente correto](#) com o que é socialmente aceitável ou politicamente palatável.
5. O relator no TJ, desembargador César Lacerda, julgou [tecnicamente correto](#) o cálculo feito pelo contador judicial, mas considerou dissociado da realidade o valor obtido.

Para esses casos, portanto, é possível afirmar que a co-ocorrência do item lexical *correto* com a palavra *tecnicamente* permite prever que o sentido adequado para realizar a desambiguação é o mesmo detectado previamente na análise do corpus NILC.

Sentido 3: isento de erros

N-gramas	Número de ocorrências encontradas no corpus NILC	Número de ocorrências encontradas na Web
Absolutamente correto	2 ocorrências	153 ocorrências
Caminho correto	15 ocorrências	293 ocorrências
Diagnóstico correto	7 ocorrências	232 ocorrências
Endereço correto	2 ocorrências	163 ocorrências
Entendimento correto	3 ocorrências	246 ocorrências
Nome correto	6 ocorrências	306 ocorrências
Raciocínio correto	3 ocorrências	133 ocorrências
Sentido correto	5 ocorrências	173 ocorrências
Sintaticamente correto	3 ocorrências	69 ocorrências

Absolutamente correto:

O WebCorp selecionou 153 ocorrências de *absolutamente correto* encontradas na WWW. Assim como na maioria dos n-gramas analisados até então, nas cinco primeiras ocorrências apresentadas na concordância, o item lexical *correto* assumiu o mesmo sentido definido a partir dos exemplos encontrados no corpus NILC, que nesse caso é o sentido *isento de erros*. Os exemplos em questão encontram-se listados abaixo.

1. Está [absolutamente correto](#), mas os nutrientes que ingerimos quando comemos têm mais funções do que simplesmente matar a fome.
2. Janilo e Arquimedes pelo excelente desempenho no desenvolvimento [absolutamente correto](#) de alguns problemas foram agraciados com Menção Honrosa e, o Yuri, que dentre eles alcançou maior pontuação, foi premiado com Medalha de Bronze.
3. Segundo, não é exatamente equivocado, mas também não é [absolutamente correto](#) dizer que a Internet é o virtual (junto a outros 24 meios, é parte constitutiva do espaço virtual).
4. Observa-se que, ultimamente, a discussão vem girando em torno da insegurança - o que nos parece [absolutamente correto](#) - visto que o ambiente que nos cerca é, e sempre foi, de insegurança.
5. Não faria o mesmo tiroteio contra o procurador, mas no conteúdo ele está [absolutamente correto](#).

Para esses casos, portanto, também é possível afirmar que a co-ocorrência do item lexical *correto* com a palavra *absolutamente* permite prever que o sentido adequado para realizar a desambiguação é o mesmo que foi previamente definido a partir da análise das ocorrências encontradas no corpus NILC.

Caminho correto:

Foram encontradas 293 ocorrências de *caminho correto* na WWW. Nas cinco primeiras ocorrências apresentadas na concordância, o item lexical *correto* assumiu o sentido de *isento de erros*, como pode ser observado nos exemplos listados abaixo.

1. Verifique qual é o [caminho correto](#) do servidor do seu site (eu costumo usar a função `phpinfo` do PHP).
2. Ao longo do percurso há postos de controle (PCs) que verificam se as equipes estão no [caminho correto](#), anotando o tempo de passagem pelo local.
3. O refúgio nos desvia do caminho errado para o [caminho correto](#), e deste modo é efetivamente uma prática do Pratimoksha.
4. Então dentro do [caminho correto](#), temos o caminho maior e o caminho menor, apesar, é claro, de ambos serem o caminho correto.
5. Aqui, a gente aconselha e mostra o [caminho correto](#) aos futuros talentos.

Para esses casos, portanto, também é possível afirmar que a co-ocorrência do item lexical *correto* com a palavra *caminho* permite prever que o sentido adequado para realizar a desambiguação é o mesmo que foi previamente identificado a partir da análise das ocorrências encontradas no corpus NILC.

Diagnóstico correto:

Foram encontradas 232 ocorrências de *diagnóstico correto* na busca na WWW via WebCorp. Nas cinco primeiras ocorrências apresentadas na concordância, o item lexical *correto* assumiu o mesmo sentido definido a partir dos exemplos encontrados no corpus NILC, que nesse caso é o sentido *isento de erros*. Os exemplos em questão encontram-se listados abaixo.

1. Exames físicos regulares que incluem verificação de alergias e da função pulmonar podem ajudar no [diagnóstico correto](#).

2. Quando o tumor é lembrado na primeira consulta a radiografia freqüentemente conduz ao [diagnóstico correto](#), imediatamente.
3. A conseqüente "falsa segurança" causa um atraso adicional de 2 a 40 semanas, antes de uma segunda radiografia que revele o [diagnóstico correto](#).
4. Ele pode fazer um [diagnóstico correto](#), após exame visual da pele e também exames laboratoriais (sangue) e de apoio diagnóstico (biópsia da pele).
5. Acredito que é muito importante que nós façamos um [diagnóstico correto](#) da situação mundial, internacional, para que possamos construir alternativas reais.

Nesses casos, portanto, também é possível afirmar que a co-ocorrência do item lexical *correto* com a palavra *diagnóstico* permite prever que o sentido adequado para realizar a desambiguação é o mesmo que foi previamente definido a partir da análise das ocorrências encontradas no corpus NILC.

Endereço correto:

O WebCorp selecionou 163 ocorrências de *endereço correto* encontradas na WWW. Nas cinco primeiras ocorrências apresentadas na concordância, o item lexical *correto* assumiu o mesmo sentido definido a partir dos exemplos encontrados no corpus NILC, que nesse caso é o sentido *isento de erros*, como pode ser observado nos exemplos abaixo.

1. Adiciona meu banner e me passa o [endereço correto](#) do seu que eu adiciono lá.
2. Assinar a folha e anexar uma foto apropriada do pretendente para envia - la ao [endereço correto](#) via correio.

3. OPCIONAL , mas para que possamos incluí-lo na listagem nós precisamos ter o [endereço correto](#) de sua página (URL).
4. Eu encontrei o [endereço correto](#).
5. Faça um novo cadastro, agora com seu [endereço correto](#) (ou altere o e-mail já cadastrado).

Para esses casos, portanto, também é possível afirmar que a co-ocorrência do item lexical *correto* com a palavra *endereço* permite prever que o sentido adequado para realizar a desambiguação é o mesmo que foi previamente estabelecido a partir da análise das ocorrências encontradas no corpus NILC.

Entendimento correto:

Foram encontradas 246 ocorrências de *entendimento correto* na busca feita na WWW via WebCorp. Assim como na maioria dos n-gramas analisados até então, nas cinco primeiras ocorrências apresentadas na concordância, o item lexical *correto* assumiu o mesmo sentido definido a partir dos exemplos encontrados no corpus NILC, que nesse caso é o sentido *isento de erros*. Os exemplos em questão podem ser observados na lista abaixo.

1. Vamos adquirindo um [entendimento correto](#) da vida e vamos vivendo com mais sabedoria e despreocupação.
2. Para mulheres, em particular, sempre foi mais importante serem aprovadas do que terem um [entendimento correto](#) do mundo a sua volta.
3. Este é o [entendimento correto](#) desta orientação que orienta o homem a seguir a revelação de Deus.

4. The Adventist Review, Hellen White é a chave par ao [entendimento correto](#) da Bíblia, já que ela é a intérprete infalível da Bíblia e a corte final de apelação para o povo de Deus.

5. Matemática pode propiciar, tanto a professores quanto a alunos, um entendimento mais claro dos aspectos sociológicos da educação matemática, além de propiciar um [entendimento correto](#) da matemática como um produto cultural, um corpo dinâmico do conhecimento e não como uma coleção estática de regras e métodos desvinculados do contexto sócio-cultural

Nesses casos, portanto, também é possível afirmar que a co-ocorrência do item lexical *correto* com a palavra *entendimento* permite prever que o sentido adequado para realizar a desambiguação é o mesmo que foi previamente definido a partir da análise das ocorrências encontradas no corpus NILC.

Nome correto:

O WebCorp selecionou 306 ocorrências de *nome correto* encontradas na WWW. Nas cinco primeiras ocorrências apresentadas na concordância, o item lexical *correto* assumiu o mesmo sentido definido a partir dos exemplos encontrados no corpus NILC (*isento de erros*). Os exemplos em questão encontram-se listados abaixo.

1. Há quem diga que, no passado, a 'letra de médico' funcionava como um código para evitar que o paciente aprendesse o [nome correto](#) do medicamento, dificultando futuras automedicações.

2. Configure uma Conexão do ActiveSync no Windows Mobile 2003 Second Edition Entre em contato com o seu departamento de TI para encontrar o [nome correto](#) do servidor, o nome de usuário, a senha e o domínio de acesso.

3. Caso o campo de mensagem de Entrada esteja em branco ou incorreto, informe o [nome correto](#) do servidor IMAP4 ou POP3 (você obtém essa informação com o seu ISP).

4. Oficial de Registro Civil retifique o erro encontrado, fazendo constar na certidão de nascimento da Requerente o [nome correto](#) da avó paterna como ADRIANA ENGRÁCIA AZEVEDO.

5. Além disso, algumas músicas que tinham apenas o nome do artista, ele completou c/o [nome correto](#) do álbum, da capa do mesmo e inclusive a faixa desta música neste álbum.

Para esses casos, portanto, também é possível afirmar que a co-ocorrência do item lexical *correto* com a palavra *nome* permite prever que o sentido adequado para realizar a desambiguação é o mesmo que foi previamente determinado a partir da análise das ocorrências encontradas no corpus NILC.

Raciocínio correto:

Foram encontradas 133 ocorrências de *raciocínio correto* na WWW. Assim como na maioria dos casos analisados até então, nas cinco primeiras ocorrências apresentadas na concordância, o item lexical *correto* assumiu o mesmo sentido definido a partir dos exemplos encontrados no corpus NILC, que nesse caso é o sentido *isento de erros*, como pode ser observado nos exemplos listados abaixo:

1. Não aja sob os impactos da emoção, confundindo capricho com [raciocínio correto](#).
2. Um sofisma, para quem não sabe, consiste de um [raciocínio correto](#) erigido sobre bases erradas, feito em geral maliciosamente, por indivíduos que estão cientes da falsidade de suas premissas.
3. Acredito que seja esse o [raciocínio correto](#).

4. Pela lógica do [raciocínio correto](#) e do bom senso, conclui-se que, se este ponto não foi focado, é porque não merecia ao menos consideração.
5. O [raciocínio correto](#) seria aquele, que baseado em premissas válidas, leva necessariamente a uma determinada conclusão.

Para esses casos, portanto, também é possível afirmar que a co-ocorrência do item lexical *correto* com a palavra *raciocínio* permite prever que o sentido adequado para realizar a desambiguação é o mesmo que foi previamente definido a partir da análise das ocorrências encontradas no corpus NILC.

Sentido correto:

O WebCorp selecionou 173 ocorrências de *sentido correto* encontradas na WWW. Nas cinco primeiras ocorrências apresentadas na concordância, o item lexical *correto*, mais uma vez assumiu o mesmo sentido definido a partir dos exemplos encontrados no corpus NILC, que nesse caso é o sentido *isento de erros*. Os exemplos em questão encontram-se listados abaixo.

1. Tanto a palavra Hôji quanto Butsu-ji são de fundamental importância para compreendermos o [sentido correto](#) dessa celebração, ou seja, dessa cerimônia.
2. Mas quando está em voga uma pseudodoutrina, que mistura o [sentido correto](#) de um texto bíblico com uma aplicação errônea, ou vice-versa, muitos crentes sucumbem.
3. Nas moto bombas monofásicas 6 (seis) fios, trifásicas, ou nas bombas mancalizadas, observe, logo na partida, pelo lado traseiro do motor, se este gira no [sentido correto](#) (sentido horário, exceto modelo BCA-43).
4. Pode ser interpretada como "grátis" ou "livre", e este é o [sentido correto](#).

5. Pensar que cada ato do usuário é uma tentativa de avançar no [sentido correto](#), pois um erro não é mais que um ato que se tem especificado incompleto ou incorreto.

Pode-se concluir que, nesses casos, a co-ocorrência do item lexical *correto* com a palavra *sentido* permite prever que o sentido adequado para realizar a desambiguação é o mesmo que foi previamente determinado a partir da análise das ocorrências encontradas no corpus NILC.

Sintaticamente correto:

Foram encontradas 69 ocorrências de *sintaticamente correto* na busca na WWW via WebCorp. Nas cinco primeiras ocorrências apresentadas na concordância, o item lexical *correto* assumiu o mesmo sentido definido a partir dos exemplos encontrados no corpus NILC, que nesse caso é o sentido *isento de erros*, como pode ser observado nos exemplos abaixo.

1. Escreva de modo claro, gramatical e [sintaticamente correto](#).
2. Note que os mínimos detalhes são observados pelo compilador, assim temos que escrever o programa [sintaticamente correto](#) para ser aceito pelo compilador.
3. Observem que é diferente do gabarito, mas compreensível, não torce o sentido, e gramatical e [sintaticamente correto](#).
4. 'Membro novato' está [sintaticamente correto](#)???
5. Assinale a opção que contém trecho [sintaticamente correto](#).

Para esses casos, portanto, também é possível afirmar que a co-ocorrência do item lexical *correto* com a palavra *sintaticamente* permite prever que o sentido

adequado para realizar a desambiguação é o mesmo que foi previamente estabelecido a partir da análise das ocorrências encontradas no corpus NILC.

Sentido 4: válido

N-gramas	Número de ocorrências encontradas no corpus NILC	Número de ocorrências encontradas na Web
É correto dizer que	2 ocorrências	105 ocorrências
Como correto	1 ocorrência	172 ocorrências

É correto dizer que:

Foram encontradas 105 ocorrências de *é correto dizer que* na busca realizada na WWW via WebCorp. Nas cinco primeiras ocorrências retiradas da concordância, o item lexical *correto*, mais uma vez, assumiu o mesmo sentido definido a partir dos exemplos encontrados no corpus NILC, que nesse caso é o sentido denominado *válido*. Os exemplos em questão encontram-se listados abaixo.

1. Por este motivo, [é correto dizer que](#) o identificador são todos os caracteres antes do primeiro traço.
2. Por isso não [é correto dizer que](#) a escala idealizada pelo americano Charles Richter, em 1935, que leva seu nome e indica a magnitude de um sismo, vai até 9.
3. Com relação a preço, [é correto dizer que](#) ele é alto, baixo, módico, insignificante, exorbitante.

4. Portanto, não [é correto dizer que](#) a origem pagã da santificação domingo nada tem a ver com a sua rejeição pelo povo de Deus.

5. Dessa forma [é correto dizer que](#) o repórter gonzo altera o objeto de sua reportagem da mesma forma que o objeto altera o próprio repórter.

Pode-se concluir que, nesses casos, a co-ocorrência do item lexical *correto* com a palavra *válido* permite prever que o sentido adequado para realizar a desambiguação é o mesmo que foi inicialmente determinado a partir da análise das ocorrências encontradas no corpus NILC.

Como correto:

Foram encontradas 172 ocorrências de *como correto* na busca realizada na WWW. Nas quatro primeiras ocorrências retiradas da concordância, o item lexical *correto*, assumiu mesmo sentido de *válido*. No entanto, no exemplo 5 o sentido mais apropriado seria *adequado*. Esse foi o único exemplo que fugiu à regra. Os exemplos em questão encontram-se listados abaixo.

1. O agente da Administração, no momento em que está exercendo sua função, deve perseguir e tomar [como correto](#) os conceitos éticos e constitucionais, e deixar de lado suas convicções político-partidárias.

2. Mesmo que qualquer nome seja considerado [como correto](#), nas normas da língua portuguesa existe uma forma para cada nome, que pode ser utilizada no BrOffice.

3. Seu exame histórico da economia, após o mercantilismo, é correto, [como correto](#) é o enfoque que oferta à globalização atual, com suas virtudes e seus defeitos, estes, ultimamente, maiores do que aquelas.

4. Tanto no primeiro quanto no segundo caso, não raro o dono do veículo tem o diagnóstico [como correto](#) e atende religiosamente as solicitações feitas.

5. Porém, às vezes as formas de "como" chegar a seus clientes pode ir contra do que você percebe [como "correto"](#).

Na ocorrência 5, o sentido mais adequado não parece ser *válido*, como nos demais casos, tendo em vista a sintaxe alterada do texto e as aspas. Nesse caso o sentido do item lexical *correto* é *adequado*. No entanto, pode-se concluir que, na maioria dos casos, a co-ocorrência do item lexical *correto* com a palavra *como* pode ser utilizada para realizar a desambiguação do mesmo.

Sentido 5: justo

N-gramas	Número de ocorrências encontradas no corpus NILC	Número de ocorrências encontradas na Web
Politicamente correto	204 ocorrências	501 ocorrências
Achar correto	20 ocorrências	145 ocorrências
Considerar correto	32 ocorrências	85 ocorrências
Administrativamente correto	1 ocorrência	3 ocorrências
Constitucionalmente correto	2 ocorrências	14 ocorrências
Princípio correto	3 ocorrências	44 ocorrências

Politicamente correto:

Foram encontradas 458 ocorrências de *politicamente correto* na busca realizada na WWW. Em todas as cinco ocorrências retiradas da concordância, o item lexical *correto*, assumiu o sentido de *justo*. Os exemplos em questão encontram-se listados abaixo.

1. Então precisa duplicar os cuidados com a história do [politicamente correto](#).
2. Porque nem tudo que é [politicamente correto](#) é necessariamente moralmente correto, tecnicamente correto, socialmente correto, cientificamente correto...
3. Então precisa quintuplicar os cuidados com o [politicamente correto](#).
4. Ao falar, seja em público ou para poucos interlocutores, saiba que ser [politicamente correto](#) não é sempre conversa de gente chata que não tem o que fazer na vida.
5. Dagoberto Dalsasso Publicitário - Floripa - SC Júlio, 21-02-07 - Maravilhoso e hilariante artigo sobre o "carnaval [politicamente correto](#)".

Pode-se concluir que, nesses casos, a co-ocorrência do item lexical *correto* com a palavra *politicamente* permite prever que o sentido adequado para realizar a desambiguação é o mesmo que foi inicialmente determinado a partir da análise das ocorrências encontradas no corpus NILC.

Achar correto:

Foram encontradas 135 ocorrências de *achar correto* na busca realizada na WWW. Nas três primeiras ocorrências retiradas da concordância, o item lexical *correto* assumiu o sentido de *justo*. No entanto, no exemplo 5 o sentido mais apropriado parece ser *isento de erros*. Os exemplos em questão encontram-se listados abaixo.

1. Eu disse para ele que a minha independência significa torcer para o governo dar certo e votar com o governo sempre que a minha consciência [achar correto](#).
2. Bom, se este é o caminho que a comunidade [achar correto](#) tomar, estarei aqui para apoiar no que for capaz.
3. Não há como [achar correto](#) alguém "roubar" o trabalho remunerado de alguém.
4. Cada um tem o direito de [achar correto](#) ou incorreto o que quiser, mas jamais de tirar oportunidade de outras pessoas.
5. Respeito, e não poderia ser diferente, a liberdade do jornal em editá-lo como [achar correto](#), mas também tenho o direito de reclamar da forma escolhida.

Tendo em vista a variação de sentidos mencionada acima, é esperado que nos demais exemplos encontrados via WebCorp sejam encontrados outros sentidos para o n-grama em questão. Isso sinaliza que esse talvez esse n-grama não seja adequado para realizar a desambiguação do item lexical *correto*. O ideal seria observar todas as 135 ocorrências encontradas na WWW para poder verificar se existe uma variação de sentidos muito grande para esses casos e buscar alternativas para a desambiguação dos mesmos.

Considerar correto:

Foram encontradas 83 ocorrências de *considerar correto* na busca realizada na WWW. Nos exemplos 1 e 2 o item lexical *correto*, assumiu o sentido de *justo*, como pode ser observado abaixo.

1. Sempre que [considerar correto](#) e necessário, o jornalista resguarda a origem e identidade das suas fontes de informações.

2. b) que o fisco, fazendo a verificação do pagamento na fiscalização, se o [considerar correto](#), o homologue e o tenha como realizado.

3. Mas o principal é que não se pode [considerar correto](#) alguém que quer passar sem bilhete e que bate num policial.

4. Novamente - quanto à expressão 'Colégio de cardeais e os Compatriotas': o 'Colégio de Cardeais' não tem nenhuma obrigação de consultar 'os Compatriotas' nessa questão; ele pode, em qualquer caso, fazê-lo, se por alguma razão [considerar correto](#). JP11.

5. Segundo os estudiosos, "não se pode [considerar correto](#) que uma tradução, sob o ponto de vista de gênero, coloque o feminino onde o texto fala em masculino.

No exemplo 3, o sentido mais apropriado seria *justo*. Já nos exemplos 4 e 5, o sentido mais adequado seria *isento de erros*. Pode-se concluir que, nesses casos, a ocorrência do item lexical *correto* com o verbo *considerar* não parece ser suficiente para realizar a desambiguação do item lexical em questão. Seria necessário, portanto, analisar para todas as ocorrências encontradas na WWW via WebCorp para tentar encontrar outros padrões de colocações que pudessem fazer a desambiguação desses casos.

Administrativamente correto:

Foram encontradas apenas 2 ocorrências de *administrativamente correto* na busca realizada na WWW via WebCorp. Nos exemplos encontrados o item lexical *correto* assumiu o sentido de *justo*, como pode ser observado abaixo.

1. Infelizmente as equipes do futebol sergipano não detêm uma estrutura de empresa e entra ano, sai ano, continuam a depender de alguns dirigentes abnegados, que muitas vezes sacrificam seus negócios pessoais em favor dos clubes que comandam. Além de não ser

[administrativamente correto](#), financeiramente as equipes acabam vivendo na berlinda o tempo todo.

2. Mas, temos que convir, nem sempre o politicamente correto é o [administrativamente correto](#)!

Pode-se concluir que, nesse caso, a co-ocorrência do item lexical *correto* com a palavra *administrativamente* permite prever que o sentido adequado para realizar a desambiguação é o mesmo que foi inicialmente determinado a partir da análise das ocorrências encontradas no corpus NILC.

Constitucionalmente correto:

Foram encontradas 16 ocorrências de *constitucionalmente correto* na busca realizada na WWW via WebCorp. Nos cinco primeiros exemplos encontrados o item lexical *correto* assumiu o sentido de *justo*.

1. Com a aplicação desse rito procura-se conseguir um resultado [constitucionalmente correto](#) através de um procedimento racional e controlado, criando, desse modo, certeza e previsibilidade jurídica.

2. Mas, se reclamada judicialmente, o [constitucionalmente correto](#) é proteger o cidadão que se sente usurpado em seus direitos fundamentais.

3. Por isso que o plebiscito é o instrumento [constitucionalmente correto](#), porque essa idéia nós temos que abolir aqui de querer diminuir o município de Ribamar para atender a uma meia dúzia de lideranças.

4. Se na legislação complementar à Constituição há súmula de efeito vinculante, mas a ela nunca foi dada esta alcunha, nada mais [constitucionalmente correto](#) que designar a nova súmula pelo seu status Constitucional.

5. Mandato Eletivo Procura-se conseguir um resultado [constitucionalmente correto](#) através de um procedimento racional e controlado, criando, desse modo, certeza e previsibilidade jurídicas.

Pode-se concluir que, nesse caso, a co-ocorrência do item lexical *correto* com a palavra *constitucionalmente* permite prever que o sentido adequado para realizar a desambiguação é o mesmo que foi inicialmente determinado a partir da análise das ocorrências encontradas no corpus NILC.

Princípio correto:

Foram encontradas 44 ocorrências de *princípio correto* na busca realizada na WWW via WebCorp. Nos cinco primeiros exemplos encontrados o item lexical *correto* assumiu o sentido de *justo*.

1. Todavia, nesse caso, meu pai fez algo que era escriturístico, a saber, ele me ensinou esse [princípio correto](#) (e que parece para muitos um princípio adulto) desde a minha tenra idade.
2. Trata-se de um [princípio correto](#) que contempla o bem público.
3. Se o imposto progressivo for instituído com o intuito de distribuir renda, é um [princípio correto](#)", disse ele, que terá de pagar R\$ 3.771,67 de IPTU.
4. Votar em candidatos locais para a Assembléia estadual e a Câmara federal é seguir um [princípio correto](#).
5. A verticalização tem um [princípio correto](#) a meu ver.

Pode-se concluir que, nesse caso, a co-ocorrência do item lexical *correto* com a palavra *princípio* permite prever que o sentido adequado para realizar a

desambiguação é o mesmo que foi inicialmente determinado a partir da análise das ocorrências encontradas no corpus NILC.

Sentido 6: aceitável

N-gramas	Número de ocorrências encontradas no corpus NILC	Número de ocorrências encontradas na Web
Apenas correto	8 ocorrências	135 ocorrências

Apenas correto:

O WebCorp selecionou 135 ocorrências de *apenas correto* na WWW. Assim como na maioria dos casos analisados até então, nas cinco primeiras ocorrências apresentadas na concordância, o item lexical *correto* assumiu o mesmo sentido definido a partir dos exemplos encontrados no corpus NILC, que nesse caso é o sentido *aceitável*. Os exemplos em questão encontram-se listados abaixo.

1. O visual é [apenas correto](#), com bom nível de detalhes, mas as texturas são genéricas.
2. Os locais Mad Dogs introduziram pitadas de diversão e humor alcoólico à sua apresentação, enquanto o Seu Zé decepcionou um pouco, com um show [apenas correto](#).
3. Clique aqui para ver a letra do samba 5B - JACAREZINHO – Samba sem sal, [apenas correto](#).
4. Acabamento: É [apenas correto](#) para um carro da categoria, ainda mais se levando em conta que é a versão sedã mais completa da linha Civic.
5. O Arcade Fire, incensado pela crítica (inclusive este repórter), realizou um show [apenas correto](#) e acabou "engolido" pelo Wilco, que driblou o som embolado do espaço para ganhar os corações e mentes do público majoritariamente indie.

Para esses casos, portanto, também é possível afirmar que a co-ocorrência do item lexical *correto* com a palavra *apenas* permite prever que o sentido adequado para realizar a desambiguação é o mesmo que foi previamente definido a partir da análise das ocorrências encontradas no corpus NILC.

Sentido 7: com a razão

N-gramas	Número de ocorrências encontradas no corpus NILC	Número de ocorrências encontradas na Web
Correto em	9 ocorrências	226 ocorrências
Correto ao	7 ocorrências	138 ocorrências

Correto em:

Foram encontradas 226 ocorrências de *correto em* na busca realizada na WWW via WebCorp. Nos cinco primeiros exemplos encontrados o item lexical *correto* assumiu o sentido de *com a razão*, assim como nos exemplos encontrados previamente no corpus NILC. Esses exemplos encontram-se listados abaixo:

1. O diretor do BC está [correto em](#) afirmar que o eixo do crescimento se desloca agora da poupança forçada (imposto inflacionário e fundos compulsórios) e dos investimentos das estatais para a poupança privada voluntária e os investimentos privados.
2. Realmente, o Elton meu colega de sala, está [correto em](#) sua resposta com certeza prestou mais atenção na aula de química do que eu.
3. O Orkut é, portanto, a experiência prática e a prova cabal de como Marshall McLuhan, no longínquo ano de 1964, estava [correto em](#) suas afirmações.
4. Nassif está [correto em](#) dizer que a privatização é um meio, e não um fim.

5. EDITORIAL O ministério público está [correto em](#) promover a ofensiva.

Pode-se concluir que, nesse caso, o n-grama *correto em* pode ser utilizado para desambiguar o item lexical *correto* em determinados contextos.

Correto ao:

Foram encontradas 138 ocorrências de *correto ao* na busca realizada na WWW via WebCorp. Nos exemplos 1, 2, 4 e 5, o item lexical *correto* assumiu o sentido de *com a razão*. No exemplo 3, o sentido mais apropriado seria *isento de erros*.

1. Para o representante da sociedade civil no Consea, Altermir Tortelli, o governo está [correto ao](#) tratar a questão dos biocombustíveis como elemento estratégico, no momento em que o mundo atravessa uma crise energética e ambiental.

2. Eu estava [correto ao](#) perceber pelo estilo do texto que foi nosso amigo quem escreveu a última matéria sobre o Robinho

3. Temos assim, um html [correto ao](#) final deste passo.

4. ASSUNTOS INTERNACIONAIS Para Pont, Lula está [correto ao](#) não agir como Bush no caso da Bolívia.

5. Com agências. Análise da notícia O governo está [correto ao](#) conferir ao tripé habitação/saneamento/infra-estrutura o status de “espinha dorsal” do pacote que está sendo elaborado pela área econômica.

Pode-se concluir que, apesar de uma das ocorrências o item lexical não ter assumido o sentido determinado previamente no corpus NILC, esse padrão ainda assim pode ser utilizado para realizar a desambiguação da maioria dos casos. É importante acrescentar que o bigrama *correto ao* funcionaria perfeitamente se fosse tratado como

trigrama **correto ao V-INF** (verbo no infinitivo). Na ocorrência 3 isso não funcionaria devido ao fato de que *ao final desse passo* é um sintagma preposicional e não uma oração reduzida de infinitivo.

Sentido 8: verdade

N-gramas	Número de ocorrências encontradas no corpus NILC	Número de ocorrências encontradas na Web
Seria correto	8 ocorrências	163 ocorrências

Seria correto:

O WebCorp selecionou 193 ocorrências de *seria correto* na WWW. Nas três primeiras ocorrências apresentadas na concordância, o item lexical *correto* não assumiu o mesmo sentido definido a partir dos exemplos encontrados no corpus NILC (*verdade*). O sentido mais apropriado para esses casos seria *justo*.

1. Se a Terceira Mensagem falar explicitamente do advento da 3ª Guerra Mundial, [seria correto](#) esconder dos povos esta previsão?
2. Os médicos chamam as pessoas em coma, já a muito tempo, de "estado vegetativo", mas [seria correto](#) chamar alguém, de certa forma, de "vegetal"?
3. Partindo deste princípio [seria correto](#) eleger o Led Zeppelin como o patrono do Metal Melódico, Power Metal e Epic Metal :av2: ???
4. Isto [seria correto](#) se a área correntemente irrigada tivesse seu suprimento garantido em qualquer situação hidrológica, mesmo durante as estiagens mais intensas.
5. Acho que, neste caso, não [seria correto](#) falar em fraude", afirmou o ex-senador chileno José Antonio Viera Gallo.

No entanto, nos exemplos 4 e 5, o sentido mais apropriado seria *verdade*. Como nesse caso houve uma variação de sentidos, o mais adequado seria fazer uma análise de todos os padrões encontrados através da busca do WebCorp. Só assim seria possível fazer afirmações mais concretas a respeito desse n-grama.

Sentido 9:

N-gramas	Número de ocorrências encontradas no corpus NILC	Número de ocorrências encontradas na Web
Correto,	8 ocorrências	105 ocorrências
Correto.	9 ocorrências	104 ocorrências

Não foi possível fazer uma verificação dos marcadores conversacionais utilizando o WebCorp porque o site de busca não é capaz de reconhecer a pontuação na hora de gerar as concordâncias. Ao digitar “*Correto,*” e “*Correto.*”, o programa desconsidera a vírgula e o ponto final fazendo a busca do item lexical isolado. Devido à falta de tempo já enfatizada, não foi possível buscar individualmente esses casos na WWW.

Sentido 10: honesto

N-gramas	Número de ocorrências encontradas no corpus NILC	Número de ocorrências encontradas na Web
Dinheiro correto	8 ocorrências	23 ocorrências
Funcionário correto	2 ocorrências	19 ocorrências

Homem correto	3 ocorrências	148 ocorrências
Jornalismo correto	3 ocorrências	38 ocorrências
Juiz correto	6 ocorrências	20 ocorrências
Político correto	7 ocorrências	78 ocorrências
Profissional correto	3 ocorrências	130 ocorrências
Muito correto	5 ocorrências	187 ocorrências

Dinheiro correto:

Foram encontradas 23 ocorrências de *dinheiro correto* na busca realizada na WWW via WebCorp. Nos quatro primeiros exemplos encontrados o item lexical *correto* assumiu o sentido de *honesto*, assim como nos exemplos encontrados previamente no corpus NILC. Esses exemplos encontram-se listados abaixo:

1. Nesta maneira o cliente de dinheiro associado mostra que o contrapeso de [dinheiro correto](#) e a ligação de coupon zero mostram a base correta do custo.
2. Você trabalha e eu ganho [dinheiro correto](#)?
3. Se, no jogo dos interesses materiais, quando tentações numerosas te induzem a trapacear, em nome da inteligência, com vantagens pessoais, a tua religião te mostra o caminho do [dinheiro correto](#), sem afastar-te do suor no trabalho e da responsabilidade no esforço próprio...
4. Você quer economizar tempo e [dinheiro correto](#)?" ou "Você não concorda que nossas soluções ira ajudá-lo (a), a economizar tempo e dinheiro?"
5. Estendido ATIVAMENTE agora e nós procuramos as caras novas as fotografias da amostra que de Deine (carcaça) são LIVRES e ANONYMOUS e DISCRETAS! E você faz o [dinheiro correto](#).

Apenas no exemplo 4, o sentido de *dinheiro correto* não pode ser considerado *honesto*, pois, na verdade, parece tratar-se de uma ocorrência de “...*dinheiro, correto?*”, apesar de não haver a vírgula no texto original. Mesmo assim, esse caso deveria ser tratado como um marcador de discurso.

Para esses casos, portanto, também é possível afirmar que a co-ocorrência do item lexical *correto* com a palavra *dinheiro* permite prever que o sentido adequado para realizar a desambiguação é o mesmo que foi previamente definido a partir da análise das ocorrências encontradas no corpus NILC.

Funcionário correto:

Foram encontradas 19 ocorrências de *funcionário correto* na busca realizada na WWW via WebCorp. Nos cinco primeiros exemplos encontrados o item lexical *correto* assumiu o sentido de *honesto*, assim como nos exemplos encontrados previamente no corpus NILC. Esses exemplos encontram-se listados abaixo:

1. O [funcionário correto](#) merece o maior respeito por parte de todos.
2. E por que só o futebol, se Kofi Annan, na ONU, é o exemplo do [funcionário correto](#)?
3. Um magistrado sem nome toca adiante sua rotina de [funcionário correto](#) a serviço de uma ordem que não lhe cabe questionar.
4. Vou ser um [funcionário correto](#), leal e trabalhador até o último dia.
5. Eu vou continuar sendo um [funcionário correto](#) e trabalhador e, se isso não for suficiente, eles que tomem decisões”, emendou o técnico, que viu a pressão aumentar sobre as suas costas.

Nesses casos é possível afirmar que a co-ocorrência do item lexical *correto* com a palavra *funcionário* parece ser suficiente para realizar a desambiguação do mesmo.

Homem correto:

Foram encontradas 148 ocorrências de *homem correto* na busca realizada na WWW. Nos cinco primeiros exemplos encontrados o item lexical *correto* assumiu o sentido de *honesto*, assim como nos exemplos encontrados previamente no corpus NILC, como pode ser observado nos exemplos abaixo:

1. Não devo nada à Justiça, sou um [homem correto](#), porém, pelas milhares de histórias que ouvimos no noticiário, a abordagem por policiais nos dá medo.
2. Ao contrário, considero-o um [homem correto](#) e digno.
3. O procurador-geral de Justiça, Francisco das Chagas Barros de Sousa, disse que confia no trabalho do novo desembargador, lembrando que ele é um [homem correto](#) e desejou que seu trabalho seja voltado para o fortalecimento da democracia.
4. O assassinato do prefeito, que era [homem correto](#) mas aparentemente mal acompanhado, foi crime comum.
5. Um [homem correto](#), rico, famoso, amigo sincero, pai dedicado, marido fiel...

Para esses casos, portanto, também é possível afirmar que a co-ocorrência do item lexical *correto* com a palavra *homem* permite prever que o sentido adequado para realizar a desambiguação é o mesmo que foi previamente definido a partir da análise das ocorrências encontradas no corpus NILC.

Jornalismo correto:

Foram encontradas 38 ocorrências de *jornalismo correto* na busca realizada na WWW. Nos cinco primeiros exemplos encontrados o item lexical *correto* assumiu o sentido de *honesto*, como pode ser observado nos exemplos abaixo:

1. Nosso público é exigente, quer a melhor informação, o [jornalismo correto](#) e analítico.
2. Meu pai sempre me ensinou o [jornalismo correto](#), a serviço do leitor", comentou ela, que, além de filha, foi aluna dele.
3. Fazendo um [jornalismo correto](#), de bom senso, de equilíbrio, a gente consegue melhorar a sociedade.
4. 12 de Agosto de 2006 @ 19:33 Por que Esses senhores não fazem apenas o [jornalismo correto](#) e imparcial que aprenderam em suas faculdades de Jornalismo.
5. É assim que se faz o [jornalismo correto](#).

Para esses casos, portanto, também é possível afirmar que a co-ocorrência do item lexical *correto* com a palavra *jornalismo* permite prever que o sentido adequado para realizar a desambiguação é o mesmo que foi previamente definido a partir da análise das ocorrências encontradas no corpus NILC.

Juiz correto:

Foram encontradas 20 ocorrências de *juiz correto* na busca realizada na WWW. Nos cinco primeiros exemplos encontrados e analisados o item lexical *correto* assumiu o sentido de *honesto*, como pode ser observado nos exemplos abaixo:

1. Registre-se que conhecemos o Juiz Flávio Batista Leite há muito anos e que se trata de [juiz correto](#), cumpridor dos seus deveres.

2. Assim, não há [juiz correto](#) que lhe dê ganho de causa.
3. Segundo Requião, Camargo é seu amigo pessoal e “é sempre bom saber que um [juiz correto](#) galga o posto mais alto na Organização Judiciária do Paraná”.
4. Achei aquela observação correta porque o [juiz correto](#), que trabalha, trabalha de madrugada, nos fins de semana, não tem direito a hora extra.
5. O Corinthians só seria beneficiado com a remarcação das partidas se vencesse com um [juiz correto](#) apitando e isso aconteceu.

Para esses casos, portanto, também é possível afirmar que a co-ocorrência do item lexical *correto* com a palavra *juiz* permite a desambiguação e confirma a informação que foi previamente definida a partir da análise das ocorrências encontradas no corpus NILC.

Político correto:

Foram encontradas 78 ocorrências de *político correto* na busca realizada na WWW. Nos três primeiros exemplos da concordância, quando o N-grama *político correto* aparece precedido de um artigo, o item lexical *correto* assumiu o sentido de *honesto*, como pode ser observado abaixo:

1. Aquele que passa a vida achando que um [político correto](#) tem que ser do povão, que tem que comer farofa com eles, e cumprimentar todo mundo, ou aquele que sabe que isso é besteira e vota num populista demagogo dessa categoria, e defende sua permanência fechando os olhos pra todas as provas de que este corrompeu os cofres públicos?
2. Se fosse tão simples escolher o [político correto](#), aberrações não seriam mantidas no poder por décadas.

3. Infelizmente, nosso povo ainda precisa saber o que é uma eleição democrática e diferenciar [o político correto](#) do mentiroso.

4. É um posicionamento [político correto](#) que eles assumem nessas horas

5. Com o ambiente [político correto](#), o setor de negócios apresentará tecnologias e serviços com baixas emissões mais rapidamente do que muitos agora acreditam possível.

No entanto, nos exemplos 4 e 5, o sentido do item lexical poderia ser *justo* ou até mesmo *isento de erros*. Para esses casos, portanto, é possível afirmar que a co-ocorrência do item lexical *correto* com a palavra *político* só assume o sentido de *honesto* se estiver precedida de um artigo.

Concluindo, *político* é um substantivo nos três casos em que o sentido é *honesto*, e um adjetivo naqueles em que é diferente de *honesto*. Portanto, um etiquetador de classe de palavras, geralmente uma etapa que antecede a desambiguação lexical no processamento computacional, resolveria o problema com a especificação do bigrama como **político_SUBS correto** = honesto; **político_ADJ correto** = outro sentido a ser especificado.

Profissional correto:

Foram encontradas 130 ocorrências de *profissional correto* na busca realizada na WWW via WebCorp. Nos cinco primeiros exemplos observados na concordância, o item lexical *correto* assumiu o sentido de *honesto*, como pode ser observado nos exemplos abaixo:

1. Segundo Gil, "um bom gestor é aquele capaz de reconhecer os problemas e indicar o método e o [profissional correto](#) capaz de sanar a questão".

2. Um [profissional correto](#), meio-campista técnico e inteligente, esteve nas Copas do Mundo de 1986 e 1990, jogou em Portugal, na França e no Japão.
3. Para jogar no meu time, o jogador deve ser um [profissional correto](#), obediente taticamente", avisou.
4. Foi lamentável ver um [profissional correto](#) como o Viera ser hostilizado daquela maneira.
5. Um [profissional correto](#) que amava a profissão que abraçou.

Para esses casos, portanto, também é possível afirmar que a co-ocorrência do item lexical *correto* com a palavra *profissional* permite a desambiguação e confirma a informação que foi previamente definida a partir da análise das ocorrências encontradas no corpus NILC.

Muito correto:

Foram encontradas 187 ocorrências de *muito correto* na busca realizada na WWW via WebCorp. Nos exemplos 2 e 3 observados na concordância, o item lexical *correto* assumiu o sentido de *honesto*. Os exemplos em questão podem ser observados abaixo:

1. Tenho meus momentos de ser [muito correto](#) moralmente, como o Bernardo, e outros de voracidade e ousadia, como ficou flagrante no ensaio.
2. A idéia de haver uma espécie de Fifa do joquempô pode parecer ridícula, a princípio, mas o site é [muito correto](#), muito eficiente, muito profissional.
3. Ivanete conhecida por 'cruela magrela' em parceria com empresário [muito correto](#), teria denunciado o ex-chefe do executivo estadual pela pratica de atos lesivos as cofres públicos.

4. Chimango é uma excelente alternativa na criação, um cão [muito correto](#), de excelentes apurmos, excelentes proporções, masculino, forte, de correto tamanho (grande mas não em demasia), muito boa movimentação em todos os passos.

5. Sex Jun 10, 2005 1:11 am Oi Edu, Sou Alemão, assim que me desculpa se o português não for [muito correto](#), mas eu acho que lhe posso ajudar um pouco.

Já no exemplo 1 e 4, o sentido mais apropriado parece ser *adequado*. No exemplo 5, o sentido do item lexical em questão é, claramente, *isento de erros*. Como foram encontradas 138 ocorrências, é esperado que haja variações de sentido. Provavelmente *muito*, pela sua característica de intensificador, não serve para a desambiguação por si só. Outros elementos teriam que ser acrescentados e seria necessário olhar para todos os exemplos encontrados na WWW para poder fazer afirmações mais precisas sobre esse caso.

Sentido 11: ideal

N-gramas	Número de ocorrências encontradas no corpus NILC	Número de ocorrências encontradas na Web
Aprendizado correto	3 ocorrências	94 ocorrências
Desfecho correto	4 ocorrências	13 ocorrências
Papel correto	2 ocorrências	108 ocorrências
Correto dimensionamento	2 ocorrências	178 ocorrências

Aprendizado correto:

Foram encontradas 94 ocorrências de *aprendizado correto* na busca realizada na WWW via WebCorp. Nos cinco primeiros exemplos observados na concordância, o item lexical *correto* assumiu o sentido de *ideal*, como pode ser observado nos exemplos abaixo:

1. PARA JOGAR BEM Obra dirigida aos principiantes e amadores que buscam aprender ou aprimorar a forma de jogar, abrangendo todos os pontos vitais para um [aprendizado correto](#) e lapidação da técnica empregada pelos amadores.
2. Em razão disto, coloca em nós a vontade de pensar, refletir, meditar, pois não há nada melhor para um [aprendizado correto](#) e para uma comunicação perfeita do que a introspecção.
3. Nesta idade a criança possui a cartilagem da garganta ainda bastante flexível, o que facilita muito o [aprendizado correto](#) de vários idiomas, podendo até mesmo tornar-se poliglota.
4. Estou meio perdido nesse novo mundo e gostaria de ajuda para orientar-me na direção de um [aprendizado correto](#) e dinâmico.
5. O [aprendizado correto](#) de técnicas da mão direita, com diversas modalidades de palhetada (swiping, speed picking, clássica, etc..) além de técnicas de dedilhado, constituem numa das mais importantes disciplinas para se tocar bem Violão & Guitarra - é bom ressaltar que estamos falando da "mão direita", portanto, da parte mais difícil numa execução musical.

Para esses casos, portanto, também é possível afirmar que a co-ocorrência do item lexical *correto* com a palavra *aprendizado* permite a desambiguação e confirma a informação que foi previamente obtida a partir da análise das ocorrências encontradas no corpus NILC.

Desfecho correto:

Foram encontradas 13 ocorrências de *desfecho correto* na busca realizada na WWW. Nos cinco primeiros exemplos observados na concordância, o item lexical *correto* assumiu o sentido de *ideal*, como pode ser observado nos exemplos abaixo:

1. Diego Rambo. Ainda estou procurando saber o [desfecho correto](#) do filme...
2. Perscrutando as peças processuais, reconheço não haver necessidade de retocar a sentença monocrática, pois o julgador singular, com muita propriedade apontou um [desfecho correto](#) para o caso em comento, a ponto de o próprio atuante acatá-la.
3. O diretor Bronwen Hughes não é Deus, mas conduz a trama por linhas tortas até chegar num [desfecho correto](#) e interessante.
4. Diante disso, não pode mais sofrer adiamentos e postergações, sob pena de frustrar o justo clamor daqueles que aguardam um [desfecho correto](#) para esse processo onde não deve pairar a menor sombra de dúvida sobre a independência do Poder Judiciário.
5. Ontem fui ver Matrix Revolutions, e eu gostei. Não responde tudo que todo mundo queria saber, mas também não demais no ar. Foi um [desfecho "correto"](#)...

Para esses casos, portanto, também é possível afirmar que a co-ocorrência do item lexical *correto* com a palavra *desfecho* permite a desambiguação e confirma a informação que foi previamente obtida a partir da análise das ocorrências encontradas no corpus NILC.

Papel correto:

Foram encontradas 108 ocorrências de *papel correto* na busca realizada na WWW via WebCorp. Nos cinco primeiros exemplos observados na concordância, o item lexical *correto* assumiu o sentido de *ideal*, como pode ser observado nos exemplos abaixo:

1. Hugo Chávez nos embates do presidente venezuelano com a oposição, a diplomacia brasileira assumiu o [papel correto](#) de intermediária entre o populista Chávez e desafetos internos e externos.
2. Se você definir o personalizado tamanho página para o sistema como um todo, Visio usa o tamanho de [papel correto](#).
3. Cores vivas e alto rendimento para seus clientes A HP Color LaserJet série 4700 vem equipada com um sensor especial, que seleciona o [papel correto](#) a ser usado nas bandejas e escolhe automaticamente a configuração de impressão ideal para o tipo de papel escolhido.
4. Silverstein acreditou que os trabalhos escritos necessitaram ser lidos no papel, e com o [papel correto](#) para o trabalho.
5. Como concessionária de um serviço público essencial, deve cumprir seu [papel correto](#) junto à população, seja ela pobre ou não.

Para esses casos, portanto, também é possível afirmar que a co-ocorrência do item lexical *correto* com a palavra *papel* permite a desambiguação e confirma a informação obtida a partir da análise das ocorrências encontradas no corpus NILC.

Correto dimensionamento:

Foram encontradas 178 ocorrências de *correto dimensionamento* na busca realizada na WWW via WebCorp. Nos cinco primeiros exemplos observados na

concordância, o item lexical *correto* assumiu o sentido de *ideal*, como pode ser observado nos exemplos abaixo:

1. A otimização do rendimento do equipamento inicia com o conhecimento de suas limitações, o seu [correto dimensionamento](#) (área filtrante), a mais econômica e eficiente especificação das mangas e a correta operação do filtro.

2. No caso de compressores, a eficiência da compressão e a vida útil do equipamento estão intimamente relacionadas ao [correto dimensionamento](#) e à confiabilidade das válvulas.

3. Para o [correto dimensionamento](#) do condicionador de ar é fundamental considerar o tamanho do ambiente, o seu isolamento, a sua exposição ao sol ou sombra e o número de pessoas que freqüentarão o ambiente.

4. Somente a partir do [correto dimensionamento](#) do déficit previdenciário seria possível conhecer seu efetivo déficit, visando a um estudo atuarial capaz de indicar o esforço financeiro a longo prazo (...).

5. O [correto dimensionamento](#) de um sistema de alívio de ar deve levar em conta além do modelo do moinho, a área perfurada de tela, o tamanho da partícula moída e o sistema utilizado para descarga do produto.

Para esses casos, portanto, também é possível afirmar que a co-ocorrência do item lexical *correto* com a palavra *dimensionamento* permite a desambiguação e confirma a informação obtida a partir da análise das ocorrências encontradas no corpus NILC.

Sentido 12: bem feito

N-gramas	Número de ocorrências encontradas no corpus NILC	Número de ocorrências encontradas na Web
Acabamento correto	2 ocorrências	23 ocorrências
Desenho correto	5 ocorrências	64 ocorrências
Desfile correto	2 ocorrências	83 ocorrências
Show correto	3 ocorrências	93 ocorrências
Trabalho correto	4 ocorrências	138 ocorrências

Acabamento correto:

Foram encontradas 23 ocorrências de *acabamento correto* na busca realizada na WWW via WebCorp. Nos cinco primeiros exemplos observados na concordância, o item lexical *correto* assumiu o sentido de *bem feito*, como pode ser observado nos exemplos abaixo:

1. Esta posição fica por conta de um veículo bem equilibrado, [acabamento correto](#), com desempenho de bom tamanho e que atende plenamente os seus usuários, apesar do preço de R\$ 65.630,00 para o sedan.
2. Para iniciantes e amadores de todas as modalidades, selas nacionais de qualidade acabam por ser a melhor escolha, aliando bom artesanato e [acabamento correto](#) a um preço intermediário.
3. Desta forma, vai estar sempre com tracinhos verticais no verso, que dão o [acabamento correto](#).

4. Cada caso é um caso e o [acabamento correto](#), bem seco e bem feito é fundamental na composição de um produto eficiente.

5. Examine a folga entre a agulha e a lançadeira Linha de má qualidade Utilize uma linha de qualidade melhor, de [acabamento correto](#).

Para esses casos, portanto, também é possível afirmar que a co-ocorrência do item lexical *correto* com a palavra *acabamento* permite a desambiguação e confirma a informação obtida a partir da análise das ocorrências encontradas no corpus NILC.

Desenho correto:

Foram encontradas 64 ocorrências de *desenho correto* na busca realizada na WWW via WebCorp. Nos cinco primeiros exemplos observados na concordância, o item lexical *correto* assumiu o sentido de *bem feito*, como pode ser observado nos exemplos abaixo:

1. A pintura de Cicarelli caracteriza-se por um [desenho correto](#) mas desprovido de sensibilidade e por um colorido pálido, sem vida.

2. A crítica elogiava-lhe o [desenho correto](#), o colorido sóbrio, cheio de verdade.

3. Mesmo com [desenho correto](#) do bolsa alimentação, onde a renovação do benefício não estava condicionada ao crescimento e ganho de peso da criança, o incentivo perverso do programa anterior

4. São vários fatores que determinam a durabilidade dos pneus reconstruídos entre eles a calibragem correta, o alinhamento, o emparelhamento, o [desenho correto](#) para o segmento e a quantidade de carga não ultrapasse os limite indicados nos pneus.

5. A composição é bem cuidada e o [desenho correto](#), mas algumas figuras são duras e despojadas de vida - bonecos, e não seres de carne e osso. ...

Para esses casos, portanto, também é possível afirmar que a co-ocorrência do item lexical *correto* com a palavra *desenho* permite a desambiguação e confirma a informação obtida a partir da análise das ocorrências encontradas no corpus NILC.

Desfile correto:

Foram encontradas 83 ocorrências de *desfile correto* na busca realizada na WWW via WebCorp. Nos cinco primeiros exemplos observados na concordância, o item lexical *correto* assumiu o sentido de *bem feito*, como pode ser observado nos exemplos abaixo:

1. A Mocidade, última do Grupo Especial a se apresentar, fez um [desfile correto](#) sobre a prevenção de acidentes de trânsito.
2. Mas segundo o presidente da Vila, Olícia dos Santos, a agremiação fez um [desfile correto](#), estilo "Kizomba".
3. Painéis contaram a história com imagens da marca e serviram como cenário para um [desfile correto](#) e enxuto.
4. É um enredo que permite que a escola faça um [desfile correto](#) e brigue pelo título.
5. Salgueiro na avenida, tornando o enredo folhetinesco, materializado com soluções conhecidas, carecendo de ousadia e criatividade, apresentando um [desfile correto](#), porém linear e previsível.

Para esses casos, portanto, também é possível afirmar que a co-ocorrência do item lexical *correto* com a palavra *desfile* permite a desambiguação e confirma a informação obtida a partir da análise das ocorrências encontradas no corpus NILC.

Show correto:

Foram encontradas 93 ocorrências de *show correto* na busca realizada na WWW via WebCorp. Nos cinco primeiros exemplos observados na concordância, o item lexical *correto* assumiu o sentido de *bem feito*. Os exemplos encontram-se listados abaixo:

1. A brincando de deus, uma das bandas mais respeitadas no underground nacional, fez um [show correto](#).
2. O Black Morro fez um [show correto](#), sem erros, com uma boa disposição no palco e uma excelente qualidade instrumental.
3. Fizeram um [show correto](#), mas àquela altura da maratona, pegaram um público cansado, que já deixava o circo em bandos. Afinal, domingo tinha mais.
4. Foi um [show correto](#) que agradou aqueles que ainda não os conhecia ao vivo.
5. Gárgula subiu ao palco às 16 horas e mostrou muita eficiência, fazendo um [show correto](#) e com muito feeling.

Para esses casos, portanto, também é possível afirmar que a co-ocorrência do item lexical *correto* com a palavra *show* permite a desambiguação e confirma a informação obtida a partir da análise das ocorrências encontradas no corpus NILC.

Trabalho correto:

Foram encontradas 138 ocorrências de *trabalho correto* na busca realizada na WWW via WebCorp. Nos cinco primeiros exemplos observados na concordância, o item lexical *correto* assumiu sistematicamente o sentido de *bem feito*, como pode ser observado nos exemplos abaixo:

1. SKF tem as pessoas certas para fazer o [trabalho correto](#).
2. O Brasil sempre esteve muito atrasado, mas hoje apesar do nível e do aumento de cães estar baixo, estamos nos primeiros passos de um [trabalho correto](#), e acredito que brevemente seremos competitivos na Europa.
3. A gente trabalha sob muita pressão da Justiça e temos que fazer um [trabalho correto](#).
4. Um [trabalho correto](#) do fortalecimento muscular abdominal, melhora o peristaltismo (digestão) e contra as cólicas menstruais.
5. Ele foi contratado pelo time pelo seu [trabalho correto](#) no seu time anterior.

Para esses casos, portanto, também é possível afirmar que a co-ocorrência do item lexical *correto* com a palavra *trabalho* permite a desambiguação e confirma a informação obtida a partir da análise das ocorrências encontradas no corpus NILC.

CONCLUINDO O CAPÍTULO

Foi constatado que, na maioria dos casos, os n-gramas encontrados a partir da busca via WebCorp apresentavam o mesmo sentido que as ocorrências dos bigramas do corpus NILC. Portanto pode-se concluir que os n-gramas definidos a partir do corpus NILC podem ser usados para desambiguar o item lexical *correto* satisfatoriamente.

É importante destacar que há um grau de arbitrariedade inevitável na vinculação de bigramas a sentidos únicos, uma vez que a polissemia não é apenas a possibilidade de uma palavra ter mais de um sentido, desambiguado pelo contexto, mas também a possibilidade de uma palavra, a despeito do contexto, reter um grau de polissemia, que nem mesmo seres humanos são capazes de eliminar integralmente. Parece, portanto, razoável optar por uma abordagem que especifique um determinado sentido que possa ser pelo menos considerado como correto por um número aceitável de seres humanos.

Também é importante ressaltar que esse modelo de processamento poderia ser aprimorado e revisado, o que não foi possível devido às limitações de prazo e recursos. O ideal seria fazer uma verificação plena de todas as colocações encontradas no WebCorp para aperfeiçoar o perfil das colocações e desenvolver um sistema mais sofisticado e detalhado. Para isso, seria necessário ter uma equipe de pesquisadores que pudessem trabalhar sistematicamente fazendo uma análise aprofundada de todas as ocorrências do item lexical polissêmico em questão.

No entanto, a utilização desta metodologia foi testada em um número limitado de colocações consideradas mais importantes, e é um aspecto importante da especificação do modelo de processamento. Finalizando, pode-se afirmar que o objetivo desse trabalho foi alcançado, já que o conjunto de n-gramas inicial foi definido e pode

ser facilmente utilizado por um etiquetador semântico dentro de um sistema que utiliza tecnologia de análise lingüística.

CONCLUSÃO

Disambiguation is possible. (WordSmith)

Todas as ocorrências do item lexical *correto* encontradas no corpus foram analisadas para definir os sentidos que elas assumiam dentro do contexto em que estavam inseridas. As definições destas estratégias de processamento constituem, portanto, hipóteses que pretendem captar o que acontece no processo de compreensão textual no que diz respeito à identificação dos significados desta palavra polissêmica.

Como já foi mencionado na seção anterior, foram especificados doze grupos de sentidos, tendo em vista o tipo de informação fundamentalmente utilizada, segundo a interpretação do analista, para resolver a polissemia do item lexical em questão.

A análise inicial deste item lexical de polissemia complexa revelou estruturas que precisam necessariamente da inteligência humana para ser desvendadas. Em muitos momentos, foi difícil definir qual o sinônimo mais adequado para definir o sentido do item lexical em determinadas ocorrências. Alguns grupos de significados como *adequado*, por exemplo, parecem ter uma grande flexibilidade e geram margem para questionamentos. Por exemplo, na ocorrência abaixo, o sentido assumido pelo item lexical *correto* foi definido como *justo*, mas poderia também ser substituído por *adequado* sem grande prejuízo para o conteúdo semântico da sentença.

Essa deve ser a diferença de preço do produto no Brasil, mas Carrara diz que ainda não dá para fixar um **preço correto**.

Talvez isso aconteça porque “a sinonímia é uma questão de grau e que todo conjunto de unidades lexicais pode ser organizado numa escala de semelhança e de diferença de sentido” (Lyons, 1979).

Essa aparente flexibilidade faz com que os grupos de sentidos sejam considerados de certa maneira frágeis e facilmente contestáveis, o que leva o analista a perceber que o conceito de sinonímia é útil no processo de desambiguação, mas, por não ser exatamente um conceito preciso, pode trazer problemas na definição das etiquetas.

Os padrões de uma palavra podem ser definidos como todas as palavras e estruturas com as quais são regularmente associadas e que contribuam para seu significado. Um padrão pode ser identificado se uma combinação de palavras ocorre com relativa frequência, se é dependente de uma palavra específica, e se há um significado claro associado. (Berber Sardinha, 2004).

A área que se ocupa da descrição de padrões se preocupa com questões como descobrir quais os padrões lexicais dos quais a palavra faz parte; investigar se a palavra se associa regularmente com outros sentidos específicos; determinar em quais estruturas ela aparece e estabelecer se há uma correlação entre o uso/sentido da palavra e as estruturas das quais ela participa e verificar se a mesma está associada com uma certa posição organizacional no texto (Berber Sardinha, 2004).

Após esta observação detalhada dos casos do item lexical *correto* encontrados no corpus, podemos concluir que existe uma variação significativa na utilização desta palavra na língua portuguesa, ou seja, ele pode realmente ser definido como um item lexical de polissemia complexa. Também foi observado que, na maioria dos casos, o quadro de polissemia pode ser resolvido a partir do princípio colocacional,

ou seja, a partir de padrões regulares desse item lexical com outras palavras em um contexto imediato, tanto à esquerda quanto à direita do mesmo.

Pode-se dizer que o famoso slogan de Firth “diga-me com quem a palavra anda e eu te direi quem ela é” tem sua validade, mas não é suficiente para explicar o fenômeno complexo da polissemia. Apesar de ter ficado claro que o sentido de um determinado item lexical é, sim, construído a partir de sua colocação com outras palavras, parece que existem outros elementos que permeiam o contexto e nem sempre podem ser determinados apenas pelo princípio colocacional.

Já ficou evidente que seria necessário fazer um estudo mais aprofundado e detalhado desse item lexical nos casos em que ele é precedido por formas do verbo *ser*, já que esse foi um padrão presente em praticamente todos os grupos de sentido.

Nos casos em que não ficou evidente nenhuma possibilidade imediata de desambiguação via colocação, seria necessário fazer um estudo aprofundado do contexto e procurar sofisticar a abordagem para encontrar novas soluções que venham a solucionar o problema da polissemia e sugerir possíveis maneiras de desambiguar esses sentidos.

Portanto, este trabalho deixa em aberto a questão de como refinar os padrões das etiquetas que foram esboçadas e encontrar soluções para os casos que não puderam em princípio ser resolvidos através do princípio colocacional.

BIBLIOGRAFIA

- AARTS, J. e MEIJS, L. (1984) *Corpus Linguistics*. Amsterdam: Rodopi.
- BERBER SARDINHA, T. B., (2004) *Linguística de Corpus*. Editora Manole Ltda.
- BORBA, (2004) “*Dicionário UNESP do Português Contemporâneo*”. Editora UNESP.
- FRANCIS, N. (1992) Language corpora B.C. IN: *Directions in Corpus Linguistics. Proceedings of Nobel Symposium*.
- GARSIDE, R.; LEECH, J. e MCENERY, T. (1997) *Corpus Annotation*. Longman: London & New York.
- LYONS, J. (1979) *Introdução à Linguística Teórica*. Companhia Editora Nacional: São Paulo, SP.
- MARRAFA, P. (2001) *WordNet do Português: uma base de dados de conhecimento linguístico*. Instituto Camões.
- MIHALCEA, R e EDMONDS, P. Proceedings of SENSEVAL - 3. Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text (2004).
- MILLER, G. A. (1993) *Introduction to WordNet: An On-Line Lexical Database*
- NILC (Núcleo Institucional de Linguística Computacional).
- PARTINGTON, A. (1998) *Patterns and Meanings. Using Corpora for English Language Research and Teaching*. John Benjamins Publishing Company: Amsterdam/Philadelphia.
- PEDERSEN, T. (2001) *A Decision Tree of Bigrams is an Accurate Predictor of Word Sense*.
- PREISS, J. e YAROWSKY, D. Proceedings of SENSEVAL - 2. Second International Workshop on Evaluating Word Sense Disambiguation Systems (2001).
- PUSTEJOVSKY, J e Boguraev, B. *Lexical Semantics: The problem of Polysemy*, Oxford University Press, 1996.
- ROCHA, M. E., (2001) “O uso de corpora computadorizados no ensino de língua portuguesa: metodologia e avaliação”, In. Grimm Cabral, L. et all (orgs). *Linguística e ensino: novas tecnologias*, Blumenal: Nova Letra, 2001.
- SAMPSON, G. (1996) *Evolutionary Language Understanding*.

SCOTT, M. *WordSmith Tools*, Oxford University Press, 1995.

SILVA, B. C. D. da (2003) *Todas as Trilhas: Perfilando Pesquisas e Projetos*. Editora Cultura Acadêmica: Araraquara -SP.

www.grammatics.com/appraisal/ Acessado em: fevereiro de 2006.

<http://www.nile.icmc.usp.br/> Acessado em: setembro de 2005.

<http://www.webcorp.org.uk/> Acessado entre fevereiro e junho de 2007.

APÊNDICE
LISTA DE OCORRÊNCIAS DO ITEM LEXICAL
***CORRETO* NO CORPUS NILC CLASSIFICADAS POR**
SENTIDO

Sentido 1: exato

(01) Esse resultado não é contraditório com o preâmbulo do argumento EPR, que aceitava o **caráter correto** da teoria, mas recusava sua completeza: é com efeito num nível ulterior de refutabilidade que o teorema de Bell se coloca e, se ele fosse verificado, os resultados adquiridos da mecânica quântica permaneceriam, mas na condição de uma aproximação, mostrando-se o domínio de validade da teoria doravante limitado.

(02) Auditoria: é uma outra aplicação natural para hipertextos, porque é baseada na geração e análise de relatórios de informação originários de várias fontes. A tarefa de uma auditoria inclui reunir e produzir um grande número de documentos, e interconectá-los para comprovar o **conteúdo correto** das informações.

(03) Quanto ao aproveitamento, o Inter é o que possui a pior marca: 30,12% dos passes do time não chegam ao **destino correto**.

(04) A CET informou que os funcionários só fazem checagem dos bloqueios (feitos com correntes e barris vazios). Mesmo que eles não estejam no **local correto**, a CET não sobe para verificar se houve invasão.

- (05) Muitos usuários abandonam o tratamento ou se esquecem de tomar a droga no **horário correto**.
- (06) A CET informou que os funcionários só fazem checagem dos bloqueios (feitos com correntes e barris vazios). Mesmo que eles não estejam no **local correto**, a CET não sobe para verificar se houve invasão.
- (07) Cada jogador recebe as cartas com as capitais dos Estados, bandeiras ou pontos turísticos. Os jogadores colocam as cartas no tabuleiro. Ganha que acertar o **lugar correto** para o maior número de cartas.
- (08) O sistema também pode usar sensores acoplados à espaçonave para varrer o solo abaixo do veículo, comparar os dados encontrados com os dados armazenados no computador e então dar automaticamente aos pilotos o **mapa correto**, baseado no ponto em que os sensores indicam que o avião está localizado.
- (09) “O impacto produzido por ele será um show espetacular para um observador que esteja no lugar certo e no **momento correto**”, diz Zdenek Sekanina. Esse é o problema. Como a explosão se dará na face oculta de Júpiter, pouco se poderá ver da Terra, mesmo com telescópios potentes, como o Hubble.
- (10) A manutenção do **nível correto** e a troca do óleo no prazo determinado pelo fabricante ajuda a prolongar a vida útil do motor.
- (11) “Encontramos certa dificuldade em localizar a inscrição de Aied, pois o nosso cadastro é organizado por CPF. O **número correto** do CPF foi localizado e, contornado o problema, o reclamante já está sendo convocado para se habilitar no sistema”, informa Inaba.

- (12) Os obesos com aplicações no **ponto correto** perderam apetite e peso nas três semanas que durou o estudo. O outro grupo não teve alterações significativas.
- (13) “O treino compartimentalizado facilita o trabalho de botar na cabeça da equipe qual é o **posicionamento correto** de cada jogador.”
- (14) Ela responde com o **resultado correto** desta e observa que, incapaz de multiplicar, decorara os resultados de todas as possíveis (e infinitas) multiplicações.
- (15) A assinatura de carteira de trabalho é difícil, e raras vezes pelo **valor correto**, a rotina de trabalho é estafante e indigna, e o roteiro de folgas, estúpido.
- (16) Quem sabe se os tamanhos fossem grande, enorme e extra? Ninguém ia sentir-se constrangido em escolher o **tamanho correto**. Mas ninguém vai poder querer usar camisinha à moda grunge.
- (17) O “novo samba” simboliza a degeneração dos costumes. Musicalmente beira o charlatanismo. Possui arranjos de teclados e guitarras, letras de romantismo amanteigado e imaginação melódica zerada. O **termo correto** para ele deveria ser “sambrega”.
- (18) Silva voltou ao **traçado correto** e ainda teve forças para perseguir e superar seu compatriota. Foi a terceira vez que um mexicano venceu a Maratona de Nova York, uma das provas de rua de maior prestígio em todo o mundo.

(19) Por consequência de tudo isso, o poder da comunicação deve estar nas mãos daqueles que querem o bem de um país ou de uma região, que saiba encaminhar as pessoas no **trajeto correto**; pois sabe-se que será muito difícil um meio de comunicação imparcial, mas independentemente disto, espera-se um planeta e uma sociedade propícios para se viver.

(20) O candidato do PDT à Presidência, Leonel Brizola, disse que seu adversário do PT, Luiz Inácio Lula da Silva, omitiu da Justiça Eleitoral o **valor correto** de sua casa em São Bernardo do Campo (SP).

Sentido 2: adequado

(21) **Armazenamento correto** também contribui com a conservação dos alimentos e figura entre as orientações anti-desperdício.

(22) Produzir um bem ou serviço com qualidade, com a entrega no prazo determinado, com o **atendimento correto** (inclusive na assistência pós-venda) ao menor custo (logo, menor preço), são as novas regras de mercado.

(23) - O sr. acha um **comportamento correto** sair da sala para não votar?

(24) Não é um **conceito correto**, como mostra reportagem sobre a Companhia de Telefones do Brasil Central (CTBC) publicada hoje por esta Folha.

(25) FHC, o ministro, marcou sua passagem na Fazenda pela construção de um **discurso correto** e pela incomparável maestria com que soube invocar razões e pretextos para não agir.

- (26) Nesta hipótese, o juiz, na sentença, pode corrigir o erro, condenando o acusado nos termos do **dispositivo correto**.
- (27) O princípio que inspira a legislação é sem **dúvida correto**.
- (28) Segundo ele, o **equipamento correto** tem que ter manômetro (aparelho que mede a pressão) e válvulas para controlar o fluxo de gás. "O uso de um cilindro maior que o recomendável, como é o caso do extintor, produz uma pressão exagerada", disse Lopes.
- (29) Pois só desse jeito, o brasileiro se tornará cidadão, despertando para o país e exigindo seus direitos a educação e trabalho e sobretudo diluindo o "mar-de-lama" brasileiro, pela vigilância e pelo **exercício correto** da cidadania.
- (30) Utilizando o **filme correto** você também garante o bom resultado do seu trabalho.
- (31) Isso pode ser ilustrado através do gráfico apresentado na Figura 2.6 [HET84], onde à medida em que aumenta o número de instruções modificadas no programa, aumenta a probabilidade dessas modificações afetarem o **funcionamento correto** do programa.
- (32) Segundo o presidente da CPI do Orçamento, senador Jarbas Passarinho (PPR-PA), o decreto não é o **instrumento correto** para este caso.
- (33) Um amigo que chegou da França, país de origem desse aparelho sanitário, me disse que o **jeito correto** de usar o bidê é sentar-se de frente para as torneiras.

- (34) O objetivo da empresa é fornecer uma cartilha sobre a aplicação de vacinas e , após o nascimento, instruções para **manejo correto**.
- (35) O coordenador regional de vendas da Antarctica, fabricante do Chope, Ademar Silvestre, disse que a empresa fornece uma bomba manual e um folheto com instruções sobre o **manuseio correto**.
- (36) Executivos de recursos humanos consultados pela Folha consideram critério da **média correto** porque, dizem, preserva a realidade dos salários.
- (37) “Esse não é o **método correto**, mas é tradicionalmente usado no Brasil”, afirma.
- (38) Do rosa-framboesa pálido usado por Daniela Mercury, ele não gostou. Achou uma cor ingrata para um **modelo correto**.
- (39) Segundo Pereira, quando a porcelana é colada de **modo correto**, ela tem resistência comparável às “jaquetas” (proteção em forma de dente postiço) tradicionais.
- (40) **Paliativo correto**. Há diferentes tipos de seguro desemprego em praticamente todos os países desenvolvidos e em várias nações do Terceiro Mundo.
- (41) O aeroclube exige apenas um treinamento antes do voo para instruir o pára-quedista principiante sobre o **procedimento correto** no ar.
- (42) É válido ressaltar que como a idéia era apenas estimar custo com base em número de casos de teste foi utilizado um **programa correto** e

portanto não podemos ainda nada concluir sobre o poder de revelar erros de cada critério.

(43) A Intec considerou o **teste correto**.

(44) Poder-se-ia sugerir, com o devido respeito à comunidade psicológica, que “psicologia teórica” realmente é o **título correto** para classificar os estudos da percepção e da cognição da inteligência artificial...

(45) Para ele, a carta de apresentação também não foi escrita em **tom correto**.

(46) Sem o **tratamento correto**, as dores persistem, chegando ao nível do insuportável.

(47) Outro fator que contribui para que o banco dure é o **uso correto**.

(48) O prezado professor Bresser Pereira, que defendia a opção de integração comercial das Américas para o Brasil, embora **conceitualmente correto**, deve anotar que o clima político aqui é contrário.

(49) “Fui escolhido para fazer este filme porque sei fazer algo que não seja muito sério nem institucional. Eu posso fazer um filme **culturalmente correto** e agradável do ponto de vista do espetáculo”, conta Farina.

(50) Produto tem normas internacionais de segurança; saiba como criar um ambiente **ergonomicamente correto**.

(51) Para o ex-presidente do BNDES e do programa, Eduardo Modiano, embora seja “**estatisticamente**” **adequado** incluir essas empresas entre as privatizações do atual governo, “eticamente” é inadequado.

- (52) Eros Grau havia sustentado que o processo estava **formalmente correto**.
- (53) Talvez isso seja **historicamente correto**.
- (54) Mas não creio que seja **inteiramente correto**.
- (55) Nem como obviedade é **plenamente correto**.
- (56) O merchandising do 25 aniversário do festival de Woodstock também terá seu lado **sexualmente correto**.
- (57) Não sei se alguma atriz serviu de modelo para a princesa Jasmine, mas há que se louvar a manutenção de seu perfil oriental, **“sheerazadianamente” correto**.
- (58) “A ação proposta contra a Fazenda Nacional equivale à proposta contra a União Federal. Embora não **tecnicamente correto**, a expressão Fazenda Nacional equivale a União Federal”.
- (59) “O princípio da apatia, isto é, de que o homem prudente não deve em tempo algum se emocionar, nem mesmo naquele pesar pelos infortúnios de seu melhor amigo, é um preceito moral **totalmente correto** e sublime da escola estóica, porque a emoção nos cega”.
- (60) Desde então, a teoria dos pulsares se desenvolveu tão rapidamente que parece **virtualmente correto** que os impulsos rádios e ópticos emitidos pelo pulsar tenham origem na própria energia proveniente de uma estrela de nêutrons em rotação.

(61) **Ecologicamente correto**, o barco, na falta de vento, será impulsionado por energia solar.

(62) Nesse sentido, talvez fosse **mais correto** dizer que a prática do zapping não faz senão potencializar, ou mesmo tirar conseqüências daquilo que é o funcionamento normal da televisão, (...).

(63) “Há muitas oportunidades no Brasil, basta estabelecer **um correto** planejamento estratégico”.

Sentido 3: isento de erros

(64) E a jurisprudência se tem revoltado contra ela (RJTJESP 93/237), inclusive sob o fundamento, **absolutamente correto**, de que o cumprimento da precatória sai mais barato que a publicação de editais.

(65) O governo do Estado considera que as atribuições federais “ditadas pela Constituição Federal” são o **caminho correto** “para superar as decepções” resultantes dos esforços das polícias estaduais, que travam combate com os traficantes só na venda final dos tóxicos.

(66) “Isto demonstra que estamos no **caminho correto**”, diz Borges.

(67) E também mais romântica e apaixonada, sem deixar de lado o virtuosismo e o **andamento correto**.

(68) Neste caso, o modelo deve identificar a diferença existente entre o que o estudante entendeu e o **conhecimento correto** do assunto ensinado.

(69) Médicos, pesquisadores e profissionais da área de saúde encontrarão aqui tanto informações para a consecução de todos os passos indispensáveis a um **diagnóstico correto** das manifestações alérgicas, quanto um abrangente levantamento de dados que devem ser tomados como objeto de análise para quem quer compreender o próprio processo de produção da alergia.

(70) Ocorrência 344 Não é o CAC mero lugar de recepção, mas órgão que soluciona as questões apresentadas ou lhes dá o **encaminhamento correto**, com as instruções adequadas e, na maioria das vezes, fornecendo roteiros e formulários adequados às pretensões do contribuinte.

(71) Basta ter o **endereço correto**.

(72) Extraído de **Ensino Correto**, da Atlante, Rev. Do Instituto Geográfico de Agostini.

(73) Mas também não há o único **entendimento correto** do pensamento de Nietzsche em um sentido definitivo e conclusivo.

(74) Não fosse pelo **acento correto** sobre o “e” da última palavra, o leitor desavisado poderia imaginar que o redator decidira abolir esses sinais de duas outras, “têm” e “pé”.

(75) **Português correto**, organização e método são atributos indispensáveis a toda secretária que trabalhe em nível executivo.

- (76) "Já perdi anunciantes por ter mudado para o **inglês correto** uma frase de anunciantes", afirma.
- (77) Alguém poderia sugerir que "pega o teu", **gramaticalmente correto**, manteria a graça e afastaria a crítica de chatos intolerantes.
- (78) De " " é analisada durante a descrição, o tradutor sempre gera código **sintaticamente correto** para uma descrição sintaticamente correta da função F e as informações adicionais são geradas pelo (SM)2 com sintaxe correta.
- (79) Eu acho que pode e deve; é bonito **falar correto**.
- (80) "Aguardo um **informe correto** e que se faça a correção necessária."
- (81) A Divecar não teria informado o **nome correto** do policial.
- (82) "É o melhor ano para os confinadores que fizeram um **planejamento correto** em fevereiro passado", diz o pecuarista Sylvio Lazzarini Neto, do Sindicato Nacional dos Pecuaristas de Gado de Corte.
- (83) Para a estruturação e **preenchimento correto** dos cartões, necessário conhecer certos detalhes, pois às vezes se recebe, especialmente de radioamadores novos, confirmações de comunicados que, à vista das convenções internacionais entre associações de Radioamadorismo, carecem de validade para quaisquer diplomas ou concursos.
- (84) Mocsányi afirma que a escolha do **processo correto** depende do tipo de empresa e de sua situação no mercado.

- (85) “Quero meu **projeto correto** justamente agora, no Carnaval”, diz ele.
- (86) Não se discute se a afirmação de Lula embute um **raciocínio correto** ou equivocado.
- (87) Expôs também um conjunto ordenado de padrões de adequação: adequação observacional, onde a gramática faz um **relato correto** de dados lingüísticos observados; adequação descritiva, onde o relato também capta a competência intrínseca do falante nativo (...).
- (88) Mas está no **rumo correto** se deseja combater a gritante desigualdade social do país.
- (89) Não só nesta parte gramatical, mas também quanto a formação de frases que possuam um **sentido correto**.
- (90) O **título correto** seria, então, “Crianças em pé têm mais diarreia”.
- (91) (...) 2) também a fronteira entre o **correto** e **incorreto** fica na dependência das forças em luta e da conjuntura.
- (92) Examino o envelope, podia ser um equívoco postal. Não, está tudo coreto, nome, endereço, até mesmo CEP que até hoje não consegui decorar.

Sentido 4: válido

- (93) Acho que **é correto dizer que** a dança, embora tomando diferentes formas, caminha em ciclos de dança pura e dança expressiva.

- (94) **É correto dizer que** o PDT tem procurado se colocar à esquerda do PT?
- (95) Dentro do que aceita **como correto** a terapia de vidas passadas, uma pessoa que tem medo de altura ou de fogo pode ter adquirido a fobia em função de acidentes anteriores, como, por exemplo, ter sido empurrada de um precipício ou ter morrido queimada.
- (96) O movimento da QT diz que medindo a qualidade estatisticamente e extraíndo todos os fatores antiqualidade aparece um grau de qualidade superior. Em certa medida isso é **correto**.
- (97) Ainda assim, o **cerne** do argumento de Nassif é **correto**.
- (98) Desta vez os invasores reclamam da demora da Justiça em concluir a vistoria da fazenda para confirmar se o **laudo** do Incra, que a classifica de improdutiva, é **correto** ou não.

Sentido 5: justo

- (99) Os dois autores, num tom alegre e **politicamente correto**, dizem que ir bem no QI não é o mais importante.
- (100) O senhor **acha correto** que a Universidade ofereça uma série de serviços gratuitamente, por ser uma universidade pública? Isso não poderia ser explorado comercialmente, para gerar recursos para a própria Universidade?
- (101) “Eu não uso e não **acho correto** utilizar religião”, disse o tucano.

(102) - Mesmo dentro da lei, o senhor **considera correto**?

(103) Ele disse não **considerar correto** remover ou sacrificar o golfinho que está atacando turistas em Caraguatatuba. “Ele está no seu habitat natural. Os banhistas não. Além do mais, os golfinhos só atacam quando incomodados”, afirmou.

(104) É honesto, é engenheiro, é democrata, é tucano, foi cassado em 68; não é apenas politicamente correto, pior, é **administrativamente correto**.

(105) Nestes primeiros dias do ano ganhei uma folhinha do Posto São Luis do Paraitinga que é verdadeiro exercício de cidadania, **constitucionalmente correto**.

(106) **Impoliticamente correto**, RICARDO SEMLER Tantas ruas têm no nome uma homenagem a políticos desonestos.

(107) ‘Acertando as contas’ peca por ser óbvio e **moralmente correto**.

(108) Enquanto nos países ricos os pênis são cortados em nome da “mais repressão” do **feminismo correto**, aqui as vaginas mudam a vida brasileira. É apavorante a fragilidade de nossa história política.

(109) Essa deve ser a diferença de preço do produto no Brasil, mas Carrara diz que ainda não dá para fixar um **preço correto**.

(110) É um **princípio correto**, mas ainda insuficiente para afastar dúvidas cruciais, especialmente no caso da conversão dos salários.

(111) Ele informou que outros oficiais do CSM/MM de que **não era correto** eles transferirem recursos do Estado para a conta dos Zaborski.

(112) **Não é correto** também o presidente atribuir culpa exclusiva aos líderes.

(113) “É um cinismo. **Não é correto** vender uma idéia de que o saneamento está uma maravilha, além de gastar recursos públicos em publicidade na mesma semana que a Sabesd cancela uma concorrência que já durava 14 meses”, diz o empresário Antônio Carlos Germano Gomes, 36.

(114) **Não é correto** gerar falsas esperanças em um povo tão ansioso de progresso, de trabalho e de viver em um país moderno e justo.

Sentido 6: aceitável

(115) O tenor Peter Dvorsky faz um Renato Des Grieux **apenas correto**. Não é um Plácido Domingo.

(116) Não é **correto** liberar o uso de drogas assim como é liberado o uso de bebidas alcoólicas ou cigarros comuns, pois sem dúvida sua intensidade de uso aumentaria.

(117) É **correto** mostrar-se indeciso em relação à Iugoslávia, quando populações democráticas não se dispõem a apoiar medidas que talvez atolassem seus combatentes num impasse de longa direção, e que a longo prazo teriam pouca probabilidade de resolver a partilha étnica.

(118) É claro que isto é um problema e que não é **correto**, de qualquer ponto de vista o uso das drogas.

- (119) “Eles estão convertendo em URV preços que embutem uma grande margem de segurança. Estão querendo se cobrir de alguma eventualidade. Isso não é **correto**.”
- (120) O sr. acha que o sistema atual de partidos e eleições é **correto** ou exige alterações?
- (121) Esta discriminação parece-me resquício de entulho autoritário. Discriminar não é **correto**; basta regulamentar o uso político de programas, sem cassar o direito do candidato ao seu ganha-pão.
- (122) Todos esses grupos são manipuladores contumazes de informações - da CUT a Maluf. O que não é **correto** é a imprensa embarcar nesse jogo.
- (123) Ela também poderá argumentar que os indicadores sociais e econômicos dos brancos nos EUA são superiores aos dos negros, o que é **correto**. Ainda assim, é inegável que o negro dos EUA vive melhor do que o da Nigéria.

Sentido 7: com a razão

- (124) Os recorrentes aumentos acima da inflação de mercadorias produzidas por oligopólios demonstram que o governo **está correto em** reduzir as tarifas de importação.
- (125) É um técnico completo. **Está correto ao** pensar na figura do técnico como uma espécie de chefe de Estado de um clube.

Sentido 8: verdade

(126) Mas não **seria correto** dizer que Freud “gostava” de Shakespeare, como gostava de Goethe ou Milton.

(127) Folha - Seria **correto** dizer que a sra. Saiu da Prefeitura mais pragmática e menos ideológica?

“Sentido” 9: marcador discursivo

(128) A genética não é um monte de pequenos projetos Manhattan prontos a explodir nas mãos da sociedade? Ayala - **Correto**, e eu começo minha resposta lembrando a primeira explosão atômica, o teste em Alamogordo.

(129) - Nesta conta, o deságio no câmbio e os juros vão gradualmente cair, **correto?**

(130) Ao mesmo tempo, em suas coreografias você mistura técnicas clássicas, dança moderna e movimentos cotidianos ... Cunningham - **Correto**, eu não rejeito o balé clássico.

Sentido 10: honesto

(131) Em realidade, alcançou, como sempre, o **contribuinte correto**, que já suportava uma pressão tributária elevada.

(132) Para Amato, ágio não é **dinheiro correto**.

- (133) Existe ainda um empreiteiro que acumula as funções de dono de bordel, um **funcionário correto** - que tenta pôr ordem no coreto - e uma prostituta, Lola, por quem ele se apaixona.
- (134) Como filha única, ela é herdeira de sua fortuna e, na avaliação dele, só um **homem correto** pode cuidar de ambas a filha e a fortuna.
- (135) O que aconteceu na Folha é grave, merece providências da Redação - mesmo porque o assunto ainda promete render muito - mas não tira do jornal o mérito de estar fazendo um **jornalismo correto** na cobertura do assassinato de Cruz.
- (136) Os italianos reclamaram do que acharam ter sido pênalti sobre Dino Baggio e os mexicanos não gostaram de um impedimento marcado ao final do primeiro tempo, o que impediu a validação de um gol feito na seqüência. Mas Lamolina foi um **juiz correto**.
- (137) “Estamos aqui para pressionar para que haja um **juizamento correto**. Ninguém está aqui para condenar inocentes”, disse a diretora do grupo, Flora Abreu.
- (138) Não é possível, em pleno final do segundo milênio, que uma entidade privada, embora poderosa, impeça o poder público de cumprir o seu dever de zelar pela correta observância da norma tributária, obstando que o **pagamento correto** de um tributo seja feito.
- (139) Estará baseada na sua condição de **político correto**, um político com passado, intelectual, uma homem de inteligência.

(140) Loyola afirmou que Mauch não vem sofrendo pressões e o qualificou como um **profissional correto** e competente.

(141) Nesse momento em que está todo mundo se delatando, a lembrança dele é muito boa para mim. Era um **sujeito correto**.

(142) Sabes que eu sempre fui **muito correto** nos meus amores.

(143) Quem tentou opor-se a Covas se deu mal. **Covas é correto**, não há nada contra ele, está do lado do bem.

Sentido 11: ideal

(144) A posição do PSDB é a visão contemporânea **do correto** papel da intervenção governamental.

(145) Provavelmente é isto que está acontecendo, e **o correto** seria reduzir a dose. Quando a tristeza normal se transforma em tristeza patológica, muitas vezes existe uma depressão.

(146) (...): pelotões de pesquisadores pensavam que se aproximavam mais e mais **do correto** ao elaborarem montanhas de papéis; ou, mais ainda, acreditavam que, apresentando números e tabelas, assemelhavam-se à ciência.

(147) Reintroduzir o Fusca foi uma idéia na direção certa, mas **o correto** seria introduzir a bicicleta com motor.

(148) Isto provoca um êxodo prematuro dos pilotos e assim não acontece um **aprendizado correto** de todos os fundamentos.

(149) Mas é, por outro ponto de vista, o **desfecho correto** e lógico para todos esses jogos inúteis que marcaram e foram a tônica desse torneio que deveria ser demonstrativo do panorama atual do futebol tetracampeão.

(150) O **papel correto** da imprensa seria a instrução e não o tapeamento.

(151) Além da escolha certa do tipo de antena, é importante o seu **correto dimensionamento**.

Sentido 12: bem-feito

(152) A perua tem **acabamento correto** e bons bancos, com apoio central de braços.

(153) Um **buffet correto**, bebidas idem, noivos embalados a “Unforgettable”; com Nat King Cole.

(154) A abertura comercial alcançada até aqui é fruto de um **cronograma correto**, mas, para que a estabilidade de preços se consolide e o câmbio não se valorize ainda mais, é preciso estimular as importações.

(155) Representava uma grande dama adorando um Menino Jesus e o fundo era uma paisagem de montanhas cobertas de neve, tratado com aquele **desenho correto**, firme e aquela sábia perspectiva aérea que tanto lhe gabam os críticos.

(156) A escola fez um **desfile correto** e luxuoso.

- (157) Robert Cray volta com um **disco correto**, onde acaba se redimindo, talvez sem saber, das tolices que pode ter cometido.
- (158) “Anos Incríveis” é um **seriado correto**. Diante da pobreza da programação para adolescentes, como o “Programa Livre” e o “Fanzine”, faz sucesso de público e de crítica.
- (159) Com a maioria das canções em inglês, o grupo formado em 92 fez sua segunda participação no festival. O Weapon fez um **show correto**, conseguindo até tocar uma cover de “Jump”, do Van Halen.
- (160) Folha - Com isso o futebol alemão não fica muito esquematizado e perde seu encanto? Matthaus - Seria assim se não fizéssemos o **trabalho correto**.
- (161) Mas não importa quantos elementos estrangeiros o grupo use. O resultado é sempre o mais puro pop **melodicamente correto**.