

3. Análise dos *Corpora*

Neste capítulo procuraremos dar continuidade à já citada discussão assinalada pelos demais autores: o **não-x** deve ser classificado como composição ou prefixação? Para tal, apresentaremos previamente a parte da metodologia referente à lingüística de corpus que orienta nossa pesquisa empírica sobre o **não-** anteposto a nome, demonstraremos como foram extraídos os dados e, a seguir, refletiremos sobre as análises.

A pesquisa foi feita com base na observação de dados coletados em um *corpus* formado, por sua vez, a partir de dois *corpora* informatizados: O CORPOBRAS-PUC RIO e o CORPUS NILC, sendo o primeiro constituído por ocorrências do português falado e o último, por registros escritos, extraídos principalmente do jornal Folha de São Paulo. Na medida em que investigamos somente os subcorpora orais do CORPOBRAS, apenas duas ocorrências foram encontradas, revelando aspectos interessantes do fenômeno estudado. Quanto ao *corpus* NILC, composto por 2.335 ocorrências correspondentes às formas **não-x**, foram analisados aproximadamente 25% de seu total.

Serão apresentadas as noções essenciais sobre a lingüística de corpus, bem como os direcionamentos adotados por esta pesquisa no que tange aos *corpora* escolhidos, sua representatividade, conteúdo e coleta dos dados. Discutiremos, então, nossas reflexões sobre os aspectos semânticos e lexicais do objeto pesquisado, e por fim, demonstraremos as conclusões definitivas referentes à análise.

3.1. O não-X e a lingüística de corpus

3.1.1. “CORPUS”, uma palavra polissêmica

Conceituar a lingüística de *corpus* não é tarefa fácil, tendo em vista que se trata de ramo relativamente recente do saber e suas conceituações, portanto, são por vezes ainda controversas. Atualmente, a lingüística de corpus é mais comumente vista como sendo uma área interdisciplinar, já que suas aplicações podem envolver saberes distintos

tais como engenharia computacional, lexicografia, tradução, lingüística aplicada, dentre outros.

Todavia, embora a lingüística de corpus seja uma área científica relativamente recente, a preocupação quanto à compilação e ao uso de *corpora* dentro dos estudos da língua remonta a um período histórico não tão atual. Conforme notam ALUÍSIO E ALMEIDA (2006: 156), data ainda do século XVII a idealização de um dicionário de nossa língua que se valeu de um *corpus* como exemplário, obra que só sairia do prelo um século depois sob o título *Vocabulário Portuguez e Latino*, de autoria do Padre Rafael Bluteau. Desde então, outras obras lexicográficas fizeram uso de *corpora*, porém a idéia sobre o que seja um *corpus* muito se alterou até o tempo presente.

ROSSINI (2005:09) lembra ainda que nos anos 30, Firth já apontava para a importância do contexto para os estudos lingüísticos, seguido pouco depois por Stubbs que questionava os estudos baseados em sentenças isoladas, coletadas aleatoriamente e segundo métodos caóticos, cuja confiabilidade era suspeitável. Estes pesquisadores já atestavam que o estudo de um idioma fora de contexto gerava uma série de resultados pouco científicos sobre a língua. O uso de *corpora* mais confiáveis demoraria ainda para ocorrer de forma concreta e Stubbs esperaria algumas décadas até que os trabalhos nesta área se desenvolvessem tal como ele previra.

Assim, tal mudança apenas seria efetivamente propulsionada após o avanço tecnológico, e se no passado a palavra “*corpus*” possuía somente a acepção referente a qualquer conjunto de dados de uma língua ou textos escritos, reunidos muitas vezes de forma imprecisa e aleatória, hoje esta concepção convive lado a lado com outro conceito um tanto mais complexo. Com o desenvolvimento da lingüística de corpus, abordagem metodológica que se caracteriza pela ênfase às novas tecnologias e às aplicações dinâmicas que as mesmas podem oferecer sobre os estudos lingüísticos, nascem novos *corpora* eletrônicos e a palavra “*corpus*” passa igualmente a receber um tratamento semântico distinto daquele que recebia até então.

Neste sentido, Sardinha reconhece a dificuldade em apresentar a melhor definição para o termo “*corpus*” dentro desta ótica e, em meio a tantas digressões sobre o tema, conclui que de acordo com a lingüística de corpus, *corpus* deve ser um corpo de linguagem natural, isto é, real e espontânea, coletado, selecionado e organizado em conformidade com critérios lingüísticos e tendo por finalidade servir como amostra da língua. Nas palavras do próprio autor: “Assim, se por um lado os textos devam ser naturais

(autênticos e independentes do *corpus*), o *corpus* em si é artificial, um objeto criado com fins específicos de pesquisa”. (SARDINHA, 2000: 336)

Seguindo esta linha de raciocínio, Sardinha assume, finalmente, que os *corpora* devem: a) ser autênticos, isto é, devem ser criados a partir de dados que existem naturalmente na língua. Dados criados com o objetivo de figurarem no *corpus* não são válidos, portanto, e se o *corpus* for composto por falantes não-nativos, isso deve ser especificado; b) os *corpora* devem ser construídos com objetivo de constituírem fonte de estudos linguísticos; c) devem ser legíveis em computador, ou seja, formatados; d) quanto à composição, o conteúdo deve ser criteriosamente escolhido, ou seja, os *corpora* devem ser selecionados de modo a possuírem uma característica específica. Por exemplo – *corpus* do português brasileiro escrito deve conter textos de variados gêneros, produzidos em grande parte da extensão territorial nacional; e) devem ser representativos de uma língua e f) devem ser vastos para serem representativos.

As duas últimas questões particularmente ensejam maiores detalhes. Sobre estes aspectos nos deteremos mais adiante e buscaremos demonstrar, ainda, como tais requisitos se aplicam ou não ao nosso trabalho. Procuraremos apresentar os recursos computacionais utilizados para a manipulação dos dados. Importa, ainda, determinar por que optamos pelos *corpora* escolhidos para este trabalho, em detrimento de dois outros *corpora*, inicialmente selecionados e posteriormente descartados.

3.1.2.

A Escolha dos *Corpora*

Nossa pesquisa sobre o **não-** anteposto a nome se iniciou a partir de um *corpus* montado através de um método simplificado de buscas pelo site Google. Buscamos àquela época palavras antecedidas pela partícula não por meio da ferramenta search, disponível no próprio sítio. Para tal, digitamos a palavra não com hífen e em seguida sem hífen, porém seguida a artigo feminino ou masculino. Foram recolhidas 49 palavras, então analisadas conforme o contexto na qual se encontravam inseridas.

Embora os resultados fossem promissores, este método, porém, mostrou-se desde o início bastante confuso, visto que não havia um controle possível sobre o *corpus*, isto é, não era possível determinar com precisão os gêneros recolhidos, o grau de formalidade da língua, tamanho dos textos, dentre outras informações relevantes.

Cogitou-se, assim, da formação de um *corpus* de linguagem jornalística, que desse conta dos fatos sem o risco de produzir resultados viciosos. Seguindo este raciocínio, montou-se um *corpus* a partir de palavras coletadas no jornal O Globo Online durante 17 dias em duas amostragens distintas espaçadas por um intervalo de 8 dias cujo objetivo era evitar a redundância de termos, fato comum em notícias que versam sobre o mesmo tema. As coletas eram feitas através do programa de navegação Mozilla Firefox 3.0.11 por meio da ferramenta search (busca) do próprio navegador que pode ser acessada através da barra de ferramentas Editar > Localizar ou por meio da tecla de atalho Ctrl+F. Digitou-se a palavra “não” a fim de reconhecê-la em cada página acessada e coletaram-se as ocorrências encontradas junto com os contextos (parágrafo) nos quais estavam inseridos.

A delimitação de um gênero – o jornalístico – conferiu maior segurança à pesquisa e mais uma vez os resultados foram positivos no que tange à produtividade da formação: encontrou-se uma média de 6 ocorrências únicas por dia⁸. Uma breve análise que se ateve somente aos substantivos encontrados também foi capaz de revelar a pluralidade de significados contextuais possíveis. Contudo, um método de coleta assaz manual se mostrava dispendioso e pouco confiável, na medida em que se tornava vulnerável às falhas humanas. Permanecia, portanto, a possibilidade de erro e falta de consistência e, por outro lado, um *corpus* formulado sob este método jamais alcançaria um tamanho razoável, o que atentaria contra a sua representatividade e contra a representatividade dos próprios resultados.

Desta forma, decidiu-se por fim utilizar dois *corpora* distintos, de modo que tivéssemos um *corpus* confiável e que abrangesse tanto o registro oral como o escrito. A primeira parte desse *corpus* se constitui de ocorrências que seguem o modelo não-X, encontradas no *corpus* NILC/São Carlos (Corpus de Extratos de Textos Eletrônicos NILC/Folha de S. Paulo, www.linguateca.pt/CETENFolha⁹). Conforme GARRÃO e

⁸ Esta média aritmética refere-se ao número de ocorrências inéditas a cada dia, ou seja, foi contabilizada cada nova formação que aparecia a cada novo dia. Para encontrar a média, dividiu-se o total de ocorrências pelo número total de dias, isto é, 16.

⁹ O citado *corpus* foi cedido pelo NILC à linguatca, iniciativa esta que permitiu que os dados fossem processados, documentados e se tornassem acessíveis. A Linguatca é um centro distribuído de recursos para o processamento computacional da língua portuguesa. Foi a continuação do projeto *Processamento computacional do português* decorrido no SINTEF de Maio de 1998 a Maio de 2000. Objetiva: 1) facilitar o acesso aos recursos já existentes, através do desenvolvimento de serviços de acesso na rede, e mantendo um portal com informação útil, 2) desenvolver, de forma harmoniosa, em colaboração com os interessados, os recursos considerados mais prementes, 3) organizar avaliações conjuntas que envolvam a comunidade como um todo. Fonte: <http://www.linguatca.pt/>

DIAS (2003: 01), o *corpus* NILC é formado por textos brasileiros do registro jornalístico (predominante), didático, epistolar e redações de alunos, contendo cerca de 39 milhões de palavras de português brasileiro. A compilação deste *corpus* se iniciou em 1993 e objetivava atender à necessidade de efetuar os testes da ferramenta de revisão com textos reais, e fornecer a base empírica do uso de algumas formas gramaticais do português (PINHEIRO E ALUÍSIO, 2003). Prestava-se, portanto, a investigar formas de melhorar a revisão ortográfico-gramatical das ferramentas computacionais. Com o passar do tempo, o NILC- São Carlos ganhou novas características, tornando-se prestigiado e reconhecido entre os pesquisadores. Dentre os diversos gêneros textuais presentes no *corpus*, destaca-se o jornalístico.

No relatório do NILC no ano de 2003, assumiu-se que a escolha por textos de tal gênero encerrava a facilidade oferecida por este tipo de *corpora*:

A compilação desse material, no entanto, orientou-se pelo critério da facilidade na aquisição dos textos, o que terminou gerando grupos muito extensos de textos ao lado de conjuntos quantitativamente muito reduzidos. Este é o caso do *corpus* jornalístico que, na época de construção, teve garantida a exportação de um cd-rom inteiro, contendo todos os textos de 1994 do jornal Folha de São Paulo. (PINHEIRO E ALUÍSIO, 2003: 06)

Neste mesmo relatório, retornamos ao problema da representatividade dos *corpora*, já mencionado brevemente em linhas anteriores. Sardinha (2000: 342) seguindo os estudos de Sinclair e Halliday, confirma que, em essência, um *corpus* representativo deve ser sobretudo um *corpus* extenso e que abarque a maior variedade possível de gêneros e registros de um dado sistema linguístico. Isso porque em uma língua, certos traços predominam em detrimento de outros, o que pode ser verificado nos estudos lexicais, onde existem palavras ou mesmo características morfológicas mais ou menos frequentes. Para estudar palavras de frequência rara, é necessário ter uma amostragem que contenha uma alta carga lexical. O mesmo se aplica aos sentidos menos frequentes de alguns itens lexicais. A representatividade de um *corpus*, portanto, se relaciona a três aspectos principais:

O número de palavras é uma medida da representatividade do *corpus* no sentido de que quanto maior o número de palavras maior será a chance do *corpus* conter palavras de baixa frequência, as quais formam a maioria das palavras de uma língua. A segunda é o número de textos, a qual se aplica a *corpora* de textos específicos. Um número de textos maior garante que este tipo textual, gênero, ou registro, esteja mais adequadamente representado. A terceira é o número de gêneros, registros ou tipos textuais. Esta dimensão

se aplica a corpora variados, criados para representar uma língua como um todo. Aqui, um número maior de textos de vários tipos permite uma maior abrangência do espectro genérico da língua. (SARDINHA, 2000: 344)

O mesmo autor acaba assumindo que cada *corpus* deve ser representativo de “uma certa variedade” (IDEM: 345), ou seja, a representatividade seria sempre relativa e nunca absoluta. Ora, este problema pertinente à representatividade, encontra consonância em uma realidade já há muito percebida dentro dos estudos da linguagem: a de que a língua é um sistema infinito de meios finitos, isto é, desde Chomsky notou-se que o sistema de regras lingüísticas permite engendrar um conjunto infinito de frases ou de itens lexicais. Nenhum estudo científico, no entanto, pode ser feito com um número infinito de dados, sendo preciso, portanto, estabelecer uma amostragem limitada. Na medida em que a língua é um sistema de caráter probabilístico, o estudo de um “recorte” já poderia dar conta de explicar certos fenômenos “macros”, podendo o lingüista, quando preciso, se valer de cálculos estatísticos a fim de estabelecer a quantidade mínima de palavras necessárias para definir um *corpus* que seja representativo para sua pesquisa.

O *corpus* NILC-São Carlos utilizado para a primeira etapa de nossa pesquisa, é considerado um *corpus* representativo do gênero jornalístico, pois conforme lemos no relatório do NILC do ano de 2003 “o jornal Folha de São Paulo e o Jornal do Brasil, ambos com 1 ano de publicação compilado, são as únicas fontes significativas. Falta representar as demais categorias previstas no *corpus*, por exemplo, as revistas” (ALUÍSIO E PINHEIRO, 2003: 25), ou seja, o NILC não pode ser considerado um *corpus* representativo do português brasileiro escrito, na medida em que não disponibiliza de forma suficiente os mais variados gêneros possíveis dentro desta modalidade (cartas, e-mails, redações, relatórios, teses, artigos diversos, etc), porém no que diz respeito ao gênero jornalístico, este *corpus* comporta uma quantidade de dados considerada significativa. A partir de dados disponibilizados pelos lingüistas que atuam frente ao *Corpus* NILC São Carlos, calculamos que os textos jornalísticos correspondem a 85% do *corpus* total, conforme gráfico (anexo 3).

O outro *corpus* adotado foi o CORPOBRAS, cedido pela PUC-Rio para uso de nossa pesquisa que se ateu aos subcorpora referentes ao registro oral da língua. Este subcorpus integra o projeto de compilação de *Corpus Representativo do Português do BRASIL*, em fase de desenvolvimento, e que pretende fornecer dados e subsídios para uma análise multidimensional da variação entre gêneros discursivos. No momento desta pesquisa, o CORPOBRAS conta com 1.172.821 palavras, 1385 textos, 26 gêneros dis-

cursivos. O subcorpus de língua oral do CORPOBRAS¹⁰ é composto por todos cinco gêneros do discurso oral e dois gêneros do discurso escrito para ser falado, a saber: a) Discurso oral: conversas cariocas, conversas de crianças, entrevistas (acadêmicas), grupos de enfoque e atendimento ao cliente e b) Discurso escrito para ser falado: discursos políticos e roteiros cinematográficos.

Ignoramos os textos pertencentes ao discurso escrito para ser falado e nos ativamos aos demais gêneros pertencentes a este subcorpus. Utilizamos, portanto, dados autênticos, legíveis pelo computador, criteriosamente escolhidos, e compilados com o propósito de serem objetos lingüísticos. Fica, assim, comprovada a autenticidade de nosso *corpus*, sua representatividade, extensão e conformidade aos requisitos exigidos pela lingüística de *corpus*.

3.1.3. Extração dos dados

Os dados referentes ao CORPUS NILC foram retirados da seguinte forma: acessamos o sítio Linguatca (www.linguatca.pt) em 05 de setembro de 2009 às 20:37h. Através da janela localizada à esquerda do sítio, clicamos em “acesso a recursos”. Uma nova janela se abriu e clicou-se em AC/DC – acesso/ disponibilização de *corpora*. Nesta nova página, ao final, encontra-se o link para o *corpus* NILC/ São Carlos. O processo pode ser visualizado abaixo, para melhor compreensão:

¹⁰ Para maiores detalhes sobre o CORPOBRAS, verificar anexo 1.

Linguateca

[Estrutura](#)
[Equipa](#)

Apresentação
Acesso a recursos

- **AC/DC**
- Procura
- Frequência
- Corpos
- Anotação
- Exemplos
- CETEMPúblico
- CETENFolha
- CHAVE
- COMPARA
- Composição
- CorTrad
- Esfinze
- Floresta Santa/Citica
- METRA
- PAPEL
- ReEscreve
- REPENTINO
- Repositório
- WebSpell
- WPT

[Catálogo de recursos](#)
[Catálogo de ferramentas](#)
[Catálogo de actores](#)
[Catálogo de publicações](#)
[Informação interessante](#)
[Forum](#)
[Avaliação conjunta](#)
[Perguntas já respondidas](#)

[English](#)

Acesso a corpos de português: Projecto AC/DC

[Information in English](#)

O projecto AC/DC (Acesso a corpos/Disponibilização de corpos), iniciado em 1999, surgiu da necessidade de juntar os poucos recursos disponíveis num único ponto na rede e dessa forma facilitar a comparação e a reutilização do material, permitindo ao mesmo tempo acesso a uma ferramenta poderosa de interrogação de corpos, o sistema [IMS corpus workbench](#), para o qual desenvolvemos esta interface.

Desde 2000, a anotação dos corpos tem sido feita automaticamente pelo [PALAVRAS](#) de Eckhard Bick, e convertida para o "formato AC/DC", descrito pormenorizadamente na página de [Anotação](#).

Uma descrição quantitativa inicial dos corpos servidos presentemente pelo AC/DC encontra-se na tabela abaixo. **Clique num dos corpos para o interrogar.** Para cada corpo, pode pedir concordâncias, distribuição e frequências simples e complexas, veja os [exemplos](#). Se é a primeira vez que visita o AC/DC e quer apenas experimentar, [procure no corpo clássicos da língua portuguesa Porto Editora](#).

Breve descrição dos corpos

Corpo	Tamanho (unidades)	Tamanho (palavras)	Tamanho (frases)	Variante (s)	Breve descrição
AmostrA-NILC	124.836	98.505	4.965	BR	AmostrA-NILC
ANCIB	1.690.376	1.258.764	80.992	BR	Correio electrónico correspondente ao tráfego na lista ANCIB
Avante!	7.766.338	6.501.266	204.414	PT	Semanário político Avante!, 1997-2002
CD HAREM	222.407	147.077	8.185	PT BR	Colecção dourada do HAREM
CETEMPúblico	232.543.379	189.575.095	7.665.410	PT	Jornal PÚBLICO, dividido em extractos, 1991-1998
CETEMPúblico (primeiro milhão)	1.202.938	912.294	38.251	PT	CETEMPúblico (primeiro milhão)
CHAVE	132.868.732	90.478.064	4.710.448	PT BR	Jornais PÚBLICO e Folha de São

Figura 1

Linguateca

[Estrutura](#)
[Equipa](#)

Apresentação
Acesso a recursos

- **AC/DC**
- Procura
- Frequência
- Corpos
- Anotação
- Exemplos
- CETEMPúblico
- CETENFolha
- CHAVE
- COMPARA
- Composição
- CorTrad
- Esfinze
- Floresta Santa/Citica
- METRA
- PAPEL
- ReEscreve
- REPENTINO
- Repositório
- WebSpell
- WPT

[Catálogo de recursos](#)
[Catálogo de ferramentas](#)
[Catálogo de actores](#)
[Catálogo de publicações](#)
[Informação interessante](#)
[Forum](#)
[Avaliação conjunta](#)
[Perguntas já respondidas](#)

[English](#)

Clássicos LP/Porto Editora	1.922.433	1.304.284	74.690	PT	Clássicos da literatura portuguesa, séc. XV e XIX
CONDIVport	7.117.746	5.570.215	328.214	PT BR	Jornais desportivos e revistas de moda e saúde
CoNE	925.228	685.231	31.561	PT BR	Mensagens de correio electrónico não-endereçadas
DiaCLAV	7.757.579	6.651.564	232.152	PT	Diário de Coimbra, Diário de Leiria, Diário de Aveiro, Visu Diário
ECI-EBR	916.253	724.010	44.381	BR	Texto do corpo Borba-Ramsey, compilado pelo ECI
ECI-EE	31.863	27.149	780	PT	Texto de chamada do programa europeu ESPRIT
ENPCPUB (parte portuguesa)	92.683	72.389	4.371	PT BR	Literatura traduzida do inglês proveniente do ENPC
FrasesPB	22.730	17.745	651	BR	Frases em português do Brasil
FrasesPP	20.050	16.234	594	PT	Frases em português de Portugal
Museu da Pessoa	517.733	375.158	27.288	PT BR	Entrevistas realizadas pelo Museu da Pessoa
Natura/Minho	2.156.187	1.749.083	68.910	PT	Jornal regional Diário do Minho, antes da revisão
Natura/Público	7.369.349	6.274.542	225.752	PT	Jornal PÚBLICO, dois parágrafos por notícia, 1991-1994
NILC/São Carlos	42.605.624	32.342.767	1.963.795	BR	Texto do corpo NILC, contendo maioritariamente texto jornalístico, mas também cartas comerciais e textos didácticos
Vercial	8.968.057	8.376.956	383.805	PT	Clássicos da literatura portuguesa, século XIX
Total	447.842.514	362.159.282	16.129.609	PT BR	todos os corpos

Para uma contabilização detalhada de todos os corpos, veja a página dos [Corpos](#).

Existe também um serviço que permite obter frequências de itens lexicais nos variados corpos, o [Ordenador](#).

Figura 2

A ferramenta AC/DC (Acesso/ Disponibilização dos Corpora) foi desenvolvida dentro do projeto linguateca com o objetivo de tornar todos os *corpora* já existentes em

língua portuguesa¹¹ acessíveis através de uma interface na rede, de forma simples e intuitiva¹². Utilizamos este filtro para buscarmos ocorrências de nomes anteceditos de **não-**. Optamos por buscar somente as ocorrências hifenizadas, pois uma pesquisa que também abrangesse o problema da hifenização em língua portuguesa¹³ seria inviável. A busca foi feita por meio do formalismo {**não-.***}, conforme visualizado a seguir:

Linguateca

Projecto AC/DC: corpo NILC/São Carlos

AC/DC : Linguateca

O corpus NILC/São Carlos do Núcleo Interinstitucional de Linguística Computacional, sediado no Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo em São Carlos, contém textos brasileiros do registro jornalístico, didático, epistolar e redações de alunos (Nunes et al., 1996a, 1996b). Para uma descrição mais abrangente, veja-se a página de [descrição do corpus NILC e sua descendência](#).

Procurar:

Resultado:

- ☒ Concordância
- ☐ Distribuição das formas
- ☐ Distribuição dos lemas
- ☐ Distribuição da categoria gramatical (PoS)
- ☐ Distribuição do tempo verbal e/ou do caso pronominal
- ☐ Distribuição de pessoa e/ou número
- ☐ Distribuição do gênero morfológico
- ☐ Distribuição da função sintática
- ☐ Distribuição por gênero de texto
- ☐ Distribuição por campo semântico
- ☐ Distribuição por grupo (de cor, roupa, etc.)

Opções

☐ Resultados por ordem alfabética (só distribuições)

Tipo	Jornalístico e outros
Variante(s)	BR
Tamanho (unidades)	42,6 milhões
Tamanho (palavras)	32,3 milhões

Página principal

Procure outros corpos:

- [AmosRA-NILC ANCIIB Avante!](#)
- [CD HAREM CETEMPúblico](#)
- [CETEMPúblico \(primeiro milhão\)](#)
- [CHAVE Clássicos LP Porto](#)
- [Editora CONDIVport CoNE](#)
- [DiaCLAV ECI-EBR ECI-EE](#)
- [ENPCPUB \(parte portuguesa\)](#)
- [FrasesPB FrasesPP](#)
- [Museu da Pessoa Natura/Ninho](#)
- [Natura Público NILC/São Carlos](#)
- [Vercial](#)

Estrutura do corpo

Marcadores estruturais: s, p, texto, t (anteriormente à versão 4.0, estava marcado como título), subtítulo, assinatura. Para os textos da folha de São Paulo, a (autor), artigo, caixa, situacao, li (elemento de lista).

Figura 3

¹¹ Quando autorizados por seus proprietários, naturalmente. Para maior detalhamento ver SARMENTO e SANTOS (2003: 707): “No projecto AC/DC, usamos o sistema Corpus Workbench desenvolvido pelo IMS da Universidade de Estugarda, o IMS-CWB, descrito em Christ et al. (1994) e Evert (2002). Finalmente, um terceiro componente é a interface aos corpora propriamente ditos, ou seja, a adaptação e personalização do sistema a um contexto de uso particular (no nosso caso, na rede, com um dado conjunto de corpora portugueses, adicionando um conjunto de funcionalidades). Este último trabalho foi feito pela Linguateca, e em geral deverá ser feito por aqueles que desejam fornecer um dado serviço”.

¹² O projeto AC/DC (Acesso a Corpora / Disponibilização de Corpora), iniciado em 1999, surgiu da necessidade de juntar os poucos recursos disponíveis num único ponto na rede e dessa forma facilitar a comparação e a reutilização do material, permitindo ao mesmo tempo acesso a uma ferramenta poderosa de interrogação de corpora, o IMS-CWB. Logo a partir de 2000, os corpora passaram também a ser anotados pelo PALAVRAS, o anotador sintático para o português desenvolvido por Eckhard Bick. (Fonte: <http://acdc.linguateca.pt/index.html/>).

¹³ Aqui nos deparamos com uma questão técnica, uma vez que pelos mecanismos de busca utilizados no mapeamento de nosso corpus, a leitura dos itens não hifenizados demandaria um trabalho mais braçal e menos automatizado. A listagem de todos os itens precedidos por **não** sem hífen apresenta também o **não** atuando como advérbio, sendo impossível a discriminação dos elementos. No caso dos *corpora* orais, havia particularmente a impossibilidade de controlarmos os critérios de transcrição adotados quanto ao uso ou não do hífen. Optamos, por este motivo, pela análise exclusiva dos itens hifenizados.

Projecto AC/DC: corpo NILC/São Ca

AC/DC : [Linguateca](#)

O corpus NILC/São Carlos do [Núcleo Interinstitucional de Linguística Computacional Matemáticas e de Computação](#) da [Universidade de São Paulo em São Carlos](#), contém jornalístico, didático, epistolar e redações de alunos (Nunes *et al.*, 1996a, 1996b). veja-se a página de [descrição do corpus NILC e sua descendência](#).

Procurar:

Resultado:

- ☒ Concordância
- ☐ Distribuição das formas
- ☐ Distribuição dos lemas

Figura 4

A concordância é uma listagem das ocorrências de um item específico, dispostas de tal modo que a palavra de busca aparece centralizada na página, acompanhada do seu contexto original. (SARDINHA: 2004). Assim, a ferramenta “concordância em contexto” (figs. 4 e 5) do sítio LINGUATECA indicou que **não-x** ocorre 2.771 vezes dentro do *corpus* NILC, sendo este o primeiro dado quantitativo apresentado. Esta contabilização é geral e inclui repetições.

Ao realizarmos a mesma pesquisa por meio de “distribuição das formas”¹⁴ (fig. 6), outra ferramenta disponível no mesmo sítio, contabilizamos 1.139 ocorrências para o formalismo {não-.*}. Neste caso, x é sempre palavra que não se repete. A título de economia, a ferramenta detalha somente mil das 1.139 ocorrências contabilizadas, mostrando ainda o número de repetições de cada ocorrência.

Tendo em vista que a análise de todas as 2.771 ocorrências em contexto seria impossível, optamos por utilizar como guia as mil ocorrências listadas, buscando verificar as mesmas em seus contextos. No entanto, notamos que em alguns casos, havia erros de digitação, tal qual observado em par=2548 (fig. 5). Por outro lado, não analisamos **não-** quando sucedido de infinitivos ou de gerúndios. Na medida em que foram ignoradas as ocorrências deste gênero, bem os de erros de digitação e formatação quando detectados, o número de ocorrências foi reduzido de mil para 696, o que corresponde

¹⁴ “Concordância em contexto” e “distribuição das formas” correspondem aos termos *token* (ocorrência no texto) e *type* (lexema referido pela ocorrência formal), mais corriqueiramente empregados.

a aproximadamente 25% dos dados apresentados pela ferramenta “concordância em contexto” (2.771 ocorrências).

Vale especificar ainda que no caso de palavras em que x apresentava múltiplas ocorrências, tal como “não-governamentais” (figura 6), buscou-se fazer a análise em apenas três contextos diferentes, a fim de tornar o estudo mais prático e eficiente. As figuras a seguir ilustram parte do processo descrito.

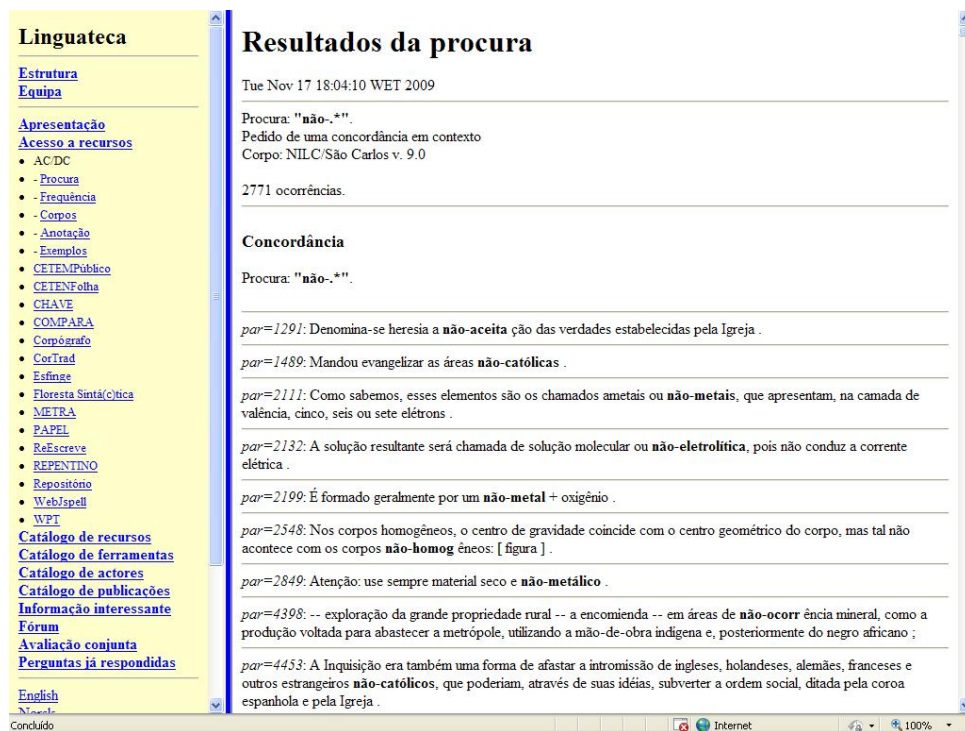


Figura 5

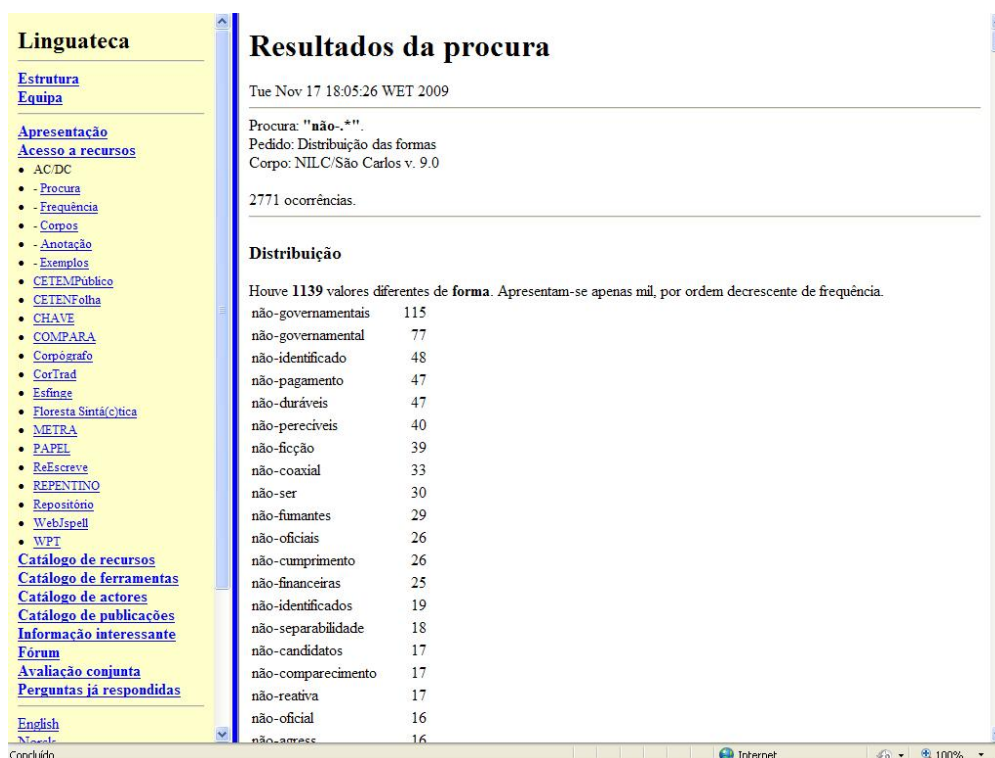


Figura 6

Por fim, procuramos mapear o *corpus* da seguinte forma: durante a análise em contexto, foram feitas anotações codificadas para os casos de natureza semelhante. Feitas as discriminações, utilizamos outro programa, o Atlas.ti versão 5.0. Atlas.ti é um software para análise de dados qualitativos criado por Thomas Muhr e desenvolvido pela Universidade de Berlim, sendo amplamente utilizado por pesquisadores na análise de dados, gráficos, áudios e vídeos. Através do mecanismo de busca, foram criadas seis variáveis ([ausência de; [que não; [binário; [humano; [estilístico; [advérbio) correspondentes a cinco códigos idênticos às marcações realizadas no *corpus* durante a análise. A partir daí, foram contabilizadas as incidências destas variáveis. Os resultados numéricos encontram-se em anexo.

O outro *Corpus* utilizado, o CORPOBRAS, foi manipulado através do programa WordSmith Tools versão 4.0, desenvolvido por Mike Scott da Universidade de Oxford há aproximadamente 10 anos. Sardinha reconhece a eficiência e importância do programa para os estudos lingüísticos: “O programa coloca à disposição do analista uma série de recursos que, bem usados, são extremamente úteis e poderosos 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” (SARDINHA, 2004: 86).

O WordSmith Tools disponibiliza diversos sistemas de busca diferentes, possibilitando a listagem de palavras individuais, multipalavras, agrupamentos lexicais, padrões, palavras-chave, dentre outras possibilidades. Dentre as diversas ferramentas existentes no programa, utilizamos a chamada “*concord*” ou “concordanciador” que gera concordâncias ou lista as ocorrências de um item isolado, a palavra de busca (que pode ser formado por mais de uma palavra) junto ao seu contexto. A palavra buscada foi **não-**seguido de hífen. Os resultados impressos foram posteriormente analisados (anexo 2).

3.1.4. Manipulação dos dados

A pesquisa realizada sobre o *corpus* NILC por meio da ferramenta AC/DC disponibilizado pelo linguatca apresentou 2.771 ocorrências. Foram detectados 1139 valores diferentes para esta forma das quais mil foram verificadas em nossa análise. Esta profusão de ocorrências em um *corpus* composto por textos da modalidade escrita, majoritariamente jornalísticos, demonstra a produtividade do **não-** neste tipo de registro da língua.

Já a análise feita pelo programa Word Smith Tools no dia 18/09/ 2009 às 14:45h a partir dos dados disponíveis no CORPOBRAS, mais especificamente nos subcorpora referentes à língua oral, apresentou resultados bastante distintos na medida em que se detectaram somente dois casos de nomes seguidos da partícula **não-**: não-engenheiro e não-universitário (conforme anexo 2). Não-universitário ocorreu 10 vezes e não-engenheiro apenas uma vez. As duas palavras apareceram no subcorpus Entrevistas Acadêmicas e sugerem que as formas [não-x] são mais profusas em falas do ambiente acadêmico, caracterizadas por uma maior formalidade e pelo uso de termos técnicos.

Lembrando dos dizeres de Basílio (1990: 02), verificamos que:

(...) não faz sentido pensar na produtividade de processos lexicais na língua escrita ou na língua falada, mas sim, em condições de produção. O estudo das condições de produção de construções morfológicas lexicais em diferentes tipos de discurso é tão importante quanto o estudo das condições de produtividade lexical.

Tal assertiva nos leva a concluir que nos discursos jornalísticos, a formação [não-x] mostra-se bastante produtiva. Dentro dos gêneros orais estudados, no entanto, [não-x] apenas é recorrente dentro de uma linguagem mais acadêmica. Neste sentido, a análise das ocorrências presentes no *corpus* oral não se mostrou dificultosa.

Mais trabalhosa, no entanto, foi a percepção detalhada dos dados presentes no *corpus* NILC Folha de São Paulo. Utilizamos como base de análise o recorte de mil ocorrências apresentadas através do mecanismo “distribuição das formas”. A ferramenta, porém, não releva os erros de digitação bastante comuns nas formas não-X que foram eliminadas nestes casos, visto que corrigi-las uma a uma era tarefa que demandava tempo, além de desnecessária já que o tamanho do *corpus* é significativo. Devido aos fatos relatados, optamos por realizar um recorte e analisamos aproximadamente cerca de 25% do *corpus* total.

Durante a análise dos dados em questão, notou-se ainda que muitos dos vocábulos ocorrem repetidas vezes no *corpus*, o que talvez se explique pela reedição de matérias jornalísticas em dias distintos. Esta repetição não é prevista quando da contabilidade feita na distribuição de formas, o que nos leva a concluir que os números apresentados são aproximados e não exatos.

3.2. Análise dos dados

3.2.1. Aspectos Lexicais

Vimos anteriormente que quase todos os autores estudados definem que as formas **não-x** são de natureza derivacional. Apenas PEREIRA (2006) e CUNHA (1983) associam a forma aos casos de composição. A prefixação, por sua vez, é apontada por diversos autores (HOUAISS, 2001; PANTE E MENEZES, 2003; CAMPOS, 2001; AZEREDO, 2009; MOURA NEVES, 1999; MORENO E MARTINS, 2006, ALVES, 2004) como classificação plausível para o fenômeno ou para algumas ocorrências (BASÍLIO, 2000; DUARTE, 1999-A). Os aspectos sintáticos por sua vez, ora são notados (BASÍLIO, idem; DUARTE, idem; ALVES, idem) ora seguem imperceptíveis.

Seria incauto validar a hipótese derivacional somente por meio de sua preponderância acadêmica. Importa, assim, apontar as reais dificuldades referentes à classificação do **não-**, e estudar a questão a partir de exemplos concretos e verificáveis. Seguindo esse caminho, notamos que enquanto pensar o **não-x** à luz dos conceitos referentes à composição gera dúvidas, a prefixação tampouco se mostra uma classificação despida de problemas. Um dos equívocos presentes nas análises já citadas refere-se justamente às justificativas que se apóiam em uma interpretação parcial feita sobre princípio da economia lexical. O uso do advérbio **não-** diante de adjetivos e substantivos é apontado como um recurso econômico do ponto de vista do falante. De fato a idéia procede, porém o critério não é exclusivo de nenhum dos dois processos apontados e não serve para invalidar nenhuma das hipóteses correntes, ao contrário do que sugerem algumas análises.

Um ponto que desfavorece a defesa da prefixação como classificação plausível é o já citado critério fonológico. A existência na fala de pausa entre o **não** e o nome que antecede é característica marcante da composição. A defesa da composição por vezes é feita a partir da tentativa de enquadrar o **não-** junto aos casos de composição em que um dos elementos é uma forma fixa, tal como ocorrem com os já citados exemplos iniciados pelo termo porta (porta-lápis, porta-aviões) que poderia atuar como base para diversas formas inéditas tais como porta-treco, porta-puzzle, porta-bebê, facilmente compreensíveis. Esta análise não é impossível, porém cabe lembrar que o **não-** é partícula que nesses casos pode se adjungir a inúmeros tipos de base, isto é, em nosso estudo notamos que, na fórmula **não-x**, **x** comporta um número praticamente irrestrito de possibilidades. Há mesmo casos em que o **não-** se antepõe a outros advérbios tais como o exemplo: não-magicamente, presente em nosso *corpus*.

A quase universalidade da partícula **não-** quando diante de nomes, adjetivos e participípios também se mostra, à primeira vista, um obstáculo à interpretação que elenca o **não-** junto aos demais prefixos. Vale notar, porém, que os mesmos se distinguem dos elementos composicionais justamente por possibilitarem uma produtividade mais ampla, consoante já dito. Mesmo as chamadas formas composicionais constantes encerram possibilidades restritas quando comparadas aos prefixos da língua portuguesa. Pode se argumentar, por um lado, que prefixos são morfemas e não formas livres. A crítica procede. Mas por outro lado, os prefixos atuaram no passado como preposições ou advérbios, e algo análogo poderia estar acontecendo também no momento atual. O **não-** se assemelharia, assim, a casos como **bem-**querer, **contradizer**, **mal-**acostumado, **sobrea-**

viso, em que as formas classificadas ora como prefixo ora como elementos composicionais, também são formas livres.

Após esta digressão, não soa plausível considerar cada formação a partir de **não-x** como constituindo uma nova entrada lexical. Mais estranho ainda seria considerar o **não-** uma espécie de prefixo universal, capaz de se unir a praticamente todos os nomes da língua portuguesa, inclusive nomes próprios e até advérbios¹⁵, fato que em si mesmo seria suficiente para descartar a hipótese, uma vez que tal característica se mostra incompatível com os demais prefixos da língua pátria, suscitando uma revisão no conceito de prefixação até agora vigente. O mesmo problema se repete quanto à proposição que defende o **não-** como um elemento composicional fixo, pois o enquadramento do **não** requer uma nova definição das composições por formas presas.

O tema é tão controverso que alguns autores assumem a dificuldade: “Qual é a natureza deste processo? Não importa. Sejam formadas por derivação prefixal ou por composição (para muitos, aliás, dois nomes para um mesmo fenômeno), (...) o não, aqui, não é um advérbio de negação, mas sim um elemento da composição do vocábulo”. (MORENO, *sítio virtual*).¹⁶ Vimos que o mesmo autor define o **não-x** como um caso de prefixação, mas aqui deixa claro que a questão não se encontra resolvida.

O grande problema, a nosso ver, está em ignorar os aspectos sintáticos desse processo, que julgamos deveras relevantes, e analisar todos os casos como um grande monólito lexical. BASÍLIO (2000:13) afirma que: “Existem casos, por outro lado, em que o uso do não é mais adequadamente analisado como sintático, apesar do escopo ser nominal”. Efetivamente esta observação, apoiada por Duarte (1999-A), procede nos exemplos constituídos por nominalizações de verbos em que o aspecto sintático fica evidente, bem como nos adjetivos e participípios empregados como adjetivos conforme nossa análise. Esta característica também é ressaltada por ALVES (2004: 28), consoante já citado anteriormente: “a negação lexical permite frases como ‘policiais não-violentos’ e ‘entidades ligadas aos sem-terra’, ao invés de frases sintaticamente mais complexas do tipo ‘policiais que não são violentos’ e ‘entidades ligadas a aqueles que não possuem terra’”. Verificamos a facticidade deste argumento na medida em que a ocorrência **não-X** pode ser facilmente desenvolvida em uma oração, quando x é adjetivo:

¹⁵ Encontramos algumas ocorrências, notadamente com efeito estilístico: “par=Brasil-94b-pol-1: Se considerado o conceito de liquidez (caixa mais algumas aplicações **não-imediatamente** disponíveis), o número batia os US\$ 38,28 bilhões na mesma data”.

¹⁶ Em: <http://wp.clicrbs.com.br/sualingua/2009/10/01/o-nao-como-prefixo/>. Acesso: 20 de novembro de 2009, 14h08min:26seg.

- a) par=61130: A crítica Absurdista realiza o seu distanciamento na cultura, na arte e na literatura modernas invertendo a suposição, até agora não questionada, de que a civilização vale o preço pago em sofrimento humano, angústia e dor pelos **não-civilizados** (os povos primitivos, as culturas tradicionais, as mulheres, as crianças, os proscritos ou párias da história do mundo) e afirmando os direitos dos não-civilizados contra os civilizadores.
- b) par=103563: A Internet é uma entidade **não-comercial**, e seus objetivos principais são a pesquisa (pois a Internet facilita a pesquisa em grupo de elementos dos mais diversos países) e as comunicações em geral, ainda oferecendo uma série de recursos para o usuário comum .
- c) par=Brasil-94a-pol-1: O presidente da 3ª Junta do Trabalho de Belém, juiz Edílzimo Elizário Bentes, deu ontem ganho de causa em primeira instância para mais 16 profissionais **não-diplomados** em jornalismo e sem registro profissional

Nos exemplos citados, o **não-x** pode facilmente ser substituído pela oração [**que não é/está x**]

- d) “(...) afirmando os direitos dos que não são civilizados contra os civilizados.
- e) “A Internet é uma entidade que não é comercial”.
- f) “(...) Edílzimo Elizário Bentes, deu ontem ganho de causa em primeira instância para mais 16 profissionais que não são diplomados em jornalismo e sem registro profissional”.

Percebe-se, assim, que o **não** nesses casos continua operando tal qual um advérbio, modificando o elemento que antecede. BECHARA (2004) chama atenção para o fato de que os advérbios são, geralmente, modificadores do verbo, podendo, no entanto, se referir a adjetivo, a outro advérbio ou a uma declaração inteira. No exemplo “Felizmente, Maria está bem melhor”, o advérbio “bem” modifica o advérbio “melhor” que por sua vez modifica o verbo. Já o advérbio “felizmente” modifica toda a declaração.

Embora não possamos explicar detalhadamente o escopo sintático dessas construções, notamos que em diversos casos a posição de um advérbio atuando sobre o adjetivo soa razoável. Tomemos os exemplos retirados de nosso *corpus*:

- a) A Internet é uma entidade **não-comercial**, e seus objetivos principais são a pesquisa (pois a Internet facilita a pesquisa em grupo de elementos dos mais diversos países) e as comunicações em geral (par=103563).
- b) Problemas com a formação docente, instalações precárias, falta de equipamentos, escolas destituídas de práticas de pesquisas, com hospitais **não-adaptados** ao ensino, tornaram-se freqüentes (par=Cotidiano-94a-soc-2).

Tanto em a) como em b) as formas podem ocorrer com outros advérbios:

- c) A Internet é uma entidade **quase** comercial, embora seus objetivos principais sejam a pesquisa e a comunicação em geral.
- d) Problemas com a formação docente, instalações precárias, falta de equipamentos, escolas destituídas de práticas de pesquisas, com hospitais **mal** adaptados ao ensino, tornaram-se freqüentes.

Certamente a profusão de construções utilizando o advérbio **não** diante de adjetivos parece mais recente na língua. É presumível, por outro lado, que nos casos apontados esteja ocorrendo um enfraquecimento das relações sintagmáticas. O fenômeno merece ser pesquisado a fim de se verificar mais profundamente seus reais desdobramentos e, nesse sentido, o estudo sob o ponto de vista sintático se faz necessário.

Enquadramos nessa mesma interpretação, os casos em que **não-** é seguido de advérbio x-mente:

- a) Se considerado o conceito de liquidez (caixa mais algumas aplicações **não-imediatamente** disponíveis), o número batia os US\$ 38,28 bilhões na mesma data. (par=Brasil-94b-pol-1)

- b) A permanência de Amato até as próximas eleições da entidade seria também a primeira vez desde 1968 que um empresário de fora do Nordeste assumiria a presidência da CNI **não-interinamente**. (par=Brasil-94b-pol-2)

c) Um evento só acontece, dizia Emerson, quando se deixa que aconteça: vale dizer, quando a negação se cala e o evento é visto **não-magicamente**. (par=Mais-94a-nd-2)

Nos casos não-xmente, o sentido adverbial de modo é apenas negado: “vale dizer, quando a negação se cala e o evento é visto de forma não mágica (**não-magicamente**)”.

Quando o **não** aparece frente a casos de nominalizações, sempre com o sentido de ausência de/ falta de, conforme os exemplos:

- a) par=Brasil-94b-pol-2: Para o jornal, o **não-aproveitamento** das verbas se deve em parte à alta rotatividade de ministros do governo Itamar Franco.
- b) par=Brasil-94a-pol-2: Julgava-se responsável pela **não-eleição** do aliado no 1º turno.

O atributo é mais uma vez de ordem sintática, pois embora a nominalização possa ter função designadora, nomeando um ente a partir do significado verbal, a mesma, sob outro ângulo, se assemelha a um predicado nominal, se caracterizando por certo hibridismo, uma vez que dotada de valores verbais e nominais:

Uma forma nominalizada permite referência neutra ao processo verbal, sendo, portanto, uma estratégia básica de estruturação textual, em especial por causa da possibilidade de referência anafórica. Na função de mudança categorial, que obedece sobretudo a motivações de estruturação textual, a nominalização é uma construção transparente e sem objetivos designadores. (BASÍLIO, 1999-B: 63)

A nominalização de formas verbais está muitas vezes perpassada pela motivação gramatical referente à construção do texto, o que se manifesta na estrutura sintática, visto que a forma nominalizada é utilizada para substituir todo um período. A constatação de Basílio (2000) e Duarte (1999-A) é acertada, pois não parece lógico considerar a negação do fato verbal em nominalizações como formadora de novo item lexical. Quanto aos demais substantivos, citamos mais uma vez Bechara, pois este afirma que os advérbios podem ser modificadores de substantivos “quando este é entendido não tanto como substância, mas como qualidade que esta substância apresenta” (2004: 274). O exemplo apresentado por Bechara – “Gonçalves Dias é *verdadeiramente* poeta [grifo do autor]” (idem) é bastante ilustrativo e, a nosso ver, ajuda a explicar as ocorrências **não-x** em que x é substantivo. A construção “José de Alencar é não poeta”, soa similar à frase

apresentada pelo citado gramático. No entanto, a partir da análise realizada, constatamos que o advérbio **não** atuava diante do substantivo também nos casos em que este é entendido como substância e não apenas como qualidade.

Os casos em que **não** antecede substantivo correspondente à forma não nominalizada são minoritários e possivelmente correspondem a essa definição, tais como em:

a) Julgava-se responsável pela **não-eleição** do aliado no 1º turno.
[par=Brasil-94a-pol-2]

b) Outras reclamações comuns dizem respeito a aumentos abusivos, propaganda enganosa, **não-cobertura** de doenças ou serviços e qualidade de atendimento. [par=Dinheiro-94b-eco-2]

c) Ouve-se muito que uma das razões pelas quais a legislação de patentes, aprovada na Câmara dos Deputados, não agradou a comunidade internacional foi a **não-adoção** de princípios que serão acolhidos na legislação do Gatt (Acordo Geral sobre Tarifas e Comércio) , incorporando, dentre outros, a pesquisa no setor farmacêutico. [par=Dinheiro-94b-eco-2]

Porém, notamos que em alguns casos de substantivos modificados por **não**, predomina o aspecto estilístico. Citamos o exemplo: “Uma coisa era matar ou morrer pelas Diretas-Já, pelos quatro ou cinco anos para Sarney, pelo impeachment ou **não-impeachment** (par=Opinião-94b-opi-1)”. Nessa frase o **não-** tem sentido claro de contrariedade e a escolha pelo **não-** parece válida para realçar a oposição e tornar o texto expressivo. A vivacidade textual perpassa os demais exemplos:

a) Agora que já passou a onda maior, podemos reler e avaliar o que Ricupero disse no ar, ou melhor, o que não disse no ar ou, melhor ainda, o que disse no **não-ar**. (par=Ilustrada-94a-nd-1)

b) Porque os são-paulinos do Santos é que colaboraram: o Gilberto colaborou com o gol e o Macedo colaborou com o seu **não-gol** escandaloso. (par=Esporte-94b-des-1)

c) Se o STF adotar o critério político no julgamento de Collor, se não resistir à pressão da mídia e da opinião que ela plantou na sociedade, teremos a **não-pizza** que foi guindada a supremo valor moral da nação. (par=Opinião-94a-opi-2)

- d) A mudança, o **não-jeitinho**, não estaria levando a nada melhor.
(par=Revista-94b-nd-1)

CARDOSO (2006) já havia percebido o potencial literário do **não-** anteposto às palavras quando versa sobre as questões estilísticas na obra de Carlos Drummond de Andrade. O autor percebe o uso de duas formações assaz interessantes: não-domingo e o não-Laforgue: “Que aventura doida/ no domingo livre/ estarão desfiando/ enquanto eu sozinho/ contemplo escorrer a lesma infundável/ do meu **não-domingo**?” (CARDOSO, 2006: 21). Em outro exemplo, Drummond utiliza um nome próprio: “Era preciso? / fazer um verso **não-Laforgue**/ à base desse novo sentimento” (Idem). O recurso estético é óbvio e a semântica fica totalmente atrelada ao conhecimento de mundo requerido pelo leitor. O não-domingo é certamente o domingo que por algum motivo ficou desprovido das características comumente associadas a este dia da semana, tornando-se, assim, um pseudo-domingo¹⁷, ou um domingo atípico, que foge às características ou sentidos esperados. O verso não-Laforgue¹⁸, da mesma forma, é o verso oposto à estética poética de Laforgue, autor francês.

Percebemos que nessas formações os significados são apreendidos com certa facilidade, embora por vezes se construam de forma mais abstrata e se refiram a situações bastante específicas. Retornando aos exemplos extraídos de nosso *corpus*, a não-pizza não é a negação da preparação culinária feita à base de trigo, queijo e tomate, mas sim a negação com base na expressão popular “acabar em pizza”. Da mesma forma, o não-gol se refere a possibilidades limitadas: um gol contra, um gol que acabou sendo anulado ou um gol esperado que não ocorreu. Os sentidos, portanto, flutuam entre as acepções pseudo, anti- ou contra e são mais ou menos intuitivos. Caso excepcional é o não-ar

¹⁷ O sentido de “pseudo x” ou “proximidade” foi encontrado em outros recortes de nosso corpus com o sentido estilístico. Citamos mais um exemplo, em que tal sentido é auto-explicável a partir do contexto: “par=43427: (11) Os acontecimentos da primeira metade do pacto são, na verdade, não-acontecimentos ou acontecimentos apenas virtuais, anulação e presença potencial de tudo o que poderia acontecer”.

¹⁸ Vale notar que encontramos no corpus outras duas ocorrências de não- anteposto a nome próprio, ambas com o nome Glauber, ou seja, Glauber Rocha. Citamos uma delas: O título tem dois significados: descrente em alemão e não-Glauber, o anti-Glauber [par=Ilustrada-94b-nd-1:]. A questão dos nomes próprios encerra grande polêmica dentro dos estudos de semântica. Este assunto obviamente extrapola a temática de nossa pesquisa, porém, consideramos de grande valia citá-la. Assim, lemos em Frege: “O sentido de um nome próprio é entendido por todos que estejam suficientemente familiarizados com a linguagem ou com a totalidade de designações a que ele pertence; isto, porém, elucida a referência, caso ele tenha uma, mas de uma maneira sempre parcial. Para um conhecimento total da referência, exigir-se-ia que fôssemos capazes de dizer, imediatamente, se um dado sentido pertence ou não a essa referência. Isto, porém, nunca conseguiremos” (FREGE, Gottlob. *Lógica e filosofia da linguagem*. São Paulo: Cultrix, 1978: página 73).

(exemplo a). Observando a frase, percebemos que o “não-ar” parece ser uma entidade de difícil definição, a não ser para os que conhecem exatamente a situação em que o então ministro da fazenda Rubens Ricupero se encontrava.¹⁹ O não-ar neste caso é o mesmo que “fora do ar”. A dificuldade possivelmente advém da polissemia presente na palavra “ar” que gera uma ambigüidade provavelmente intencional, visto que a própria situação a que o texto se refere se caracterizou pela falta de clareza dentro da realidade política do país.

A criação de palavras com intuitos estilísticos em criações literárias e jornalísticas possui particularidades relevantes na medida em que as escolhas, feitas de modo consciente pelos autores, têm geralmente a intenção de surpreender e chamar a atenção dos leitores. Embora alguns teóricos não reconheçam como sendo relevantes as formações dessa natureza, na medida em que muitas vezes não são incorporados pelos falantes, MARTINS (2003: 111) por outro lado, reitera que:

A idéia de que vocábulos que não se incorporam na língua não têm interesse estilístico é bem discutível. Primeiramente porque não podemos antever o destino dos vocábulos forjados por um escritor (...). Ainda que as novas palavras tenham existência efêmera, elas revelam um meio de o falante realizar o seu desejo de expressividade. Muitas delas são realmente de emprego restrito, e não poucas se limitam a uma ou outra ocorrência, da mesma forma que as metáforas que se criam para um único enunciado. Mas, pela sua novidade, causam um inegável efeito expressivo que não se pode menosprezar.

Deste modo, assumimos que as ocorrências estilísticas **não-x** nos textos escritos, embora minoritárias e alheias aos dicionários e ao discurso oral, são formas inovadoras, porém ocasionais, de caráter fugaz e sem permanência no idioma.

Do mesmo modo, constatamos a existência de ocorrências de cunho binário quando **não** se combina com adjetivos, tornando-os negativos em criações de termos matemáticos, jurídicos ou pertencentes ao jargão científico. Nesses casos, o **não-** adjunge a especificidade da negação ao significado original do adjetivo. Enumeramos em nossa pesquisa algumas situações de clássica dicotomização de uma realidade. Citamos aqui os exemplos não-A, em contexto referente à teoria matemática, não-hodking (um tipo de tumor), não-beneficiado (tipo de arroz também conhecido como integral). No entanto, nem sempre é intuitiva a percepção do antagonismo, sendo preciso conhecer a

¹⁹ Ora, é sabido que no referido ano o personagem citado, Rubens Ricupero, esteve envolvido naquilo que ficou conhecido como o “escândalo da parabólica”, em que declarações do então ministro foram transmitidas ao vivo pela televisão, sem o que o mesmo tivesse conhecimento do fato. Sem a posse destes elementos, contudo, a compreensão do jogo de linguagem ficaria incompreensível.

realidade científica a que se refere. Nos exemplos “jogos não-cooperativos”²⁰ e “não-localidade”²¹, a importância do conhecimento de mundo fica patente.

Em diversos momentos, portanto, foi preciso recorrer à busca em dicionários e em pesquisas através da internet, mesmo em casos aparentemente simples, objetivando alcançar o sentido do elemento que estava sendo analisado. A fim de melhor compreender o significado da palavra “não-tecido”, por exemplo, recorreu-se à procura pela palavra em contexto diverso através do site de buscas Google onde encontramos a forma *nãotecido*. O sítio Met@lica²² hospeda a seguinte definição:

‘Nãotecido’ é uma estrutura plana flexível e porosa constituída de véu ou manta de fibras ou filamentos, orientados direcionalmente ou ao acaso, consolidados por processo mecânico (fricção) e/ou químico (adesão) e/ou térmico (coesão) ou a combinação destes. Ou seja, as fibras não são tecidas pelo modo convencional através de tecelagem ou malharia. Com produtos direcionados para diversas áreas e utilizações, como agricultura, decoração, ecologia, construção, indústria, embalagem, confecção, medicina, proteção, geotêxtil, transporte e esportes, entre outras, os ‘nãotecidos’ já podem ser encontrados em vários produtos, desde forrações para urnas funerárias, passando por alto-falantes, fitas indicadoras de gravidez, panos de limpeza e roupas especiais e de segurança, até papéis de parede, travesseiros e fitas impressoras para máquinas, entre outras utilizações.

Entendemos, após esta explicação, que não-tecido constitui uma nova formação designadora de um tipo específico de objeto que contém características similares a um tecido em sua aparência, mas que devido a diversas outras peculiaridades, não pode ser enquadrado como tal, sendo classificado em uma categoria própria pela indústria têxtil. Embora a palavra se assemelhe às já citadas formas estilísticas, ela parece melhor classificada quando analisada junto aos termos de cunho taxonômico.

Ressalte-se que nesses casos de classificações duais, o **não-** passa a ser, em certo sentido, utilizado como um operador lógico. A polarização possui o escopo de transformar uma variável (econômica, sociológica, jurídica, estatística) em uma variável bi-

²⁰ Atrelado à teoria dos jogos, organizada por John von Neumann e Oskar Morgenstern em 1944. O equilíbrio em jogos não-cooperativos foi desenvolvido pela matemática posteriormente, tendo se destacado o teórico John Nash, após criação do que ficou conhecido como “equilíbrio de Nash”.

²¹ Não-localidade em mecânica quântica, se refere à propriedade de estados quânticos emaranhados no qual dois estados emaranhados “colapsam” simultaneamente no ato de medição de uma das componentes emaranhadas, independente da separação espacial entre os dois estados. Essa “estranha ação-a-distância” é o conteúdo do Teorema de Bell e do paradoxo EPR.

²²

http://www.metalica.com.br/pg_dinamica/bin/pg_dinamica.php?id_jornal=6467&id_noticia=195&id_pag=929

Último acesso: 00:24 dia 10 de janeiro de 2010.

nária, marcada por sua neutralidade axiológica. O aspecto binário do **não-** foi também previsto por Moreno e Martins:

O uso do não como prefixo foi uma das grandes novidades com que a língua nos brindou no fim do século XX, permitindo que tudo possa ser dividido em duas categorias complementares, X e não-X — o que constitui uma ferramenta muito útil no discurso argumentativo. Com esse providencialíssimo não, podemos criar divisões binárias de praticamente tudo o que quisermos: os votantes e os não-votantes, os alfabetizados e os não-alfabetizados, os hispânicos e os não-hispânicos, os marxistas e os não-marxistas. Ele até nos permite falar no não-eu ou no não-ser, vocábulos que seriam impensáveis com nossos prefixos negativos clássicos, o in- e o des- (...).(MORENO e MARTINS. 2006: 169)

Também nos casos em que x é substantivo de referentes humanos, gentílicos e nomes que denotam profissões, há formações de novos itens lexicais designadores de seres. Essa acepção é também marcante quando diante de formações que designam filiações políticas, partidárias ou ideológicas. É o caso de termos como não-marxistas, não-carlistas²³, não-sexista, não-petista, não-stalinista, etc.²⁴ Quanto a essas ocorrências, há que se levar em conta a mútua conversão entre substantivos e adjetivos muito presente em nosso idioma, bem como a extensão de propriedades de uma classe para outra. É comum, então, que nomes de agente sejam usados como adjetivos, em recurso morfológico conhecido como flutuação. O mesmo ocorre em nomes pátrios que ora agem como adjetivos de caracterização de proveniência/origem, ora denotam indivíduos, sendo, portanto, substantivos. Nesses casos, ressaltamos que quando antecidos de **não-**, somente aqueles que se referem a elemento humano possuem valor lexical predominante.

As diferenças são notáveis:

- a) Mandou evangelizar as áreas **não-católicas** [par=1489].
- b) Interessados em arte e arquitetura, ainda que **não-católicos** ou mesmo ateus, dispõem nesta Semana Santa de um bom pretexto para conhecer o variado patrimônio eclesiástico da cidade [par=Cotidiano-94b-soc-1].

²³ Todos os exemplos foram retirados do corpus analisado. Carlismo: De Antônio Carlos Magalhães, o ACM. No Brasil, Carlismo é a designação dada ao movimento político nascido na Bahia a partir do poder criado no Estado pelo político Antonio Carlos Magalhães, morto em 2007, ainda hoje lembrado como um dos mais influentes “caciques políticos” da história do nordeste.

²⁴ Muitos dos termos citados terminam com o sufixo –ista. Sobre este sufixo, lê-se no dicionário eletrônico Houaiss: “ainda, do s.XIX e no s.XX, seu uso se disseminou para designar movimentos sociais, ideológicos, políticos, opinativos, religiosos e personativos, através dos nomes próprios representativos, ou de nomes locativos de origem, e se chegou ao fato concreto de que potencialmente há para cada nome próprio um seu der. em –ismo.

Diante do exposto, concordamos com a proposta de que o **não-** quando anteposto a substantivo ou adjetivo nem sempre constitui lexema. No entanto, a pergunta que nos propusemos a responder segue sem definição: o **não-** quando formador de palavra atua como prefixo ou como elemento composicional?

Retomando a discussão, concluímos que as duas propostas resguardam suas limitações. Não descartamos a possibilidade de classificar as formas como composicionais, porém visualizamos que a prefixação, por suas características, é a categoria mais adequada para abrigar o fenômeno. Sabemos que a alta produtividade verificada também justifica a natureza prefixal pleiteada por alguns autores: formações regulares de natureza não lexical também se caracterizam pela frequência de uso, não servindo esse fato como parâmetro para a constituição de novas unidades, o que explica a profusão de formas **não-** diante de nominalizações ou de adjetivos em construções sintáticas. Porém, conforme demonstrado, **não-** quando diante de elemento humano ou quando possui intuito de gerar sistemas binários ou formas estilísticas, se comporta como um formador de item lexical. Ainda nesses casos, o **não-** se mostra mais sistemático e produtivo do que as formas composicionais presas. Note-se ainda que o **não** possui um significado bastante generalizante, o que destoaria dos demais elementos composicionais, que se caracterizam por serem mais particulares. Nesse sentido, estaria ocorrendo uma lexicalização do advérbio nos casos em que atua como prefixo.

3.2.2. Aspectos Semânticos

Após os dados apresentados, resta ainda abordar as questões semânticas pertinentes a nossa pesquisa. Vimos que quanto aos possíveis significados gerados pelas novas formas, não existe concordância plena. Alguns autores afirmam que o **não** encerra um único sentido – oposição perfeita (MORENO E MARTINS, *idem*), privação/negação/falta (CAMPOS, *idem*), negação (AZEREDO, *idem*) -, enquanto outros vêem nessas construções casos de polissemia (PANTE e MENEZES, *idem*).

Conforme nossa análise, a ideia de que o **não-x** possui sentido único de falta/ ausência elucida somente os casos em que x é substantivo sem objetivo estilístico, enquanto os casos em que x é adjetivo ou particípio possuem viés sintático. Quando x atua

como forma substantiva não nominalizada, em que o caráter estilístico é sobressalente, os sentidos podem ser de anti-/ pseudo-/ ou contrariedade. Nos casos em que não- se antepõe a nomes humanos ou quando engendra classificações binárias, o sentido que encerra é o de oposição. Arrolamos a seguir os possíveis significados exemplificando as ocorrências.

a) Quando **não-x** possui sentido de negação simples:

Ocorre quando x é adjetivo ou advérbio.

- Erpen, corregedor do Tribunal de Justiça do Rio Grande do Sul, diz que o atual Certificado de Registro e Licenciamento de Propriedade de Veículo, expedido pelo Detran (Departamento Nacional de Trânsito) , é apenas um meio, **não-óbvio**, para comprovar a propriedade. [par=Cotidiano-94a-soc-1]

- A sentença foi anulada porque Souza foi defendido por um leigo, **não-inscrito** na Ordem dos Advogados do Brasil (OAB). [par=Cotidiano-94a-soc-2]

E se x é substantivo nominalizado ou substantivo sem cunho estilístico.

- Julgava-se responsável pela **não-eleição** do aliado no 1º turno [par=Brasil-94a-pol-2]

- Vamos culpar o governo pelo **não-atendimento**, afirmou o subsecretário executivo da Federação, Danilo Maciel [par=Cotidiano-94a-soc-2]

b) Quando **não-x** pode significar contra/ pseudo/ anti- / ou outros.

Nos casos em que há formação de palavras com objetivos estilísticos. São os únicos casos em que pode ser possível algum tipo de polissemia, ainda que rara.

- É o último herdeiro do som pelo **não-som**, da economia de frases musicais. [par=Ilustrada-94b-nd-2]

c) Quando expressa oposição:

Se formador de classificações duais ou se diante de elemento humano.

- Entre os exemplos de jogos **não-cooperativos** estão o xadrez e o pôquer [par=Mundo-94a-pol-2].
- As taxas de matrícula do Sesc custam entre Cr\$ 650,00 e Cr\$ 2.600,00, para comerciários e **não-comerciários**. [par=Cotidiano-94b-soc-1]

Resta ainda, elucidar a questão da antonímia, tratada por Moreno e Martins (2006) e Campos (2001). Moreno e Martins definem o **não-x** como sendo uma possível tentativa de formar antônimos perfeitos. Já Campos afirma que: “A frequência de uso do não com FB’s²⁵ destituídas de antônimos lexicais ou formados com PTs, a exemplo de violência/não-violência, poderia ter gerado, por analogia, a prática da substituição dos prefixos tradicionais pelo não prefixal”. (CAMPOS, 2001:151). O autor pretendeu, assim, associar as formações em **não-x** à necessidade de criação de antônimos até então inexistentes em nossa língua e neste sentido, demonstrou que muitas das ocorrências **não-x** em seu *corpus* não possuíam antônimos formados a partir dos prefixos tradicionais²⁶. O não- segundo essa acepção, estaria preenchendo uma lacuna, o que parece razoável. Fato, contudo, é que algumas formas possuem antônimos constituídos por tradicionais prefixos de negação e ainda assim comportam a anteposição do **não-**. Citamos como exemplo:

- a) par=Imóveis-94b-eco-2: O professor Walderez Gambale, 47, do Instituto de Ciências Biomédicas da Universidade de São Paulo, alerta que a concentração de fungos atinge níveis **não-aceitáveis** se a umidade relativa do ar for superior a 60 % [inaceitáveis].
- b) par=Brasil-94b-pol-1: O Serviço Social da Indústria (Sesi) de Minas Gerais mantém um programa de educação formal e **não-formal** que atinge cerca de 40 mil pessoas, entre adultos e crianças. [informal]

²⁵ Sobre as siglas: Forma Base (FB) e prefixos tradicionais (PT).

²⁶ Esta nomenclatura, assim como “antônimos lexicais”, foi adotada pelo autor e achamos de bom tom segui-la.

Nesses casos, segundo o autor, as construções lexicais estariam atuando de maneira análoga às formas desprovidas de antônimos.

A conclusão do autor de que “o não prefixal tem sido empregado, satisfatoriamente, junto a FBs que não possuem antônimos, nem lexicais nem constituídos pelos processos tradicionais de antonímia”²⁷ (Campos, 2000: 114) é válida para justificar este novo emprego do **não-** em alguns casos, porém não é suficiente para caracterizá-lo como um processo de formação de palavras e tampouco um caso solucionado de prefixação. Por outro lado, a idéia defendida por CAMPOS não se sustenta diante da questão produtividade-listagem pela ação do bloqueio. O bloqueio é uma ausência de aplicação de uma forma devido à existência prévia de outra forma que exerceria a mesma função. A redundância sempre seria evitada a partir do conhecimento da lista lexical.

Nesse sentido, julgamos válido notar que dentre as ocorrências listadas pelo mecanismo de distribuição de formas, um número expressivo e majoritário ocorre de maneira esparsa e exclusiva. A análise numérica revela que 86,3% das palavras listadas ocorreram no máximo três vezes no *corpus*, o que demonstra que não-x é bastante produtivo; adicionalmente, **não-** pode ser adicionado à maioria dos vocábulos da língua, negando-os. Em não-x, x pode ser tal que talvez constitua ocorrência única e não mais se repita. No entanto, as mesmas ocorrências não são correntes ou possivelmente são evitadas, havendo a preferência pelas construções tradicionais, o que se confirma na língua oral com a análise do CORPOBRAS PUC-RIO, em que as formações não-x somente aparecem em contexto acadêmico.

A questão da antonímia é demasiado complexa, vista a dificuldade em estabelecer a existência de antônimos perfeitos. Perini, por exemplo, é contundente quando aborda o assunto. O autor defende a impotência lingüística do falante e mesmo do pesquisador em delimitar com exatidão as fronteiras da antonímia. Ao analisar os pares normal-anormal, velho-novo, bom-mau, emigrante-imigrante e pré-nupcial - pós-nupcial, Perini conclui a inexistência de qualquer relação lógica entre as contrariedades expressas em cada caso. Este sentimento fica óbvio em sua conclusão: “Aqui a situação e de tal heterogeneidade que me parece conveniente abandonar, pelo menos por ora, a noção de ‘antonímia’. Enquanto não se obtiver uma conceituação muito melhor do que a atual, essa será uma noção sem utilidade nenhuma” (PERINI, 2007: 250)

²⁷ Investigamos as ocorrências no corpus NILC Folha de São Paulo e concluímos que 580 ocorrências não possuíam antônimos formados a partir de prefixos tradicionais, ou seja, aproximadamente 84,82% das palavras que compunham o corpus escrito.

Não adentrando nas possíveis definições e problemáticas pertinentes quanto à caracterização dos antônimos, lembramos que a análise do **não-** sob essa ótica deve levar em conta que o léxico em seu funcionamento atua conforme o já citado princípio da economia. Quando da constituição de um novo item lexical, haveria, o bloqueio heterônimo pelo qual uma palavra não se forma pela existência de outra, completamente diferente (um heterônimo), com o mesmo significado. Assim, “aviãozeiro” não se formaria pela existência de “piloto”. Haveria também o bloqueio homofônico em que uma palavra não se forma com determinado significado porque já existe na língua com outro significado. Ou seja, “cobreiro” com o significado de “trabalhador em minas de cobre” não se forma pela existência prévia de “cobreiro”, que nomeia uma doença. Esta é uma visão de um léxico ativo, circundando em torno da produtividade, do possível e do impossível dentro do âmbito da criação. Seguindo essa lógica, o uso do prefixo **não-** nos casos em que o nome posposto já possui antônimo formado por prefixos tradicionais, demonstra a atuação dessa partícula em nova formação denota uma necessidade inédita e diferente.

Diante do que foi exposto, é viável que o **não-** esteja atuando como antônimo, porém, não acreditamos que a forma **não-x** esteja competindo com os prefixos tradicionais, mas sim que possivelmente constitua uma forma dotada de novo sentido. Não sabemos se a idéia de “antonímia perfeita” seria linguisticamente possível em nosso idioma, na medida em que enseja discussões de ordem semântica aprofundadas, porém em uma escala gradativa, assumimos que não-x esteja mais próximo a 0 (zero) x do qualquer outro prefixo. (acho complicada esta afirmação, na medida em que você colocou vários significados, de modo que as posições na escala seriam um tanto diferentes – e há possibilidades abaixo de zero (aliás, seria bom dar uma olhada na negação com “zero”, mas deixe isto para o doutorado).

Outros fatores²⁸ não estudados podem ser significativos para explicar a escolha de **não-x**. De todo modo, a fim de ilustrarmos as possíveis diferenças que motivam a escolha pelo **não**, citamos o termo “não-judeu”, encontrado em nosso *corpus*. Notamos a existência do antônimo lexical “gentio”, ou o vocábulo estrangeiro “goyin”. As duas

²⁸ Esta questão chamou minha atenção em uma situação acadêmica real. Durante um congresso de sociologia, ouvi um dos debatedores utilizar o termo “não-competência”. Após o fim do debate, dirigi-me aos participantes e perguntei o que entendiam por não-competência e os mesmos assumiram ser o termo sinônimo à incompetência. Perguntados sobre o porquê de suas escolhas, responderam que “incompetência” trazia um sentido também extrajurídico e que o mesmo era notadamente negativo significando inabilidade, ignorância. A escolha pelo correspondente não-competência parecia mais neutra.

palavras, nota-se, encontram-se carregadas de sentido pejorativo que adquiriram ao longo da história, sendo por isso atualmente preteridas em detrimento de “não-judeu”²⁹. Já em “idosos não-dependentes”, termo de cunho jurídico, “não-dependente” difere de “independente” por seu caráter científico, referindo-se a pessoa que não carece das condições financeiras necessárias para custear sua subsistência e que, para efeitos legais, não depende de outra.

Deste modo, percebemos que a compreensão abrangente sobre a questão semântica requereria um estudo mais pormenorizado sobre as diversas formas de negação léxica, e nesse sentido, nossa análise pretende dirimir apenas brevemente alguns dos aspectos pertinentes ao tema. Fica perceptível, contudo, que o **não-** carrega maior neutralidade em relação ao prefixo de negação **in-**, que possivelmente remete o falante a interpretações mais negativas ou ao **des**³⁰ - já marcado pela idéia de reversão. Esse aspecto possivelmente justifica a profusão de formas **não-x** nomeando realidades científicas e acadêmicas, porém assumimos que existe a necessidade de estudos mais aprofundados sobre o tema.

²⁹ A palavra gentio designa um não-israelita e deriva do termo Latim "gens" (significando "clã" ou um "grupo de famílias") e é muitas vezes usada no plural. Os tradutores cristãos da Bíblia usaram esta palavra para designar colectivamente os povos e nações distintos do povo Israelita. A palavra é especialmente importante em relatos sobre a História do Cristianismo, para designar os povos Europeus que gradualmente se converteram à nova religião, sob a influência do apóstolo Paulo de Tarso (São Paulo) e outros. O próprio São Paulo nascera na actual Turquia mas tinha sido educado no Judaísmo. A partir do século XVII o termo é mais normalmente usado para referir não-judeus. Com o mesmo sentido de gentio existe o termo goy. Em tempos recentes, ambos os termos deixaram de ser bem vistos, preferindo-se muitas vezes usar a expressão "não-Judeu" como substituto.

Outras acepções: 1 - Quem segue o paganismo; 2 - Quê ou o que não é civilizado; 3- S. m. Pop. Grande porção de gente; (Dicionário Completo da Língua Portuguesa Folha da Tarde) <http://pt.wikipedia.org/wiki/Gentio> Acessado às 11h22min do dia 26 de outubro de 2009.

³⁰ O prefixo des- é, segundo Houaiss, um dos mais prolíficos uma vez que o mesmo pode denotar não apenas oposição, mas também a cessação de um estado anterior ou a má execução de algo (exemplo – um desgoverno) quando diante de substantivos. Outro sentido possível ao des- é o meramente expletivo.