

AONDE VAMOS EM RELAÇÃO A AONDE*
Where to Find the Portuguese Non-Preferred Form *Aonde* (Where/Where to)

Diana SANTOS
(Linguatca, SINTEF ICT, Oslo, Noruega)

Abstract

*What are the issues underlying the use of *aonde* versus *onde* in Portuguese (both translated by *where*), a favourite target for prescription-oriented philologists? In this paper, we investigate this question by (1) performing an overview of the reference literature; (2) creating a set of hypotheses about the proper use of *aonde*; (3) trying to check them in corpora. In addition to giving some insight in the question, the paper illustrates the possibilities (and shortcomings) of corpora studies.*

Key-words: *corpus linguistics; grammar; Portuguese; grammatical words.*

Resumo

*Este artigo versa sobre a palavra *aonde* e sua relação com *onde*, um dos assuntos favoritos da gramática normativa. Começamos por apresentar uma panorâmica da literatura sobre o uso dessas formas, seguida da formulação de algumas hipóteses e sua verificação em corpora e na rede WWW. É também brevemente discutida a questão da variante do português e algumas vantagens e problemas de estudos baseados em corpora.*

Palavras-chave: *lingüística com corpora; gramática; língua portuguesa; palavras gramaticais.*

1. Introdução

Neste artigo, debruço-me sobre a palavra *aonde* para investigar que padrões regem a sua utilização e a da palavra *onde*: é um erro?¹ Um regionalismo? Uma simples variante de *onde*? Ou em variação paradigmática com esta? Como segundo objectivo, pretendo identificar o uso de acervos textuais (*corpora*) e sobretudo da rede (World Wide Web) como ferramenta para o estudo de uma palavra ou construção em português.

2. Pesquisa em material de referência

A consulta que fiz a dicionários, prouários, gramáticas e manuais de estilo para ver o que diziam sobre a palavra *aonde* foi muito pouco conclusiva, como passo a descrever: como seria de esperar, a maior parte dos dicionários monolíngues que consultei (Costa, s/d; *Lello Ilustrado*, 1990; Figueiredo, s/d; *Lello Escolar*, 1978; Vilela, 1990) limitavam-se a indicar o sentido, não o contexto do uso. Por vezes, apresentavam também a origem etimológica (que ainda é transparente, a contracção da preposição *a* com a palavra *onde*). Como nota interessante, observe-se que *aonde*, assim como *donde*, encontram-se ausentes de Ramalho (1985), dicionário *estilístico* da língua portuguesa. Em dicionários bilíngues, a palavra *aonde* não era simplesmente problematizada, aparecendo como uma entrada plena nos poucos que consultei (Ferreira, 1983; Fernandes, 1991; *Dicionário Português-Ingês*, 1989).

*Este trabalho foi parcialmente executado no âmbito da Linguatca, financiada pela Fundação para a Ciência e Tecnologia portuguesa, projecto POSI/PLP/43931/2001, e co-financiada pelo POSI. O presente artigo foi apresentado em 1999 no 10º INPLA – LAEL - PUCSP, e revisto para publicação em Agosto de 2000.

¹ Ou, de forma menos prescritiva, identificar se é geralmente considerado como erro.

As gramáticas, por outro lado, em geral não se dedicam a palavras específicas (sendo que *aonde* e *onde* parecem encontrar-se numa zona cinzenta entre as palavras gramaticais e as palavras de sentido pleno). Note-se, aliás, que se a palavra *aonde* for um regionalismo ou um erro, ou tiver um emprego muito específico, é natural que não seja sequer mencionada. De facto, os exemplos encontrados nas gramáticas – ou livros sobre gramática – são escassos ou mesmo inexistentes. Em Miguel (1989:147-149), não consegui encontrar referência a, ou uso de, *aonde*, apenas exemplos de *onde*. Martins, por outro lado, não cita *onde* nem *aonde* quer como pronomes interrogativos ou relativos nem como advérbios, mas nas suas abonações é possível encontrar ocorrências de ambos (1982:160, 166, 261, etc.). Em Bechara (1988:71-72, exs: 117-118) é possível encontrar exemplos de *onde* e de *aonde*, assim como algumas notas sobre o seu emprego (1988:121). Por outro lado, Mateus et al. (1989:241) não só não fazem distinção entre as duas formas, como apresentam, como exemplo, precisamente aquilo que parece ser objecto de crítica acérrima nas outras obras. Cunha & Cintra (1984:351), mais cautelosos, afirmam que a distinção – “o contraste que a disciplina gramatical procura estabelecer, na língua culta contemporânea, entre *onde* (= o lugar em que) e *aonde* (= o lugar a que)” – está praticamente anulada na linguagem coloquial.

Em obras por sua vez consagradas aos problemas de linguagem (Nogueira, 1989:270; Estrela, 1984:86; Freire, 1983:186), foi possível encontrar referência a usos errados e certos de *onde* vs. *aonde*, referência essa, contudo, não particularmente elucidativa em relação ao que nos preocupava. Os trechos encontrados ilustram, aliás, claramente os limites desse tipo de obras quando não se está em presença dos verbos *ir*, *vir*, *estar* ou *morar*. Veja-se, por exemplo, Nogueira (1989:270): “Diga-se “*Onde* moras?” (quietação); “*Donde* vens?” (proveniência); “*Aonde* vais?” (destino sem demora); “*Para onde* vais?” (destino com demora); “*Por onde* passaste?” (passagem através de)”. Idêntico resultado obtemos da consulta (indirecta) ao Ciberdúvidas². O próprio uso de *aonde* pelos especialistas desse serviço fornece, contudo, material bem mais interessante, como os leitores poderão confirmar se o visitarem.

Finalmente, consultámos uma obra recente, Peres & Mória (1995), que analisa material jornalístico autêntico – construída, pois, em moldes muito diferentes dos das anteriores – mas que, apesar de conter um capítulo dedicado às orações relativas, não analisa essa questão específica. O livro compreende, contudo, uma crítica à utilização do pronome relativo *onde* em casos que não envolvem valor locativo – por exemplo, envolvendo os predicados *envolver* ou *participar*, ainda que considere perfeitamente válidos casos de outros “espaços” conceptuais tais como os exemplificados pelas frases *Onde o Paulo se desembaraça melhor é na matemática* e *A parte do discurso onde a Ana foi mais convincente foi [...]* (Peres & Mória, 1995:302).

3. Formulação das perguntas

São três as questões a que queremos responder com esta investigação sobre o uso de *onde* e *aonde* e sobre a sua possível correcção ou incorrecção. A primeira é: Com que verbos é lícito (ou habitual) o uso de *aonde*? Qual a delimitação – sintáctica, semântica, sintáctico-semântica ou outra – a que é possível chegar?

Do (pouco) que é dito na literatura consultada, algumas hipóteses podem ser formuladas. Passemos à sua explicitação: o uso de *aonde* é lícito nas frases que incluam:

- Hip 1. o verbo IR (incluindo todas as construções com o verbo IR como auxiliar);
- Hip 2. verbos pedindo preposição *a* mais lugar;
- Hip 3. verbos incorporando a noção de destino;
- Hip 4. verbos de movimento em que o agente/sujeito se move ("automovimento");

² <http://www.ciberduvidas.com/>

Hip 5. verbos de movimento;

Hip 6. verbos de movimento em que o movido é humano ou animado.

Como é evidente, as condições expressas nas várias hipóteses não são mutuamente exclusivas. De facto, todas elas são generalizações do exemplo *Aonde vais?*, o que é, evidentemente, uma condição necessária para que representem a realidade lingüística. De qualquer forma, também não pretendemos chegar à conclusão de que apenas uma dessas delimitações é correcta. Pelo contrário, acreditamos que quanto mais condições desse tipo forem verificadas, mais aceitável será o uso de *aonde*, dando razão à teoria da categorização lingüística advogada por Dahl (1985), Lakoff (1987) e Sweetser (1990).

Desenvolvendo em pormenor cada uma dessas hipóteses, observamos que a Hip. 1 nos leva a esperar encontrar *aonde* junto a construções com o verbo auxiliar IR, tais como IR BUSCAR, IR FAZER, IR ARRANJAR, IR PUBLICAR, etc. E, se estendermos a noção de IR ao futuro (consonante com o facto de que é muito mais habitual dizer *Aonde vais?* do que *Aonde foste?* – de facto a versão não marcada é *Onde estiveste?*), a Hip. 1 assim alargada prevê que seja mais usual encontrar *aonde* em frases que se referem ao futuro do que em frases sobre o passado.

De acordo com a Hip. 2, por outro lado, somos levados a supor que encontraremos *aonde* com qualquer verbo que peça *a* com um complemento de localização concreta ou abstracta (ou seja, em que a preposição *a* indique posição e não, por exemplo, benefício), tais como LEVAR, ACORRER, VOLTAR, VIR, TRAZER, etc.

Se, por sua vez, a Hip. 3 estiver correcta, e a palavra *aonde* se encontre associada a verbos com destino semântico (independentemente da preposição associada), também a encontraremos com PÔR, ENTRAR, DESAGUAR, etc., que coocorrem com *em*, e com ATIRAR, EMPURRAR, MANDAR, etc. que são utilizados com *para*. Admitindo, além disso, *donde* como o paralelo de *aonde* em relação à preposição *de*³, poderíamos considerar (nova hipótese Hip. 3') que *donde* seria usado com verbos pressupondo uma origem semântica, donde deveríamos encontrá-lo coocorrendo com os verbos TIRAR, RETIRAR, TRADUZIR, SAIR, PROVIR, etc., todos requerendo a preposição *de*.

Finalmente, as hipóteses Hip. 4, 5 e 6 foram inicialmente formuladas em termos da categoria "verbos de movimento" embora essa expressão (= verbos que contêm informação sobre a maneira como o seu sujeito se move) não é, no meu entender, aquilo a que os autores das obras acima citadas se referem. Verbos como DANÇAR, VOAR, MOVER-SE, MEXER-SE, etc. não coocorrem seja com *aonde*, seja com qualquer complemento circunstancial de lugar exprimindo destino (temporário ou não). A categoria que me parece relevante em termos de uma categorização portuguesa de verbos portugueses é a dos verbos denotando uma mudança de posição (em termos aspectuais, mudanças (Santos, 1996a; 1996b)), classe de que IR é um membro de pleno direito⁴.

Reformulando, pois, a Hip. 4 em termos de verbos de mudança de posição, confirmá-las iam exemplos em que *aonde* coocorresse com FICAR, ENTRAR, CHEGAR, COLOCAR-SE, PÔR-SE, APARECER, etc. Da mesma forma, a hipótese paralela Hip. 4' prevê a existência de *donde* com

³ Ainda que haja duas diferenças importantes entre *aonde* e *donde*: subsiste a possibilidade de usar *de onde*, mas não *a onde*; *de* pode marcar origem ou ponto de vista, ambos aceitando uma localização como complemento (respectivamente *ele veio de França* e *eu vi-o da janela*). Além disso, *donde* utiliza-se também como advérbio retórico: *Donde, a conclusão não é justificada*, sem paralelo em *aonde*.

⁴ Os únicos verbos que escapam a essa generalização e que se poderiam chamar de movimento são PASSAR e ANDAR nos exemplos seguintes: *Passas ao merceiro quando vieres para casa?* e *Anda à praia comigo!* É, contudo, possível defender que o verbo ANDAR nessa acepção é sinónimo de VIR, e que PASSAR, apesar de não exprimir uma mudança de posição, também não exprime uma forma de movimento. O que exprime, de facto e sem qualquer dúvida, é temporariedade intrínseca.

RETIRAR-SE, SAIR, DESAPARECER, etc.

O desdobramento dos verbos de mudança de posição em duas hipóteses distintas, Hip. 4 e Hip. 5, prende-se com o facto, deveras interessante, de que em português europeu não parece existir diferença entre *a* e *para* quando é o objecto que é movido em vez do sujeito, pelo menos no caso dos verbos que não indicam "automovimento". Ou seja, a maioria dos verbos transitivos directos exprimindo uma mudança de lugar usam *em* como preposição indicando o destino: PÔR, DEIXAR, COLOCAR, PENDURAR, PUBLICAR, AFIXAR, INTRODUIR, etc. Como excepções (ou seja, verbos que movem o objecto e usam *a*) apenas conseguimos identificar LEVAR e TRAZER (mas note-se que ambos são verbos de três lugares, ao contrário dos anteriores) e IR BUSCAR (em que *a* indica a origem e não o destino). A Hip. 5 seria assim reforçada se *aonde* fosse encontrado com esses verbos, enquanto que a hipótese análoga Hip. 5' implica a utilização de *donde* em frases com verbos como APAGAR, COPIAR, DESVIAR, EXPULSAR, DESLOCAR, etc.

Uma observação interessante é a de que é possível, em português de Portugal, *ir pôr a criança a casa*, embora PÔR em geral requeira a preposição *em*. Isso poderia levar a crer que a característica humana (ou animada) do que é movido (ou mudado de posição) possa ter influência na forma de expressão, e daí a Hip. 6, em que *aonde* apenas seria permitido em contextos que implicassem a mudança de posição de um ser animado, capaz de locomoção. (Por outras palavras, num caso dir-se-ia *Aonde levaste a avó?* e noutro *Onde levaste o carro?*, o que reputo de falso.) Penso, aliás, que essa hipótese é desnecessária, visto que a maior coocorrência da distinção entre *a* e *para* com pessoas e seres animados se deve precisamente ao facto de ser em relação a estes que mais sentido faz distinguir um destino ou localização temporário de um permanente⁵. Além disso, note-se que a mesma conotação de temporário está presente no uso de PÔR como aspectualizador: *ele pôs-se a rir, pus o programa a correr, pôr as roupas a secar*, etc. sem qualquer relação com objectos humanos ou animados. Por outro lado, a distinção *a/para* também está presente em empregos metafóricos, cf. *passou-se para o inimigo* (não tem retorno) e *passou à reserva* (mas pode voltar a ser chamado). Finalmente, não sinto qualquer diferença de aceitabilidade entre *levar a criança ao médico* ou *levar o carro ao mecânico*, ou entre *deixar a mala em casa* ou *deixar a avó no cabeleireiro*, etc., embora o traço animado exista na nossa língua (unicamente?) na forma do pronome interrogativo sujeito (*Quem é que levaste ao médico?* vs. *O que é que levaste ao mecânico?*).

A segunda questão prende-se com a influência da classificação sintáctica das palavras gramaticais: *onde* e *aonde* têm ou não características diferentes quando usadas com diferentes funções sintácticas?

Começo por fixar uma nomenclatura que permita distinguir os usos de *onde*, *aonde*, e *donde* nos seguintes quatro casos:

Caso 1, Pronome relativo com antecedente expresso:

*Debaixo da escada havia um grande armário **onde** costumavam guardar todas as coisas; É uma espécie de albergue espanhol **aonde** tudo e todos vão parar.*

Caso 2, Pronome relativo sem antecedente expresso:

*Levo-os **aonde** há madressilvas e poderão colher um ramo para a vossa mãe; Então, já sabemos **onde** procurar!*

Caso 3, Pronome interrogativo em oração interrogativa directa:

***Onde** o terá enterrado?; Mas **aonde** foram os Portugueses buscar tal palavra?*

⁵ Claro que "permanente" significa, em termos de mudança de posição, ir para ficar algum tempo, não para sempre.

Caso 4, Pronome interrogativo em oração interrogativa indirecta:

*Rodrigo perguntava a si mesmo **aonde** iria o irmão; Você sabe sempre **onde** existem as melhores coisas, não é verdade, Luís?*

Dada a profusa diversidade terminológica, bem patente na consulta às obras referidas, escolhi a usada por Peres & Mória (1995) nos dois primeiros casos por me parecer bem fundamentada e concordar com a minha visão da língua. Penso que é lícito generalizar a palavra “pronome” – etimologicamente “em vez do nome” – de forma a dar-lhe o sentido “em vez de outra palavra” (nesse caso o de “em vez de advérbio”).

Interessa-nos saber se existe alguma diferença entre os usos e o contraste entre *onde* e *aonde* quando são usados nesses contextos. Na literatura normativa apenas são mencionados os casos 1 e 3.

A terceira questão pode ser posta do seguinte modo: Quando temos dois verbos e *onde* ou *aonde* é usado como pronome relativo sem antecedente expresso, qual é o verbo/oração que decide? Ou seja, qual das duas formas se deve empregar: *vou aonde eles estão* ou *vou onde eles estão*?

Mais especificamente em relação ao caso 2, quando *onde* ou *aonde* são argumentos de duas orações diferentes com padrões e características diferentes/contraditórias, fica-nos a dúvida. O material que conseguimos encontrar leva-nos a concluir, aliás, que entre as duas "ordens" o coração dos falantes do português balança: o exemplo *Se assim for, só o pai da criança é que pode saber o que significa e qual a origem, porque só ele sabe **aonde** o foi buscar* obedece ao verbo da oração subordinada, enquanto o exemplo *A imprensa tem que ir **aonde** a notícia está* obedece ao da subordinante.

Ainda que, de um ponto de vista estritamente normativo, a solução mais fácil seja desaconselhar o uso dessa construção em casos semelhantes, o certo é que, em muitas ocasiões de utilização da língua, esse problema se põe e é inconscientemente resolvido de uma forma automática pelos falantes.

4. Variantes portuguesa e brasileira

É possível que haja diferença entre as variantes brasileira e portuguesa a esse respeito:

- São conhecidas as diferenças aspectuais entre as duas variantes, exactamente na distinção entre *a* e *para*, central em português de Portugal e não operativa (ou marginal) no português do Brasil.
- O exemplo de Bechara (1988) *A praia **aonde** te diriges parece perigosa* não concorda com a minha intuição: considero preferível *A praia **a que** te diriges parece perigosa*.
- Precisamente em relação a alguns morfemas gramaticais, a norma ortográfica é diferente dos dois lados do Atlântico, como é o caso de *porque* vs. *por que*.

Como a minha competência nativa não me permite julgar a qualidade de um texto em português brasileiro, ao contrário da de textos escritos em português de Portugal, pensei inicialmente delimitar a minha investigação à norma portuguesa. Contudo, após observar alguns dados do português brasileiro, concluí que não havia diferença imediata entre as duas variantes a esse respeito, e que seria interessante debruçar-me sobre um possível contraste, dado que, como Davies (1999) menciona em relação às semelhanças na evolução do português e do castelhano, não é de excluir a hipótese de que mudanças paralelas, mas independentes, ocorram nas duas variantes do português.

5. Descrição das experiências concretas com base nos textos

Não é já inovador mencionar que a rede (WWW) e os corpora linguísticos podem constituir matéria-prima muito útil para o estudo e ensino de uma língua. Contudo, a sua existência não é uma panaceia, e convém também focar as limitações e os perigos de ter acesso a material tão diversificado como jornais *on-line*, textos literários, *news*, mensagens de correio electrónico, páginas pessoais escritas por um leque variado de falantes, documentação técnica e legal, etc.

De entre as experiências que é possível fazer, o significado que lhes é lícito atribuir tem de ser avaliado sob três ângulos: procura de exemplos autênticos; obtenção de uma estimativa quantitativa; e avaliação de autoridade.

Para o presente estudo, segui a seguinte metodologia, dividida em três fases:

1. Procura de ocorrências de *aonde* e *donde* na Web e em corpora.
2. Quantificação dos casos em domínios claros.
3. Refinamento das perguntas e novas procuras.

Para obter exemplos autênticos, é possível utilizar os variados motores de procura na Web (visto que não consideram as palavras gramaticais portuguesas como *stopwords*, ou seja, não as excluem da indexação), além de utilizar todos os corpora que tenhamos à nossa disposição: quer corpora linguísticos (no sentido de terem sido compilados com o fim de facilitarem a investigação linguística) quer corpora literários ou técnicos (todos os textos em português a que temos acesso em formato electrónico, sejam eles colecções de obras literárias, teses, artigos, mensagens de correio electrónico, etc.).

Nesta primeira fase, a intenção é apenas coligir um conjunto significativo de ocorrências encontradas em várias fontes, sem preocupações de completude ou de repetibilidade. É, no entanto, sempre importante atender à fonte, visto que esta confere graus diversos de autoridade. Em particular, na Internet convém distinguir o caso de páginas pessoais ou conversas por meio electrónico, por um lado, de páginas institucionais ou com conteúdos científicos ou técnicos, por outro, visto que das últimas se espera ter havido maior cuidado na sua redacção e mesmo, eventualmente, revisão. Por outro lado, consideramos que obras publicadas (e especialmente traduzidas) devem corresponder a um grau elevado de competência linguística por parte dos seus autores (tradutores), donde os exemplos nelas encontrados deverão ter um peso superior.

Quanto a procuras na Internet, veja-se um breve apanhado:

- Procurámos através do Altavista, a 24 de Maio de 1999, páginas com *aonde*: 15352 em 6914 páginas. Infelizmente para os nossos objectivos, existe um motor de procura brasileiro denominado precisamente AONDE, o que falseia bastante os dados quantitativos que pretendemos obter.
- No servidor EUROPA⁶, numa pesquisa feita a 23 Julho de 1999, foi possível encontrar 17 documentos com *aonde* contra 17990 com *onde*⁷.
- Numa edição do Público on-line (uma semana é pesquisável, resultados de 18/5 a 24/5/99), encontramos 3 ocorrências de *aonde* e 11 de *onde*.
- Nos arquivos do Expresso on-line, procurando todos os artigos de 19 de Julho de

⁶ http://europa.eu.int/geninfi/query_pt.htm permite procurar no servidor da União Europeia EUROPA.

⁷ Infelizmente, onde e donde são palavras gramaticais também em castelhano e em italiano, donde tentámos excluir encontros em documentos nessas línguas através da procura de "onde NOT es NOT el NOT il".

1997 a 17 de Julho de 1999, encontrámos 115 documentos com *aonde*, 300 documentos com *onde*.

De forma a tentar extrapolar alguns valores quantitativos em relação ao uso de *onde* (pese embora o facto de que esse item lexical também tenda a ser considerado sobregerado, quer por lingüistas brasileiros quer por portugueses, nomeadamente substituindo indevidamente outros pronomes relativos), foi preciso, numa segunda fase, cingir-nos a colecções fechadas cujos resultados fossem repetíveis ou pelo menos simuláveis. Assim, contámos as ocorrências de *aonde*, *donde*, *onde*, *para onde* e *de onde* nos corpora a que temos acesso (veja-se Santos & Sarmiento, 2003, Rocha e Santos, 2000), assim como no conjunto de seis livros infantis traduzidos para português europeu, caracterizados na Tabela 1⁸.

Origem	Tamanho (páginas)	<i>onde</i>	<i>aonde</i>	<i>donde</i>	<i>para onde</i>	<i>de onde</i>
Outra vez os 6	217	55	1	2	0	0
6 primos	270	48	5	1	2	1
A casa do Rio	138	11	0	4	0	0
A ilha dentro	135	40	1	3	0	1
A montanha B	138	44	4	1	4	0
Seta dupla	149	63	3	3	9	3
Total		261	14	14	15	5

Tabela 1: Caracterização pormenorizada do corpus de livros infantis

Criámos esse pequeno corpus para tentar ter uma ideia da distribuição dessas entidades *por narrativa* (visto que os outros corpora a que temos acesso contêm simplesmente fragmentos/extractos de livros e não a sua totalidade, e sugerimos em Santos e Oksefjell (1999) que a distribuição ao longo de um livro podia ser significativa).

A Tabela 2 apresenta os valores encontrados durante a fase 1. De notar que o total de ocorrências de *onde* inclui as de *para onde* e *de onde*, e que em alguns casos (marcados por asterisco) o número de encontros se refere a documentos diferentes e não a ocorrências.

Origem	Tamanho (mil palavras)	<i>onde</i>	<i>aonde</i>	<i>donde</i>	<i>para onde</i>	<i>de onde</i>
Natura/Público	6.242	6354	15	41	136	243
ENPCport	231	338	2	7	11	8
Natura/Minho	2.111	114	1	15	39	47
ECI-BR	718	707	5	5	18	37
ECI-EE	26	1	0	1	0	0
MLCC-DEB	8.887	3354	18	148	71	93
NILC/S.Carlos	33.903	23047	163	204	415	905
Infantis		261	14	14	15	5
CETEMPúblico	90.000	204356	258	534	2453	3779
Expresso*		300	115			
EUROPA*		17990	17			

Tabela 2: Número de ocorrências

Como seria de esperar, além de indicar que o item lexical *aonde* existe em vários registos, mas com uma frequência muito inferior a *onde*, essas contagens não nos dão muito mais informação. Para responder às perguntas formuladas acima, precisamos de investigar as

⁸ De Will Scott, editados pelos Estúdios Cor, Lisboa, anos 70: *A casa do rio*, trad. Luisa Ducla Soares; *A ilha dentro de casa*, trad. Fernanda Pinto Rodrigues; *A montanha balancé*, trad. Fernanda Pinto Rodrigues; *O mistério da seta dupla*, trad. Luisa Ducla Soares; de Enid Blyton, editados pela Editora Meridiano: *Seis primos numa quinta*, trad. Helena Valente, 1970; *Outra vez os seis primos*, trad. Helena Valente, 1968.

ocorrências uma por uma, classificando-as por função sintáctica, e por verbo(s) com que coocorrem. Eventualmente, através de uma metodologia iterativa para refinar as várias procuras sobre questões possivelmente relevantes para a descrição do uso de *aonde*, poderemos então chegar a conclusões apenas permitidas por uma abordagem de "linguística com corpora".

Visto que a separação entre os casos 2 e 4 resultou mais problemática do que parecia à primeira vista, decidimos classificar o material em apenas 3 casos (juntando pois o 2 com o 4), identificados respectivamente por R (relativa), I (interrogativa directa) e D (dupla dependência). Apresentamos os resultados na Tabela 3 para *aonde*, *donde* e *de onde*. Note-se que *donde* usado como retórico não foi incluído, ao contrário do que acontece na tabela anterior.

Contexto sintáctico	R			I			D		
	<i>aonde</i>	<i>donde</i>	<i>de onde</i>	<i>aonde</i>	<i>donde</i>	<i>de onde</i>	<i>aonde</i>	<i>donde</i>	<i>de onde</i>
Natura/Público	9	23	217	0	1	12	5	2	16
ENPCport	1	4	6	0	2	2	1	0	0
Natura/Minho	1	12	33	0	0	1	0	1	5
ECI-EBR	3	2	18	2	0	2	0	2	15
MLCC-DEB	4	44	57	7	11	5	7	29	30
NILC/São Carlos	43	68	606	10	3	31	75	20	244
Infantis	0	8	0	7	1	2	7	5	3
Total por forma	61	161	937	26	18	55	95	59	213
Total	1159 (71%)			99 (6%)			367 (23%)		

Tabela 3: Distribuição sintáctica

É interessante notar a elevada percentagem de casos com dois verbos (ou predicadores) relacionados com *aonde* ou *donde*. É também interessante verificar que *de onde* e *donde* não são, distribucionalmente, muito distintos (ou seja, que são usados mais ou menos nas mesmas percentagens no que se refere a I, R e D). Por outro lado, resulta claro da Tabela 2 que corpora diferentes têm preferências diferentes, tanto por *aonde* como por *donde* vs. *de onde*. Inesperadamente, é no corpus de debates do Parlamento Europeu que se encontra maior predominância dessas unidades, enquanto que o Diário do Minho é o mais parco.

Os verbos da oração subordinada que coocorrem com *aonde* são os seguintes (seguidos da frequência): *ir* (46), *chegar* (19), *estar* (10), *ser* (9), *levar* (7), *conduzir* (6), *querer* (5), *deslocar* (4), *ir_parar haver* (3), *aparecer ir_levar dirigir-se encontrar querer_chegar ter vir ir_buscar desejar* (2), *aplicar-se existir coletar passar chamar explorar residir ir_abrigar esperar ir_estudar acontecer lançar trazer estudar tocar vangloriar-se levar_a_serio surgir dar-se ocultar-se chover dormir regressar pressentir quebrar assistir ir_chegar enviar dirigir roubar atacar ir_dar sentir cuidar pousar estender ir_morar escrever avisar crescer lavrar ver acorrer passar-se influenciar encontrar-se utilizar homenagear voltar* (1). Os verbos da oração principal correspondente (no caso D) são: *saber* (24), *chegar ir* (13), *levar* (6), *perguntar* (4), *mostrar ter ver entender* (3), *guiar* (2), *transportar enviar lembrar carregar esperar decidir dar_o_tom haver acudir voltar indicar acompanhar interrogar defender descobrir dizer informar estar achar* (1).

Para *donde*, temos os seguintes resultados: *vir* (64), *sair* (25), *provir* (18), *partir* (13), *retirar fugir* (5), *estar* (4), *passar proveniente atirar querer emergir voltar* (3), *aparecer derivar esperar natural trazer excluir tirar seleccionar regressar contar ver importar resultar dimanar* (2), *transferir colocar espiar arrancar brotar banir manar espalhar emanar morrer fazer soltar-se dizer cair conhecer avexar transparecer ser dependurar desempenhar extirpar filtrar proceder expulsar canalizar espalhar-se desdobrar baixar expelir avistar começar ser_natural resvalar reinar ausentar roubar atacar N principiar originário recolher decorrer voar pender*

iniciar oriundo arredar seguir rebentar precipitar tornar descer ir sobressair escapar (1). Os verbos da oração principal correspondente (no caso D) são: *saber* (19), *vir* (13), *perguntar* (6), *expulsar* (3), *cogitar ver* (2), *fazer_fogo navegar averiguar perceber perto estar_claro partir inquirir olhar pôr explicar esclarecer ir surgir descobrir dizer* (1).

Alguns casos são demasiado complexos para que esta análise ilumine a sua constituição. Essa constatação é, aliás, típica do trabalho com corpora. Por muito simples que qualquer classificação nos pareça, aparecem sempre casos de charneira sobre os quais é preciso documentar as opções tomadas na marcação. Apresentamos a título de exemplo alguns casos em que tivemos de fazer opções em relação aos dados apresentados: O exemplo *A história é como um criatura que toma posse do escritor e guia o livro aonde quer* foi codificado como tendo o verbo QUERER na oração subordinada; *arrogante Romário faz questão de ostentar sua condição financeira mesmo quando trata de questões mais do que cabíveis como levar familiares aonde quer que seja* como tendo o verbo SER (em ambos os casos outra alternativa teria sido o implícito LEVAR). Quando nos encontramos em presença de várias orações encaixadas, marcamos o verbo principal da última, ou seja TIRAR em *E se são pródigos em estatísticas (de onde o leitor acha que a Folha tirou a idéia?), são também pródigos em histórias humanas*. Por outro lado, em *Um dia apresentaram ao imperador um topetudo não sei donde, que perguntou, mui concho...* escolhemos marcar *N* (um nome) em vez de TOPETUDO, enquanto em *Eu penso que este caso poderia ser, de facto, um bom exemplo de onde poderíamos usar este novo instrumento* consideramos EXEMPLO como o predicador (aqui denominado "verbo") da oração principal. Noutros casos podemos também ter adjectivos, tal como ORIUNDO ou PROVENIENTE.

Finalmente, convém notar que em alguns casos foram consideradas locuções verbais e não o tradicional verbo principal: *ir buscar, ir parar* etc. Infelizmente, por limitações de espaço e tempo não me posso debruçar aqui sobre essa questão, muito interessante.

Da leitura das listas de verbos anteriores pode observar-se que os verbos IR e VIR são, de facto, os verbos mais utilizados com *aonde* e *donde* respectivamente, seguidos por outros verbos que com eles partilham várias propriedades (CHEGAR, LEVAR são próximos de IR porque implicam mudança de lugar e exprimem destino com a preposição *a*; SAIR, PARTIR são próximos de VIR porque implicam mudança de lugar e exprimem origem com a preposição *de*). Outros verbos surgem, contudo, que são suficientemente diferentes para serem interessantes, como é o caso de IR PARAR e ESTUDAR, coocorrendo com *aonde*, ou IMPORTAR e ROUBAR, coocorrendo com *donde*.

6. Observações finais

A questão de *onde* vs. *aonde* (e o consequente estudo de *donde*) revelou-se mais interessante do que esperava quando a ela decidi dedicar alguma atenção, por várias razões: aparentemente é um caso de novas normas por oposição às antigas, dado que, no século passado, segundo alguns autores, tal distinção não se tinha ainda imposto. Não nos encontramos, portanto, perante um fenómeno em declínio, mas sim em expansão. Além disso, contribuiu para iluminar algumas características aspectuais do português (o caso dos verbos de movimento e o da dicotomia permanente/temporário) sobre as quais, em princípio, o estudo não versava. Finalmente, demonstra que existem áreas cuja investigação e aprendizagem é pertinente tanto para falantes de português língua materna como de português língua estrangeira.

Contudo, não pretendo de forma alguma ter esgotado o assunto. Em particular, as questões da ortografia e da morfologia não foram aqui abordadas, nem as características prosódicas. Deste estudo sobressai, aliás, a necessidade de prosseguir estudos concretos com base em corpora, a saber: (1) Observar a diferença sistemática em relação a *onde*, e em relação a *para onde*; (2) Investigar a correlação com os referentes das relativas; (3) Indagar se o tipo de

texto (narrativo, expositivo, argumentativo) é relevante; (4) Investigar os casos em que, no mesmo texto, se encontram as duas possibilidades (seja *aonde* e *onde*, seja *donde* e *de onde*, com o mesmo verbo).

Em última análise, o presente artigo demonstra claramente que, se é possível, dado suficiente engenho, escrever um texto em que várias hipóteses são propostas e a sua discussão apresentada com base em opiniões subjectivas do autor sobre a sua própria língua (o que faço na secção 3), algo muito diferente é estabelecer e produzir um estudo baseado em corpora que leve ao mesmo resultado, ou melhor, que permita confirmar ou induzir um dado conjunto de hipóteses. Um estudo em linguística com corpora implica muito mais trabalho e muito menos garantia de obter conclusões definidas em relação ao problema que se quer tratar; contudo, permite um "banho de língua" que o torna muito mais útil do que a escrita de um artigo tradicional, quer para o próprio linguista, quer para os resultados do seu trabalho. E com "resultados" abranjo quer a produção de ferramentas computacionais para lidar com a língua, quer a produção de materiais de descrição e explicação dessa mesma língua, assim como simplesmente uma maior compreensão desta.

Agradecimentos

Agradeço a Tony Berber Sardinha a apresentação deste artigo no Simpósio Redescobrir a Linguagem: Pesquisa em Linguística de Corpus, por ocasião do 10.º InPLA, e a Elisabete Ranchhod as críticas feitas a uma versão preliminar do artigo.

Recebido em: ??????; Aceito em: ??????.

Referências bibliográficas

- BECHARA, E. 1988 *Lições de português pela análise sintáctica*. Padrão – Livraria Editora Ltda. 14ª edição [1ª edição, 1960].
- CUNHA, C. & CINTRA, L. 1984 *Nova gramática do português contemporâneo*. Edições João Sá da Costa.
- COSTA, J.A. & MELO, A.S. s/d *Dicionário da língua portuguesa* [Dicionários Editora]. Porto Editora, 6ª edição corrigida e aumentada.
- DAHL, Ö. 1985 *Tense and aspect systems*. Basil Blackwell.
- DAVIES, M. 1999 The historical development of subject raising in Portuguese: a corpus-based approach. *Neuphilologische Mitteilungen* 100: 95-110.
- DPI 1989 *Dicionário de português-inglês* [Dicionários escolares]. Porto Editora.
- LELLO, E. 1978 *Lello Escolar: novo dicionário ilustrado da língua portuguesa, com um epítome de gramática e regras ortográficas*. Lello & Irmão Editores.
- LELLO, I. 1990 *Dicionário prático ilustrado: novo dicionário enciclopédico luso-brasileiro publicado sob a direcção de Jaime de Séguier*. Lello & Irmão Editores.
- ESTRELA, E. 1984 *Dúvidas do falar português: consultório da língua portuguesa*, 3 vols., Editorial Notícias. 2ª edição [1ª edição, 1983].
- FERNANDES, A. 1991 *Portugisisk Svensk ordbok*. Logica Verlag, Bona.
- FERREIRA, P.J.A. & MORAIS, A. 1983 *Dicionário português-inglês*. Editorial Domingos Barreira. Nova edição.
- FIGUEIREDO, C. s/d *Grande dicionário da língua portuguesa*. Bertrand Editora, Venda Nova. 23ª edição [1ª edição, 1939].
- FREIRE, A. 1983 *Lições de filologia e língua portuguesa*. Publicações da Faculdade de Filosofia, Braga.
- LAKOFF, G. 1987 *Women, fire, and dangerous things: what categories reveal about the mind*. The University of Chicago Press.
- MARTINS, M.T.H.-S. 1982 *Portugiesische Grammatik*. Niemeyer, Tübingen.

- MATEUS, M.H.M; BRITO, A. M.; DUARTE, I. & FARIA, I. H. 1989 *Gramática da língua portuguesa*, Caminho. 3ª edição [1ª edição, 1971].
- MIGUEL, J. 1989 *Estudos de língua portuguesa*. Editora Harbra Ltda.
- NOGUEIRA, R. de S. 1989 *Dicionário de erros e problemas de linguagem*. Clássica Editora. 3ª edição [1ª edição, 1969].
- PERES, J. A. & MÓIA, T. 1995 *Áreas críticas da língua portuguesa*. Caminho.
- RAMALHO, E. 1985 *Dicionário estrutural, estilístico e sintático da língua portuguesa*. Livraria Chardron de Lello & Irmão Editores.
- ROCHA, P.A. & SANTOS, D. 2000 CETEMPúblico: um *corpus* de grandes dimensões de linguagem jornalística portuguesa. *Actas do V Encontro para o processamento computacional da língua portuguesa escrita e falada (PROPOR'2000)*, Atibaia, São Paulo, Brasil, 19 a 22 de Novembro de 2000.
- SANTOS, D.M. de S.M.P. 1996a Tense and aspect in English and Portuguese: a contrastive semantical study. Tese de doutoramento, Instituto Superior Técnico, Lisboa.
- _____. 1996b Para uma classificação aspectual portuguesa do português. IN I. CASTRO (ed.) *Actas do XII Encontro da Associação Portuguesa de Linguística* (Braga-Guimarães, 1-3 de Outubro de 1996).
- _____ & OKSEFJELL, S. 1999 Using a parallel corpus to validate independent claims. *Languages in Contrast*, **2.1**: 117-132.
- _____ & SARMENTO, L. 2003 Projecto AC/DC: Acesso a corpora/disponibilização de corpora. IN A. MENDES & T. FREITAS (ed.) *Actas do XII Encontro da Associação Portuguesa de Linguística*, Porto, 2-4 de Outubro de 2002.
- SWEETSER, E.E. 1990 *From etymology to pragmatics: metaphorical and cultural aspects of semantic structure*. Cambridge University Press.
- VILELA, M. 1990 *Dicionário do português básico*. Edições Asa. 1.ª edição.

Diana Santos is a researcher at SINTEF Information & Communication Technology, Oslo, Norway, and leads Linguateca, a distributed resource center for the processing of the Portuguese language. She graduated in 1985, took her Msc in 1988 and PhD in Computer Science (Natural Language Processing) in 1996, all at Instituto Superior Técnico (Lisbon). Main research areas: corpora, translation, evaluation.