Romansk Forum **XV Skandinaviske romanistkongress**
Nr. 16 – 2002/2 **Oslo 12.-17. august 2002**

Diana Santos
SINTEF

Com e *med*: um estudo contrastivo português-norueguês

Este estudo tenta verificar duas hipóteses: a primeira, linguística, de que há uma diferença significativa entre as preposições *med* e *com*, não obstante a semelhança entre os valores a elas tradicionalmente atribuídos. A segunda, metodológica, de que o uso de grandes corpora, e sobretudo o recurso a dados de distribuição (e não apenas de frequência), ultrapassa em muito um estudo baseado apenas em casos isolados.

Sabíamos a priori que *med* e *com* são usados de maneira diferente nas duas línguas. Por um lado: 1. Há contextos sintácticos de *med* sem paralelo em português, tais como *Det er godt med... Det er stygt med...*, em que podemos dizer que o sintagma preposicional (SP) iniciado por *med* é o sujeito da predicação. 2. *Med* indicando um meio de transporte também não tem nenhum paralelo que envolva *com* em português, em que *med* em *Jeg kom med buss* teria como equivalente a preposição *de*: *Vim de autocarro*. 3. Formas de exprimir, em norueguês, o papel com que uma dada acção é desempenhada, como em *Jeg var på tur med skolen* ou *Jeg var i Nantes med jobben*, têm de ser indicados de forma muito diferente em português: *Na escola hoje fomos dar um passeio* ou *Estive em Nantes numa viagem de trabalho*. 4. Existem contextos sintácticos em norueguês que não têm, pura e simplesmente, paralelo em português: partículas verbais (preposições pertencentes ao grupo verbal), cf. *Jeg var med på å skrive brevet* ou *Jeg var med på at du skulle få slippe...*, traduzíveis respectivamente por *Participei na escrita da carta* e *Concordei em te deixar não ...* 5. Finalmente, há uma panóplia de locuções e expressões fixas cujos "equivalentes" em português não contêm *com*: *til og med (até), med mindre (a não ser que), med vilje (de propósito), med en gang (de repente), med ett (de súbito), med flere (entre outros)...*

Por outro lado, também há vários contextos de *com* sem paralelo em norueguês, tais como: 1. a sensação ou sentimento "em companhia", ou em relação a outros seres humanos (a que poderíamos chamar "verbos experienciais acompanhados"), ilustrados por frases como: *Eu zanguei-me com ele*, *Nós divertimo-nos imenso com os palhaços*, cuja expressão em norueguês não envolve *med*. 2. *Com* é, além disso, frequentemente usado para exprimir estados temporários, em casos em que as línguas germânicas empregariam um adjectivo, como é o caso de *estar* (ou *ficar*) *com frio, fome, medo, pena*, respectivamente traduzidos por *være/bli kalt*, *sulten*, *redd*, *lei*... 3. Existem, também, muitas locuções ou expressões fixas de português que não são expressas em norueguês com *med*, veja-se *de acordo com* (*i følge*), *para com* (*for*), *com certeza* (*sikkert*).

Outras formas de expressão, cuja sistematização é mais complicada, mas que ilustram diferenças palpáveis no uso destas preposições são: *jobbe med saken* (*trabalhar no assunto*; *tratar do caso*), *ble med barn* (*ficar grávida*), *fazer com que* (*gjøre slik at*, *få X til å*).

Para terminar, mencionamos os empregos mais gramaticais (com menos conteúdo associado) das duas preposições, nomeadamente quando são subcategorizadas por verbos, em que é fácil encontrar casos em que apenas uma das duas preposições é empregue: *perder com* (*tape mot*), *sonhar com* (*drømme om*), *espantar-se com* (*ble overrasket over*), e, conversamente, *ta med* (*levar*), *følge med* (*pertencer*; *fazer parte de*), *bli med* (*acompanhar*), *få med seg* (*perceber*).

Observemos o que nos dizem os dicionários quanto a *com* e *med*: o DLPC apresenta 19 "sentidos" para *com*, e o Bokmålsorboka 11 "sentidos" para *med*: 7 sentidos de *com* em 19 também foram (independentemente) identificados em *med*; dos 11 de *med*, 9 também figuram como atributos de *com*. Dada a diferença entre os estilos e formato dos dois dicionários, e analisando com atenção cada um dos "campos semânticos" atribuídos, é inegável que a maior parte dos sentidos atribuídos pelos dicionários monolingues são partilhados pelas duas línguas, ainda que especificados de forma bem diferente.

Os dicionários bilingues dão mais atenção às diferenças entre as línguas, claro. É interessante observar que dois dos cinco "campos de tradução" identificados no dicionário norueguês-português de Nilsson (meios de transporte e *e*, a que chamarei "companhia semântica") não se encontrem sequer dicionarizados no Bokmålsorboka (por fazerem parte de categorias mais amplas lá identificadas). De qualquer maneira, e apenas pela entrada da preposição *com* em Nilsson (1994), nada indica que este par seja especialmente problemático.

Contudo, como já referimos, uma preposição é muitas vezes determinada por outras palavras (verbos, substantivos e adjectivos), devendo, nesse, caso estar dicionarizada sob esses outros itens lexicais. Por outras palavras, nem toda a informação sobre *com*, seja em que dicionário for, se encontra na entrada *com*: todas as entradas que contêm *com* são relevantes. Para as obter de forma expedita, porém, é preciso poder manipular o dicionário de forma electrónica. Para a presente investigação, usámos o dicionário português-norueguês (Engh, em prep.), de 16.000 entradas, correspondendo a 35.000 pares de tradução (tabelas 1 a 2).

Tabela 1a: De *com* para *med* no dicionário português-norueguês Engh (em prep.)

Ocorrências de *com*		sem *med*	com *med*
Total	335	261 (78%)	74 (22%)
Verbos	150	116	34
Subst. m+f	72+85	60+68	12+17
Adjectivos	20	13	7
Adv., prep.	2+3	1+3	1+0

Tabela 1b: De *med* para *com* no dicionário português-norueguês Engh (em prep.)

Ocorrências de *med*		sem *com*	com *com*
Total	451	377 (84%)	74 (16%)
Verbos	160	126	34
Subst. m+f	111+71	99+54	12+17
Adjectivos	77	70	7
Adv., prep.	19+3	18+3	1+0

Se as tabelas 1a e 1b apontam para que a sobreposição das preposições nas duas línguas é muito reduzida, a tabela 2 indica, contudo, que tal não é apanágio das preposições *med* e *com*, que se encontram no meio do espectro em relação à sua intertradutibilidade. Aliás, devido a estarmos em presença de um dicionário bilingue, calculamos que estejam mais explícitas as diferenças do que as semelhanças, embora também possa incluir as semelhanças inesperadas.

Tabela 2: Vários pares de preposições no dicionário Engh (em prep.)

X	Freq. X	Freq. de X e Y	Freq. Y	Y
em	520	144 (28%)	1213	*i*
para	184	25 (14%)	391	*for*
para	184	37 (20%)	668	*til*
sem	93	41 (44%)	128	*uten*
sob	7	4 (57%)	47	*under*
com	335	74 (22%)	451	*med*

Mas uma preposição é muito mais do que uma palavra (uma entrada no léxico). É, antes de mais, uma componente importantíssima da gramática e do sistema de uma língua, e é portanto francamente discutível considerar os dicionários como o melhor local para obter informação sobre o seu uso e "sentido" (contudo, cf. Carvalho (2001) para uma proposta nesse sentido). Um estudo baseado em corpora permite-nos avaliar a distribuição e o peso relativo da informação já presente nos dicionários, eventualmente descobrindo características adicionais.

Tabela 3a. Corpora monolingues criados pela Linguateca e pelo Tekstlab (UiO)

Nome e língua	Tamanho em milhões de palavras	Tamanho em milhões de frases	URL
AC/DC, português	65	2,5	http://acdc.linguateca.pt/acesso/
CETEMPúblico,port.	192	7	http://cetempublico.linguateca.pt/
OBTK (Bokmål)	18/14,5		http://www.tekstlab.uio.no/Norwegian/
ONTK (Nynorsk)	3,7/3,37		http://www.tekstlab.uio.no/Norwegian/

Tabela 3b. Corpora paralelos (TO: texto original; TT: texto traduzido)[1]

Nome, línguas, número de textos originais	Tamanho em palavras em TO	Tamanho de palavras em TT	Número de frases em TO
ENPC Fiction En-No (30)	415091	410783	29578
ENPC Fiction No-En (30)	403571	438438	16642
ENPC Non-Fiction En-No (20)	256211	250897	12369
ENPC Non-Fiction En-No (20)	226495	261097	14642
OMC En-Po (16)	200000	230000	15500
COMPARA En-Po (5)	146113	148306	8239
COMPARA Po-En (11)	136274	150797	9400
OMC/En-No-Po (15)	203827	202731 Nor 213289 Port.	14557

Visto que temos acesso a uma quantidade considerável de corpora electrónicos, cf. tabelas 3a e 3b, grande parte dos quais criada no âmbito da Linguateca, iniciámos um estudo

[1] Todos os corpora paralelos excepto o COMPARA, http://acdc.linguateca.pt/COMPARA/Conteudo.html, foram desenvolvidos pelo IBA (UiO), http://www.hf.uio.no/german/sprik/english/corpus.shtml. A informação quantitativa foi-nos gentilmente cedida por Jarle Ebeling. O COMPARA, sendo um corpus em permanente expansão, já inclui neste momento mais textos do que quando esta investigação foi levada a cabo [Abril 2002].

contrastivo das duas preposições para tentar esclarecer a relação complexa entre *med* e *com*, e também para contribuir para a metodologia de estudos contrastivos baseados em corpora.

Tabela 4a. Ocorrências de *com* em minúsculas nos corpora AC/DC

Corpus	Frequência absoluta	Frequência relativa por 10000 palavras	Frequência relativa (por 100 frases)[2]
eci-ee	114	43,0	14,6
eeanot	114	42,2	14,8
ancib	2.717	59,9	14,4
ancibanot	1.776	60,6	14,7
natura	45.397	72,6	20,1
natpanot	45.389	72,1	20,1
avante	47.438	72,3	22,2
minho	13.410	77,1	25,2
minhanot	12.346	77,3	26,0
cetempublicoprmi	7.643	76,6	20,0
cetempanot	7.603	76,9	20,0
diaclav	54.412	83,9	25,8
saocarlos	201.233	75,4	12,3
scanot	204.196	74,8	15,1
frasespp	118	72,7	19,9
fppanot	118	72,8	19,9
frasespb	170	88,8	26,1
fpbanot	170	88,7	26,1
eci-ebr	5.678	78,7	12,5
ebranot	5.672	78,5	12,7
enpcpub	679	93,7	15,5
enpcanot	679	94,0	15,5
classlppe	11.310	86,5	15,2*
» sem poesia	10.981	93,1	14,8
compara-port	2.414	94,7	16,3
» original	757	71,0	11,6
» traduzido	1.657	111,7	20,1*

Usando primeiro os corpora monolingues, começámos por comparar simplesmente a frequência de *com* e de *med* em vários corpora. Desde logo se nos deparam as primeiras decisões: o que contar como palavra, e o que contar como *com*? Devem locuções como *de acordo com* ou *para com* ser contadas como uma unidade? E devem contar, ou não, como ocorrências de *com*? As nossas opções foram fazer uma separação por palavras gráficas – ou seja, contar 3 e 2 palavras nos casos acima, cf. Santos & Bick (2000) – e contar todos como ocorrências de *com*. No caso dos corpora noruegueses, contudo, *til og med*, por exemplo, está codificado como uma palavra e não conta, assim, como ocorrência de *med*.

Tabela 4b. Ocorrências de *med* em minúsculas nos Oslo tagged corpora

Corpus	Tamanho	Frequência	Frequência relativa

[2] A noção de frase gráfica pode ser problematizada. Além disso, em alguns casos – indicados por asterisco na tabela 4a – é preciso ter cuidado nos cálculos: descontar a poesia dos corpora que a incluem, e reconhecer que o número de frases utilizado para a contagem de ocorrências de *com* na tradução portuguesa dos textos ingleses não é o de frases em português, mas o de frases em inglês: portanto, um número aproximado, apenas.

		absoluta de *med*	por 10000 palavras
Bokmål	18.302.845	201.688	110,2
» texto jornalístico	9.626.760	108.628	112,8
» texto não-literário	6.961.489	75.125	107.9
» texto literário	1.714.596	17.935	104,6
Nynorsk	3.771.506	45.727	121,2
» texto jornalístico	991.449	12.688	128,0
» texto não-literário	665.802	6.738	101,2
» texto literário	2.114.255	26.301	124,4

Os resultados, apresentados nas tabelas 4a e 4b, permitem-nos observar que *com* é mais frequente em texto literário traduzido do inglês (a frequência duplica em relação a texto técnico), aproximando-se, nesse caso, da frequência de *med* (que parece ser maior em Nynorsk do que em Bokmål). Note-se desde já, contudo, que nem todas as medidas podem, com os recursos a que temos acesso, ser comparadas: por exemplo, não existe informação quanto ao número de frases constantes nos textos noruegueses. E não só: Seria mais elucidativo olhar para a distribuição sintáctica de *com* usando a informação presente nos corpora anotados do projecto AC/DC (veja-se Bick, 2000), apresentada a título de ilustração na tabela 5. Não temos, contudo, forma de a contrastar com informação equivalente para o norueguês, visto que os corpora utilizados não possuem essa informação.

Tabela 5. Ocorrência de *com* por função sintáctica em português de Portugal (AC/DC)

	1.<PIV	2.A<	3.<ADVL	4.N<	5.<SC	6.A<PIV	7.N<PRED	3 ou 7	outros
minhanot	1439	953	3504	2090	235	48	759	1213	2330
(em 12.571)	11,4%	7,6%	27,8%	16.6%	1.87%	.38%	6.04%	9.6%	18.5%
eeanot	13	11	20	26	2		6	5	36
(em 135)	9.6%	8.14%	14.8%	19.2%	1.48%		4.44%	3.70%	26.7%
fppanot	15	8	41	24	4		3	4	26
(em 115)	13.0%	6.96%	35.6%	20.9%	3.48%		2.60%	3.48%	22.6%
cpprmianot	907	215	2585	1318	222		435	653	1592
(em 7.927)	11.4%	2.7%	32.6%	16.6%	2.80%		5.49%	8.2%	20.1%
natpanot	5113	2308	14224	7820	1026		2681	3978	9725
(em 46.875)	10.9%	4.92%	30.3%	16.7%	2.19%		5.72%	8.49	20.7%
cetempanot	99633	25866	281382	157189	23751	6054	53451	79220	209390
em 935.963	10.6%	2.76%	30.1%	16.8%	2.54%	.64%	5.72%	8.46%	22.4%

Assim, recorremos ao seguinte expediente: observar apenas o contexto sintáctico (em termos de categoria gramatical) imediatamente anterior (ou posterior) – tabelas 6a e 6b. De notar, contudo, que não estamos a falar do mesmo tipo de informação, visto que a primeira se referia à ligação entre o SP e o resto da frase, e a segunda se refere apenas a uma questão de ordem linear (que não significa sequer a mesma coisa nas duas línguas, como iremos apreciar em seguida).

As maiores diferenças são devidas ao uso de preposições ou partículas como parte do grupo verbal em norueguês, reflectidas na maior frequência da sequência de duas ou mais preposições (em português apenas em *para com*) e na ocorrência de *med* antes de pontuação (ponto final, de exclamação, interrogação ou vírgula), sequência quase inexistente em

português. Note-se que esta última (e apenas esta última) categoria não é independente das outras listadas na tabela 6b, já que uma mesma ocorrência de *med* pode ocorrer após um nome e antes de uma vírgula, por exemplo.

Tabela 6a. Ocorrências de *com* no NATPANOT (número total de *com*: 45.389)

Contexto	frequência absoluta	frequência relativa ao total de *com*	frequência relativa (10000 palavras)
Após um nome	15594	34,5	24,8
A seguir a um verbo	13549	29,9	21,5
Após um adjectivo	3672	8,1	5,8
Após um nome próprio	2905	6,4	4,6
A seguir a um advérbio	2263	5,0	3,6
Depois de preposição	**234**	**0,5**	**0,37**
Início de frase	1719	3,8	2,7

Tabela 6b. Ocorrências de *med* no OTBK (número total de *med*: 106.032)

Contexto	frequência absoluta	frequência relativa (%) ao total de *med*	frequência relativa (10000 palavras)
Após um nome	46727	44,1	33,0
A seguir a um verbo	29114	27,4	20,6
Após um adjectivo	9413	8,9	6,7
A seguir a um advérbio	8445	8,0	6,0
Após um nome próprio	4211	4,0	3,0
Antes de pontuação *	**3127**	**2,9**	**2,2**
Depois de preposição	**5515**	**5,2**	**3,9**
Início de frase	2825	2,7	2,0

Esta comparação, contudo, não nos permite observar diferenças nítidas entre as duas preposições. Em primeiro lugar, algumas diferenças aparentes são imediatamente atribuíveis a diferenças na estrutura sintáctica das duas línguas: O facto de uma preposição seguir um adjectivo em português não indica que não se refira ao substantivo que esse adjectivo por sua vez também modifica, como em *rapaz loiro com camisola vermelha*, em que o sintagma com *com* refere-se a *rapaz*, embora *com* se encontre imediatamente a seguir ao adjectivo. Da mesma forma, as construções do tipo *det er adj med X*, embora tendo *med* a seguir a um adjectivo, não são por ele subcategorizadas, nem a ele referentes.

Além disso – e esta observação é válida para todos os estudos que comparem corpora compilados por grupos diferentes – não é garantido que designações ou conceitos sejam entendidos, e processados, da mesma maneira. Sem uma documentação cuidadosa, os contrastes podem ser enviesados por diferentes maneiras de ver a realidade, em vez de indicarem realidades diferentes. Da mesma forma, e tanto mais importante quanto mais pormenorizada for a análise que estamos a efectuar, é preciso não desprezar a possibilidade real de existirem erros na anotação dos corpora, não apenas por descuido humano mas, por maioria de razão, por falhas da análise automática. É preciso, pois, fazer algumas medições (Santos & Gasperin, 2002) da fiabilidade das anotações das características que nos interessam: Se os advérbios que também podem ser adjectivos em norueguês se encontram, por exemplo,

marcados erradamente em 50% dos casos, pouco ou nenhum sentido farão os estudos que comparam a diferente distribuição de *med* com advérbios vs. adjectivos! De igual modo, não se poderá confiar na distribuição de funções sintácticas, se se observar que os SP com *com* funcionando como advérbio de modo (adjuntos adverbiais) estão na sua maioria erradamente identificados como objectos preposicionais nos corpora que utilizámos...

Convencidos, contudo, do interesse de comparar, nas duas línguas, uma eventual distribuição sintáctica em traços largos, resolvemos recorrer à análise humana de alguns casos obtidos aleatoriamente: Os resultados encontram-se tabelados em 7a e 7b. Não é, evidentemente, garantido a priori que o número de ocorrências inspeccionado produza uma distribuição representativa; contudo, o facto de termos tidos resultados semelhantes nas várias vezes que fizemos a contagem parece indicá-lo.

Tabela 7a. 100 casos de *med*, obtidos aleatoriamente, no OTBK, duas vezes

Ligação sintáctica do SP com *med*	1.a vez	2.a vez	Total (%)
verbo	44	32	38,0
nome	30	41	35,5
adjectivo	6	4	5,0
adverbial livre	12	18	15,0
sendo sujeito	3	2	2,5
discurso	3	3	3,0
expressão fixa	2	1	1,5

Tabela 7b. 100 casos de *com*, obtidos aleatoriamente, em três corpora diferentes

Ligação sintáctica do SP com *com*	CETEMP	DiaCLAV	Minho	Total (%)
verbo	36	37	34	35,7
nome	29	31	37	32,3
adjectivo	5	2	3	3,3
adverbial livre	14	22	18	18,0
sendo sujeito	1	2	1	1,3
discurso	10	6	7	7,7
expressão fixa				

Convém indicar o que significa aqui "ligação sintáctica": "ligação ao verbo" corresponde a objecto preposicional ou associação com o verbo; "ligação ao nome" corresponde a modificação ao nome (ou argumento); idem para adjectivo; "sendo sujeito" (exemplos 1-2) indica que o sintagma indicado pela preposição funciona como sujeito da predicação; e "discurso" foi utilizado quando a preposição funciona discursivamente como uma conjunção subordinativa (exemplos 3-4).

(1) *UP-sjef Einar Holand mener landet må deles inn i soner hvor det er tillatt* ***med*** *piggdekk* (AV/VG96/01)

(2) *Ninguém se realiza* ***com*** *este tipo de tormentas.* (DL-N25558-2)

(3) *Heftet er tredelt,* ***med*** *EU og Gatt som én sekvens, med næringsutvikling, miljøarbeid og samvirke som andre del, for så ...* (AV/Bb95/01)

(4) *Entretanto, o comandante da INTERFET, Peter Cosgrove, anunciou ontem que as tropas da força internacional vão ser reforçadas hoje, quarta-feira,* ***com*** *a chegada de outro contingente a Timor Leste.* (DC-N1133-1)

De facto, "adverbial livre" e "ligação ao verbo" são, paradoxalmente, as categorias cuja distinção é menos clara, visto que a segunda classificação foi usada para casos como *bateu com a mão*, enquanto que a primeira foi sobretudo utilizada para SPs que modificam ou classificam a oração completa, como *bateu com violência* ou *com efeito, bateu.*

Note-se, de qualquer maneira, que as duas preposições se portam, curiosamente, de maneira muito semelhante nas duas línguas. Apenas em duas categorias se podem apontar diferenças talvez relevantes: há mais contextos de *med* associados a adjectivos; por outro lado, *com* usa-se mais discursivamente que *med.* Mas note-se que não descontámos os casos em que *med* é partícula em norueguês, o que poderá prejudicar a comparação nas contagens.

Tabela 9a. Os verbos mais frequentes com *com* (MINHANOT e NATPANOT)

verbo	frequência do verbo com argumento SP (*com*)		frequência do verbo com *com* na mesma frase		frequência do verbo com *com* a distância <10		frequência absoluta do verbo		frequência relativa do verbo *com* (<10)	
contar	537	908	608	1262	583	1052	742	2096	72%	50%
acabar	76	286	144	836	120	468	377	2850	32%	16%
prender*	72	126	76	209	75	158	95	601	**79%**	26%
confrontar*	28	227	60	268	61	241	70	299	**87%**	**81%**
encontrar*	27	186	229	962	101	449	985	4047	10%	11%
reunir*	50	174	151	507	95	289	513	1765	18%	16%
concordar	35	131	38	172	39	152	61	314	**64%**	48%
terminar	31	181	119	558	82	378	316	1761	26%	21%
coincidir	22	179	24	187	23	185	30	209	**77%**	**89%**
trabalhar	32	139	118	373	62	223	398	1242	16%	18%
falar	27	167	176	615	69	329	797	2072	8.7%	16%
comparar	16	112	32	148	29	126	41	260	71%	48%
lidar	31	46	34	63	37	55	41	86	**90%**	**64%**
confundir	20	93	31	118	25	107	38	166	66%	64%
preocupar*	20	90	103	407	132	374	206	672	64%	56%
viver	16	75	120	466	61	221	438	1988	14%	11%
relacionar*	14	54	172	498	179	496	185	513	**97%**	**97%**
deparar*	13	71	20	83	15	75	21	104	**71%**	**72%**
conversar	11	71	11	71	16	76	28	131	57%	58%
tratar	11	53	169	489	68	205	588	2241	12%	9.1%

Em seguida, resolvemos passar para estudos "menos" sintácticos, e observar com que itens lexicais (primeiro, verbos) as preposições ocorriam nas duas línguas. Para o português, fizemos um estudo simultâneo em dois corpora, apresentado na tabela 9a, para confirmar se a coocorrência verbal era dependente do estilo do jornal. Visto que a frequência absoluta de um verbo (quantas vezes ele ocorre) tem, claramente, importância na definição de quais os verbos mais frequentes com *com*, calculámos também a frequência relativa do verbo com a preposição (a percentagem de vezes que esse verbo coocorre com *com*). Naturalmente, verbos

mais frequentes terão necessariamente mais ocorrências com *com*, mas confiar apenas na frequência relativa também tem perigos: Não podemos dizer que esse é o único parâmetro relevante, visto que verbos muito frequentes com um número significativo de ocorrências com *com* podem, de facto, indicar que existem sentidos específicos desse verbo associados a essa preposição; tal como em *dar com*, *dar-se com*. (Os verbos marcados com asterisco incluem também a sua forma reflexiva.) Ao traduzi-los para norueguês, chegamos, contudo, à conclusão de que, na sua esmagadora maioria são, ou podem ser, traduzidos também por verbos com *med*: contar: regne med; acabar: ende med; **gjøre ende på**; slutte med; prender(-se): ha å gjøre med; confrontar(-se): **stille overfor**; konfrontere med; encontrar(-se): **møte, treffe;** reunir(-se): ha ett møte med; concordar: være enig med; terminar:ende med; **gjøre ende på;** coincidir: falle sammen med; trabalhar: arbeide (sammen) med; falar: snakke med; comparar: sammenlikne med; lidar: **ta seg av**, hanskes med; confundir: forveksle med; preocupar(-se): **bekymre seg over;** viver: leve (sammen) med; bo sammen med; relacionar(-se): har å gjøre med; deparar(-se): **støte på; komme over**; conversar: snakke med; tratar: snakke med; ha med å gjøre.

Para comparar com o norueguês, e visto que os corpora não estão anotados com função sintáctica nem é possível delimitar o contexto a uma frase, usámos a estratégia mais primitiva de procurar sequências de «verbo seguido de *med*» – ver tabela 9b.

Tabela 9b. Verbos mais frequentes seguidos por *med* no OTBK

være	2673	være, er, var, vært	990,760,519,404	ser, estar com
regne	1034	regner, regne, regnet, regnes	572, 291, 158, 13	contar com
snakke	947	snakke, snakket, snakker	505, 365, 77	falar com
ha	698	har, ha, hadde, hatt	357, 151, 138, 52	ter com, estar com
komme	1106	kom, komme, kommer, kommet	335, 329, 234, 208	vir com
ta	719	ta, tatt, tok, tar	315, 238, 84, 82	**levar**; tirar com
arbeide	612	arbeidet, arbeider, arbeide, -s	289, 180, 127, 16	trabalhar com
bli	496	bli, blir, ble, blitt	280, 104, 83, 29	**acompanhar; ficar grávida**
gjøre	357	gjøre, gjort, gjør, gjorde	239, 52, 44, 22	fazer com, ter com
gå	731	gå, går, gikk, gått	234, 228, 173, 96	ir com; **ser preciso**
leve	274	leve, levd, lever, levde	220, 28, 22, 4	viver com
få	460	få, fått, fikk, får	192, 110, 104, 54	**perceber; levar**
følge	397	følge, følger, fulgte, fulgt, følg, -s	171,109,56,37, 20,4	**pertencer**
begynne	416	begynne, begynte, begynner, -nt	199, 135, 51, 31	começar com, por
sitte	244	sitter, satt, sitte	128, 71, 45	estar com; **dedicar-se a**
øke	328	økt, økte, øker, øke, økes	128, 91, 45, 44, 20	**aumentar**
starte	215	startet, starter, starte, start	105, 57, 47, 6	começar com
skje	310	skjer, skje, skjedd, skjedde	101, 90, 61, 58	**acontecer a**, com

Procedimento idêntico foi realizado para adjectivos, substantivos e pronomes pessoais, não tendo em nenhum dos casos obtido diferenças flagrantes. Tentámos, por isso, comparar, então, não a distribuição sintáctica, mas aquilo a que poderíamos chamar os "campos semânticos" associados com as preposições. O facto de em quase metade dos casos não ter sido fácil

identificar um sentido não nos deverá surpreender, visto que nem sempre a preposição tem sentido. Na tabela 10 temos os resultados obtidos por uma análise subjectiva dos contextos, em 200 casos de *med* no OBTK e em 300 casos, respectivamente cem no CETEMPúblico, no DiaCLAV e no Diário do Minho:

Tabela 10. 100 casos de *med* no OTBK duas vezes e 100 casos de *com* três vezes

Campo semântico	1.a vez	2.a vez	CETEMP	DiaCLAV	Minho
companhia/parceria	25	18	27	18	21
acontecimento	8	6	15	8	11
estado de espírito/atitude	6	7	12	8	6
instrumento/meio	4	7	3	2	2
característica física	17	7	15	18	15
característica moral				2	2
transporte	1	2			
tempo	1	1			
grau/modo				3	
outros	42	50	25	40	43

Nota-se, mais uma vez, uma concordância inesperada nas distribuições obtidas, tendo apenas sido verificada uma tendência ligeiramente maior de uso de *com* com acontecimentos e *med* com instrumentos. Estes campos estão relacionados, mas não de forma trivial, com o seu comportamento sintáctico (veja-se Binot & Jensen, 1987; Ravin, 1993).

Ainda convencidos de que eventuais diferenças poderiam surgir comparando directamente o "mesmo texto" nas duas línguas, calculámos o grau de compatibilidade entre as duas línguas em relação à tradução do inglês *with*. A tabela 11 mostra que apenas 30% dos casos acusam diferença entre as estratégias de tradução norueguesa e portuguesa.

Tabela 11. Relação entre *med* e *com* no OMC En-No-Po

with traduzido por *com* e por *med*	919		58,0%
with traduzido por *com* e não *med*	225	-19%[3]	14,2%
with traduzido por *med* e não *com*	251	-18%	15,8%
with traduzido por não *med* e não *com*	190		12,0%
Total de *with*	1585		

Estes dados contrastivos, contudo, só podem ser interpretados se tivermos um conjunto muito maior de dados sobre a fidelidade sintáctica na tradução, que nos permitam verificar se as diferenças reflectem o procedimento natural dos tradutores ou são indicadoras de maior ou menor parecença entre as preposições.

A tabela 12 apresenta as frequências absolutas de *com*, *with* e das suas combinações, e a frequência relativa em termos de ocorrências da preposição. A tabela 13 apresenta o mesmo tipo de resultados, agora para o par *med* e *with*. Os números não são muito diferentes dos apresentados em Schmied (1998) para a relação tradutiva entre *with* e o alemão *mit*. Estes valores são interessantes por confirmarem a maior frequência de *med* do que de *com*, mas também demonstram que *with* está igualmente distanciada das duas preposições sobre as quais

[3] Casos em que a presença de *com* ou de *med* não estava relacionada com o *with* inglês.

o nosso estudo incidia. A introdução do inglês, assim, em vez de ser esclarecedora, provavelmente apenas porá em evidência as diferenças das duas línguas em relação à língua inglesa. Não tendo no entanto corpora bilingues português-norueguês, efectuámos um estudo mais fino do tipo de diferenças na tradução entre *with* e *com*, que infelizmente por falta de espaço não podemos incluir aqui.

Tabela 12. Relações entre *with* e *com* no COMPARA e no ENPC (En-Po)

Padrão	COMPARA	ENPC (En-Po)	Total
com em texto original	757		
com em tradução do inglês	1657	2313	3970
with em texto original	1063	1585	2648
with em tradução do português	882		
com traduzido por *with*	480 (63.4%)		
com não traduzido por *with*	277 (36.6%)		
with traduzido por *com*	804 (75,6%)	1144 (72,2%)	1948 (73,5%)
with não traduzido por *com*	259 (24%)	441 (27,8%)	700 (26,4%)
com em tradução não correspondendo a *with*	751 (45,3%)	1028 (44,4%)	1779 (44,8%)
with em tradução não correspondendo a *com*	387 (43,9%)		

Tabela 13. Relações entre *with* e *med* no ENPC (Fiction e Non-Fiction)

Padrão	Fiction	Non Fiction	Total ENPC
med em texto original	4799	2431	7230
med em tradução do inglês	5319	2903	8222
with em texto original	3258	1799	5057
with em tradução do norueguês	3557	1782	5339
med traduzido por *with*	2678 (58,0%)	1280 (52,6%)	3958 (54,7%)
med não traduzido por *with*	2014 (42,0%)	1151 (47,3%)	3165 (43,8%)
with traduzido por *med*	2338 (71,7%)	1280 (71,1%)	3618 (71,5%)
with não traduzido por *med*	920 (28,2%)	511 (28,4%)	1431 (28,2%)
with em tradução não correspondendo a *med*	879 (24,7%)	567 (31,8%)	1446 (27%)
med em tradução não correspondendo a *with*	2761 (51,9%)	1461 (50,3%)	4222 (51,3%)

Em conclusão, a nossa procura da diferença entre *med* e *com* ainda não terminou. Uma observação mais fina de casos específicos impõe-se, agora. Neste artigo, tentámos ilustrar várias formas de abordar o problema numa macroperspectiva, usando dicionários e sobretudo corpora, salientando as opções, sempre exigidas, em qualquer investigação.

Agradecimento

Este artigo foi originalmente pensado em colaboração com Jan Engh, que, por falta de tempo, acabou por não participar. Toda a recolha de material contrastivo, assim como a tradução para norueguês das palavras portuguesas e a análise dos contextos noruegueses foi feita ou revista por ele, a quem muito agradeço. Agradeço também a Stig Johansson, Signe Oksefjell, Jarle Ebeling e ao resto da audiência do IBA os comentários pertinentes aquando da apresentação de uma versão preliminar deste trabalho.

Bibliografia

Bick, Eckhard. 2000: The Parsing System "Palavras": Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework. Tese de doutoramento, Aarhus University. Aarhus: Aarhus University Press.

Binot, Jean-Louis & Karen Jensen. 1987: A semantic expert using an online standard dictionary. *Proceedings of IJCAI'87*: 709-14.

Carvalho, Orlene Lúcia de Saboia. 2001: *Lexicografia Bilingüe Português/Alemão: Teoria e Aplicação à Categoria das Preposições*. Brasília: Thesaurus Editora.

DLPC. 2001: *Dicionário da Língua Portuguesa Contemporânea.* Academia das Ciências de Lisboa e Fundação Calouste Gulbenkian, 2 vols. Lisboa: Verbo.

Engh, Jan. Em preparação: *Portugisisk-norsk ordbok.*

Landrø, Marit I. & Boye Wangensteen et al 1986: *Bokmålsordboka*. Oslo: Universitetsforlaget

Nilsson, Kåre. 1994: *Norsk-portugisisk ordbok*. Oslo: Universitetsforlaget.

Ravin, Yael. 1993: Disambiguating and Interpreting Verb Definitions. In Karen Jensen, George E. Heidorn & Stephen D. Richardson (eds.). *Natural Language Processing: The PLNLP Approach*: 176-189. Dordrecht: Kluwer Academic Press.

Santos, Diana & Eckhard Bick. 2000: Providing Internet access to Portuguese corpora: the AC/DC project. In M. Gavriladou et al (eds.). *Proceedings of the Second International Conference on Language Resources and Evaluation LREC'2000*: 205-210. Atenas: ELRA

Santos, Diana & Caroline Gasperin. 2002: Evaluation of parsed corpora: experiments in user-transparent and user-visible evaluation. In M. G. Rodríguez & Carmen Paz Suárez Araujo (eds.). *Proceedings of LREC 2002, the Third International Conference on Language Resources and Evaluation*: 597-604. Las Palmas de Gran Canaria: ELRA.

Schmied, Josef. 1998: Differences and similarities of close cognates: English *with* and German *mit*. In Stig Johansson & Signe Oksefjell (eds.). *Corpora and Cross-linguistic Research: Theory, Method, and Case Studies*: 255-275. Amsterdam: Rodopi.