



**MAKE IT SIMPLE WITH PARAPHRASES:
AUTOMATED PARAPHRASING FOR AUTHORIZING AIDS
AND MACHINE TRANSLATION**

Anabela Barreiro



PHD DEFENSE IN LINGUISTICS

25 May 2009

ADVISORS:

PROFESSOR BELINDA MARY HARPER SOUSA MAIA
FACULDADE DE LETRAS DA UNIVERSIDADE DO PORTO

PROFESSOR ADAM MEYERS
NEW YORK UNIVERSITY

GENERAL MOTIVATION

- 1. MT changed the world of translation**
it can no longer be ignored or underestimated
- 2. It is increasingly more used and useful**
it often replaces human translation when the client requires gisting
- 3. However, good quality MT is still an ambitious goal**
after +50 years, MT still eludes researchers and developers who are constantly challenged to provide better translations

How can MT move forward?

MACHINE TRANSLATION CHALLENGES

1. homography and common-noun nuance
2. anaphora + distant referential associations
3. ellipsis
4. extra-sentential and extra-textual information + extra-linguistic knowledge
5. lexical divergences, idioms, etc.
6. named entities
7. long sentences and wordiness
8. unusual word order
9. **multiword expressions (including support verb constructions)**

SUPPORT VERB CONSTRUCTION

= multiword or complex predicate consisting of a semantically weak verb (the support verb), and a predicate noun, a predicate adjective, or a predicate adverb.

Predicate nouns can be:

➤ **morphologically related to a verb**

PT: fazer uma apresentação de N = apresentar N

EN: pay a visit to N = to visit N

➤ **autonomous**

PT: fazer um mestrado - *mestrar

EN: have fun - *to fun

WHY SUPPORT VERB CONSTRUCTIONS?

- **Abundant in language and indispensable to communicate**

ASSESSMENT TEST: Search pattern in corpora (COMPARA)

<dar.V+inf> + <tomar.V+inf> + <pôr.V+inf> + <fazer.V+inf> + <ter.V+inf> <Modif?> <N>

Selected the 1st 100 sentences for each verb and manually annotated the 500 sentences as SVC or NO SVC

PT	dar	tomar	pôr	fazer	ter
transl.	give	take	put	make/do	have
NO SVC	11%	12%	33%	53%	80%
SVC	89%	88%	77%	47%	20%

RESULTS:

These verbs occur very frequently in SVCs

Overall, in **64.2%** of their occurrence, these verbs are support verbs

- **Extensively and systematically studied within the Lexicon-Grammar Theory**

[M. Gross and followers], also in Portuguese [Ranchhod, 1983, 1990] [Baptista, 2005] [Chacoto, 2005] and in contrastive studies [Salkoff, 1990, 1999]

- **Often can be replaced by stylistic variants** - paraphrases
- **Present several degrees of variability** (variable - invariable - idiomatic)
- **Semantically weak and ambiguous**
- **They cannot be translated literally**

MT OF SUPPORT VERB CONSTRUCTIONS

METRA3

Tradução de "Eles querem dar uma vista de olhos ao museu"

E qual a melhor tradução na sua opinião ?

- They are going to give a sight of eyes to the museum FreeTranslation
- They want to give an eye sight to the museum E-Translation Server
- They want to give a sight of eyes to the museum WorldLingo
- Nenhuma destas traduções

METRA3

Tradução de "Eles querem ver o museu"

E qual a melhor tradução na sua opinião ?

- They want to see the museum WorldLingo
- They are going to see the museum FreeTranslation
- They want to see the museum E-Translation Server
- Nenhuma destas traduções

MT SYSTEMS
MISHANDLING SUPPORT
VERB CONSTRUCTIONS

AH... THAT'S
BETTER!

WE PROMPTLY VERIFY THE SUPERIOR QUALITY
OF THE RESULTS IN THE SECOND CASE!

THESIS STATEMENT

Linguistic knowledge of paraphrases can improve the quality of machine translation

➤ **The conversion of support verb constructions into morphosyntactically and/or semantically related verbs produces a controlled language that is applicable both to general and domain specific contexts and makes MT more reliable**

1. Reduces complexity (= paraphrasing by simplification)
2. Reduces lexical ambiguity (paraphrase conveys the correct interpretation)
3. Facilitates interpretation when the support verb construction is idiomatic
4. Reduces wordiness (sometimes, text quality)
5. Improves translation quality

OBJECTIVES

1. Build a body of lexical, syntactic and semantic knowledge around support verb constructions
2. Apply this linguistic knowledge to paraphrasing
3. Improve machine translation

THEORETICAL BACKGROUND - LINGUISTICS

- **Meaning-Text Theory** [Mel'čuk, 1988, 1996, 2003] – Support verb constructions are described and formalized in terms of lexical functions (LF). LF can be used as an interlingua to facilitate lexical transfer in MT projects. [Milićević, 2007] establishes a typology of paraphrases based on lexical-syntactic rules.
- **FrameNet** [Fillmore et al., 2002, 2003] – records the information necessary for the representation of argument mapping relations between a **support verb** and a nominalization.
- **NomBank** [Meyers et al., 2004b, 2004b] - maps syntactic positions in nominalizations to verbal arguments and identify the allowed complements for a nominalization, relating the nominal complements to the arguments of the corresponding verb, including information about **support verbs**.
- **Lexicon-Grammar Theory** [M. Gross and followers] - theoretical framework adopted in this research

THEORETICAL FRAMEWORK

Lexicon-Grammar, founded by [M. Gross \(1975, 1981, 1996\)](#) established in the Harrisian framework of the transformational grammar

the basic unit of analysis and description of meaning is the **simple or elementary sentence** (predicate or head of the sentence and its compulsory arguments)

predicate nouns select both arguments and support verbs

transformations are **non-oriented** relations of equivalence between sentences

systematic and **detailed** description of predicates

formal representation of descriptions

suitable for **automated text processing** (NLP)

applicable to **machine translation** - LG studies contemplate many languages

RELATED WORK ON PARAPHRASING

- **question answering** – discovering paraphrased answers provides additional evidence that an answer is correct - [Ibrahim et al., 2003], [Paşca, 2003], [Duboué & Chu-Carroll, 2006]
- **information extraction** and **text mining** - paraphrases help text categorization tasks or mapping to texts with similar characteristics - [Ibrahim et al., 2003], [Shinyama et al., 2002] [Shinyama & Sekine, 2003], [Sekine, 2005] [Paşca, 2005], [Paşca & Dienes, 2005]
- **summarization** - the identification of paraphrases allows information across documents to be condensed and helps improve the quality of the generated summaries [McKeown et al., 2002], [Barzilay, 2001, 2003], [Hirao et al., 2004] [Zhou et al., 2006b]
- **natural language generation** – the generation of paraphrases allows the production of more varied and fluent text - [Iordanskaja et al. 1991]
- **machine translation** – paraphrases help create a more fluent translation and are valuable in the evaluation of MT results - [Zhou et al., 2006], [Callison-Burch et al., 2006a, 2006b, 2007 and 2008]

RELATED WORK ON CONTROLLED WRITING AND STYLISTIC EDITORS/AUTHORING AIDS

- **Controlled language** - application of human language techniques in an industrial environment. It is used to support technical writers in producing high quality technical documentation by checking spelling, grammar, style and terminology in technical documents (text optimisation). Automated rewriting for controlled language translation
- **Stylistic editing** - focuses on clarity and expression, re-ordering of paragraphs, sentences and words; elimination of verbiage and jargon; simplification of untangle complicated clauses; elimination of inconsistencies, use of stylistic paraphrases (reductions/extensions)
- **Authoring aids** - contemplate mostly spell checking, simple grammar checking (correction of text), writing and morphological variants and synonyms [LREC 2008](#)

Some available tools:

MULTIDOC [[Haller, 2000](#)]

KANT CE Checker [[Mitamura and Nyberg, 2001](#)] [[Mitamura et al., 2003](#)] [[Rascu, 2006](#)]

CLAT [[Schmidt-Wigger, 1998](#)]; [[Carl et al., 2002](#)]; [[Hernandez and Rascu, 2004](#)]

CLOUT - the Controlled Language Optimized for Machine Translation.

METHODOLOGY

Task 1

Analysis, extraction from corpora and formalization of support verb constructions by using information described in dictionaries and grammars – NooJ linguistic environment was the tool used

Task 2

Paraphrasing of predicate noun constructions

Phase 1

Development of resources and pre-processing

Construction of simple grammars to recognize and extract SVCs from texts

Definition of properties in the dictionary and establishment of semantic links between verbs and SVCs + establishment of classes of paraphrases

Application of grammars with dictionary information to texts

Refinement of grammars

Finer recognition of SVCs in texts -> paraphrasing + translation

Phase 2

Evaluation experiments

FORMALIZATION IN THE DICTIONARIES

Explicit marking of derivation and semantic verb associations

Adjective entries:

- Identification of derivational paradigms for **predicate adverbs - adverbializations** (annotation *AVDRV*)

literal,A+FLX=PRINCIPAL+IN+symb+EN=literal+**DRV=AVDRV00**:LITERALMENTE

Autonomous predicate nouns:

- Identification of **autonomous predicate nouns** (annotation *Npred*)
- Identification of a **semantically related verb** (annotation *VRB*)
- Link to the **support verb** (annotation *VSUP*)

curso,N+FLX=ANO+**Npred**+IN+inst+EN=course+**VSUP=tirar**+**VRB=estudar**+NPrep=de+Det=um

FORMALIZATION IN THE DICTIONARIES

Explicit marking of derivation and support verb associations

Verb entries:

- Identification of derivational paradigms for **predicate nouns - nominalizations** (annotation *NDRV*) and **predicate adjectives** (annotation *ADRV*)
- Link to the **support verbs** (annotation *VSUP*) occurring with the derived **predicate nouns** and **adjectives**

adaptar,V+FLX=FALAR+Aux=1+INOP57+Subset=132+EN=adapt+**VSUP=fazer**+**DRV=NDRV00**:CANÇÃO
azedar,V+FLX=LIMPAR+Aux=1+OBJTRundif98+Subset=740+EN=sour+**VSUP=estar**+**DRV=ADRV00**:ALTO

NEW RESOURCES

➤ Port4NooJ

- a set of open source Portuguese linguistic resources representing ontological relations, and paraphrasal relations. This resources integrates a bilingual extension for PT-EN MT.

Semantic relations	
Verb - Predicate Nouns	8472
Predicate Adjectives - Adverbs	222
Total Links	8694

- Publicly available resources at:

<http://www.nooj4nlp.net>

<http://www.linguateca.pt/Repositorio/Port4Nooj/>

➤ DicTUM (Dicionário de Termos e Unidades Multipalavra)

- a Dictionary of Multiword Expressions

NEW TOOLS

➤ ReWriter

- a monolingual standalone paraphraser to pre-edit texts, using paraphrasing capabilities
- Portuguese version **ReEscreve** - publicly available service at:
<http://www.linguateca.pt/ReEscreve/>

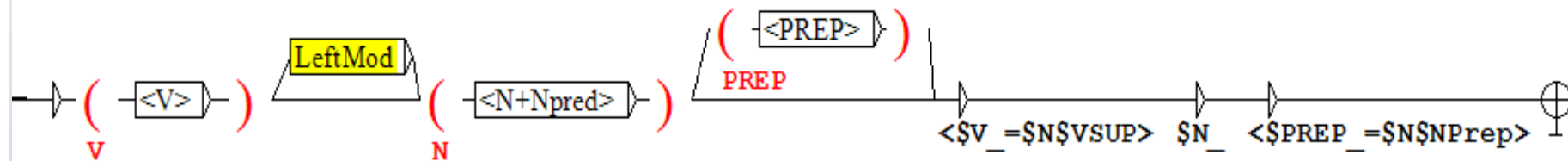
➤ ParaMT

- a bilingual/multilingual paraphraser to be integrated in machine translation systems

USEFULNESS OF THE AFOREMENTIONED PARAPHRASING TOOLS

- Controlled writing & text pre-editing
- Machine translation
- Text production and stylistics (authoring aids)

REWRITER: A STANDALONE PARAPHRASER



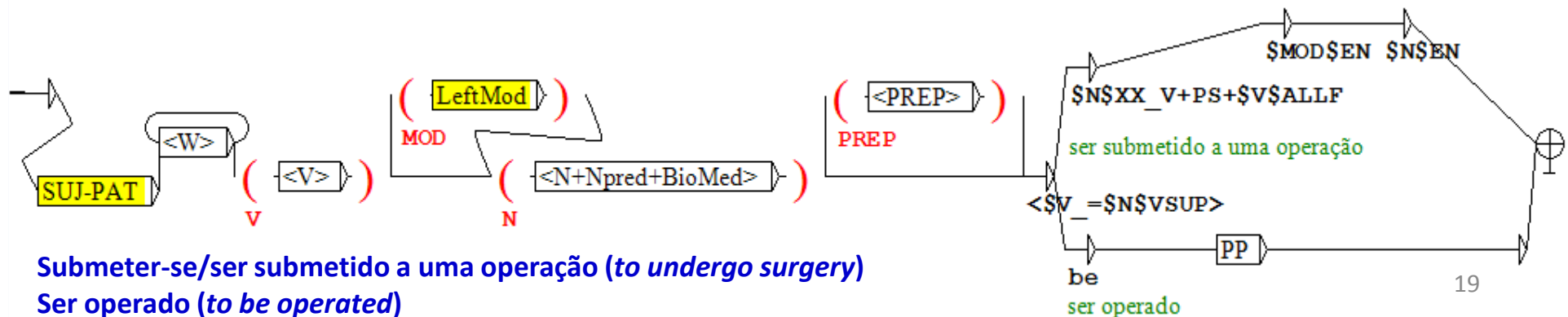
gosto de ver o comboio a	fazer corridas /correr	à velocidade máxima ao long
o de cheque especial para	fazer doações /doar	às entidades que escolher. A
cores e, quando é preciso ir	fazer filmagens/filmar	fora do estúdio, às vezes fic
o que queria trocar de pares e	fazer um jogo /jogar	ao melhor de três sets , mas
o dra deu-me um papel para	fazer uma lista de/listar	todas as coisas boas que ex
res foram à caracterização	fazer uns retoques/retocar ,	outros estão a descansar n
ça, o cirurgião Faivre, ao	fazer uma amputação amputar	
ça, o cirurgião Faivre, ao	fazer uma amputação efectuar uma amputação	
ça, o cirurgião Faivre, ao	fazer uma amputação realizar uma amputação	

CONTROLLED WRITING
TEXT PRE-EDITING

CONTROLLED WRITING
TEXT STYLISTICS

Elementary SVC > Lexical Verb – fazer uma amputação = amputar (to amputate)

Elementary SVC > non-elementary SVC - realizar/efectuar uma amputação (to perform an amputation)



Submeter-se/ser submetido a uma operação (to undergo surgery)

Ser operado (to be operated)

REWRITER – APPLICATION INTERFACE

Ficheiro Editar Ver Histórico Marcadores Ferramentas Ajuda

http://poloclup.linguateca.pt/Reescreve/re.pl

TEXT PRODUCTION



Google

Mais Visitados Começar Aqui Últimas Novidades http://www.google.pt...

& STYLISTICS

Foram encontradas 26 sugestões para o texto apresentado.

Também gostava de ir [dar aulas](#) [leccionar](#) [ensinar](#) na universidade, mas não posso dar-me ao luxo de ser colocado agora que estou casado e com um bebé a caminho.

Escute, os administrativos da MLA vão [dar uma festa](#) [festejar](#) hoje à noite, na suite do último andar.

Resolvi [fazer a experiência](#) [experimentar](#) e passei cuidadosamente um dedo pela face que se me apresentava;

Catherine corou porque se sentiu como se estivessem a [fazer troça](#) [troçar](#) dela.

Estava preso há dois meses e, embora continuasse fraco e em perigo de uma recaída, tive de [fazer uma viagem de](#) [viajar](#) algumas léguas para chegar à sede do tribunal.

Quanto a Catherine, nem sequer fingia querer [fazer conversa](#) [falar](#) [conversar](#);

Disse que Frances nem sequer estava a [pôr a hipótese de](#) [considerar](#) [fazer um aborto](#) [abortar](#) e que provavelmente ir ter o bebé em Espanha e depois a sua voz esmoreceu.

William largou o rolo de cordel e avançou em passo estugado para o caminho de acesso, a acenar e a [fazer gestos](#) [gesticular](#).

Dizia que as escadas eram uma maçada, que o senhorio estava a [fazer planos para](#) [planejar](#) instalar um elevador, o que seria uma vantagem quando Frances fosse lá ficar com o bebé.

Ultimamente comecei a [dar ainda mais valor](#) [valorizar](#) à privacidade e ao anonimato desta casa.

Só tenho pena de a minha mãe e o meu pai já não serem vivos e eu não poder levá-los a [dar uma volta](#) [passear](#) nele.

Pede-me sempre desculpa se uma ou outra reacção particularmente forte a uma agulha me faz [dar um salto](#) [saltar](#);

Dois espasmos no joelho, enquanto estive a escrever isto, um dos quais forte ao ponto de me fazer [dar um grito](#) [gritar](#).

Não querendo [dar a impressão de](#) [parecer](#) ser racista, engoli a reclamação.

Na verdade, Gordon Masters tinha decidido [dar apoio a](#) [apoiar](#) Philip para o intercâmbio em euforia

Adrian, é perfeitamente capaz de [dar glória a](#) [glorificar](#) Deus.

Dediquei-me então ao objectivo de [pôr termo](#) [terminar](#) ao racismo e à injustiça, objectivo que prossegui e prosseguirei enquanto viver.

Mas a criada que Miss Kenton deixara a [tomar conta](#) [tratar](#) [cuidar](#) dele levantou-se ao ver-me e começou a sacudir-lhe o ombro.

O movimento do braço, a [tomar balanço](#) [balançar](#), desequilibróu-a.

não há de que [ter medo](#) [recear](#).

Sai da loja antes de ele [ter tempo de](#) [poder](#) chamar a brigada de narcóticos.

Decidi tentar me levantar dessa vez e deslizei os pés para o chão apalpando à procura dos chinelos, antes de [ter chance de](#) [poder](#) mudar de ideia.

A noite parecia não [ter fim](#) [acabar](#).

Quando se dirigiu ao gabinete exterior e estendeu o envelope a Jean, esta pareceu [ter dificuldade em](#) [evitar](#) olhá-lo nos olhos.

Por agora não era possível olhar para trás e tudo o que podia fazer era seguir em frente e [ter esperança de](#) [esperar](#) acabar antes da semana seguinte.

Ver detalhes:

- #1 ([dar aulas](#))
- #2 ([dar uma festa](#))
- #3 ([fazer a experiência](#))
- #4 ([fazer troça](#))
- #5 ([fazer uma viagem de](#))
- #6 ([fazer conversa](#))
- #7 ([pôr a hipótese de](#), [fazer um aborto](#))
- #8 ([fazer gestos](#))
- #9 ([fazer planos para](#))
- #10 ([dar ainda mais valor](#))
- #11 ([dar uma volta](#))
- #12 ([dar um salto](#))
- #13 ([dar um grito](#))
- #14 ([dar a impressão de](#))
- #15 ([dar apoio a](#))
- #16 ([dar glória a](#))
- #17 ([pôr termo](#))
- #18 ([tomar conta](#))
- #19 ([tomar balanço](#))
- #20 ([ter medo](#))
- #21 ([ter tempo de](#))
- #22 ([ter chance de](#))
- #23 ([ter fim](#))
- #24 ([ter dificuldade em](#))
- #25 ([ter esperança de](#))

Copiar texto

20

2477 / 2500 caracteres

PARAMT: A BILINGUAL/MULTILINGUAL PARAPHRASER FOR MT

a fazer um estágio para	dar aulas de/teach	religião, mas não se import
m -- os filhos -- juntos e	fizeram a mudança para/change	Johannesburg, e ensinaram
. Necessitava apenas de	ter a certeza de/know	que não escapara à sua
ente hipotética. -- Deves	ter alguma ideia/know	. Dorothy andava a fazer um
. não podemos deixar de	ter cautela/beware	. Pobre Caro, pensou Lync
ra dos chinelos, antes de	ter chance de/can	mudar de idéia. Como pos
ope a Jean, esta pareceu	ter dificuldade em/avoid	olhá-lo nos olhos. Deixou
ao Kiss dela. Apesar de	ter falta de/lack	amor-póprio, isso não sign
igos e imprensa estava a	ter lugar /occur	numa longa galeria com car
guiu ter filhos. -- Tens de	ter mão /control	nessa confusão toda. Sam
spondi, minha mãe deve	ter medo de/fear	cobras. Eu disse no Gabin
da loja antes de ele	ter tempo de/could	chamar a brigada de narcó
a triste aventura havia de	ter um fim/finish	.
Ela ouvira a tia Velma	ter uma discussão com/argue	Jack acerca de mostarda r
de olhos fechados para	ter uma ideia de/know	como seria ser cego e
ter paciência.» «Voltei a	ter uma imensa vontade de/want	viver. A conversa parecia :

**MACHINE
TRANSLATION**

**Recognition of Portuguese SVC
and translation into English verbs**

EVALUATION EXPERIMENT 1

5 support verbs

500 baseline sentences - manually annotated

100 for each elementary support verb

Evaluation of
recognition
and
paraphrasing
of SVC with
ReWriter

	SVC Recognition Precision	SVC Recognition Recall	SVC Paraphrasing Precision
Pôr	73/73 - 100%	73/100 - 73%	72/73 - 98.6%
Tomar	75/75 - 100%	75/100 - 75%	68/73 - 93.1%
Ter	65/65 - 100%	65/100 - 65%	59/65 - 90.7%
Dar	57/60 - 95%	57/100 - 57%	46/51 - 90.1%
Fazer	43/45 - 95.5%	43/100 - 43%	40/45 - 88.8%
	98.4%	62.6%	93.4%

EVALUATION EXPERIMENT 2

Impact of
SVC
paraphrases
in translation

5 different support verbs

100 sentences with SVC (for each language PT and EN)

20 sentences for each support verb – manually assigned a paraphrase for each one of them
> 20 SVC sentences + 20 paraphrases of the original SVC sentences

TOTAL: 200 sentences for each language
100 with SVC + 100 with paraphrases of the SVC

10 testers translated and evaluated translations
students of the MA in Translation and Linguistic Services

	Equal quality result	SVC translation is better	Paraphrase translation is better
EN-PT MT	17%	26%	57%
PT-EN MT	19%	30%	51%

For both EN-PT and PT-EN, testers considered that the paraphrase translations presented a better quality. The evaluation results clearly confirm that these paraphrases helped the systems produce better MT.

MAIN CONCLUSION

➤ Linguistic knowledge formalized in this research when applied to a machine translation system improves its output quality.

When SVC were identified and replaced with semantically equivalent or similar verbal expressions as a pre-processing step to translating:

a **21% improvement** was observed in the evaluated quality of the results of PT-EN machine translation

a **31% improvement** in the results of EN-PT machine translation.

FUTURE RESEARCH SUGGESTED BY THIS THESIS

- Enlargement of syntactic-semantic relations between predicates, and other morphosyntactically and semantically related elements

Exs: *apresentar – apresentação – apresentado*

literal – literalmente

- Paraphrasing of various linguistic phenomena for ReWriter extensibility (comprising stylistic variance and controlled language)
- Development and enhancement of ParaMT
- Enhancement of the machine translation model

PARAPHRASING FOR REWRITER

EXTENSIBILITY

A formal linguistic study of paraphrases as these ones would represent a significant contribution to NLP in general, and to MT in particular.

Linguistic Phenomenon	Expression or sentence	Paraphrase
Adverbial	à volta da órbita <i>around the orbit of the eye</i> de forma interactiva <i>in an interactive way</i>	periorbital <i>periorbital</i> interactivamente <i>interactively</i>
Relative Clause	N0 que têm sido escritos <i>N0 that have been written</i> A velocidade a que se move a luz <i>F *The speed to which light moves</i> O papel que a Europa tem <i>The role that Europe plays/has</i> As dificuldades que temos <i>The difficulties we have</i>	N0 que foram escritos > N0 escritos <i>N0 that were written > N0 written</i> A velocidade da luz <i>The speed of light</i> O papel da Europa <i>The role of Europe ≡ Europe's role</i> As nossas dificuldades <i>Our difficulties</i>
If clause	se for necessário <i>if it is necessary</i>	se necessário <i>if necessary</i>
Named Entity	A rainha de Inglaterra <i>The queen of England</i>	A rainha inglesa <i>The British queen</i>
Noun Phrase	O heróico povo português <i>The heroic Portuguese people</i>	Os heróicos portugueses <i>The heroic Portuguese</i>

RELEVANT PUBLICATIONS

- Belinda Maia & Anabela Barreiro "Uma experiência de recolha de exemplos classificados de tradução automática de inglês para português". In Diana Santos (ed.), *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa*. Lisboa, Portugal:IST Press, 2007, pp. 205-216.
- Luís Sarmiento, Anabela Barreiro, Belinda Maia & Diana Santos "Avaliação de Tradução Automática: alguns conceitos e reflexões". In Diana Santos (ed.), *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa*. Lisboa, Portugal:IST Press, 2007, pp. 181-190.
- Anabela Barreiro. "ParaMT: a Paraphraser for Machine Translation". In António Teixeira, Vera Lúcia Strube de Lima, Luís Caldas de Oliveira & Paulo Quaresma (eds.), *Computational Processing of the Portuguese Language, 8th International Conference, Proceedings (PROPOR 2008) Vol. 5190*, (Aveiro, Portugal, 8-10 de Setembro de 2008), Springer Verlag, pp. 202-211.
- Anabela Barreiro. "Port4NooJ: Portuguese Linguistic Module and Bilingual Resources for Machine Translation". In *Proceedings of the 2007 International NooJ Conference (Barcelona, Spain, June 7-9, 2007)*, Cambridge Scholars Publishing.
- Anabela Barreiro & Elisabete Ranchhod. "Machine Translation Challenges for Portuguese". *Linguisticæ Investigationes* 28.1 (2005), pp. 3-18. (Machine Translation, Controlled Languages and Specialised Languages). Amsterdam/Philadelphia: John Benjamins Publishing Company. ISSN: 0378-4169.
- Anabela Barreiro "Novas Ferramentas e Recursos Linguísticos para a Tradução Automática: Por ocasião d'Ó Fim do Início de uma Nova Era no Processamento da Língua Portuguesa". In Luís Costa, Diana Santos & Nuno Cardoso (eds.), *Perspectivas sobre a Linguatca / Actas do encontro Linguatca : 10 anos. Linguatca*, 2008, pp. 13-23.
- Anabela Barreiro. "Port4NooJ Linguistic Resources Overview". 2008.
<http://poloclup.linguatca.pt/Port4NooJ/ResourcesOverview.pdf>
- Anabela Barreiro. "Tutorial introdutório do Port4NooJ". 21 de Novembro de 2008.

REFERENCES

- Max Silberztein. "NooJ: A Cooperative, Object-Oriented Architecture for NLP". *INTEX pour la Linguistique et le traitement automatique des langues*. Presses Universitaires de Franche-Comté. Cahiers de la MSH Ledoux, 2004. Besançon, France.
- Adam Meyers, Ruth Reeves, Catherine Macleod, Rachel Szekely, Veronika Zielinska, Brian Young & Ralph Grishman. "Annotating noun argument structure for NomBank". In Maria Teresa Lino, Maria Francisca Xavier, Fátima Ferreira, Rute Costa & Raquel Silva (eds.), *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC'2004)* (Lisboa, Portugal, May 26-28, 2004).
- Morris Salkoff. *A French-English Grammar: A Contrastive Grammar On Translational Principles*. Amsterdam, Philadelphia: John Benjamins Publishing Company. 1999.
- Maurice Gross. "Les bases empiriques de la notion de prédicat sémantique". *Formes Syntaxiques et Prédicat Sémantiques, Langages* **63** (1981), pp. 7-52. Paris, França: Larousse.
- Maurice Gross. "Simple sentences. Discussion of Fred W. Householder". *Text Processing* (1982), pp. 297-315. Almqvist Wiksell.
- Elisabete Ranchhod (ed.). *Sintaxe dos predicados nominais com Estar*. Lisboa, Portugal: INIC. 1990.
- Jorge Baptista. *Sintaxe dos Nomes Predicativos com verbo-suporte SER DE*. Lisboa: Fundação para a Ciência e a Tecnologia/Fundação Calouste Gulbenkian. 2005.
- Lucília Chacoto. *O Verbo Fazer em Construções Nominiais Predicativas*. PhD dissertation. Universidade do Algarve, Portugal. 2005.
- Diana Santos. "Lexical gaps and idioms in Machine Translation". In Hans Karlgren (ed.), *Proceedings of the 14th International Conference on Computational Linguistics (COLING'90) 2*, (Helsinki, 20-25 August 1990), pp. 330-335.

REFERENCES

- Maurice Gross. "Simple sentences. Discussion of Fred W. Householder". *Text Processing* (1982), pp. 297-315. Almqvist Wiksell.
- Zellig Harris. "Transformations in Linguistic Structure". *Proceedings of the American Philosophical Society* (1964), pp. 418-422. Repr. in 1970a.472-481
- Catherine Fuchs. *La Paraphrase*. Presses Universitaires de France. 1982. ISBN: 9782130371038.
- Bud Scott. "The Logos Model: An Historical Perspective". *Machine Translation* **18** (2003), pp. 1-72.
- Belinda Maia & Anabela Barreiro "Uma experiência de recolha de exemplos classificados de tradução automática de inglês para português". In Diana Santos (ed.), *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa*. Lisboa, Portugal: IST Press, 2007, pp. 205-216.
- Chris Callison-Burch. Paraphrasing and Translation. PhD Thesis. University of Edinburgh. 2007.
- Fillmore, C. J., Johnson, C. R. & Petruck, M. R. L. "Background to Framenet". *International Journal of Lexicography* **16.3** (2003), pp. 235-250.
- Igor A. Mel'čuk "Lexical Functions: A Tool for the Description of Lexical Relations in the Lexicon". In L. Wanner (ed.), *Lexical Functions in Lexicography and Natural Language Processing*. Amsterdam/Philadelphia: Benjamins, 1996, pp. 37-102.
- Regina Barzilay & Lillian Lee. "Learning to Paraphrase: An Unsupervised Approach Using Multiple-Sequence Alignment". In *Proceedings of Human Language Technology conference (HLT- NAACL 2003)* 2003, pp. 16-23.
- Jasmina Milićević. *Semantic Equivalence Rules in Meaning-Text Paraphrasing*. Amsterdam and Philadelphia: Benjamins. 2007. ISBN: 978-90-272-3094-2. In Honour of Igor Mel'čuk
- Jasmina Milićević. *La paraphrase. Modélisation de la paraphrase langagière*. Bern: Peter Lang. 2007. ISBN: 978-3-03911-197-8.

ACKNOWLEDGMENTS

This research work was funded by *Fundação para a Ciência e a Tecnologia* (doctoral scholarship SFRH/BD/14076/2003), co-financed by POSI and partly supported by *Fundação para a Computação Científica Nacional*.

In addition to the doctoral scholarship, I had a ½ month contract from Linguateca to help make Port4NooJ and ReEscreve public.





**MAKE IT SIMPLE WITH PARAPHRASES:
AUTOMATED PARAPHRASING FOR AUTHORIZING AIDS
AND MACHINE TRANLATION**

Anabela Barreiro



PHD DEFENSE IN LINGUISTICS

MAY, 25TH 2009

ADVISORS:

PROFESSOR BELINDA MARY HARPER SOUSA MAIA
FACULDADE DE LETRAS DA UNIVERSIDADE DO PORTO

PROFESSOR ADAM MEYERS
NEW YORK UNIVERSITY