ParaMT: a Paraphraser for Machine Translation

Anabela Barreiro
Faculdade de Letras da Universidade do Porto & CLUP-Linguateca
New York University
barreiro_anabela@hotmail.com

Abstract. In this paper we present ParaMT, a bilingual/multilingual paraphraser to be applied in machine translation. We select paraphrases of support verb constructions and use the NooJ linguistic environment to formalize and generate translation equivalences through the use of dictionary and local grammars with syntactic and semantic content. Our research shows that linguistic paraphrasal knowledge constitutes a key element in conversion of source language into controlled language text that presents more successful translation results.

Keywords: ParaMT; paraphraser; paraphrase; support verb construction; multiword expression; machine translation, controlled language, NooJ, Lexicon Grammar Theory

1 Introduction

The benefits of paraphrasal knowledge to *Natural Language Processing* have been quantified in areas such as summarization [1]-[3], question-answering [4]-[5], information extraction [6], and machine translation [7]-[8], among others. Recent ACL workshops dedicated exclusively to paraphrasing reveal the growth in this field of knowledge. However, most published works describe statistics-based approaches to gather paraphrases. Statistical methods to acquire paraphrases are based on word co-occurrences and word combinations and have little or no linguistic knowledge. They also apply algorithms to corpora that may be inadequate or insufficient.

In this paper, we claim that effective results from linguistically based research on paraphrases can save substantial effort and resources employed by statistically based machine translation systems, by providing the opportunity to improve linguistic precision as a means to drive machine translation, rather than statistics. We argue that science in general is founded on direct analytical observation and believe that good quality machine translation relies on intimate language knowledge, not on probabilistic calculations. We have taken one important linguistic phenomenon, a particular type of phrase, a support verb construction, and built a body of lexical, syntactic and semantic knowledge around this phrasal type. Then, we applied this knowledge to a bilingual/multilingual paraphraser, which we intend to integrate in machine translation systems. Our hypothesis was that linguistic knowledge applied to a machine translation system would improve its output quality. We verified that support verb constructions is an area where statistics tend to "trap" systems. If statistical systems are not sensitive to these constructions, the consequence may be misleading translations. We argue that our linguistic system provides a statistical system with special training data that could correct this problem.

2 Support Verb Constructions

Support verb constructions are predicate noun constructions (noun + arguments) where the main verb has a weak semantic value, such as in *make a promise*. Semantically weak verbs are called support verbs in Lexicon-Grammar theory, but they are also known as light verbs [9]-[10]. Support verb constructions are multiword expressions that, within the area of corpus linguistics, have been subcategorized as collocation phenomena [11]-[14]. The term "*collocation*" is generally used to define words or terms that 'go together' with a precise meaning. Most works on collocations consist mainly in identifying collocations within a corpus, with the goal of including them in extended dictionaries. Even though it is widely known and used, collocation is a 'sort of' statistical related term (co-location means positioning side by side or close together) that is too broad for linguistic analysis. We look at collocations as multi-layered linguistic phenomena which, in our opinion, must be identified and studied individually, as proposed in [15]. We consider that the more we know about multiword expressions, the more sophisticated their descriptions are in the electronic dictionaries or the more accurately they are formalized in computer grammars, the better the quality of machine translation output and of natural languages applications in general.

Identifying source language multiword expressions such as support verb constructions is not a trivial task, but it is the starting point for paraphrasal knowledge, as it is for translation. As early as 1988, as demonstrated inter alia by [16]-[18], the suggestion of conceptually separating monolingual paraphrasing from translation in machine translation has been put forward by the insertion of a "style transfer" module which selects the "best or chosen translation" from multiple "possible" translations. The idea of dynamically invoking monolingual grammars to perform translation of multiword expressions was raised by developers on the working prototype built by the IBM-INESC Scientific Group back in the late eighties [ibidem]. Our approach uses monolingual grammars for the identification of support verb constructions and bilingual/multilingual grammars for translation and bilingual/multilingual paraphrasing. We can use paraphrasing in a monolingual text as a pre-editing procedure for controlled language writing and generate and translate paraphrases allowing their insertion directly in machine translation. We will load the paraphraser with Portuguese to English data and use the NooJ linguistic environment [19] to formalize and translate support verb constructions through finite-state transducers (local grammars) for bilingual/multilingual purposes. The theoretical framework behind this study is the Lexicon-Grammar [20]-[21], which stands on the principles of the transformational grammar of Harris, [22]-[23]. According to the Lexicon-Grammar, simple sentences (predicate and its arguments), also known as elementary sentences, and not the individual words, represent basic syntactic-semantic units. Natural language processing systems, particularly machine translation systems that take into account these linguistic units yield more opportunities for success.

3 Machine Translation Problem Evidence

Our experience with machine translation confirms that currently the results are far from perfect [24]. Translation results extracted from METRA [25], and described in [26] prove that machine translation engines are unsuccessful particularly at handling the translation of support verb constructions. A literal and unnatural translation is provided by most machines. For example, the English support verb construction *make a decision* is translated into Portuguese as *fazer uma decisão* instead of *tomar uma decisão* or even as the strong verb *decidir*, which represent its optimal paraphrase. This inaccuracy means that the English support verb *make* is directly translated into the Portuguese support verb *fazer* (default translation), instead of being recognized as part of the support verb construction which embeds semantic meaning as a whole.

We have tried to replace some support verb constructions with lexical verbs and verified that overall machine translation engines showed significantly better results. For example, machine translation engines are unanimous in choosing the Portuguese verb *decidir* as the correct translation for the English verb *decide*. This pre-editing, or more precisely controlled language writing by paraphrasing, improves translation results and makes output sentences more comprehensible overall. This proves that, if we consider pre-editing of the input sentences where support verb constructions occur, changing each instance into a lexical verb, we are not changing the meaning of the source sentence and we are giving the machine translation engine a distinctly better chance of improving the output result, by filtering out some noise, i.e., the weak verb. The support verb construction *make a decision* is a stylistic alternative to the verb *decide*, where neither the support verb *make* nor the determiner *a* add any meaning to the expression. In fact, in support verb constructions, the support verb is often void of meaning. Trying to translate them brings additional difficulties to machine translation systems, which is unnecessary until/unless they become more sophisticated. Our idea is to have several possibilities and not limit the system to only one possibility, as long as the system translates with precision. However, we believe that it is pointless to challenge one limited system with structures that we know *a priori* this system cannot translate well. For equivalent paraphrasing, the support verb must be recognized as part of a support verb construction which must be considered as a single semantic unit. The default assumption of all machine translation systems which cannot discern whether a word, in this case a support verb, adds semantic meaning to a phrase, is to assign equal semantic value to each word individually, unless, otherwise instructed. The system fails by incorrectly assigning semantic value to a support verb, resulting in a loss in equivalence of the output sentence. This is the problem of direct translation.

In sum, empirical evidence shows that application of linguistic knowledge to proper handling of support verb constructions by machine translation systems or NLP applications is effective. We believe that our methodology leads to attainable paraphrasing translation solutions. This paper demonstrates that we can create an instrument of some utility to the research community. We chose support verb constructions because they have been extensively studied from both theoretical and practical perspectives, in several different languages, by many authors, over a considerable period. They are fairly systematic, and therefore quite suitable for formalization and integration with machine approaches. Support verb constructions are abundant in language and their formalization is generally essential for machine translation. Lastly, they often represent paraphrases. For example, the English support verb construction [*pay a visit to NP*] is a phrasal alternative to the transitive construction [*visit NP*]. Both expressions have equivalent meaning and can be translated in the same way into Portuguese, [*visitar NP*].

4 ParaMT Resources and Methodology

In any language processing application, the linguistic resources represent the foundation. High-quality linguistic descriptions lead to sophisticated resources that help improve systems. In machine translation especially, the linguistic resources are the driving force that boosts the translation process. Our paraphrasing system is based on *Port4NooJ*, the open source NooJ Portuguese linguistic module, which integrates a bilingual extension for Portuguese-English machine translation. Port4NooJ is developed on two original sources: NooJ linguistic environment and OpenLogos lexical resources. The module and the linguistic resources are described thoroughly in [27] and available online in [28] and [29]. The elements that we want to emphasize here are the ones directly concerned with processing of support verb constructions. Accordingly, each dictionary entry includes, beyond the commonly used part-of-speech and inflectional paradigm, a description of the syntactic and semantic attributes (*SymSem*), as well as the associated distributional and transformational properties, such as predicate arguments, information about which determiners and prepositions occur with predicate nouns in “less variable” expressions, and derivational descriptions. Derivation is a very important issue, because it has implications not only at the lexical level, but also at the syntactic level. Derivational suffixes often apply to words of one syntactic category and change them into words of another syntactic category, while semantically they maintain their integrity. For example, the affix *–ção* changes the verb *adaptar* (*to adapt*) into the noun *adaptação* (*adaptation*) and the affix *-mente* changes the adjective *rápido* (*quick*) into the adverb *rapidamente* (*quickly*). This is extremely important for support verb constructions because it permits the establishment of equivalence grammars that map (i) support verb constructions such as *fazer uma adaptação (de)* (*to make an adaptation (of)*) to the verb *adaptar* (*to adapt*), where the predicate noun *adaptação* (*adaptation*) has a semantic and morpho-syntactic relationship with the verb *adaptar* (*to adapt*) or (ii) support verb constructions such as *ter um final rápido* (*to have a quick ending*) to the verbal expression *terminar rapidamente* (*to end quickly*), where the autonomous predicate noun *final* (*ending*) has a semantic relationship with the verb *terminar* (*to end*), and the adverb *rapidamente* (*quickly*) has a semantic and morpho-syntactic relationship with the adjective *rápido* (*quick*). Thus, our verb entries contain the identification of derivational paradigms for nominalizations (annotation *NDRV*) and a link to the derived noun's support verbs (annotation *NVSUP*), as in Fig. 1 below. Nominalizations are followed by their inflectional paradigm properties. Any other lexical constraints, such as prepositions, determiners, specific arguments, etc., will be added. Autonomous predicate nouns (non-nominalizations), such as *favor* (*favor*) are lexicalized and classified with the annotation *Npred* and have associated with them support verb and other lexical constraints, such as a preposition (*NPrep*), and a lexical verb (*VRB*) with the same semantics. We have also classified predicated adjectives and established the link between them and the corresponding verbs (*ADRV*), such as between the verb *adoçar* (*to sweeten*) and the adjective *doce* (*sweet*). We have started the assignment of corresponding support verbs (copula verbs) to these adjectives.

adaptar,V+FLX=FALAR+Aux=1+INOP57+Subset132+EN=adapt+*VSUP=fazer* +*DRV*=NDRV00:CANÇÃO +*NPrep*=de
favor,N+FLX=MAR+Npred+AB+state+EN=favor+*VSUP*=fazer+*NPrep*=a+*VRB*=ajudar
rápido,A+FLX=RÁPIDO+PV+eagerType+EN=quick+*DRV*=AVDRV06:RAPIDAMENTE
adoçar,V+FLX=COMEÇAR+Aux=1+OBJTRundif75+Subset604+EN=sweeten+*DRV*=ADRV11:VERDE+*VSUP*=tornar

Fig. 1. Sample of the dictionary

According to these linguistic constraints, we have created relationship properties at the dictionary level and then apply those properties in local grammars in order to recognize support verb constructions in corpora and generate them for applications such as controlled language writing and machine translation. In section 5, we describe how we use these resources to recognize and generate paraphrases automatically.

Our strategy to formalize idiomatic expressions and distinguish them from expressions with a more complex syntactic behavior is to lexicalize them. Therefore, semi-frozen support verb constructions, where the support verb is the only variable word in the whole expression, are lexicalized in the dictionary of multiword expressions. For example, in *dar a mão à palmatória* (*to acknowledge being wrong*) or *fazer vista grossa* (*to ignore*), the support verbs *dar* (*to give*) and *fazer* (*to make*) are assigned an inflectional paradigm and the rest of the words in the expression remain invariable. As our electronic dictionaries provide enhanced meaning of single words, including contextual significance and increasingly more valuable tagging data, we also intend to enlarge and refine the role of a bilingual dictionary to include entries for multiword expressions that consider the understanding and analysis of each type of multiword expression, by beginning with support verb constructions and their paraphrases. The ability to give the machine translation user multilingual paraphrasing ability constitutes an important step towards achieving better quality machine translation.

5 Paraphrases for Machine Translation

As we have mentioned above, in order to obtain monolingual paraphrases or to translate support verb constructions from Portuguese to English using NooJ, we combine the properties formalized in the Portuguese dictionary with local grammars. Local grammars are ways of formalizing language constructs using input and output symbols, i.e., they are language descriptions in the form of graphs containing an input entry (with linguistic information) and an output entry (with linguistic constraints to the output, or simply the binary information of the recognized or not recognized sequence). In NooJ, these local grammars are represented by finite-state transducers, and are widely applied to texts/corpora, for identification and analysis of local linguistic phenomena of a natural language, extraction of named entities from texts, recognition and tagging of words, or multiword expressions, identification of syntactic constituents such as noun phrases and completives, extraction of semantic relations, and disambiguation. Among these possible applications, we extended local grammars to recognize, paraphrase and translate support verb constructions, creating ParaMT, a bilingual/multilingual automatic paraphraser. In order to establish relations of equivalent morpho-syntactic predicates in the same language (Portuguese) or between two languages (particularly between Portuguese and English), we use the dictionary properties. Since we have classified all predicate nouns in the dictionary as [*NPred*], we can now use this lexical information in a syntactic grammar to identify the predicate in a support verb construction and apply this grammar in corpora. Fig. 2 represents a simple local grammar used to recognize and generate support verb constructions and transform them into their verbal paraphrases.

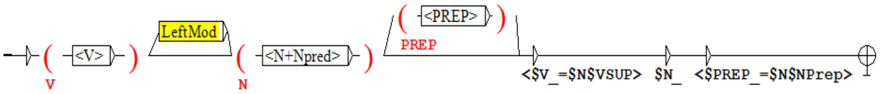
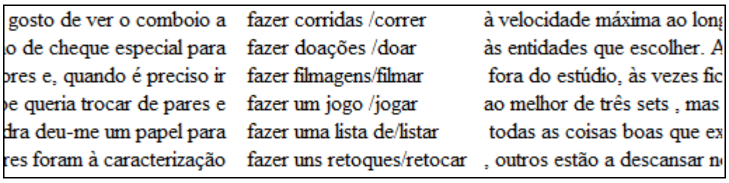


Fig. 2. Grammar to recognize and paraphrase support verb constructions

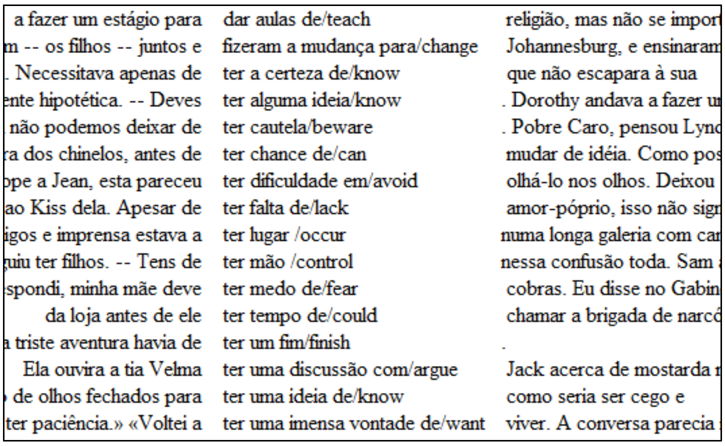
This grammar matches verbs, which are marked in the dictionary as support verbs that are followed by a left modifier (determiner, adjective or adverb or other quantifiers), a predicate noun and optionally a preposition. The elements in parentheses ( ) are stored in variables V, N or PREP. If a dictionary entry has a lexical constraint, such as NPrep=*a* in the phrase [*dar um grande abraço a*] (*to give a big hug to*], the support verb construction will be recognized by the grammar and mapped to the verb *abraçar* (*to hug*), the lemma of the noun specified in the variable $N_. The elements in bold <$V_=$N$VSUP>, and $PREP_=$N$NPrep> represent lexical constraints that are displayed in the output, such as specification of the support verb or the preposition that belongs to a specific support verb construction. The predicate noun is identified, mapped to its deriver and displayed as a verb, the other elements of the phrase are eliminated. Fig. 3 shows a concordance where Portuguese support verb constructions are recognized and paraphrased as lexical strong verbs.



gosto de ver o comboio a	fazer corridas /correr	à velocidade máxima ao lon
o de cheque especial para	fazer doações /doar	às entidades que escolher. A
ores e, quando é preciso ir	fazer filmagens/filmar	fora do estúdio, às vezes fic
e queria trocar de pares e	fazer um jogo /jogar	ao melhor de três sets , mas
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

Fig. 3. Recognition and monolingual paraphrasing of support verb constructions
(Support verb construction / corresponding verb)

ParaMT makes possible the recognition of Portuguese support verb constructions in a text and their automatic conversion into an English verb (bilingual paraphrasing), as in Fig. 4.



a fazer um estágio para	dar aulas de/teach	religião, mas não se impor
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 u
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 ca
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
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

Fig. 4. Recognition and bilingual paraphrasing of support verb constructions (Portuguese support verb construction / corresponding English verb)

6 Preliminary Quantitative Evaluation

Currently, our bilingual Portuguese-English general dictionary comprises about 60,000 entries distributed by 30,000 nouns, 11,000 verbs, 2,800 adjectives, 4,700 adverbs, and 11,500 other part of speech entries. The dictionary of proper names comprehends about 6,000 entries. Our multiword expression dictionary comprehends about 40,000 entries, 20,000 nominal; 10,000 verbal; 5,000 adjectival and 5,000 adverbial multiword expressions. We have over 8,000 derivational links between verbs and nominalizations and about 1,000 derivational links between verbs and predicate adjectives. A few general multiword expression grammars cover over 5,000 expressions of several other types. We have not yet evaluated the coverage of the multiword expressions dictionary and grammars in corpora, but we have some preliminary results for the evaluation of support verb constructions. In order to obtain these results, we selected from COMPARA [30], a parallel corpus of English-Portuguese fiction, all sentences where the infinitive form of the Portuguese verbs *fazer* (*to do*), *dar* (*to give*), *pôr* (*to put*), *tomar* (*to take*) and *ter* (*to have*) occurred with a noun or with a left modifier and a noun. First, we manually classified these combinations as to whether they corresponded to support verb constructions or not. We confirmed that these verbs occur very frequently in a support verb construction. 89% of the occurrences of *dar*, 88% of *tomar*, 77% of *pôr*, 47% of *fazer* and 20% of *ter* were in a support verb construction. This means that globally in 64.2% of the times, these verbs are used as support verbs, that corresponds to nearly 2/3 of the occurrences.

Subsequently we selected randomly a sub-corpus with 500 sentences (100 for each selected verb), containing instances of only support verb constructions. We classified them manually and compared these results with the results obtained automatically. We tried to have constraining recognition rules so that paraphrasing would be more precise. Currently, we can recognize 62.6% of the support verb constructions with high scores in precision. Furthermore, we not only recognize the support verb constructions, as we also paraphrase them with high degree of success. Fig. 5 shows the results of the support verb construction recognition (precision and recall) and the results (precision) of our automatic paraphraser.

	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%
Average	62.6/63.6 - **98.4%**	62.6/100 - **62.6%**	57/61 - **93.4%**

Fig. 5. Evaluation of simultaneous recognition and paraphrasing of support verb constructions

7 Conclusions

In this paper we have tried to answer the question of whether paraphrase information can improve machine translation output and how the analysis and formalization of paraphrases can contribute to the larger task of machine translation. We have addressed linguistic analysis and computational formalization of bilingual short paraphrases for support verb constructions using NooJ linguistic environment We have demonstrated the scope of the phenomenon as a basis for a machine translation multiword expression dictionary, which can be used both in machine translation development or machine translation evaluation and in the extension of the scope in current dictionary functionality.

The discovery process has provided results in two areas. First, it has led to the creation of a primitive multiword expression electronic dictionary that addresses monolingual Portuguese and bilingual Portuguese-English paraphrases of equivalent meaning between support verb constructions and their noun counterparts. Second, it has helped to further specify the definition of multiword expressions. The interface between user and software that is presented is not finished yet, but once the sub-task is well-understood this interface can be simplified and, hopefully, will be usable and as easily integrated into the larger task of machine translation, as the single word electronic dictionary that has already been integrated.

Our work based on support verb constructions illustrates what can be done with ParaMT for any kind of multiword expression. The method is repeatable. Furthermore, the tool is extensible to cover larger and more complex linguistic phenomena, including sentence level paraphrases that can be used for controlled language writing or translation. While this research is intended to find a place in ideal machine translation, it can be used as an electronic multiword dictionary. From a monolingual point of view, it is useful to simplify pre-translated source text (rendering the text less complex, less flowery, etc.). Converting support/weak verbs into lexical strong verbs helps to simplify and reduce the number of words in a text which has a positive impact on translation cost, in circumstances where word count or "white space" is sensitive. From a bilingual point of view, it helps reduce ambiguity and verbosity. It can be used as an on-line linguistic aid for translators so they can determine the best translation (evaluation purposes), and for automated machine translation evaluation. This knowledge is useful to machine translation development, because it permits deeper understanding of source text, and it provides a successful methodology to analyze paraphrasing, given that paraphrasal intelligence is crucial in both machine translation development and machine translation evaluation.

Acknowledgements

We would like to thank Max Silbeztein and Slim Mesfar for providing support on technical aspects of NooJ, and Sérgio Matos and Diana Santos for helpful comments on this paper.

This work was partly supported by grant SFRH/BD/14076/2003 from *Fundação para a Ciência e a Tecnologia* (Portugal), co-financed by POSI.

References

1. Barzilay, R. and McKeown, K, 2001. Extracting Paraphrases from a Parallel Corpus. In Proceedings of the ACL/EACL, 50-57, Toulouse, 2001.
2. Barzilay, R. 2003. Information Fusion for Multidocument Summarization. Ph.D. Thesis, Columbia University.
3. Hirao, T., J. Suzuki, H. Isozaki, and E. Maeda. 2004. Dependency-based Sentence Alignment for Multiple Document Summarization. In Proceedings of the COLING, pages 446–452.
4. Ibrahim, A., B. Katz, and J. Lin. 2003. Extracting structural paraphrases from aligned monolingual corpora. In Proceedings of the Second International Workshop on Paraphrasing (ACL 2003). On Text Summarization Branches Out, pages 10–17.
5. Duboué, P. A., J. Chu-Carroll. 2006. Answering the question you wish they had asked: The impact of paraphrasing for Question Answering. HLT-NAACL 2006.
6. Shinyama, Y. and S. Sekine. 2003. Paraphrase Acquisition for Information Extraction. The Second International Workshop on Paraphrasing: Paraphrase Acquisition and Applications (IWP2003) 2003; Sapporo, Japan.
7. Callison-Burch, C., P. Koehn and M. Osborne, 2006. Improved Statistical Machine Translation Using Paraphrases. In Proceedings NAACL-2006.
8. Callison-Burch, C. 2007. Paraphrasing and Translation. PhD Thesis, University of Edinburgh.
9. Kearns, K. 2002. Light verbs in English. manuscript.
10. Butt, M. 2003. The light verb jungle. Harvard Working Papers in Linguistics 9:1–49.
11. Quirk, R, S. Greenbaum, G. Leech and J. Svartvik. 1985. A comprehensive grammar of the English language. London: Longman.

12. Biber, D. 1988. Variation Across Speech and Writing. Cambridge, England: Cambridge University Press. pp.3-27.
13. Crystal, D. 1991. *A dictionary of linguistics and phonetics*, 3rd edition. Oxford: Blackwell Publishers.
14. Sinclair, J. 1991. Corpus Concordances Collocations. Oxford: Oup.
15. Meyers, A., R. Reeves, C. Macleod. 2004. "NP-External Arguments: A Study of Argument Sharing in English". In Proceedings of the ACL 2004 Workshop on Multiword Expressions: Integrating Processing, pp. 96-103, Barcelona, Spain, July 26, 2004.
16. Santos, D. 1988. "A fase de transferência de um sistema de tradução automática do inglês para o português", Tese de Mestrado, IST, UTL, Outubro de 1988.
17. Santos, D. 1990. "Lexical gaps and idioms in Machine Translation", Hans Karlgren (ed.), Proceedings of COLING'90. Helsinki, August 1990, Vol. 2, pp.330-5.
18. Santos, D. 1992. "Broad-coverage machine translation", INESC Journal of Research and Development, Vol. 3, No. 1, 1992, pp. 43-59. Reprinted in K. Jensen, G. Heidorn and S. Richardson, Natural Language Processing: The PLNLP Approach, Kluwer Academic Press, 1993, pp. 101-118.
19. Silberztein, M. 2004. "NooJ: A Cooperative, Object-Oriented Architecture for NLP". In INTEX pour la Linguistique et le traitement automatique des langues. Cahiers de la MSH Ledoux, Presses Universitaires de Franche-Comté.
20. Gross, M. 1975. Méthodes en syntaxe. Hermann.
21. Gross, M. 1981. "Les bases empiriques de la notion de prédicat sémantique". In A. Guillet and C. Leclère (eds). Formes Syntaxiques et Prédicat Sémantiques, Langages, 63: 7-52. Larousse, Paris.
22. Harris, Z. 1957. "Co-occurrence and transformation in linguistic structure". Language, 33, 293-340.
23. Harris, Z. 1968. Mathematical Structures of Language, New York: Wiley, 230p.
24. Barreiro, A. and E. Ranchhod. 2005 "Machine Translation Challenges for Portuguese". In Linguisticæ Investigationes 28.1 Amsterdam/Philadelphia: John Benjamins Publishing Company. ISSN: 0378-4169.
25. Sarmento, L. 2007. "Ferramentas para experimentação, recolha e avaliação de exemplos de tradução automática". In Diana Santos (ed.), Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa. Lisboa, Portugal: IST Press, 2007, pp. 193-203.
26. Barreiro, A. 2008 (forthcoming). Formalization of Support Verb Constructions and their Paraphrases: Applications in Machine Translation (provisory title). PhD dissertation.
27. Barreiro, A. 2008 (forthcoming). "Port4NooJ: Portuguese Linguistic Module and Bilingual Resources for Machine Translation". In Xavier Blanco and Max Silberztein (eds). Proceedings of the 2007 International NooJ Conference. Univ. Autonoma de Barcelona, June 7-9, 2007. Cambridge Scholars Publishing.
28. NooJ, http://www.nooj4nlp.net/
29. Linguateca, http://www.linguateca.pt/Repositorio/Port4Nooj/
30. Frankenberg-Garcia, A. and Diana Santos, 2003. "Introducing COMPARA, the Portuguese-English parallel translation corpus". In Federico Zanettin, Silvia Bernardini & Dominic Stewart (eds.), *Corpora in Translation Education*. Manchester: St. Jerome Publishing, 2003, pp. 71-87. http://www.linguateca.pt/COMPARA/