ReEscreve: a translator-friendly multi-purpose paraphrasing software tool

Anabela Barreiro
Linguateca - FCCN
Rua Visconde de Santarém, nº 12, R/C Dtº
1000-287 Lisboa, Portugal
barreiro_anabela@hotmail.com

Luís Miguel Cabral
Linguateca - SINTEF ICT
P.O. Box 124 BlindernN-0314
Oslo, Norway
luismiguel.cabral@sintef.no

Abstract

This paper describes ReEscreve, a multi-purpose paraphraser that uses grammar-based paraphrasing capabilities suitable for pre-editing and useful for human and machine translation. At the current stage, ReEscreve transforms, with a 93.4% precision, support verb constructions into verbs or similar expressions, but it is being used to progressively paraphrase other linguistic phenomena enabling it to be used as an authoring and stylistic aid in word processing applications. ReEscreve is freely available on the Internet.

1 Introduction

For over a decade, translators have been using translation memories and other primitive linguistic tools to support their translation tasks. The basic idea of translation memory tools is to recycle text that has already been translated, either full sentences, or parts of text (fuzzy matches), and reuse them to expedite the translation process. However, such tools do not incorporate any level of linguistic knowledge. Attempts to build bigger and better parallel corpora and more refined terminological databases have also helped the translator in recent years. Nevertheless, in the current global market, the increasing need for translation and the demand to be simultaneously quick and rigorous requires more robust and sophisticated tools. Now is the time when new avenues ought to be explored to produce more refined products. In future ventures, linguistic knowledge and linguistic quality control should be a requirement for language technology tools and paraphrasers will play an important role in the next generation of natural language processing tools, including translation aids.

This paper describes ReEscreve (ReWriter), a multi-purpose paraphraser that helps users to change, simplify or clarify their texts. While it was conceived in the long run to be a useful authoring aid for word processing in general, we present ReEscreve for the specific purpose of helping or simplifying machine translation by using it as a pre-editor. Here we describe its interface, functionalities, and application to controlled writing. However, ReEscreve's applicability goes beyond stylistic and authoring aid and it is already being applied to a prototype machine translation system proposed in (Barreiro, 2008b). The paraphrases generated by ReEscreve also help reduce ambiguity and the number of words, and eliminate unnecessary complexity from the source text, making it a useful tool for translators.

2 Linguistic phenomena currently paraphrased and why

At the current stage of development, ReEscreve recognizes Portuguese support verb constructions, and generates paraphrases of them. A support verb construction is a category of multiword expression. Support verb constructions, also known as complex predicates, consist of a semantically weak verb (the support verb), and a nominal, which is the predicate of the sentence, such as *fazer planos* (*to make plans*) or *ser aplicável a* (*to be applicable to*). Efforts are being made to integrate paraphrases of other types of multiword expression and larger and more complex linguistic phenomena, such as free syntactic constructions. Up to now, within support verb constructions, only predicate noun constructions have been formalized systematically. Predicate nouns can be morphologically related to a verb or be autonomous. For example, the Portuguese and the English predicate nouns *apresentação* (*presentation*) and *visit* in the correspondingly support verb constructions *fazer*

uma apresentação de N (*to make a presentation of N*) and *to pay a visit to* are morphologically related to the verbs *apresentar* (*to present*) and *to visit*, respectively. On the other hand, the Portuguese and the English predicate nouns *mestrado* (*master*) and *fun* are autonomous, because they are not morphologically related to any verb (there are no such verbs as **mestrar* in Portuguese or **to fun* in English).

There are several reasons why support verb constructions were chosen for the research that led to the development of ReEscreve. Firstly, support verb constructions are very frequent and indispensable to communicate in most languages, including in Portuguese and English. In addition to being abundant, support verbs have been extensively and systematically studied within the Lexicon-Grammar Theory by M. Gross (Gross, 1981) and followers, also in Portuguese by (Ranchhod, 1990), (Baptista, 2005), and (Chacoto, 2005), and in contrastive studies by (Salkoff, 1999). Often support verb constructions can be replaced by stylistic variants that are paraphrases of them; they present several degrees of variability or can be idiomatic; they are semantically weak and ambiguous; and they cannot be translated literally. Finally, they are difficult to formalize in machines, and constitute a difficulty for second language learners and human translators.

3   Machine translation problems with support verb constructions

A brief analysis of the results for freely available machine translation systems on the Internet demonstrates that the translation of multiword expressions is currently a problem area for any system, whether statistical or rule-based. The ambiguity of these expressions can lead to loss of meaning, mistakes, unclear or unsuitable translations, when there is no linguistic knowledge associated with them, particularly syntactic-semantic knowledge. Translation results obtained from the publicly available meta-translator called METRA (www.linguateca.pt/METRA) developed by Linguateca prove that machine translation engines are unsuccessful at handling the translation of support verb constructions. A literal and unnatural translation is provided by some or most machines, as illustrated in Fig. 1. When contrasting METRA results for Portuguese-English translation

of a support verb construction sentence, as in Fig. 1 with the results of a sentence with a verb with equivalent meaning, as in Fig. 2, we promptly verify the superior quality of the results in the second case.
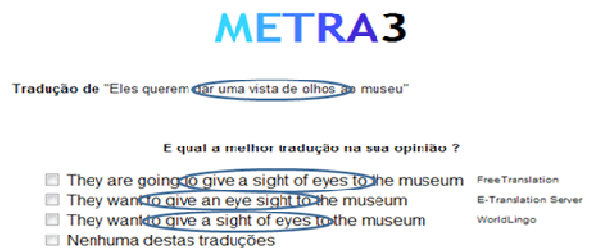

Fig. 1: METRA search results for machine translation of the Portuguese support verb constructions *dar uma vista de olhos* into English.
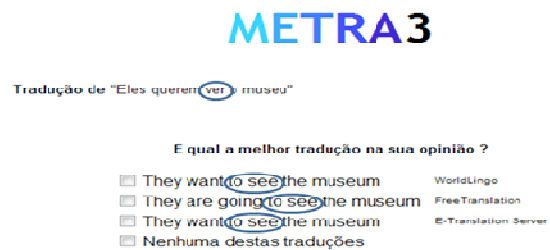

Fig. 2: METRA search results for machine translation of the Portuguese verb *ver* into English.

The inferior quality of the results of the translation in Fig. 1, shows that Portuguese idiomatic support verb construction *dar uma vista de olhos* is being translated literally into English. This inaccuracy means that in default translations, the support verbs are not recognized as part of the support verb construction which embeds semantic meaning as a whole and need to be appropriately analyzed and formalized, so that they can be properly translated. Currently, machine translation is not addressing the appropriateness or suitability issue of the translation effectively. An experiment that measured the impact of paraphrases of support verb constructions in machine translation was performed to confirm this empirical evidence. In this test, 5 different support verbs were selected, 100 sentences with support verb constructions (20 sentences for each support verb) for English and Portuguese. Then, a paraphrase for each of the support verb constructions was assigned manually by a linguist – resulting in a total of 200 sentences (100 paraphrases). The test was evaluated by 10 testers with approximately similar judgment

capabilities and good linguistic knowledge and translation skills. They were asked to translate all the sentences using METRA, choose the best translation results for both the support verb construction and the paraphrase, and then decide which one represented a better translation. Both English-Portuguese and Portuguese-English testers considered that the paraphrase translations (with verbal predicates) were of a higher quality. Table 1 shows that when support verb constructions were identified and replaced with semantically equivalent or similar verbal expressions as a pre-processing step to translating, a 31% improvement was observed in the evaluated quality of the results of English-Portuguese machine translation, and a 21% improvement in the results of Portuguese-English machine translation.

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

Table 1: Impact of paraphrases of support verb constructions in machine translation.

The evaluation results clearly confirmed that paraphrases have helped the systems produce better machine translation. In general, machine translation engines unanimously show significantly better results for the translation of verbs than for the translation of support verb constructions.

4 Importance of linguistic knowledge of paraphrases

This paper maintains that linguistic knowledge of paraphrases can help improve the quality of not only machine translation, but it can also be useful for several other purposes and applications. The conversion of support verb constructions into morphosyntactically and/or semantically related verbs produces a controlled language that is applicable both to general and technical/domain specific contexts and makes translation more reliable. This transformation reduces avoidable complexity and lexical ambiguity; it facilitates interpretation when the support verb construction is idiomatic; it reduces wordiness, often improving source text quality; and subsequently it improves translation quality. In a few cases, paraphrasing consists in replacing the semantically weak support

verb of the predicate nominal construction with lexical-syntactic and stylistic variants, such as *realizar uma análise* or *efectuar uma análise* (*to perform an analysis*), which are more frequent in technical language than the elementary support verb construction *fazer uma análise* (*to do an analysis*).

ReEscreve uses Port4NooJ lexical-ontological resources as its engine. Port4NooJ was built by (Barreiro, 2008b) using two types of source: NooJ (http://www.nooj4nlp.net/pages/nooj.html), a linguistic environment developed by (Silberztein, 2005; 2008), and OpenLogos (http://logos-os.dfki.de/) English-Portuguese dictionary, a derivative of the Logos machine translation system envisaged by (Scott, 2003). Linguateca corpora resources and tools, such as COMPARA (http://www.linguateca.pt/COMPARA/compara_en.html) and METRA were also used to support this development. In addition, Port4NooJ represented an effort to answer the demand for Portuguese public large-scale resources that could be freely used by the research community. The development of enhancing dictionaries is the foundation to build linguistically sophisticated authoring aid software. The paraphrase resource integrated in Port4NooJ contains over 10,000 semantic relations established in the dictionary. Most of these relations are between verbs and predicate nouns and are accountable for the existing paraphrases.

5 Methodology

There were two main phases in the methodology adopted. Phase 1 included two main tasks: the development of resources and pre-processing. All linguistic information was formalized in NooJ dictionaries and grammars. This linguistic information was explored for several natural language processing tasks, from both a monolingual and a bilingual perspective. This phase included the construction of simple grammars to recognize and extract support verb constructions from texts; definition of properties in the dictionary and establishment of semantic links between verbs and support verb constructions and establishment of several syntactic-semantic subclasses of paraphrases; application of grammars with dictionary information to texts; refinement of the grammars; and better recognition of support verb constructions in texts, which resulted in

paraphrasing and also translation. Phase 2 was dedicated to evaluating the experiments.

Regarding the formalization in the dictionary, derivation and semantic verb associations were explicitly marked. In the adjective entries, the derivational paradigms for adverbializations were identified. For autonomous predicate noun entries, the annotation Npred was inserted and their semantically related verbs were identified. Links were also established between Npreds and the support verbs they link with. Derivation and support verb associations were explicitly marked. In the verb entries, the derivational paradigms for the nominalizations and predicate adjectives were identified and verb entries were linked to the support verbs that occur with their derived nouns and adjectives.

6   ReEscreve functionalities

ReEscreve is an application that suggests alternative ways of expressing a given text, based on large dictionaries and a set of local grammars, developed in the NooJ environment. Currently it is limited to paraphrasing mainly support verb constructions in their most straightforward and frequent appearances: that is, when the verb precedes the object and is at most modified by intervening adverbs, although other grammars (and therefore paraphrasing capabilities) are being developed. The user can see and choose from the paraphrasing suggestions, or opt not to paraphrase, and can even propose an alternative paraphrasing solution. User interaction is logged so that the linguist can identify problems and improve the underlying resources. Even though ReEscreve is a Web application (see Fig. 3), in principle anyone with NooJ could also, independently, use the grammars and lexicons in their own environment.

7   Interaction with ReEscreve

The ReEscreve Web interface, incorporating Web technologies like Ajax to NooJ, is available at: http://www.linguateca.pt/Reescreve/.

Recognition and conversion of support verb constructions into semantically equivalent verbs or similar verbal expressions by ReEscreve follows three main steps, as represented in Fig. 3. First, ReEscreve recognizes a support verb construction in a text and extracts it. Then the support verb

construction expression is matched against the paraphrase database (developed in NooJ) and the corresponding candidates are retrieved with the inflectional features of the support verb in the original expression. Finally, the suggestions are presented to the user.
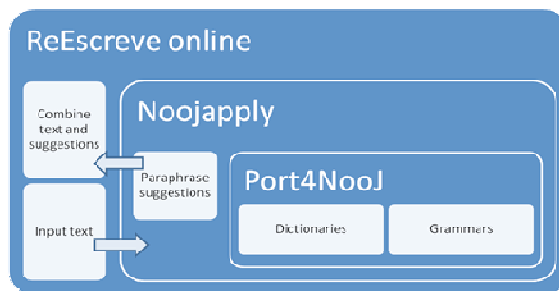

Fig. 3: Structure of ReEscreve

For example, if the recognized form of the support verb is the third person plural of "Perfeito" simple past, such as *deram um passeio* (*they went for a walk*), the suggested paraphrase, with a corresponding verb, will also be inflected with the same tense, number, and person features, *passearam* (*they walked*).

The ReEscreve homepage presents a brief description of its functionalities. A window box allows the user to insert the text to be rewritten, while it is also possible to submit a file, to date, only in plain text format. The user can try the service with an example text by clicking on the word **"exemplo"**, which displays an example text with several cases already covered by the tool. For this or any user text, paraphrasing results appear after clicking on **"Obter resultados"** (Obtain results). Fig. 4 shows the result of a rewritten text, after editing by the user: The expressions in blue are new words introduced into the text. These words were chosen to replace the original expressions provided by the user.

Fig. 4: Text rewritten after interactive use of ReEscreve

At the end of the text in Fig. 4, a box, shows the alternatives still available. The right-hand side menu shows the original expression for which paraphrasing suggestions were provided (for user approval). By clicking on any of these expressions, the user navigates to the corresponding sentence and a box opens underneath the sentence that contains the pair original | suggestion. The box then shows all the different possible paraphrases for this sentence, as illustrated in Fig. 5.
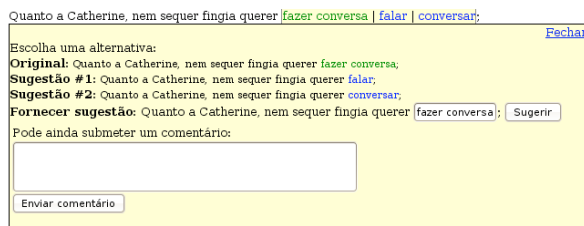

Fig. 5: Paraphrases box with suggestion paraphrase

The user can select the preferred paraphrase or contribute to the enlargement of the paraphrase database by suggesting new expressions, i.e., the user can insert one or more equivalent expressions that are suitable in the same context and do not yet exist in the system. Obviously, although this process is transparent for the user in his/her text, the integration of suggested paraphrases in the system is monitored, that means, no paraphrase is added permanently to the database without the intervention of the Port4NooJ responsible.

Through logging the suggestions that the user selects (either suggestions made by the system or suggestions made by the user), the ReEscreve system helps us build a set of user data that can later be used to improve the current resources (grammars and dictionaries). These data will also be used to allow ReEscreve to automatically rewrite text for purposes of providing a front-end to machine translation systems. If we attain a significant number of serious users by maintaining the system available on the Internet, accumulation of usage data can "grow" this resource naturally, as well as improve machine translation meta-searchers or even machine translation systems.

8 Evaluation of recognition and paraphrasing of support verb constructions using ReEscreve

In order to evaluate the automated recognition and paraphrasing of support verb constructions by ReEscreve, a test was performed on 5 different support verbs (*dar* (*to give*), *tomar* (*to take*), *pôr* (*to put*), *fazer* (*to make*/*do*), and *ter* (*to have*)). This test was run against a total of 500 baseline support verb construction sentences, which were manually annotated (100 sentences for each support verb).

	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%**

Table 2: Evaluation of recognition and paraphrasing of support verb constructions with ReEscreve

Globally, precision rates of 98.4% in recognition and a 93.4% in paraphrasing were achieved. Concerning recall, a 62.6% rate was obtained. Currently, ReEscreve can recognize and paraphrase support verb constructions with a high degree of success, as illustrated in Table 2. The low recall is due to the fact Port4NooJ dictionary is not complete and the grammar was deliberately restricted to achieve greater precision. We expect that the improvement of the resources that is taking place at the moment will considerably improve recall.

9 Related work and application to machine translation

The usefulness of paraphrasing tools that can properly handle support verb constructions and other multiword expressions is broad. Beyond machine translation, paraphrases are suitable for controlled writing and text pre-editing, text production and stylistics. Related work on controlled language, stylistic editing and authoring aids has been done, for which some tools are available. Existing tools include the MULTIDOC (Haller, 2000), KANT CE Checker (Mitamura and Nyberg, 2001); (Rascu, 2006), CLAT (Schmidt-Wigger, 1998); (Hernandez and Rascu, 2004), and

CLOUT (the Controlled Language Optimized for Machine Translation), among others.

Controlled language is a common practice in English style and writing aid manuals (many of them available on the Internet). It is also recommended by machine translation developers to their clients, because it is easier for a machine to translate. Controlled language makes use of paraphrases. This pre-editing, or controlled language writing by paraphrasing, improves translation results and makes output sentences more comprehensible overall. This proves that, if we include pre-editing of the input sentences where support verb constructions occur, changing each instance into a verb (whenever that is semantically suitable), we are not changing the core 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 or some other less meaningful elements. For example, the support verb construction *make a decision* is a stylistic alternative to the verb *to decide*, in which neither the support verb *to make* nor the determiner *a* adds any meaning to the expression. Trying to translate support verb constructions results in additional difficulties for machine translation systems. Our paraphrasing system allows several possibilities. However, paraphrasing by simplification proved to be the most suitable for publicly available machine translation systems. The ideal framework is that sophisticated machine translation systems are not limited to one possibility, and translate both support verb constructions and their paraphrases accurately, knowing that they are related. However, we believe that it is pointless to challenge one limited machine translation 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 paraphrasing capability of the output sentence.

A simple NooJ grammar, as illustrated in Fig. 4, can be suitable to recognize elementary support verb constructions and convert them into verbs, as the concordance in Fig. 5 shows.
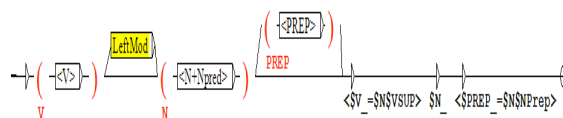

Fig. 5: NooJ grammar to recognize and paraphrase support verb constructions into verbs

Fig. 6: Concordance resulting from the application of the grammar represented in Fig. 5.

The paraphrasing represented in Fig. 6 is useful for controlled writing and text pre-editing. A similar type of paraphrasing is also useful in text stylistics with a double function: style change and controlled language. For example, it can be used for recognition and paraphrasing support verb constructions co-occurring with predicate nouns from the biomedical field, such as *fazer uma operação* (ambiguous between *to operate* and *to undergo surgery/an operation*. These support verb constructions can be paraphrased into stylistic variants, such as *efectuar* ou *realizar* (*to perform*), as the concordance in Fig. 7 shows. They could also be paraphrased by constructions such as "*submeter-se a*" or "*ser submetido a*" (*uma operação*) – *to undergo surgery* or *to be operated*, whenever the subject of the support verb construction is a patient (identified in context or by keyword). This helps eliminate the ambiguity created by the support verb and allows the construction to be translated correctly.

Fig. 7: Stylistic variants of elementary ambiguous support verb constructions, such as "*fazer uma amputação*" (*to perform or to be amputated*)

In this particular case, information about the argument structure helps resolve the ambiguity.

However, if not paraphrased and disambiguated, the elementary support verb construction may still be translated incorrectly.

Another application of paraphrases is for translation, as used by the bilingual/multilingual paraphraser ParaMT, described in (Barreiro, 2008a). ParaMT works in a similar way to ReEscreve, but it can be used directly in machine translation or translation in general. The concordance in Fig. 8 illustrates Portuguese-English bilingual paraphrasing (= translation) of support verb constructions into verbs suggested by ParaMT.

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
nte hipotética. -- Deves	ter alguma ideia/know	. Dorothy andava a fazer ur
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
pe 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 sigr
igos e imprensa estava a	ter lugar /occur	numa longa galeria com car
quiu ter filhos. -- Tens de	ter mão /control	nessa confusão toda. Sam a
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 :

Fig. 8: Recognition of Portuguese support verb constructions and translation into English verbs

To sum up, empirical evidence shows that application of linguistic knowledge to proper handling of support verb constructions by machine translation systems or other natural language processing applications is effective. We believe that our methodology leads to attainable paraphrasing translation solutions. This work demonstrates that we can create an instrument of some utility to the research community, which has good applicability in machine translation, and can be ported into translation memories and several other areas.

10   Conclusion

Machine translation results are far from perfect. It appears that linguistically well studied paraphrases increase the quality of machine translation results. A paraphrase relevance assessment test run to appraise the relevance of paraphrases of support verb constructions to machine translation quality improvement, confirmed our hypothesis that machine translation results for support verb constructions are poor. It then showed that paraphrases of support verb constructions help improve machine translation results, and confirmed that a linguistically based paraphrasing tool helps control source language so as to facilitate the translation of texts submitted to machine translation. The main conclusion derived from our work is that the body of linguistic (lexical, syntactic and semantic) knowledge formalized around support verb constructions can be applied to paraphrasing and when applied to a machine translation system improves its output quality.

The same linguistic resources that helped improve machine translation proved to be important to other natural language processing tasks and were successful employed by software paraphrasing tools, such as ReEscreve and ParaMT, for both monolingual and bilingual purposes. These new tools can help the translator discover alternative ways of looking at the source text and provide more fluent translations.

11   Future work

The resources created during this research are improving and can be added to existing similar tools and resources in order to create better natural language processing products.

Currently, the quality of the resources and applications is being improved with the enlargement of syntactic-semantic relations between predicates, and other morphosyntactically and semantically related elements, such as *apresentar – apresentação – apresentado* (*to present – presentation – presented*) or *literal – literalmente* (*literal – literally*); the paraphrasing of various linguistic phenomena for ReEscreve extensibility (comprising stylistic variance and controlled language). The development and enhancement of ParaMT multilingual paraphraser/translator can be based on the same principles as ReEscreve. In the near future, we would like to integrate our paraphrasing capabilities with translation memories, and enhance the new machine translation model presented in (Barreiro, 2008b).

One example of such a future improvement is the extensibility of ReEscreve using new paraphrasing phenomena. We are repeating the

procedure applied on support verb constructions to obtain additional categories of paraphrase, so that coverage of paraphrases increases. Paraphrasing facilitates corrections that are standard practice in (human) revision. The methodology adopted for paraphrasing support verb constructions is extensible and ReEscreve is useful for controlled writing by employing paraphrases of adverbials, relative clauses, if clauses, named entities, nouns phrases, among other linguistic phenomena, as illustrated in the sample of Table 3. Some promising outcome resulting from application of some of these phenomena are awaiting evaluation. We believe that a formal linguistic study of paraphrases such as those referred to represent a significant contribution to natural language processing in general, and to (machine) translation in particular.

Linguistic Phenomenon	Expression or sentence	Paraphrase
Adverbial	à volta da órbita *around the orbit of the eye*	Periorbital *periorbital*
	de forma interactiva *in an interactive way*	interactivamente *interactively*
Relative Clause	N0 que foram escritos *N0 that were written*	N0 escritos *N0 written*
	A velocidade a que se move a luz *T *The speed to which light moves*	A velocidade da luz *The speed of light*
	O papel que a Europa tem *The role that Europe plays/has*	O papel da Europa *The role of Europe/Europe's role*
	As dificuldades que temos *The difficulties we have*	As nossas dificuldades *Our difficulties*
If clause	se for necessário *if it is necessary*	se necessário *if necessary*
Named Entity	A rainha de Inglaterra *The queen of England*	A rainha inglesa *The British queen*
Noun Phrase	O heróico povo português *The heroic Portuguese people*	Os heróicos portugueses *The heroic Portuguese*

Table 3: Paraphrasing for ReEscreve extensibility

Acknowledgments

References

Baptista, J. *Sintaxe dos Nomes Predicativos com verbo-suporte SER DE*. Lisboa: Fundação para a Ciência e a Tecnologia/Fundação Calouste Gulbenkian. 2005.

Barreiro, A. 2008b. Make it simple with paraphrases: Automated paraphrasing for authoring aids and machine translation. PhD dissertation. FLUP. Porto, Portugal. December 2008.

Barreiro, A. 2008a. "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.

Chacoto, L. O Verbo Fazer em Construções Nominais Predicativas. PhD dissertation. Universidade do Algarve, Portugal. 2005.

Gross, M. "Les bases empiriques de la notion de prédicat sémantique". *Formes Syntaxiques et Prédicat Sémantiques, Langages* **63** (1981), pp. 7-52. Paris, França: Larousse.

Haller, J. 2000, 'MULTIDOC. Authoring Aids for Multilingual Technical Documentation'. In: Proceedings of the 1st Congress of Specialized Translation. Barcelona, pp. 143-147.

Hernandez, M. & E. Rascu. 2004. Checking and Correcting Technical Documents. In Séverine Vienney and Mounira Bioud, editors, Correction automatique: bilane t perspectives, number 29 in BULAG, pp. 69-84. Presses universitaires de Franche-Comté.

Mitamura, T. & E. Nyberg. 2001. Automatic Rewriting for Controlled Language Translation. In Proceedings of NLPRS–2001. Workshop on Automatic Paraphrasing: Theory and Application.

Ranchhod, E (ed.). *Sintaxe dos predicados nominais com Estar*. Lisboa, Portugal: INIC. 1990.

Rascu, E. (2006). A controlled language approach to text optimization in technical documentation. In *Proc. of KONVENS 2006*, pp. 107–114. Konstanz, Germany.

Salkoff, M. *A French-English Grammar: A Contrastive Grammar On Translational Principles*. Amsterdam, Philadelphia: John Benjamins Publishing Company. 1999.

Schmidt-Wigger, A. 1998, KURD in MULTILINT – A Grammar and Style Checker for German, in: Proceedings of CLAW 98.

Scott, B. "The Logos Model: An Historical Perspective". *Machine Translation* **18** (2003), pp. 1-72.

Silberztein, M. Complex Annotations with NooJ. In *Proceedings of the 2007 International NooJ Conference*. Cambridge Scholars Publishing: Newcastle, 2008.

Silberztein, M. NooJ's Dictionaries. In Proceedings of the 2nd Language and Technology Conference, Poznan: 2005.