Linguistic Evaluation of Support Verb Constructions by OpenLogos and Google Translate

A. Barreiro[1], J. Monti[2], B. Orliac[3], S. Preuß[4], K. Arrieta[3], W. Ling[1,5], F. Batista[1,6], I. Trancoso[1]

[1] INESC-ID, Rua Alves Redol 9, 1000-029, Lisboa, Portugal
[2] University of Sassari, Italy
[3] Logos Institute, USA
[4] University of Saarlandes, Germany
[5] Carnegie Mellon University - Instituto Superior Técnico, USA-Portugal
[6] ISCTE-IUL - Instituto Universitário de Lisboa, Portugal
anabela.barreiro@inesc-id.pt, johannamonti@hotmail.it

Abstract

This paper presents a systematic human evaluation of translations of English support verb constructions produced by a rule-based machine translation (RBMT) system (OpenLogos) and a statistical machine translation (SMT) system (Google Translate) for five languages: French, German, Italian, Portuguese and Spanish. We classify support verb constructions by means of their syntactic structure and semantic behavior and present a qualitative analysis of their translation errors. The study aims to verify how machine translation (MT) systems translate fine-grained linguistic phenomena, and how well-equipped they are to produce high-quality translation. Another goal of the linguistically motivated quality analysis of SVC raw output is to reinforce the need for better system hybridization, which leverages the strengths of RBMT to the benefit of SMT, especially in improving the translation of multiword units. Taking multiword units into account, we propose an effective method to achieve MT hybridization based on the integration of semantico-syntactic knowledge into SMT.

Keywords: Machine Translation, MT Evaluation, Support Verb Constructions, Multiword Units, Semantico-Syntactic Knowledge, MT Hybridization.

1. Introduction

MT researchers and developers have a common goal of creating robust translation systems that can produce high quality translation. However, MT is a work-in-progress. After six decades of research on MT models, tools and linguistic resources, translations produced by widely used MT systems show unfortunate errors which require significant post-editing effort if the result is to be used by professional translators or for purposes other than gisting. A noticeable trend is that of linguistically enhancing SMT models, to produce systems that combine linguistic resources and analysis with statistical techniques. While this is a promising direction, making significant advances towards hybrid MT (HMT) systems requires a deep understanding of different approaches, their weaknesses and strengths. Bringing different approaches together, contrasting them and measuring which modules need improvement seems to be an effective method for achieving the desired end result.

In pursuing the creation of an HMT model and improving existing translation technology, a systematic fine-grained linguistic analysis of the performance of individual models appears to us as an important first step. To our knowledge, no major effort has been made to combine the strengths of different approaches with the purpose of overcoming known weaknesses on the basis of a joint linguistic evaluation of those weaknesses, as used in this paper. In addition, state-of-the-art quality metrics and estimation have been targeting human-factors tasks such as measuring post-editing time and effort in terms of keystrokes, etc., not diagnosing fine-grained linguistic errors to improve syntactic structure and meaning. The lack of such qualitative evaluation efforts involving MT systems of different nature was our main motivation for evaluating the performance of RBMT and SMT when dealing with a very specific linguistic phenomenon.

This paper describes an evaluation exercise that consisted of the linguistic analysis of the translations of 100 support verb constructions (SVC), such as *make a presentation*. The sentences in our SVC corpus were randomly collected from the news and the Internet and hand-picked for the task. The corpus was translated into 5 languages: French, German, Italian, Portuguese and Spanish by the OpenLogos (OL) and the Google Translate (GT) MT systems. Five MT expert linguists, native speakers of the target languages, evaluated the translations of the SVC considering their meaning in context, and classified the translation errors based on a SVC typology, as illustrated in Table 1.

The remaining of the paper is organized as follows: Section 2. presents the state-of-the-art HMT. Section 3. describes the main characteristics of the OL and the GT models. Section 4. describes the corpus, datasets, and the evaluation task. Section 5. discusses the linguistic challenges found in the translations for the 5 language pairs contemplated in this research. Section 7. proposes a method to achieve MT hybridization based on the integration of semantico-syntactic knowledge into SMT. Finally, Section 8. presents the main conclusions and points to future work, namely the inclusion of linguistic expertise in MT evaluation.

2. Hybrid Machine Translation

A noticeable trend in current MT research is that of HMT models that combine linguistic knowledge with statistical

Nominal Support Verb Construction (NSVC)
make a presentation
Non-contiguous nominal (NON-CONT NSVC)
have [ADV+ADJ-particularly good] links
Prepositional nominal (PREPNSVC)
give an illustration of
Non-contiguous prepositional nominal (NON-CONT PREPNSVC)
be the [ADJ-immediate] cause of
Idiomatic nominal (IDIOM NSVC)
set in motion, place at risk, go on strike
Idiomatic prepositional nominal (IDIOM PREPNSVC)
earn an income of
Non-contiguous idiomatic nominal (NON-CONT IDIOM NSVC)
hold [NP-the option] in place, be of [ADJ-practical] value
Non-contiguous idiomatic prepositional nominal (NON-CONT IDIOM PREPNSVC)
give [PRO-us] a [bird's-eye] view of, be [ADV-clearly] at odds with, open talks [May 14] with
Adjectival Support Verb Construction (ADJSVC)
be meaningful
Non-contiguous adjectival (NON-CONT ADJSVC)
be [ADV-extremely] selective
Prepositional adjectival (PREPADJSVC)
be known as; be involved in
Non-contiguous prepositional adjectival (NON-CONT PREPADJSVC)
fall [ADV-so far] short of

Table 1: Major categories of support verb construction in our corpus

techniques. HMT systems attempt to combine RBMT systems, such as the work presented by (Scott, 2003), with data-driven MT systems, such as phrase-based SMT proposed by (Koehn et al., 2007). System combination often leads to improvements in the final translation quality, as different systems tend to address different translation challenges.

In general, SMT methods learn the generalizations of the translation process using parallel corpora, which are sets of translated sentences pairs. In language pairs where there are abundant amounts of parallel corpora, such as English-Mandarin, these models tend to perform better than RBMT approaches. However, when parallel data is scarce, such as for the Spanish-Basque language pair, it is less likely that the SMT methods will have enough data to properly learn such generalizations (Labaka et al., 2007). Morphologically rich languages require more data to learn accurate translations, since there are more word types due to different word forms that can be generated for the same stem. While more complex SMT models for morphologically rich languages have been proposed (Chahuneau et al., 2013), RBMT systems where the morphology of the language is encoded manually, represent a strong alternative for the translation of such languages.

Many methods to combine RBMT with SMT have been proposed. One straightforward way to create an HMT system is to combine the translations of the same text by two different systems (Heafield and Lavie, 2011; Eisele et al., 2008). Other approaches attempt to use data-driven techniques to improve RBMT systems. For example, (Eisele et al., 2008) use phrase pair extraction in phrase-based MT to extract phrasal translations entries that are used to improve the coverage of a RBMT system. A similar method using example-based MT for the same end has been proposed in (Sánchez-Martínez et al., 2009). Finally, the translation quality of RBMT output can also be improved by using statistical post-editing methods (cf. (Simard et al., 2007; Elming, 2006; Dugast et al., 2007; Terumasa, 2007)). The opposite has also been proposed, where RBMT systems are used to enhance data-driven approaches. (Shirai et al., 1997) use an example-based MT system (Brown, 1996) to create an initial translation template, and use a RBMT system to translate individual words and phrases according to this template. To the best of our knowledge, it is still not obvious which HMT approach will be the most efficient one and will lead to higher quality translation in the long run.

3. The OpenLogos and the Google Translate Models

OL is an open source copy of the commercial Logos System (Scott, 2003; Barreiro et al., 2011). The system addresses morphology, syntax, and semantics, it has robust parsers, sets of semantico-syntactic rules, terminology sets and tools. Unlike most other RBMT systems, the OL model is closer in spirit to the SMT approach in that both methodologies are pattern-based, with the additional advantage of including semantic understanding. The use of an intermediate language (SAL) to encode linguistic information and process text contributes to OL high quality translation and lessens one of the main problems in SMT, viz., the sparseness in linguistic examples. OL linguistic knowledge databases have not been developed for over 10 years.

GT is one of the most widely used online MT systems. It is an SMT system that benefits from the huge amount of parallel data that Google collects from the web. In March 2014, it was set to account for 80 language pairs. As is typical of SMT, the translation quality is highly dependent on the language pair, producing much better results for close language pairs (e.g. Portuguese and Spanish) and languages for which large amounts of parallel data are available. GT is a closed system, however, no knowledge of semantic understanding is known to exist in GT.

4. Corpus, datasets, and evaluation task

The corpus used in this research task contains 100 English SVC that appear in sentences collected from the news and the Internet. A SVC is a multiword or complex predicate consisting of a semantically weak verb (the support verb), and a predicate noun (most commonly), a predicate adjective, or a predicate adverb (Barreiro, 2009). For example, *make a presentation* is a SVC made of the support verb *make* and the predicate noun *presentation*, and *make it simple* is a SVC made of the support verb *make* and the predicate adjective *simple*. We have selected SVC as the object of our evaluation for two main reasons. Firstly, SVC have been studied sistematically and in depth within the Lexicon-Grammar Theory (Gross, 1987) and have been recognized and processed computationally in general and specific-purpose corpora for several languages. The scientific study of SVC eliminates subjectivity concerns for the evaluation task and allows evalution to be done by only one evaluator per language. Secondly, SVC occur abundantly in texts and most MT systems still fail at addressing the compositional aspect of multiword units. SVC When translated incorrectly, SVC have a negative impact in the understandability and quality of translations (Barreiro et al., 2013). Like other multiword units (e.g. phrasal verbs, such as *set up*, prepositional predicates, such as *look after*, SVC can be non-contiguous, i.e., the individual elements that compose the unit are placed apart in the sentence, with a smaller or greater number of inserts. We considered an insert to be any word in between elements of the multiword unit other than a definite or an indefinite article before a predicate noun. Non-contiguous SVC are extremely difficult to align in SMT, remaining one of the key cross-language challenges for MT.

In our research, each SVC under evaluation was annotated in the context of its sentence and classified according to the taxonomy presented in Table 1, prior to the translation process so to ease the translation evaluation task. The English SVC-annotated corpus was then translated into French, German, Italian, Portuguese and Spanish, using the OL and the GT systems. Native linguists, who are also MT experts, evaluated the translation quality of the SVC for each target language (one evaluator per language pair), and classified the errors found for each of the target languages, according to a binary evaluation metrics: *OK* for correct translations and *ERR* for incorrect ones. In addition to ERR, which was used for semantically incorrect translations of the SVC or serious syntactic problems within the SVC translation, evaluators have also classified errors of *Agreement* (*Agreem* in Table 2) and *Other* to distinguish morphologically-related or other problems, such as incorrect prepositions or wrong word order affecting directly the SVC under evaluation. The linguists evaluating the translation quality were also instructed to provide a more comprehensive qualitative evaluation of mistranslations according to the different types of SVC (Section 6.). None of the systems was specifically trained for the task, as the texts were not domain specific.

Lang. pair	System	OK	ERR	Agreem	Other
EN-FR	GT	64	32	4	-
	OL	51	48	1	-
EN-GE	GT	37	46	3	14
	OL	60	33	1	6
EN-IT	GT	61	31	-	8
	OL	43	52	-	5
EN-PT	GT	68	27	5	-
	OL	41	58	1	-
EN-ES	GT	51	41	6	2
	OL	25	70	3	2

Table 2: Performance of the OL and GT MT systems for 100 SVC.

5. Quantitative Results

A systematic linguistic evaluation of the translations of the SVC provided by the GT and OL MT systems showed that, for the corpus used, GT translated more SVC correctly than OL due to its richer lexical knowledge. However, OL structural analysis is a powerful feature that could turn OL performance for the Romance languages as satisfactory as that of the German language given the fact that the structural analysis is the same for all language pairs. OL use of linguistic knowledge in its structural analysis ability, such as the SEMTAB semantico-syntactic knowledge representation and its ability to translate different surface structure of a sentence, combined with GT rich word selection in the transfer powered by sophisticated SMT methods to extract knowledge from vast amounts of parallel empirical data can be a powerful combination that would certainly result in better translation quality. Table 2 presents the quantitative results for the translation of the 100 SVC in our corpus for French, German, Italian, Portuguese, and Spanish with the OL and the GT MT systems.

6. Linguistic Evaluation

In this section, we will discuss the linguistic challenges found in the translations for the 5 language pairs contemplated in our research, individually and with detailed examples. We believe that such qualitative evaluation is valuable to the research community.

6.1. English-French

Of the 100 SVC in the English test corpus, 64 were translated correctly by GT and 51 by OL. The majority of the translation errors (64% for GT, 69% for OL) involved incorrect lexical choice for some or all of the elements of the SVC, resulting in non-idiomatic expressions. Examples of literal translations of SVC include *the economist is equally correct in generalizing* translated as *l'économiste est tout aussi correct de généraliser* (GT) and *we are taking* *a growing interest in* translated as *nous prenons un intérêt croissant pour* (OL). A number of interesting errors by GT involved idiomatic translations which were not accurate in the context (e.g. *value judgments [...] come into the picture* translated as *les jugements de valeur [...] entrent en scène*) or were only partly generated (e.g. *cause-and-effect relationships are typically not self-evident* translated as *les relations de cause à effet ne sont généralement pas de soi*). Errors in source sentence analysis (36% for GT, 31% for OL) accounted for the remainder of the SVC errors. For GT, these errors showed mostly as lack of subject-verb agreement (e.g. *The mass market banks then put in place punitive measures* translated as *Les banques du marché de masse alors mettre en place des mesures punitives*) whereas the analysis of errors in OL was more varied and the result of missing dictionary entries (e.g. *It's a lot more helpful to them* translated as *Il est un lot plus utile à eux*) or incorrect noun phrase analysis, especially when involving a numerical expression (e.g. *It also posted a $1.3 billion gain* translated as *Il a également posté un $1.3 milliard le gain*).

6.2. English-German

The analysis of the 100 translations of SVC in the test corpus provided the following picture: OL has a higher number of correct SVC translations (60) than GT (46). The incorrect translations in both German corpora were categorized with respect to the error type: (i) lexical (L) for incorrect word choice, including prepositions (e.g. *has particularly good links with* was translated as *hat besonders gute Verbindungen mit* instead of *zu*, (ii) order (O) for incorrect word order, including incorrect clause segmentation (e.g. *Is it directly usable or does the user have to do additional data manipulation before one can make use of it?* where the German non-finite *tun* was incorrectly scrambled into the subordinate clause *Ist es direkt verwendbar oder hat der Benutzer zusätzliche Datenmanipulation bevor man tun kann, davon Gebrauch machen?*, (iii) morphology (M) for incorrect word form, including the choice between bare-infinitive and to-infinitive, (iv) ellipsis (E) for missing word, mainly auxiliary verb and main verb (e.g. *I hope to be worthy of it* translated as *ich hoffe, würdig*, instead of *ich hoffe, dessen würdig zu sein*). When applicable, incorrect SVC translations were categorized with more than one error type. For the 40 incorrect OL translations, there were 33 lexical errors, and 7 word order errors. For the 54 incorrect GT translations, there were 37 lexical errors, 3 agreement errors, and 14 other errors. In general, GT has more structural (word order and morphology) and missing words errors than OL. Surprisingly, the performance of GT is also poor with regards to lexical coverage of contiguous SVC. Finally, GT does not translate well the German verb split, even though there seems to be a reordering component.

6.3. English-Italian

Of the 100 English SVC, 61 were translated correctly by GT and 43 by OL. Of the 39 incorrect translations by GT, 22 (56%) are related to wrong lexical choices, which range from untranslated SVC elements such as in *Economists have had a spotty record in* translated as *Gli economisti hanno avuto un record spotty nel*, literal translation of the SVC elements, such as in *Canada was falling so far short of* translated as *Canada stava cadendo finora breve di*, to wrong selections of the meaning of the SVC elements, such as in *These specifications gave insight into space* translated as *Queste specifiche hanno spaccato lo spazio*. The second major category of mistranslations corresponds to 8 wrong agreements (e.g. *Il processo dell'economista [...] è fondato sui; quando due insiemi di dati sono direttamente correlate*). Similarly to GT, OL mistranslations are mainly due to incorrect lexical selection (70%), such as in *This is another way of poking fun at* translated as *questo è un altro modo di conficcare il divertimento in*, and wrong agreement (10%). Of the 100 SVC, 51 were non-contiguous, 23 being translated incorrectly by GT and 36 by OL. In GT, mistranslations of this typology of SVC are mainly due to wrong lexical choices (15 instances) and affect both prepositional nominal SVC (7 instances) and prepositional adjectival ones (6). OL shows translation problems mainly with non-contiguous idiomatic SVC (7), non-contiguous prepositional nominal SVC (11 instances), and finally, non-contiguous prepositional adjectival SVC (8). In OL, mistranslation due to wrong lexical choices are more frequent in comparison with other error typologies. Of the 20 prepositional adjectival SVC construction, OL and GT performed differently, i.e. OL translated 8 constructions incorrectly, whereas GT translated 6 constructions incorrectly.

6.4. English-Portuguese

Of the 100 English SVC, 68 were translated correctly by GT and 40 by OL. Of the 32 incorrect translations by GT, 5 were related to agreement between the subject and the predicate adjective of the SVC (e.g. *To be meaningful, facts must be* was translated as *Para ser significativa, os fatos devem ser* instead of *Para serem significativos, os fatos devem ser*), and between the subject and the verb of the SVC (e.g., *protests will have no effect on* was translated as *os protestos não terá nenhum efeito sobre* instead of *os protestos não terão nenhum efeito sobre*). The SVC translations obtained from OL did not present any agreement errors. Of the 100 SVC, 51 were non-contiguous. The most idiomatic SVC were translated correctly by both systems in 3 cases (*to be in a position*; *to set in motion* and *to take on duties*), and incorrectly in 5 cases (*to come to a rest*; *to open talks*, *to put in place*, *to fall short of*, *to have a spotty record*). In the remaining cases, GT performed better than OL. GT missed *to hold in place*, *to be in charge of*, *to be on guard*. OL missed *to come into the picture*, *to place at risk*, *to put under the microscope*, *to be on strike*, *to be at odds with*, *to earn an income*, and *to give a bird's-eye view*. In the case of less idiomatic SVC, both systems presented problems with (i) prepositions (e.g. *was responsible for* was translated as *foi responsável para*, instead of *foi responsável por*); (ii) literal translation of the support verb (e.g. *makes it possible* was translated as *faz possível* instead of *torna possível*); and (iii) wrong lexical choice for the predicate noun (e.g. *gave insight into* was translated as *deu uma visão para*, intead of *deram um esclarecimento sobre*). In general, SVC problems by GT were more structural, while SVC problems by OL were more lexical.

6.5. English-Spanish

Of the 100 examples of SVC, there are 41 examples in which GT and OL exhibited different behavior. Of these 41 cases, GT got 33 correct translations and 8 incorrect translations of SVC, and OL the reverse distribution. Both systems encountered the same types of problem: either the system recognizes the SVC or it does not. Beyond that they both presented problems related to translating prepositions correctly or assigning the correct choice of a determiner. One interesting example is *To be meaningful, facts must be*, which was translated as *Para que tenga sentido* (GT) and *Para ser de mucho sentido* (OL), where the translation provided by GT is correct and that provided by OL is incorrect. OL would easily translate the frozen SVC correctly, provided it added it to its dictionary. Another interesting example is the SVC *We have a lot of worldwide experience in* translated as *Tenemos mucha experiencia en* by GT, and *Tenemos mucha experiencia mundial* by OL. In this case, OL is able to resolve the SVC internal modifier adjective *worldwide* and, therefore, able to generate the intended meaning. GT drops the adjective, removing meaning from the source in the translation. Finally, the SVC *a desire to give priority to* was translated as *dar prioridad al desarrollo* by GT, and *dar la prioridad al desarrollo* by OL. While both translations are correct, the example illustrates the fact that, for a good amount of SVC, both MT systems can benefit from *free rides*, where only minor corrections of determiners would be needed.

7. Proposal for Semantico-Syntactic Knowledge Integration into SMT

The OL method for semantico-syntactic integration of multiwords, including SVC, is one of the most interesting approaches to multiword processing in MT. In OL, semantico-syntactic information is encoded in the lexicon, both in the dictionary entries and in the rules. Any linguistic element in the OL system is represented in an abstraction language with ontological properties called SAL[1]. SAL is an hierarchical taxonomy made up of supersets, sets and subsets and represents the heart of OL, accounting for its relative effectiveness in parsing and semantic understanding.

Upon entering the OL system, all natural language input sentences are converted into SAL patterns, which represent not only the semantico-syntactic features of each word, but also its morphology. SAL elements interact with semantico-syntactic rules, or tables as they are known in OL, called SEMTAB. SEMTAB rules represent the meaning of words on the basis of their association with other words (context); disambiguate the meanings of words in the source text by identifying the syntactic structures underlying each meaning; and provide the target language equivalents of each identified meaning of a source language. SEMTAB rules are conceptual and encode deep structure relations that are not sensitive to differences at the surface syntactic level (e.g. *they raised their salaries* and *salaries were raised*) or morphological level (*the raising of the salary* and *salary raising policies*). SEMTAB rules are called after dictionary look-up and during the execution of target transfer rules (TRAN rules) to solve ambiguity problems, such as verb dependencies and multiwords of different nature, overriding the default dictionary transfer. When a sentence is being parsed by TRAN, OL sends the SAL patterns to the SEMTAB database to look for a rule match. If the rule exists for a linguistic string, TRAN uses that rule and overrides the dictionary transfer for that string. In the case of a SVC, such as *to apply paint to*, a SEMTAB rule can maintain the SVC structure or paraphrase it, transforming it into the verb *to paint*. When TRAN finds a SEMTAB rule that transforms and translates the SVC into another language, such as Portuguese, it can choose to override the dictionary entries for the string [*to* + *apply* + *paint* + *to*] into the SVC *aplicar tinta a* or into the verb *pintar*. The SEMTAB rule allows for the different surface structures of the SVC and any insert specified in the rule, such as *aplicaram tinta vermelha a* (*they applied red paint to*) or *iria eventualmente pintar* (*he/she would eventually paint*). As long as the SEMTAB rule exists in the semantico-syntactic database, OL can process and translate correctly the SVC in our corpus, which were incorrectly translated in the OL and the GT systems. Therefore, the OL method can overcome the structural problems presented by SMT, not only the contiguous, but also the non-contiguous SVC, independently of how near or how remotely they occur in the sentence. The OL methodology applies to any other type of multiword unit and allows the translation of other context-sensitive challenges.

8. Conclusions and Future Work

Two main conclusions can be extracted from our evaluation. Our first conclusion refers to the importance to develop systematic linguistic quality evaluation metrics with a phased error categorization task where specific linguistic phenomena can be evaluated individually in stages by MT expert linguists. Independently of the approach, there is still significant work to be done as far as evaluation is concerned. For example, each subtype of multiword unit needs to be tested individually as we have done for support verb constructions. Fine-grained error categorization can contribute to more controlled and systematic evaluation tasks. In addition, for the task of error categorization to be successfully accomplished, grammar specific evaluation corpora need to be developed and used to evaluate each group of linguistic error and identify which system has more difficulties translating each particular type of linguistic challenge. Future comparative evaluation tasks require the construction of corpora to test grammatical correctness addressing individual linguistic phenomena, such as corpora of different types of multiword units. In this case, the translation of each particular type of multiword unit would be evaluated autonomously. Similarly, grammar targeted evaluation would be done by using corpora of relative constructions, passives, pronouns, determiners, locative prepositions, and so on and so forth. No effective hybridization can take place before linguistic evaluation of the results provided by different approaches is successfully accomplished. Therefore, we believe that the question "how effectively can rule-based and statistical MT be combined?" can

[1] freely available at `https://www.l2f.inesc-id.pt/abarreiro/openlogos-tutorial/newA2menu.htm`

only be answered after linguistic quality evaluation metrics are developed and validated by the MT community.

Acknowledgements

Anabela Barreiro was funded by FCT Fundação para a Ciência e Tecnologia (post-doctoral grant SFRH / BPD / 91446 / 2012). This research was also supported by FCT Fundação para a Ciência e Tecnologia, under project PEst-OE/EEI/LA0021/2013.

Autorship contribution is as follows: Anabela Barreiro is author of the Abstract and Sections 1., 5., 6., 6.4., 7. and 8.; Johanna Monti of Sections 4. and 6.3.; Brigitte Orliac of Section 6.1.; Susanne Preuß of Section 6.2.; Kutz Arrieta of Section 6.5.; Wang Ling of Section 2.; and Fernando Batista of Section 3..

9. References

Anabela Barreiro, Bernard Scott, Walter Kasper, and Bernd Kiefer. 2011. Openlogos rule-based machine translation: Philosophy, model, resources and customization. *Machine Translation*, 25(2):107–126.

Anabela Barreiro, Johanna Monti, Brigitte Orliac, and Fernando Batista. 2013. When multiwords go bad in machine translation. In *Proceedings of the Workshop on Multi-word Units in Machine Translation and Translation Technology, Machine Translation Summit XIV*.

Anabela Barreiro. 2009. *Make it Simple with Paraphrases: Automated Paraphrasing for Authoring Aids and Machine Translation*. Ph.D. thesis, Universidade do Porto, Porto, Portugal.

Ralf D. Brown. 1996. Example-based machine translation in the pangloss system. In *COLING*, volume 1, pages 169–174.

Victor Chahuneau, Noah A. Smith, and Chris Dyer. 2013. Knowledge-rich morphological priors for bayesian language models. In *HLT-NAACL*, pages 1206–1215. The Association for Computational Linguistics.

Loïc Dugast, Jean Senellart, and Philipp Koehn. 2007. Statistical post-editing on systran's rule-based translation system. In *Proceedings of the Second Workshop on Statistical Machine Translation*, StatMT '07, pages 220–223, Stroudsburg, PA, USA. Association for Computational Linguistics.

Andreas Eisele, Christian Federmann, Hans Uszkoreit, Hervé Saint-Amand, Martin Kay, Michael Jellinghaus, Sabine Hunsicker, Teresa Herrmann, and Yu Chen. 2008. Hybrid machine translation architectures within and beyond the euromatrix project. In J. Hutchins and W.V. Hahn, editors, *Hybrid MT Methods in Practice: Their Use in Multilingual Extraction, Cross-Language Information Retrieval, Multilingual Summarization, and Applications in Hand-Held Devices.Proceedings of the European Machine Translation Conference*, pages 27–34. European Association for Machine Translation, HITEC e.V.

Jakob Elming. 2006. Transformation-based correction of rule-based mt. In *Proceedings of EAMT 2006*, pages 219–226.

Maurice Gross. 1987. The use of finite automata in the lexical representaion of natural language. In *Electronic Dictionaries and Automata in Computational Linguistics*, pages 34–50.

Kenneth Heafield and Alon Lavie. 2011. Cmu system combination in wmt 2011. In *Proceedings of the Sixth Workshop on Statistical Machine Translation*, WMT '11, pages 145–151, Stroudsburg, PA, USA. Association for Computational Linguistics.

Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondřej Bojar, Alexandra Constantin, and Evan Herbst. 2007. Moses: Open source toolkit for statistical machine translation. In *Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions*, ACL '07, pages 177–180, Stroudsburg, PA, USA. Association for Computational Linguistics.

Gorka Labaka, Nicolas Stroppa, Andy Way, and Kepa Sarasola. 2007. Comparing rule-based and data-driven approaches to spanish-to-basque machine translation. In *Machine Translation Summit XI*, Copenhagen, Denmark.

Felipe Sánchez-Martínez, Mikel L. Forcada, and Andy Way. 2009. Hybrid rule-based - example-based mt: feeding apertium with sub-sentential translation units. In *EBMT 2009 - 3rd Workshop on Example-Based Machine Translation*, Dublin, Ireland. DORAS.

Bernard (Bud) Scott. 2003. The logos model: An historical perspective. *Machine Translation*, 18(1):1–72.

Satoshi Shirai, Francis Bond, and Yamato Takahashi. 1997. A hybrid rule and example-based method for machine translation. In *Recent Advances in Example-Based Machine Translation*, pages 211–224. Kluwer Academic Publishers.

Michel Simard, Nicola Ueffing, Pierre Isabelle, and Roland Kuhn. 2007. Rule-based translation with statistical phrase-based post-editing. In *Proceedings of the Second Workshop on Statistical Machine Translation*, pages 203–206, Prague, Czech Republic, June. Association for Computational Linguistics.

Ehara Terumasa. 2007. Rule based machine translation combined with statistical post editor for japanese to english patent translation. In *Proceedings of the MT Summit XI Workshop on Patent Translation*, volume 11, pages 13–18.