# Using Common Sense to generate culturally contextualized Machine Translation

**Helena de Medeiros Caseli** **Bruno Akio Sugiyama** **Junia Coutinho Anacleto**
Department of Computer Science (DC)
Federal University of São Carlos (UFSCar)
Rod. Washington Luís, km 235 – CP 676
CEP 13565-905, São Carlos, SP, Brazil
{helenacaseli,bruno_sugiyama,junia}@dc.ufscar.br

## Abstract

This paper reports an ongoing work in applying Common Sense knowledge to Machine Translation aiming at generating more culturally contextualized translations. Common Sense can be defined as the knowledge shared by a group of people in a given time, space and culture; and this knowledge, here, is represented by a semantic network called ConceptNet. Machine Translation, in turn, is the automatic process of generating an equivalent translated version of a source sentence. In this work we intend to use the knowledge represented in two ConceptNets, one in Brazilian Portuguese and another in English, to fix/filter translations built automatically. So, this paper presents the initial ideas of our work, the steps taken so far as well as some opportunities for collaboration.

## 1 Introduction

In this paper we describe an ongoing work concerning the studies in gathering and using Common Sense knowledge and building Machine Translation applications. Common Sense (CS) can be defined as the knowledge shared by a group of people in a given time, space and culture.[1] Machine Translation (MT), in turn, is the application of computer programs to generate a translated equivalent version of a source text, in a target language.

MT is one of the oldest and most important areas of Natural Language Processing (NLP) / Computational Linguistics (CL).[2] From its beginnings we have witnessed some changes in the proposed MT paradigms ranging from the basic level —in which MT is performed by just replacing words in a source language by words in a target language— to more sophisticated ones —which rely on manually created translation rules (Rule-based Machine Translation) or automatically generated statistical models (Statistical Machine Translation, SMT). Nowadays, the majority of the researches has being centered around the phrase-based statistical MT (PB-SMT) approach —such as (Koehn et al., 2003) and (Och and Ney, 2004). PB-SMT is considered the state-of-the-art according to the automatic evaluation measures BLEU (Papineni et al., 2002) and NIST (Doddington, 2002)[3].

Although PB-SMT models have achieved the state-of-the-art translation quality, there are strong evidences that these models will not be able to go further without more linguistically motivated features, as stated by Tinsley and Way (2009). This is already being illustrated by the recent shift of researches towards linguistically enriched models as (Koehn and Hoang, 2007) and (Tinsley and Way, 2009) among others.

Following the same idea of these most recent researches, here we are also interested in seeing

[1] This definition of Common Sense is adopted by *Open Mind Common Sense* (OMCS) and *Brazilian Open Mind Common Sense* (OMCS-Br) projects and is only one of the several possible definitions.

[2] In this paper we will use the terms NLP and CL interchangeably since this is the assumption adopted in Brazil.

[3] BLEU and NIST are two automatic measures widely applied to evaluate the target MT output sentence regarding one our more reference sentences.



*Proceedings of the NAACL HLT 2010 Young Investigators Workshop on Computational Approaches to Languages of the Americas*, pages 24–31, Los Angeles, California, June 2010. 

how it is possible to improve MT performance based on more linguistically motivated features. In our case, we intend to investigate how to apply Common Sense knowledge to generate more culturally contextualized automatic translations.

For example, considering the translation of slangs[4] as in the English sentence "*Jump, you chicken!*"[5]. In this case, the word "*chicken*" do not mean "*a kind of bird*" but "*a coward*" or "*a person who is not brave*". However, its translation to Portuguese ("*galinha*") can also be applied as a slang with a completely different meaning. In the Portuguese language, the slang "*galinha*" means a men with a lot of girlfriends. Although the problem stated in the given example could also be fixed by some dictionary entries, CS knowledge is the kind of information that varies a lot and frequently can not be found in traditional dictionaries. Thus, we believe that the CS knowledge derived from the OMCS projects is an alternative way to cope with these translation problems.

Before presenting our ideas, section 2 describes some related work on SMT and more recent linguistically motivated empirical MT. Common sense and the Open Mind Common Sense project are the subjects of sections 3. Section 4 brings some of our ideas on how to apply the common sense knowledge in the automatic translation from/to Brazilian Portuguese and to/from English. After presenting the current scenario of our ongoing work, we point out some opportunities for collaboration in section 5. Finally, section 6 finishes this paper with insights about the next steps of our research.

## 2 Machine Translation

Machine Translation (MT) has about 70 years of history and lot of its recent achievements are directly related to the advances in computer science, which enable almost everyone to have access and use MT tools. Some of these tools were traditionally developed following the rule-based approach (e.g., Systran[6] and Apertium[7]) but the statistical approach is now being widely applied at least in part (e.g., Google[8]) (Cancedda et al., 2009).

The SMT was born in the late 1980s as an effort of researchers from IBM (Brown et al., 1990). In those days, SMT was performed based on two models: a word-based translation model and a language model. While the first model is concerned with the production of target equivalent versions of the source sentences, the second one guarantees that the output sentence is a possible one (it is grammatical and fluent) in the target language. In the current PB-SMT systems, the word-based models were replaced by the phrase-based ones built based on sequences of words (the *phrases*).[9]

The translation and language models used in SMT are built from a training parallel corpora (a set of source sentences and their translations into the target language) by means of IBM models (Brown et al., 1993) which calculate the probability of a given source word (or sequences of words) be translated to a target word (or sequence of words). The availability of some open-source toolkits (such as Moses (Koehn et al., 2007)[10]) to train, test and evaluate SMT models has helping the widely employment of this MT approach to perhaps almost any language pair and corpus type. In fact, SMT is an inexpensive, easy and language independent way for detecting recurrent phrases that form the language and translation models.

However, while PB-SMT models have achieved the state-of-the-art translation quality, its performance seams to be stagnated. Consequently, there is a recent common trend towards enriching the current models with some extra knowledge as the new approaches of factored translation models (Koehn and Hoang, 2007) or syntax-based (or syntax-augmented) MT systems (Tiedemann and Kotzé, 2009; Tinsley and Way, 2009; Zollmann et al., 2008).

More related to our work are the proposals of Musa et al. (2003) and Chung et al. (2005). Both

[4]Slangs are typically cultural because they characterize the mode of a group's speech in a given space and time.

[5]Sentence extracted from Cambridge Advanced Learner's Dictionary: `http://dictionary.cambridge.org/define.asp?key=13018&dict=CALD`.

[6]`http://www.systransoft.com/`

[7]`http://www.apertium.org/`

[8]`http://www.google.com/language_tools`

[9]In SMT, a *phrase* is a sequence of two or more words even though they do not form a syntactic phrase.

[10]`http://www.statmt.org/moses/`

of them are CS-based translation tools which take the topics of a bilingual conversation guessed by a topic spotting mechanism, and use them to generate phrases that can be chosen by the end-user to follow the conversation. Since they are interactive tools, the phrases are first displayed on the screen in the end-user's native language and, then, he/she selects a phrase to be translated (by a text-to-speech engine) in the language in which the conversation is taking place.

In our work, the main goal is also investigating new ways to improve MT performance, but instead of greater BLEU or NIST values we are interested in producing more culturally contextualized translations. Similarly to (Musa et al., 2003) and (Chung et al., 2005), we intend to help two bilingual users to develop a communication. However, in our case we are not only concerned with the language differences, but also the cultural divergences. To achieve this ambitious goal we rely on common sense knowledge collected from Brazilian and North-American individuals as explained in the next section.

## 3 Common Sense

Common sense (CS) plays an important role in the communication between two people as the interchanged messages carries their prior beliefs, attitudes, and values (Anacleto et al., 2006b). When this communication involves more than one language, translation tools can help to deal with the language barrier but they are not able to cope with the cultural one. In this case, the CS knowledge is a powerful mean to guarantee that the understanding will overcomes the cultural differences.

The CS knowledge applied in our research was collaboratively collected from volunteers through web sites and reflects the culture of their communities (Anacleto et al., 2006a; Anacleto et al., 2006b). More specifically, our research relies on CS collected as an effort of the Open Mind Common Sense projects in Brazil (OMCS-Br[11]) and in the USA (OMCS[12]).

The OMCS started in 1999, at the MIT Media Lab, to collect common sense from volunteers on the Internet. More than ten years later, this project encompass many different areas, languages, and problems. Nowadays, there are over a million sentences in the English site collected from over 15,000 contributors.[13]

OMCS-Br is a younger project that has being developed by LIA-DC/UFSCar (Advanced Interaction Laboratory of the Federal University of São Carlos) since August 2005. Figure 1 illustrates the OMCS-Br architecture to collect and manipulate CS knowledge in five work fronts: (1) common sense knowledge collection, (2) knowledge representation, (3) knowledge manipulation, (4) access and (5) use. A detailed explanation of each work front can be found in (Anacleto et al., 2008a).[14]

As can be seen in Figure 1, the CS knowledge is collected in the OMCS-Br site (bottom-left) by means of templates[15]. Then, the collected fact is stored in a knowledge base (up-left) from which it is converted into graphs that form a semantic network. These semantic networks, called ConceptNets, are composed of nodes and arcs (to connect nodes) as shown in the bottom-right part of Figure 1. The nodes represent the knowledge derived from the CS base while the arcs represent relations between two nodes based on studies on the theory of (Minsky, 1986). Examples of Minsky relations extracted from the ConceptNet, in English, related to the term *"book"* are: IsA (*"book" IsA "literary work"*), UsedFor (*"book" UsedFor "learn"*), CapableOf (*"book" CapableOf "store useful knowledge"*), PartOf (*"book" PartOf "library"*) and DefinedAs (*"book" DefinedAs "foundation knowledge"*).

Figure 2 brings an extract of our Brazilian ConceptNet (Anacleto et al., 2008b) and Figure 3, a parallel extract obtained from the North-American ConceptNet (Singh, 2002). As it is possible to notice from these figures, there is a straight relationship between these ConceptNets. It is possible to find many cases in which relations in English have

[11] `http://www.sensocomum.ufscar.br`

[12] `http://commons.media.mit.edu/en/`

[13] `http://csc.media.mit.edu/`

[14] Examples of successful applications using the CS knowledge derived from OMCS-Br can be fount at `http://lia.dc.ufscar.br/`

[15] The templates are semi-structured statements in natural language with some gaps that should be filled out with the contributors' knowledge so that the final statement corresponds to a common sense fact (Anacleto et al., 2008a).

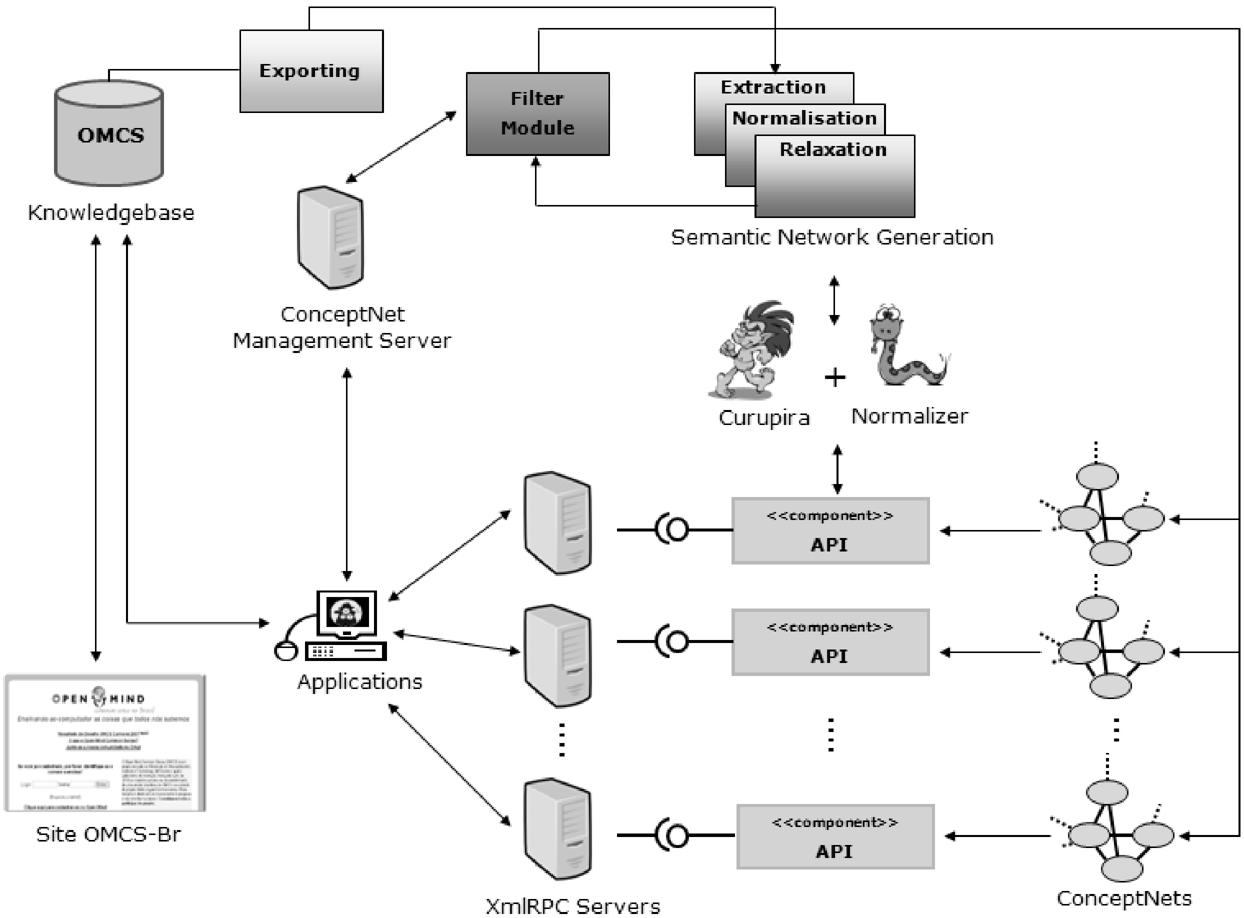


Figure 1: OMCS-Br Project architecture (Anacleto et al., 2008a)

their counterpart in Portuguese as in the example given in which "*book*" is connected with "*learn*" by the relation *UsedFor* and the *book*'s translation to Portuguese, "*livro*", is also linked with the translation of *learn* ("*aprender*") by a relation of the same type.

Different from other researches using semantic networks, such as MindNet[16] (Vanderwende et al., 2005), WordNet[17] (Fellbaum, 1998) and FrameNet[18] (Baker et al., 1998), here we propose the application of source and target ConceptNets together in the same application.

## 4 Culturally Contextualized Machine Translation

As presented in the previous sections, the main goal of our research is to investigate how CS knowledge can help MT systems to generate more culturally contextualized translations. To do so, we are working with two ConceptNets derived from OMCS and OMCS-Br projects, that represent the CS knowledge in English and Brazilian Portuguese, respectively, as presented in section 3.

In this context, we intend to investigate the application of CS knowledge in the MT process in three different moments:

1. Before the automatic translation – In this case the source sentence input is enriched with some CS knowledge (for example, context information) that can help the MT tool to choose the best translation;

[16] http://research.microsoft.com/en-us/projects/mindnet/
[17] http://wordnet.princeton.edu/
[18] http://framenet.icsi.berkeley.edu/

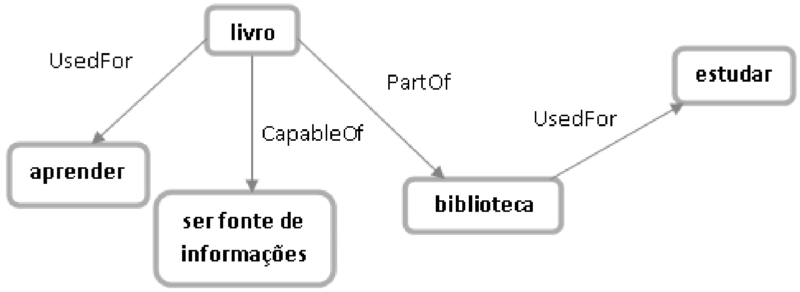


Figure 2: Graphical representation of the Brazilian ConceptNet (Meuchi et al., 2009)

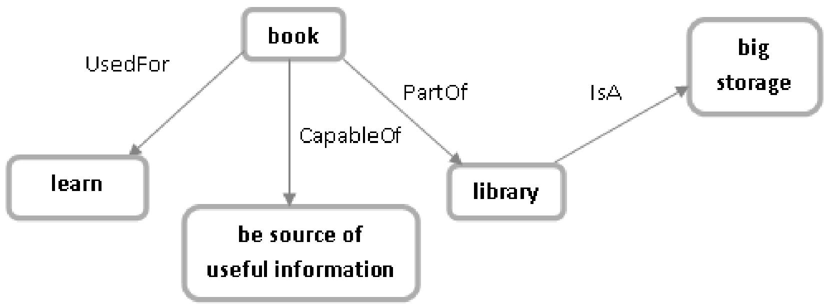


Figure 3: Graphical representation of the North-American ConceptNet (Meuchi et al., 2009)

2. During the automatic translation – In this case the CS knowledge is used as a new feature in the machine learning process of translation;

3. After the automatic translation – In this case some target words in the output sentence can be enriched with CS knowledge (for example, the knowledge derived from the "DefinedAs" or "IsA" Minsky relations) to better explain their meanings.

Currently, we are dealing with the last moment and planing some ways to fix/filter the target sentences produced by a SMT system. This part of the work is being carried out in the scope of a master's project which aims at building a bilingual culturally contextualized chat. By using a SMT tool (SMTT) and a CS knowledge tool (CSKT), this chat will help the communication between two users with different languages and cultures.

The SMTT is a phrase-based one trained using Moses and a corpus of 17,397 pairs of Portuguese–English parallel sentences with 1,026,512 tokens (494,391 in Portuguese and 532,121 in English). The training corpus contains articles from the online version of the Brazilian scientific magazine *Pesquisa FAPESP*[19] written in Brazilian Portuguese (original) and English (version) and, thus, a vocabulary that do not fit exactly the one found in chats. The SMTT trained based on this training corpus had a performance of 0.39 BLEU and 8.30 NIST for Portuguese–English translation and 0.36 BLEU and 7.83 NIST for English–Portuguese, in a test corpus composed of 649 new parallel sentences from the same domain of the training corpus (Caseli and Nunes, 2009).[20] For our experiments with culturally-contextualized MT, the option of using SMT models trained on general language in spite of building specific ones for the chat domain was taken aiming at measuring the impact that the CSKT has on the final translation.

The CSKT, in turn, will help one user to write his/her messages taking into account the

[19] `http://revistapesquisa.fapesp.br`

[20] In previous experiments carried out on the same corpora, the best online MT system was Google with 0.33 BLEU and 7.61 NIST for Portuguse–English and 0.31 BLEU and 6.87 NIST for English–Portuguese translation (Caseli et al., 2006).

cultural differences between he/she and the other user. A culturally contextualized translation will be generated by applying the knowledge derived from the two ConceptNets (see section 3) to fix/filter the automatically generated translations in a semi-automatic process assisted by both chat users.

To illustrate the use of both tools in the production of a culturally contextualized translation, lets work with slangs in the following example. Imagine a Brazilian and an American communicating through our bilingual chat supported by the SMTT and the CSKT. The American user writes the sentence:

American says: "*Hey* ***dude****, I will borrow a* ***C-note*** *from someone tomorrow!*".

Supposing that our SMTT is not able to provide a translation for the words "*dude*" and "*C-note*" —what is, indeed, a true possibility— outputting an incomplete translation in which these words remain untranslated. Consequently, the Brazilian user would not understand the American's sentence incompletely translated to Portuguese. So, since the SMTT do not know the translation of these slangs, the CSKT will be started to look for possible definitions in the CS knowledge bases. At this moment, the CSKT could provide some basic information about the untranslated words, for example that "*dude is a slang*" and "*dude (is) defined as guy*" or that "*C-note (is) defined as 100 dollars*", etc. Being aware of the untranslated words and their cultural meanings displayed by the CSKT, the American user could change or edit his/her original message by writing:

American says: "*Hey* ***guy****, I will borrow* ***100 dollars*** *from someone tomorrow!*".

The final edited sentence has a higher probability to occur in the target language than the original one and, so, to be corrected translated by the SMTT.

In addition to this master's project, we are also developing two undergraduate researches aiming at discovering useful knowledge from the "parallel" ConceptNets. The first ongoing undergraduate research (Barchi et al., 2009) aims at aligning the parallel concepts found in Brazilian and English ConceptNets. This alignment can be performed, for example, based on lexical alignments automatically generated by GIZA++[21] (Och and Ney, 2000) or the hierarchical structure of the nodes and arcs in the ConceptNets. The second ongoing undergraduate research (Meuchi et al., 2009), in turn, is involved with the enrichment of one ConceptNet based on the relations found in the other (parallel) ConceptNet and also in lexically aligned parallel texts.

## 5 Opportunities for Collaboration

The work described in this paper presents the first steps towards applying semantic knowledge to generate more culturally contextualized translations between Brazilian Portuguese and English texts. In this sense, we see some opportunities for collaboration regarding the roles that are played by: (1) our research work, (2) the semantic resources available to be used and (3) the resources and results that will be produced by our work.

First of all, this work is a joint effort of two research areas: NLP/CL (machine translation) and human-computer interaction (HCI) (common sense knowledge gathering and usage). From this fact, we see a great opportunity to bring a new "vision" to the NLP/CL applications in which we are concerned with not only to produce a correct answer to the proposed problem, but also an answer that sounds more natural and user-friendly. So, regarding our work's role, we see the opportunity to improve the collaboration between researchers from NLP/CL and HCI.

The second possibility of collaboration envisioned by us is related to other sources of semantic knowledge that could be applied to our work. Although we are using common sense knowledge to support the generation of more culturally contextualized translations, other semantic information bases could also be applied. In this case, we believe that this workshop is a great opportunity to be aware of other research projects that apply semantic knowledge to MT or are engaged with the construction of semantic resources that could be used in our work.

Finally, we also see a future source of collaboration regarding the use of the bilingual resources obtained as the product of this research.

[21] http://code.google.com/p/giza-pp/

The parallel-aligned (in Brazilian Portuguese and English) common sense base, the translation knowledge inferred from this aligned base or even the bilingual culturally contextualized chat would be useful in other research projects in MT or other bilingual applications such as information retrieval or summarization. We also believe that the methodology applied to develop these resources and the results obtained from this work could be applied to other language pairs to derive new bilingual similar resources.

## 6 Conclusions and Future Work

In this paper we have described the first ideas and steps towards the culturally contextualized machine translation, a new approach to generate automatic translations using a phrase-based SMT tool and a common sense knowledge tool.

It is important to say that this proposal involves researchers from NLP/CL an HCI and it brings an opportunity for collaboration between these related areas. Furthermore, this work aims at stimulating researchers from other countries to work with the Brazilian Portuguese and presenting its ideas in this workshop is a great opportunity to achieve this goal.

Future steps of this ongoing work are concerned with the implementation of the proposed prototypes designed for the bilingual culturally contextualized chat, the alignment and the enrichment of the ConceptNets. After the implementation of these prototypes they will be tested and refined to encompass the needed improvements.

## Acknowledgments

We thank the support of Brazilian agencies CAPES, CNPq and FAPESP and also the workshop organizers by making possible the presentation of this work.

## References

Junia Coutinho Anacleto, Henry Lieberman, Aparecido Augusto de Carvalho, Vânia Paula de Almeida Néris, Muriel de Souza Godoi, Marie Tsutsumi, José H. Espinosa, Américo Talarico Neto, and Silvia Zem-Mascarenhas. 2006a. Using common sense to recognize cultural differences. In *IBERAMIA-SBIA*, pages 370–379.

Junia Coutinho Anacleto, Henry Lieberman, Marie Tsutsumi, Vnia Neris, Aparecido Carvalho, Jose Espinosa, and Silvia Zem-mascarenhas. 2006b. Can common sense uncover cultural differences in computer applications. In *Proceedings of IFIP WCC2006, Spring-Verlag*, pages 1–10.

Junia Coutinho Anacleto, Aparecido Fabiano P. de Carvalho, Alexandre M. Ferreira, Eliane N. Pereira, and Alessandro J. F. Carlos. 2008a. Common sense-based applications to advance personalized learning. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (SMC 2008)*, pages 3244–3249, Singapore.

Junia Coutinho Anacleto, Aparecido Fabiano P. de Carvalho, Eliane N. Pereira, Alexandre M. Ferreira, and Alessandro J. F. Carlos. 2008b. Machines with good sense: How can computers become capable of sensible reasoning? In *IFIP AI*, pages 195–204.

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet Project. In *Proceedings of the COLING-ACL*, Montreal, Canada.

Paulo Henrique Barchi, Helena de Medeiros Caseli, and Junia Coutinho Anacleto. 2009. Alinhamento de Grafos: Investigação do Alinhamento de ConceptNets para a Tradução Automática. In *Anais do I Workshop de Iniciação Científica em Tecnologia da Informação e da Linguagem Humana (TILic)*, pages 1–4.

Peter F. Brown, John Cocke, Stephen A. Della Pietra, Vincent J. Della Pietra, Fredrick Jelinek, John D. Lafferty, Robert L. Mercer, and Paul S. Roossin. 1990. A statistical approach to machine translation. *Computational Linguistics*, 16(2):79–85.

Peter F. Brown, Vincent J.Della Pietra, Stephen A. Della Pietra, and Robert. L. Mercer. 1993. The Mathematics of Statistical Machine Translation: Parameter Estimation. *Computational Linguistics*, 19:263–311.

Nicola Cancedda, Marc Dymetman, George Foster, and Cyril Goutte, 2009. *A Statistical Machine Translation Primer*, chapter 1, pages 1–36. The MIT Press.

Helena de Medeiros Caseli and Israel Aono Nunes. 2009. Statistical machine translation: little changes big impacts. In *Proceedings of the 7th Brazilian Symposium in Information and Human Language Technology*, pages 1–9.

Helena de Medeiros Caseli, Maria das Graças Volpe Nunes, and Mikel L. Forcada. 2006. Automatic induction of bilingual resources from aligned parallel corpora: application to shallow-transfer machine translation. *Machine Translation*, 20:227–245.

Jae-woo Chung, Rachel Kern, and Henry Lieberman. 2005. Topic Spotting Common Sense Translation Assistant. In Gerrit C. van der Veer and Carolyn Gale, editors, *Extended Abstracts Proceedings of the 2005*

*Conference on Human Factors in Computing Systems (CHI 2005)*, Portland, Oregon, USA, April 2-7. ACM.

G. Doddington. 2002. Automatic evaluation of machine translation quality using n-gram cooccurrence statistics. In *Proceedings of the Human Language Technology Conference (HLT 2002)*, pages 128–132.

Christiane Fellbaum, editor. 1998. *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, MA.

Philipp Koehn and Hieu Hoang. 2007. Factored Translation Models. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 868–876, Prague, June. Association for Computational Linguistics.

Philipp Koehn, Franz Josef Och, and Daniel Marcu. 2003. Statistical phrase-based translation. In *Proceedings of the Human Language Technology (HLT/NAACL 2003)*, pages 127–133.

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 Association for Computational Linguistics (ACL 2007)*, pages 177–180, Prague, Czech Republic, June. Association for Computational Linguistics.

Laís Augusta Silva Meuchi, Helena de Medeiros Caseli, and Junia Coutinho Anacleto. 2009. Inferência de relações em ConceptNets com base em corpus paralelo alinhado. In *Anais do VI WorkShop de Trabalhos de Iniciação Científica (WTIC) - evento integrante do WebMedia 2009*, pages 1–3.

M. Minsky. 1986. *The Society of Mind*. Simon and Schuster, New York.

Rami Musa, Madleina Scheidegger, Andrea Kulas, and Yoan Anguilet. 2003. Globuddy, a dynamic broad context phrase book. In *CONTEXT*, pages 467–474.

Franz Josef Och and Hermann Ney. 2000. Improved statistical alignment models. In *Proceedings of the 38th Annual Meeting of the ACL (ACL 2000)*, pages 440–447, Hong Kong, China.

Franz Josef Och and Hermann Ney. 2004. The Alignment Template Approach to Statistical Machine Translation. *Computational Linguistics*, 30(4):417–449.

K. Papineni, S. Roukos, T. Ward, and W. J. Zhu. 2002. BLEU: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual meeting of the Association for Computational Linguistics (ACL 2002)*, pages 311–318.

P. Singh. 2002. The OpenMind Commonsense project. KurzweilAI.net. Available at: <http://web.media.mit.edu/˜push/OMCSProject.pdf>.

Jörg Tiedemann and Gideon Kotzé. 2009. Building a large machine–aligned parallel treebank. In Marco Passarotti, Adam Przepirkowski, Savina Raynaud, and Frank Van Eynde, editors, *Proceedings of the 8th International Workshop on Treebanks and Linguistic Theories (TLT'08)*, pages 197–208. EDUCatt, Milano/Italy.

John Tinsley and Andy Way. 2009. Automatically generated parallel treebanks and their exploitability in machine translation. *Machine Translation*, 23:1–22.

Lucy Vanderwende, Gary Kacmarcik, Hisami Suzuki, and Arul Menezes. 2005. Mindnet: an automatically-created lexical resource. In *Proceedings of HLT/EMNLP on Interactive Demonstrations*, pages 8–9, Morristown, NJ, USA. Association for Computational Linguistics.

Andreas Zollmann, Ashish Venugopal, Franz Och, and Jay Ponte. 2008. A systematic comparison of phrase–based, hierarchical and syntax–augmented statistical mt. In *COLING '08: Proceedings of the 22nd International Conference on Computational Linguistics*, pages 1145–1152, Morristown, NJ, USA. Association for Computational Linguistics.