Machine Translation Using Automatically Inferred Construction-based Correspondence and Language Models[*]

Shimon Edelman[a] and Zach Solan[b]

[a]Dept. of Psychology, Cornell University, Ithaca, NY 14853, USA
and Dept. of Brain and Cognitive Engineering, Korea University, Seoul 136-713, Republic of Korea
se37@cornell.edu
[b]Meaningo.com
zsolan@meaningo.com

Abstract. We discuss the problem of translation in the wider context of the problem of meaning in cognition and describe a structural statistical machine translation (MT) method motivated by philosophical, cognitive, and computational considerations. Our approach relies on a recently published algorithm capable of learning from a raw corpus a limited yet effective grammar that can be used to construct probabilistic parsers and language models, and on cognitively motivated heuristics for learning construction-based translation models. A pilot system has been implemented and tested successfully on simple English to Hebrew and Spanish to English translation tasks.

Keywords: machine translation, grammar inference, cognitive linguistics

1 Introduction

Machine translation (MT) is the one application domain of computational linguistics where it would seem to be impossible to achieve any progress without giving some thought to semantics. For instance, the claim that the Spanish un perro mordió a un hombre should be translated into English as a dog bit a man is merely a way of stating that these two utterances *mean* the same thing. The meaning of "meaning" is, however, notoriously elusive (Putnam, 1975; Pietroski, 2003), and so practical approaches to MT wisely skirt theoretical semantics even as they embrace various ways of defining meaning in computational terms.

In computational linguistics, including MT, both context and use serve to operationalize meaning. For instance, a set of contexts inherent in a corpus can be distilled into a generative language model, while information about the use of corresponding phrases that is implicit in a parallel corpus can assist in the translation of new source-language utterances into the target language. In the present paper, we capitalize on recent advances in structural language model inference and on some cognitively motivated heuristics concerning cross-language construction mapping to derive a simple yet viable MT method.

1.1 Computational approaches to translation

To date, there has been little interaction between studies of human translation on the one hand and algorithmic developments in MT on the other hand (Poutsma, 2000), which is why we survey these separately (and very briefly). In MT, a dominant approach in the past decade has been based on the source-channel model of communication, according to which a Spanish utterance (say) is taken to be a (very) "noisy" version of an English utterance that has passed through a distorting

---

[*] We thank David Horn and Eytan Ruppin for comments on an early version of this work and Srinivas Bangalore for the Spanish version of the ATIS corpus.

channel (Berger et al., 1994). Translation then reduces to the problem of finding the most probable original English sentence, given its Spanish form.

Statistical approaches to translation rely on *language models* (LM) that assign probabilities to phrases and entire sentences (Goodman, 2001). These are an inherent part of the new empirical basis for linguistics, according to which to know a language means to approximate the probability distribution over the utterances that define it (Goldsmith, 2007). Given the set of all strings $\sum^*$ that can be formed out of the elements of some set of symbols $\sum$ (the lexicon), a grammar is then defined as a function g with the following properties: $g : \sum^* \to [0, 1]$ such that $\sum_{s \in \sum^*} g(s) = 1$. To account for the content and not just the form of the utterances, this fundamental definition needs to be extended by conditioning the probability of an utterance on its discourse context, as well as on the extralinguistic circumstances that drive the discourse in question. This extension is beyond the state of the art in natural language engineering, where practical applications, including those in MT, rely on various approximations to g.

The translation system developed by Berger et al. (1996), for example, used a smoothed trigram LM. It ranked various hypothesized alignments between source and target sentences, performed a perturbation search in the space of possible edits of the candidate sentence, and produced a maximum-likelihood output. Later developments of this approach included augmenting a large-scale general-purpose monolingual language model with a domain-specific one derived from a parallel corpus (Zhao et al., 2004) and applying a variety of phrase transduction templates in addition to edits (Kumar et al., 2005).

Koehn et al. (2003) evaluate several phrase-based methods within a Bayesian framework, which casts the goal of finding the best target-language sentence $\mathbf{t}$ given a source-language sentence $\mathbf{s}$ as $\operatorname{argmax}_{\mathbf{t}} p(\mathbf{t}|\mathbf{s}) = \operatorname{argmax}_{\mathbf{t}} p(\mathbf{s}|\mathbf{t})p(\mathbf{t})$, thus separating the target language model $p(\mathbf{t})$ from the translation model $p(\mathbf{s}|\mathbf{t})$. They report that a heuristic combination of word alignment with lexical weighting of phrase translations outperforms both pure word-based methods and syntactic methods that insist on mapping constituents.

A recent alternative to the source-channel approach (Och et al., 2004) calls for directly modeling the posterior probability via weighted alignment templates, along with a "smorgasbord" of lexical, phrasal, and shallow syntactic features that are fed to a machine learning algorithm. A large array of features fed to a machine learning stage (specifically, an online perceptron learning algorithm) is also at the core of another alternative approach, which uses a discriminative rather than generative probabilistic model (Liang et al., 2006).

Some of the newer methods involve various kinds of syntactic information in addition to "flat" phrases do prove effective in improving translation performance, compared to the phrasal alignment. For example, the method of DeNeefe et al. (2007) uses an aligned parallel corpus where the target side has been parsed to learn translation rules; a rule consists of a sequence of words and variables in the source language, a syntax tree in the target language that has words or variables at the leaves, and a vector of feature values that attempt to capture the rule's likelihood. These latter are processes using sophisticated, typically extremely data-hungry approaches (Fossum et al., 2008).

May and Knight (2007) recently distinguished between an "idealistic" approach to MT (which calls for learning a statistical generative model from a parallel (bitext) corpus, then translating the target text using the estimated model) and a "realistic" one (in which the learned model is used only to find word alignments in the bitext, which then serve as a basis for training and interpolating a set of rules for translation). In their work, May and Knight attempted to address what they described as a disconnect between alignment and translation models in the "realistic" approach. They found that part of speech (POS) features helped, but other kinds of syntactic features either hurt performance or did not help it; the best added feature improved performance by about one percent over the baseline, as estimated automatically using the BLEU performance

index (Papineni et al., 2002). This result is quite typical of the state of the art in MT: although the involvement of syntax can lead to statistically significant performance boost, the absolute value of the improvement is usually very small: one or two percent at best.

1.2 Cognitive aspects of translation in human bilinguals

Given the persistent gap between the performance of MT systems and of human translators, it seems natural to turn to studies of human translation in search of guidance. Unfortunately, the applied linguistics literature on interpreting and translation (see Valdés and Angelelli (2003) for a review) steers clear both of computational theories and of empirical methods for performance assessment. Likewise, of the many studies of bilingualism in cognitive psychology and psycholinguistics, only very few deal with interpreting and translation (Valdés and Angelelli, 2003).

One such study is that of Christoffels and de Groot (2004), who compared subjects' performance in repeating sentences (shadowing), reformulating sentences in the same language (paraphrasing), and translating (interpreting) sentences, all delivered auditorily, in simultaneous and delayed conditions. They found that both "transcoding" (a concept that corresponds roughly to source to target alignment in MT) and target-language semantic processing were needed to account for the pattern of results.

The latter kind of processing plays a particularly important role in the Revised Hierarchical Model (RHM) of bilingual conceptual memory (Kroll and Stewart, 1994). According to the RHM, true bilinguals (people who are equally fluent in both their languages) maintain two sets of lexical semantic representations of concepts. The corresponding lexical records in the two languages are linked primarily via a common non-linguistic conceptual system, over and above any direct, asymmetrical associations at the lexical level (hence the attribution "hierarchical"). For example, in a Spanish-English bilingual, perro would be associated with dog, but, more importantly, both these words would be linked to the concept 🐎, all with different weights.

Support for RHM has been provided by a great number of behavioral studies, reviewed by French and Jacquet (2004). The asymmetrical associations posited by RHM emerged also from a recent study specifically designed to test this model (Hatzidaki and Pothos, 2008); that study also found context- and task-specific effects on translation processing. A detailed look at the time course of lexical access in bilinguals revealed parallel real-time activation of first- (L1) and second- (L2) language lexical representations (Marian et al., 2003). This and other studies of the brain basis of bilingualism reviewed by Perani and Abutalebi (2005) suggest that L1 and L2 processing is subserved by the same brain areas and likely by the same mechanisms.

Although RHM is not nearly detailed enough to qualify as a model in computational linguistics, it can help distinguish between engineering approaches that are compatible with the behavioral and neurobiological findings and those that are less so. In particular, these findings suggest that computational methods that rely exclusively on direct L1-L2 phrasal alignment cannot be very good models of human translation. This category includes extreme examples (such as the approach of Chiang (2005), which links hierarchically structured phrases in the two languages by learning a synchronous bilingual grammar from a bitext corpus), as well as the phrasal alignment approaches and the hybrid ones that are augmented with syntactic cues, as discussed in the preceding section.

2 MT based on learning structural-statistical generative language models

The RHM approach assigns a central role in the translation process to an indirect route that first maps a source utterance to its conceptual representation, which is then mapped further into a well-formed and appropriately meaningful sentence into the target language. Because building a full-scale model of the conceptual system of a human bilingual is at present infeasible, we chose to focus on modeling its most relevant subsystem, namely, the network of constructions that mediates between concepts and the channels of linguistic input (utterance comprehension) and output (utterance production).
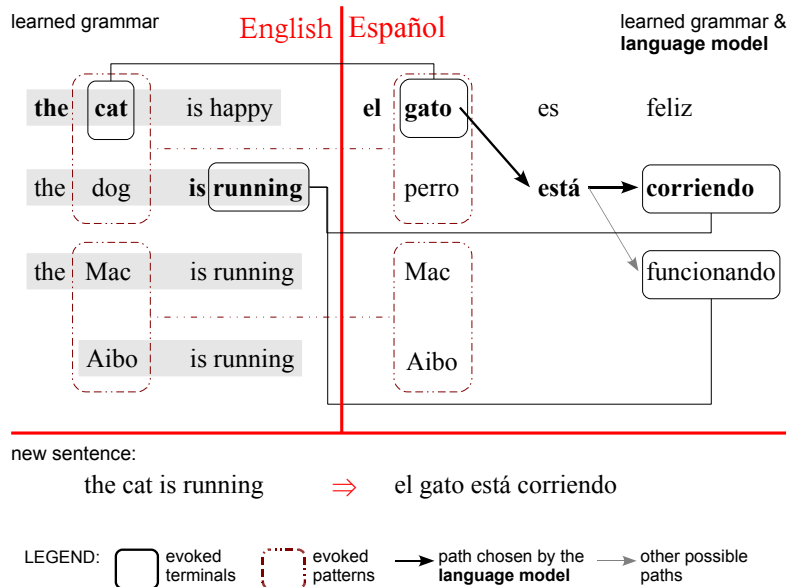
Figure 1: An illustration of the main principles behind our method, using English to Spanish translation as an example. The source-language grammar (learned from a corpus by the ADIOS algorithm; Solan et al., 2005), consists of terminals (such as cat or the), collocation patterns (which may be only partially lexicalized, as in 'the ___ is running'), and equivalence classes (such as the set {cat, dog}). The terminals evoked by the novel input (the sentence 'The cat is running') are shown in bold. These in turn activate corresponding target-language terminals on the right; patterns and equivalence classes also activate their counterparts. Due to polysemy (here, between the different senses of running) and initial structural ambiguity, a source element can activate more than one target element; the disambiguation is enforced by the context (Karov and Edelman, 1998) via the interaction between the activated target elements and the target language model. The target-language grammar and the associated structural-statistical language model (SSLM) are also learned automatically. Given a set of active elements, the target-language model constructs the most probable sentence that is consistent with the source-language original.

In the technical sense used here, constructions are the elements of a construction grammar (Fillmore, 1985; Goldberg, 2005) — partially lexicalized forms of varying phrasal complexity that carry meanings of varying specificity. Constructions are language-specific (similar meanings may be conveyed in widely different ways in different languages; Croft, 2001), and so our bilingual network of constructions is bipartite. Recently published results suggest that constructions can be learned in an unsupervised fashion from raw corpus data with a degree of success (as measured by precision and recall) that may be limited, but sufficient for certain applications (Solan et al., 2005). Consequently, one feature of our approach to MT is its reliance on the ability to acquire for each language in question a probabilistic grammar — a generative model along with a probabilistic parser — that is learned automatically from a corpus (instead of being designed by hand on the basis of syntactic analysis).

To be useful for translation, constructions belonging to the source-language grammar need to be linked to those in the target-language grammar. Establishing these associations (which are the basis for existing phrasal alignment approaches to MT) is, however, only the first step in our approach. The system that we describe below seeds the source-to-target construction map with simple lexical associations gleaned from a machine-readable dictionary (MRD). It then extends the lexical associations to encompass phrasal constructions and refines the resulting map iteratively, as a way of approximating abstract, conceptual-level associations between the representations of meaning in the two languages (cf. Figure 1), thus completing the training phase. Given an

utterance in the source language, the trained system maps its shallow parse to a set of "evoked" constructions — which is how we operationalize meaning — in the target language, then lets the target-grammar LM generate the output utterance (instead of arbitrating among or merging various phrasal output candidates; Feng et al., 2009). The training and usage phases of our system's operation are described below.

3   Training the translation model

When seeking to map a construction in the source language to its counterpart(s) in the target language, one must recursively establish correspondence between the possible fillers in any of their open slots (equivalence classes). Our approach used this recursive correspondence information to enforce the proper thematic relations between the words in the target language, as they are strung together by the output language model. Consider the transitive form of the verb bites. Structurally, any pattern that captures its thematic content must possess two slots (arguments), as in bites(actor,patient). A symbol-level translation mapping $\mathcal{T}$ would then include a correspondence between the slots in the two languages, as in

$$\mathcal{T}\,(\text{bites},\text{muerde}) \doteq (x_1 \text{ bites } x_2) \leftrightarrow (\mathcal{T}\,(x_1) \text{ muerde } \mathcal{T}\,(x_2))$$

where the correspondence is indicated by shared indices. When the two possible outputs (dog bites man and man bites dog) are ranked, the one that preserves the correspondence should be primed to win, enabling the system to figure out, given the proper context of the source utterance, who bit whom.

We base the learning of the mapping $\mathcal{T}$ on the same principle that underlies multidimensional scaling (Shepard, 1987): the layout of a metric space can be recovered from a table of pairwise distances between points sampled from that space, or, more importantly, even just from the ranks of the distances (for example, a rough map of the US can be obtained from a table of inter-city driving distances — the *distance spectrum*). As suggested by Edelman (1999), this principle can be used to compute an optimal translation mapping between two disjoint symbol systems, such as the vocabularies of two languages, as long as they refer to same conceptual network, which constitutes the common underlying metric space (Goldstone and Rogosky, 2002). Intuitively, this is the "meaning space" shared by the speakers of the two languages.

Here, we extend this idea from mapping vocabulary entries to mapping supra-word (phrasal) constructions (the ADIOS algorithm that we use (Solan et al., 2005) treats lexical items and more complex constructions on an equal footing, as do construction grammars). Specifically, we seek an optimal mapping $\mathcal{T}$ using as inter-symbol distances in the two domains — language (A) and language (B) — the probabilities of symbol co-occurrence, $P(a_{j_1}, a_{j_2})$ and $P(b_{k_1}, b_{k_2})$, where the symbols stand for the various constructions. These distance matrices are estimated from the bilingual training corpus (which need not be parsed or even aligned very precisely; co-occurrence of constructions within a certain window over utterances may suffice). The distance matrices are optionally subjected to relaxation (Al-Homidan and Wolkowicz, 2005) to alleviate sparseness effects. The construction grammars for the source and target languages are learned by the ADIOS algorithm, which can cope with large, realistic corpora but is also capable of making good use of very little data (Solan et al., 2005).

4   Using the trained translation model

The resulting translation model is used as follows. In the *first stage*, a set of candidate translations of the individual items that appear in the source sentence is formed, using a bilingual MRD in conjunction with the mapping $\mathcal{T}$ from source constructions to target constructions, learned as described in the previous section.

In the *second stage*, the target-language structural statistical language model (SSLM) is used to thread together the candidate constructions to form a properly structured sentence that is the most consistent (i) with the priors on the candidates, and (ii) with the structural knowledge of the target language embodied by the SSLM. To that end, we use a simple pruned search with backtracking: the algorithm calculates the average perplexity scores (Goodman, 2001) of all possible paths through the SSLM that start from the initial element, while pruning branches whose average perplexity exceeds that of the best path by 5% or more.[1]

5 Performance

Pending a planned full evaluation using the Moses system (Koehn et al., 2007) as a benchmark, we tested the MT method outlined above on two simple tasks. The first task was English to Hebrew translation, using corpora generated by a simple context-free grammar with 50 terminals and 28 rules (grammar TA1, whose full specification can be found in the supplementary material to Solan et al., 2005, and its Hebrew counterpart). Despite its simplicity, the grammar could produce rather involved sentences, such as 'that a cow barks and the bird jumps worries the horse', which translates into Hebrew as זה שפרה נובחת וה ציפור קופצת מדאיג את הסוס. The quality of the translation was assessed by the BLEU index (Papineni et al., 2002), calculated using a Perl script provided by NIST.[2] The system was trained on a 100-sentence parallel corpus, achieving, on a distinct corpus of the same size, a BLEU score of 0.92.

We then tested the system on a Spanish-English parallel version of the Air Traffic Information System (ATIS) corpus (Hemphill et al., 1990). Most of the corpus (11294 sentences) was used to learn the source-language (Spanish) grammar and parser, to learn the target-language (English) grammar and train the corresponding language model (using the ADIOS algorithm), and to estimate the translation candidate mapping. The remainder (1870 sentences) were translated using the resulting model. A very simple Spanish-English MRD was used to seed the translation mapping.[3]

The two examples below show the source sentence [S], the translation obtained from the Google language tools [G], and the ADIOS translation [A]:

S1   hay un vuelo de ontario california a orlando florida con una escala

G1   there is a californian flight of ontario to trimming flowery with a scale

A1   is there a flight from ontario to florida with a stopover

S2   por favor enumere los vuelos por la mañana de filadelfia a toronto

G2   please it enumerates the flights in the morning from philadelphia to toronto

A2   please list the flights for tomorrow from philadelphia to toronto

The mean BLEU score over the test corpus in this task was 0.2 for Google and 0.4 for our method. The performance of our system is comparable to the state of the art in Spanish-English MT, as exemplified by the METIS system (Melero and Badia, 2007). It should be possible to improve it by various available means, such as training with more data (Brants et al., 2007), using a better MRD, and performing state of the art word alignment (Och and Ney, 2003).

We hasten to remark that the comparisons we presented (especially the one with Google) may not be entirely fair, seeing that the Google engine and the METIS system are general-purpose, whereas the ADIOS-based model has been trained on the very small ATIS corpus, which pertains to a linguistically highly impoverished domain. Building a general-purpose translation system using

---

[1] The pseudocode for learning the required representations and for using them appears online at `http://kybele.psych.cornell.edu/~edelman/MT-pseudocode.pdf`

[2] See `ftp://jaguar.ncsl.nist.gov/mt/resources/mteval-v11b.pl`; the BLEU index ranges from 0 to 1, where 1 is the best possible score.

[3] `http://www.probertencyclopaedia.com/dic/sp-eng-a.htm`

the approach described here would require a much more powerful method for grammar inference than the ADIOS algorithm. It is important to realize, however, that the ADIOS algorithm *can* learn a highly effective (low-perplexity) language model from very little data, such as an unannotated, unparsed corpus of a few thousand sentences (Solan et al., 2005). Thus, from the standpoint of performance, we can afford to learn a specialized translation model for each narrow domain or task.

6 Conclusions

We described a viable method for MT, based on the idea that the process of translation involves (i) a cognitively motivated mapping of constructions (which pertain both to form and to meaning) from the source to the target language and (ii) a generative model for the target language that produces the output from the evoked bag of constructions. Thus, in our method, correspondences between lexical and phrasal constructions in the two languages, inferred automatically from a parallel corpus, enforce thematic consistency of the output, while an automatically learned structural statistical language model for the target language is used to enforce output grammaticality. The encouraging performance of our simple model (whose computational sophistication falls far short of the state of the art in MT; Callison-Burch et al., 2009) should serve to motivate further work in bridging the gap between cognitive characterizations of bilingualism and their application in machine translation.

References

Al-Homidan, S. and Wolkowicz, H. (2005). Approximate and exact completion problems for Euclidean distance matrices using semidefinite programming. *Linear Algebra and its Applications*, 406:109–141.

Berger, A., Brown, P., Della Pietra, S., Della Pietra, V., Lafferty, J., Printz, H., and Ures, L. (1994). The Candide system for machine translation. In *Proc. ARPA Conference on Human Language Technology*.

Berger, A., Della Pietra, S., and Della Pietra, V. (1996). A maximum entropy approach to natural language processing. *Computational Linguistics*, 22:39–71.

Brants, T., Popat, A., Xu, P., Och, F. J., and Dean, J. (2007). Large language models in machine translation. In *Proceedings of EMNLP*, pages 858–867.

Callison-Burch, C., Koehn, P., Monz, C., and Schroeder, J. (2009). Findings of the 2009 workshop on statistical machine translation.

Chiang, D. (2005). A hierarchical phrase-based model for statistical machine translation. In *Proceedings of ACL'05*, pages 263–270.

Christoffels, I. K. and de Groot, A. M. B. (2004). Components of simultaneous interpreting: Comparing interpreting with shadowing and paraphrasing. *Bilingualism: Language and Cognition*, 7:227–240.

Croft, W. (2001). *Radical Construction Grammar: syntactic theory in typological perspective*. Oxford University Press, Oxford.

DeNeefe, S., Knight, K., Wang, W., and Marcu, D. (2007). What can syntax-based MT learn from phrase-based MT? In *Proc. EMNLP-CoNLL*.

Edelman, S. (1999). *Representation and recognition in vision*. MIT Press, Cambridge, MA.

Feng, Y., Liu, Y., Mi, H., and Liu, Q. (2009). Lattice-based system combination for statistical machine translation. In *Proceedings of EMNLP*. To appear.

Fillmore, C. J. (1985). Syntactic intrusion and the notion of grammatical construction. *Berkeley Linguistic Society*, 11:73–86.

Fossum, V., Knight, K., and Abney, S. (2008). Using syntax to improve word alignment precision for syntax-based machine translation. In *Proc. Workshop on Statistical MT*.

French, R. M. and Jacquet, M. (2004). Understanding bilingual memory: models and data. *Trends in Cognitive Sciences*, 8:87–93.

Goldberg, A. E. (2005). *Constructions at work: the nature of generalization in language*. Oxford University Press, Oxford.

Goldsmith, J. A. (2007). Towards a new empiricism. In de Carvalho, J. B., editor, *Recherches linguistiques à Vincennes*, volume 36.

Goldstone, R. L. and Rogosky, B. J. (2002). Using relations within conceptual systems to translate across conceptual systems. *Cognition*, 84:295–320.

Goodman, J. T. (2001). A bit of progress in language modeling: Extended version. Technical Report MSR-TR-2001-72, Microsoft Research.

Hatzidaki, A. and Pothos, E. M. (2008). Bilingual language representation and cognitive processes in translation. *Applied Psycholinguistics*, 29:125–150.

Hemphill, C. T., Godfrey, J. J., and Doddington, G. R. (1990). The ATIS spoken language systems pilot corpus. In *Proceedings of a workshop on speech and natural language*, pages 96–101. Morgan Kaufmann, San Francisco, CA.

Karov, Y. and Edelman, S. (1998). Similarity-based word sense disambiguation. *Computational Linguistics*, 24:41–59.

Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., Cowan, B., Shen, W., Moran, C., Zens, R., Dyer, C., Bojar, O., Constantin, A., and Herbst, A. (2007). Moses: Open source toolkit for statistical machine translation. Demonstration session.

Koehn, P., Och, F. J., and Marcu, D. (2003). Statistical phrase-based translation. In *HLT/NAACL*.

Kroll, J. F. and Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, 33:149–174.

Kumar, S., Deng, Y., and Byrne, W. (2005). A weighted finite state transducer translation template model for statistical machine translation. *Natural Language Engineering*, 12:35–75.

Liang, P., Bouchard-Côté, A., Klein, D., and Taskar, B. (2006). An end-to-end discriminative approach to machine translation. In *Proceedings of ACL'06*, pages 761–768.

Marian, V., Spivey, M. J., and Hirsch, J. (2003). Shared and separate systems in bilingual language processing: Converging evidence from eyetracking and brain imaging. *Brain and Language*, 86:70–82.

May, J. and Knight, K. (2007). Syntactic re-alignment models for machine translation. In *Proc. EMNLP-CoNLL*.

Melero, M. and Badia, T. (2007). Demonstration of the Spanish to English METIS-II MT system. In *Proceedings of the 11th International Conference on Theoretical and Methodological Issues in Machine Translation*, pages 132–133.

Och, F. J., Gildea, D., Khudanpur, S., Sarkar, A., Yamada, K., Fraser, A., Kumar, S., Shen, L., Smith, D., Eng, K., Jain, V., Jin, Z., and Radev, D. R. (2004). A smorgasbord of features for statistical machine translation. In *Proc. HLT-NAACL*, pages 161–168.

Och, F. J. and Ney, H. (2003). A systematic comparison of various statistical alignment models. *Computational Linguistics*, 29:19–51.

Papineni, K., Roukos, S., Ward, T., and Zhu, W. J. (2002). BLEU: a method for automatic evaluation of machine translation. In *Proceedings of ACL'02*, pages 311–318.

Perani, D. and Abutalebi, J. (2005). The neural basis of first and second language processing. *Current Opinion in Neurobiology*, 15:202–206.

Pietroski, P. M. (2003). The character of natural language semantics. In Barber, A., editor, *Epistemology of Language*, pages 217–256. Oxford University Press, Oxford, UK.

Poutsma, A. (2000). Examining cognitive aspects of human and machine translation. Unpublished manuscript.

Putnam, H. (1975). The meaning of 'meaning'. In Gunderson, K., editor, *Language, Mind, and Knowledge*, pages 131–193. U. of Minnesota Press, Minneapolis, MN.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, 237:1317–1323.

Solan, Z., Horn, D., Ruppin, E., and Edelman, S. (2005). Unsupervised learning of natural languages. *Proceedings of the National Academy of Science*, 102:11629–11634.

Valdés, G. and Angelelli, C. (2003). Interpreters, interpreting, and the study of bilingualism. *Annual Review of Applied Linguistics*, 23:58–78.

Zhao, B., Eck, M., Vogel, S., and Waibel, A. (2004). Language model adaptation for statistical machine translation via structured query models. In *Proc. Coling 2004*, Geneva, Switzerland.