7-2006

Context-Based Machine Translation

Jaime G. Carbonell
Meaningful Machines, LLC, jgc@cs.cmu.edu

Steve Klein
Meaningful Machines, LLC

David Miller
Meaningful Machines, LLC

Mike Steinbaum
Meaningful Machines, LLC

Tomer Grassiany
Meaningful Machines, LLC

See next page for additional authors

Authors
Jaime G. Carbonell, Steve Klein, David Miller, Mike Steinbaum, Tomer Grassiany, and Jochen Frey

Context-Based Machine Translation

Jaime Carbonell, Steve Klein, David Miller, Michael Steinbaum,
Tomer Grassiany, Jochen Frey
Meaningful Machines, LLC
{jaime, steve, dave, mike, tomer, jochen}@meaningfulmachines.com

Abstract

Context-Based Machine Translation™ (CBMT) is a new paradigm for corpus-based translation that requires no parallel text. Instead, CBMT relies on a light-weight translation model utilizing a full-form bilingual dictionary and a sophisticated decoder using long-range context via long n-grams and cascaded overlapping. The translation process is enhanced via in-language substitution of tokens and phrases, both for source and target, when top candidates cannot be confirmed or resolved in decoding. Substitution utilizes a synonym and near-synonym generator implemented as a corpus-based unsupervised learning process. Decoding requires a very large target-language-only corpus, and while substitution in target can be performed using that same corpus, substitution in source requires a separate (and smaller) source monolingual corpus. Spanish-to-English CBMT was tested on Spanish newswire text, achieving a BLEU score of 0.6462 in June 2006, the highest BLEU reported for any language pair. Further testing also shows that quality increases above the reported score as the target corpus size increases and as dictionary coverage of source words and phrases becomes more complete. [1]

1 Introduction

Traditional MT paradigms require either extensive transfer-rule writing by linguists and computer scientists or very large parallel (pre-translated) training corpora. The former can take person-decades to write, debug and perfect the set of rules for reasonable quality translation (e.g., at the SYSTRAN level), and acquiring parallel text for Statistical MT and Example-Based MT in sufficient quantity for comparable or better quality MT proves to be a daunting task even for major language pairs. For less-translated language pairs, accessible parallel text is simply non-existent in sufficient quantities. To address these serious challenges, CBMT is being developed as a corpus-based method that requires neither rules nor parallel corpora. Instead, CBMT requires an extensive monolingual target-text corpus (from 50 gigabytes to 1 terabyte), a full-form bilingual dictionary, and optionally (to further improve translation quality) a smaller monolingual source-text corpus. Monolingual text can be acquired in quantities by crawling the Web, extracting pure text from HTML, and indexing such text in the manner required by the CBMT methods. The most time-consuming resource to build or acquire is the large-scale full-form bilingual dictionary for each language pair.

CBMT exhibits two advantages over traditional MT approaches: (1) higher accuracy due to the model's ability to decode long n-grams (as evidenced by the performance of CBMT's still-incomplete prototype which achieved a higher BLEU score in initial testing than any other MT system to date), and (2) the ability to extend fairly rapidly to new language pairs, including those that lack sufficient parallel text. This paper describes the fundamental new techniques underlying CBMT, and presents results from Spanish-to-English MT.

In our AMTA-2002 paper (Abir et al., 2002), we introduced some of the component concepts on a general level, but details were not provided. This paper addresses several of those concepts as well as others, and provides methodological details. Furthermore, our 2002 paper stated that we were experimenting with a corpus-based

---

[1] The authors offer special acknowledgement to Eli Abir, co-founder of Meaningful Machines, who originally conceived of and participated in the development of the methods described in this paper.

approach that does not require parallel text, which we referred to as "AIMT." That process, which is now called "flooding," is a key aspect of CBMT, and is described in sections 2.1 and 3.2 below.

2   Basic CBMT Architecture

The CBMT method consists of several modules, architected as shown in Figure 1. The principle is to produce many long n-gram candidate translations by finding – in a huge target corpus – those long n-grams that contain as many as possible of the potential word and phrase translations from the dictionary, and as few as possible (if any) other content words. Then, the large number of candidate translations are resolved against their overlapping neighbors to select the long n-grams whose suffix is also the prefix of its right neighbor, and whose prefix is also the suffix of its left neighbor. The highest scoring translation (best long n-grams and maximal overlap) is selected by the decoder. Of course, the method requires substantial back-end machinery to index and match efficiently billions of target n-grams, and other such support tasks, which are not depicted in the primary functional architecture diagram.

Figure 1. CBMT Basic Run-Time Architecture
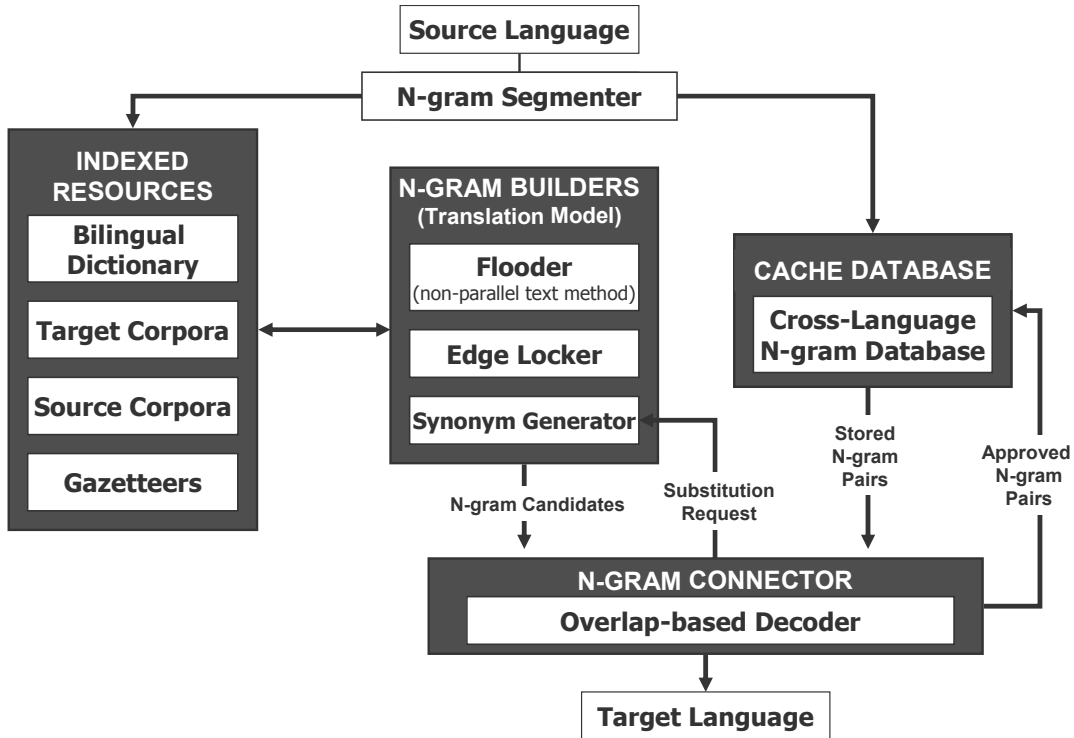
The primary functional components of CBMT are described below.

2.1   Cross-Language "Flooding"

This method generates target translation candidates for highly overlapping source language n-grams which are generally between 4 and 8 words in length (although n-grams can extend outside this range). The n-grams are highly overlapping in source; let $[w_k,w_{(k+n)}]$ denote the n-gram starting in sentence position k; we select n-grams $[w_1,w_n]$, $[w_2, w_{(n+1)}]$, $[w_3,w_{(n+2)}]$, etc. if, for example, n-grams are based on a fixed number of words. The n-gram size can also be based on the number of non-function words in the n-gram – and since the number of intervening function words is not constant – source n-gram size will vary. Regardless, the criteria for generating source overlapping n-grams from source input is an optimization parameter. Once the source n-grams are created, the

n-gram translation candidates are generated using a bilingual dictionary as a source of candidate translations on the word level (and phrase level) and we search a very large, multi-indexed target corpus for potentially corresponding target n-grams, i.e., those containing the maximal number of lexical translation candidates in context and the minimal number (preferably zero) of spurious content words. Word order in the target n-gram may differ slightly or greatly from word order in the corresponding source n-gram. The flooding process has the role of translation model in that it populates lattices with candidate translations, thus serving a similar function to the candidate translation lattice-building by SMT and EBMT systems. Each source n-gram may generate hundreds of target language n-gram translation candidates of varying length.

2.2 Source and Target Lattice Overlap

This method combines the target n-gram translation candidates by finding maximal left and right overlaps with the translation candidates of the previous and following n-grams. Thus, the retained target n-grams are those that are contextually anchored both left and right. In other words, only contextually confirmed translations are kept and scored by the overlap decoder. This process allows context well beyond the n words of the n-gram to affect the selected translation. Individual cross-language n-gram correspondences that overlap in source and target are cached in the cross-language database to be used when that source language n-gram appears in input in the future, thus improving MT speed. This step is akin to the lattice decoding process of SMT but extends the segments from trigrams to arbitrarily long n-grams (n typically ranges from 4 to 8) for greater accuracy. Furthermore, CBMT's confirmation process of requiring content word overlap on both sides of each long n-gram with preceding and succeeding n-grams is far stricter than other language-model decoding, which only requires abutting short n-grams from the translation model, scored by the target language model, which consists of transition probabilities derived from short n-grams (typically trigrams). In fact, CBMT's symbolic long-n-gram decoding outputs a fairly reliable confidence score on translated passages, which optimizes human post-editor utility by permitting him or her to focus on problematic passages.

2.3 Word and Phrasal Synonymy

This process uses an unsupervised method for contextual clustering over a monolingual corpus to generate word and phrasal synonyms and near-synonyms. If the overlap decoder fails to find coherent sequences of overlapping target n-grams, then it asks for more candidate translations. One way to obtain such is to attempt translating a synonym or near-synonym n-gram of the problematic source passage, in case such translation is more globally coherent (i.e., decodes by overlapping right and left contexts). Similarly, we can substitute target candidate n-grams with synonyms and near-synonyms and use them in flooding. In essence, replacing words or phrases with their synonyms or near-synonyms on an as-needed basis enlarges the search space of possible translations. To our knowledge, no other translation engine utilizes dynamically-generated word or phrasal synonymy to optimize translation results.

3 CBMT Under the Hood

This section provides functional details regarding the processes in operation.

3.1 Preprocessing: Preparation of Monolingual Corpus and Bilingual Dictionary

CBMT requires a full-form bilingual dictionary and a long-n-gram indexed target language corpus. The former is generated semi-automatically from a stem (citation form) bilingual dictionary, such as those available commercially, plus a set of inflection rules (conjugations, pluralizations, etc.) for source and target languages, including a cross-language inflection mapping table (e.g., "imperfait" in French maps to past progressive, simple past, and the form "used to $<V_{inf}>$" in English). In addition, multi-word entries, even non-lexicalized ones, are added to the dictionary as they too are used in flooding, and actually improve flooding efficiency and overall quality when they are matched, which is seldom at present because our phrasal coverage is minimal, but growing. Our Spanish-English dictionary has approximately 100,000 stems, which expands to 1.8M inflected forms, each with potentially multiple translations generated automatically from their stems and the cross-language inflection mapping table.

The long-n-gram indexed target language corpus requires first acquiring a large target language corpus, 50GB to 1TB via Web crawling, language identification, HTML stripping, sentence finding (vs. isolated words as in menu selections or small entries in tables), and multi-layered inverted indexing so long n-grams are quickly identified from the component words. The index search requires best-first behavior, where retrieved n-grams are those that contain the most component words at maximal proximity (minimal number spurious words). Techniques such as parallel search/merge expedite this otherwise most time-consuming step. Moreover, the index also requires left and right-context search for word and phrasal synonym finding, as described in section 3.4.

## 3.2   Process 1: Source-to-Target Flooding

The first phase of the translation process segments the source sentence into overlapping n-grams of typically 4 to 8 words each by moving a sliding window across the sentence, advancing the n-gram starting position one word at a time. Then for each n-gram, the system looks up all possible translations for every source word or known phrase in the bilingual dictionary. Then, the indexed target corpus is searched for n-grams that contain the maximal number of different potential translations of the source words (i.e., long target n-grams that contain one translation option from as many source words/phrases as possible, in any order). Essentially, an n-long conjunctive normal form (CNF) expression is generated where each conjunct is the disjunction of all possible translations for the corresponding source word or in-dictionary phrase. The flooding process is one of matching the CNF to the indexed target language corpus, taking the top matches (typically 300-1000), where "top" is defined as the highest density target language n-gram matches. The above process is repeated in a moving window left-to-right (for Spanish, English, etc.). For instance, we could start flooding with the CNF from source sentence positions 1-to-7, then 2-to-8, 3-to-9, etc. until the end of the sentence. Each n-gram search is a complex proximity query, but does not require pre-generation of permutations to find compact target n-grams without regard to target language order within the n-gram.

3.3   Process 2: Target Language Lattice Overlap Maximization

Flooding produces a lattice of n-gram translation candidates at each position within the source sentence. Then, the lattice candidates that fully overlap are selected over those that partially overlap or fail to overlap at all, as illustrated by the highlighted segments below. The final translation result is the lattice walk that globally maximizes overlaps, and that is composed of the highest match density (the most source word translation matches), longest target language n-grams. We have not yet sought the optimal parameter combination, but expect to do so based on a regression model applied to a per-language validation set. We would expect to obtain an additional modest boost in our BLEU score by so doing.

Figure 2. Lattice Overlap

Final Target Lattice   largest rebel group signed a peace deal in November
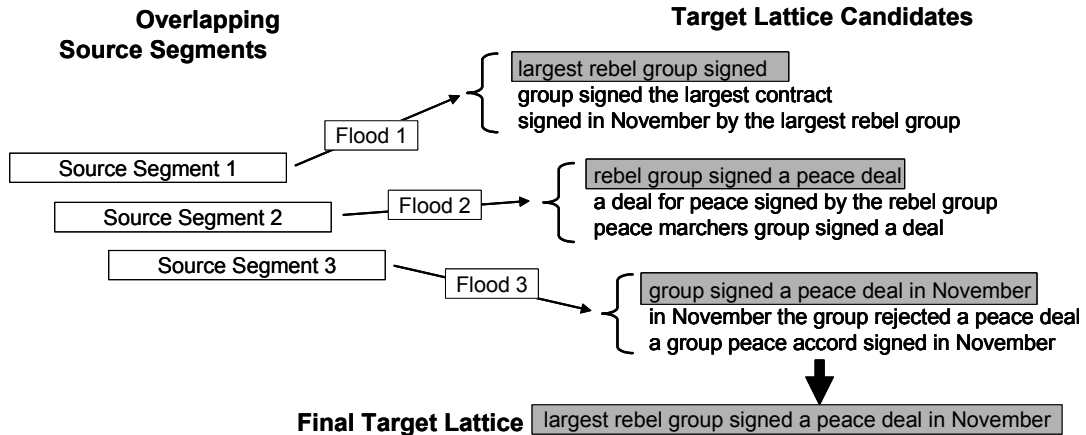
In this manner, each word in the middle of the target language sentence is confirmed by having appeared in multiple overlapping n-grams. Moreover, the coherence of the target sentence is strongly favored by the fact that all long overlapping n-grams come directly from human composed text in the target language, and the flow is at least locally coherent throughout the target sentence being composed, not just inside the flooded n-grams, but in the transitions between them, as otherwise there would not have been overlap. Strong overlap, hence, is an absolutely key process for CBMT. (Note that Figure 2 is a simplified example of the potential target candidates, with candidates comprised of only one of many possible target language translations per source word.)

3.4   Process 3: Word and Phrasal Synonym Generation

The CBMT system has a method for identifying synonyms or near-synonyms on a word or phrasal level using a monolingual corpus. This approach differs from others in that it does not require parallel resources (Barzilay and McKeown, 2001; Lin et al., 2003; Callison-Burch et al., 2006) nor does it use pre-determined sets of manually coded patterns (Lin et al., 2003). In addition, CBMT's methods work on the word as well as phrasal level. If the n-grams in the above described lattice fail to overlap fully (or at all) we could flood deeper by retrieving more than the top m flooding candidates from the target language corpus, but that would compromise computational tractability and degrade the quality of the results as we get fewer and fewer complete matches. As an alternative, the CBMT system's word and phrasal synonym (and near-synonym) generation method can identify synonyms and near-synonyms in source (or target) for the n-grams whose translations failed to resolve, thereby expanding the space of candidate translation generation in a new dimension. The challenge, of course, is to mine the source (or target) corpus for synonyms and near-synonyms dynamically and automatically.

The synonym (and near-synonym) creation process starts with a word or phrase, i.e., a short or long n-gram. Then, the process operates as follows:

Step 1: Perform a key-phrase/word-in-context index (a generalization of a KWIC index), generating paired left and right contexts that contain the desired word or phrase anywhere in the massive monolingual indexed corpus. Anywhere from 1,000 to 100,000 paired contexts, which may be of variable length, are typically generated.

Step 2: Tabulate, sort, and unify paired contexts (e.g., a long paired context that occurs multiple times ranks above one without repeat occurrences).

Step 3: Search the same massive indexed corpus, but this time with the contexts and only the contexts (not with the original phrase, a.k.a. the "middle") to find other words and phrases that fit the same contextual framework. Typically thousands are found.

Step 4: Rank the list of new middles according to several criteria, such as number of different context pairs in common with the original word or phrase, ratio of common contexts to total ones, frequency of common contexts, length of common contexts, and others. The top ranked items generally consist of synonyms and near-synonyms of the original word or phrase.

Figure 3. Synonym (and Near-Synonym) Generation

. . . of changes in **stock market** conditions based on . . .

. . . out of the **stock market** and putting it . . .

. . . stabilized as the **stock market** rebounded for most . . .

. . . changes in **stock market** wealth have become . . .

. . . from the NASDAQ **stock market** before the opening . . .

. . . to exit troubled **stock market** and banking ventures . . .

. . . out of the **stock market** and money going . . .

Find left and right contexts for
candidate phrase

. . . of changes in _____ conditions based on . . .

. . . out of the _____ and putting it . . .

. . . stabilized as the _____ rebounded for most . . .

. . . changes in _____ wealth have become . . .

. . . from the NASDAQ _____ before the opening . . .

. . . to exit troubled _____ and banking ventures . . .

. . . out of the _____ and money going . . .

Drop out the candidate phrase, leaving only
the left and right contexts

. . . of changes in **market** conditions based on . . .
. . . of changes in **industry** conditions based on . . .

. . . out of the **market** and putting it . . .
. . . out of the **stock exchange** and putting it . . .

. . . stabilized as the **economy** rebounded for most . . .
. . . stabilized as the **national market** rebounded for most . . .

. . . changes in **market** wealth have become . . .
. . . changes in **personal** wealth have become . . .

. . . from the NASDAQ **stock exchange** before the opening . . .
. . . from the NASDAQ **stock market®** before the opening . . .

. . . to exit troubled **major stock market** and banking ventures . . .
. . . to exit troubled **federal government** and banking ventures . . .

. . . out of the **national market system** and money going . . .
. . . out of the **market** and money going . . .

Results for **stock market**

market
stock exchange
national market
national market system
stock market ®
exchange
major stock market

Results are tabulated
and ranked

Then search for "new middles" that fit the same contextual "signature".
Thousands of new middles can be found, and of course, multiple new
middles can be found for each signature.

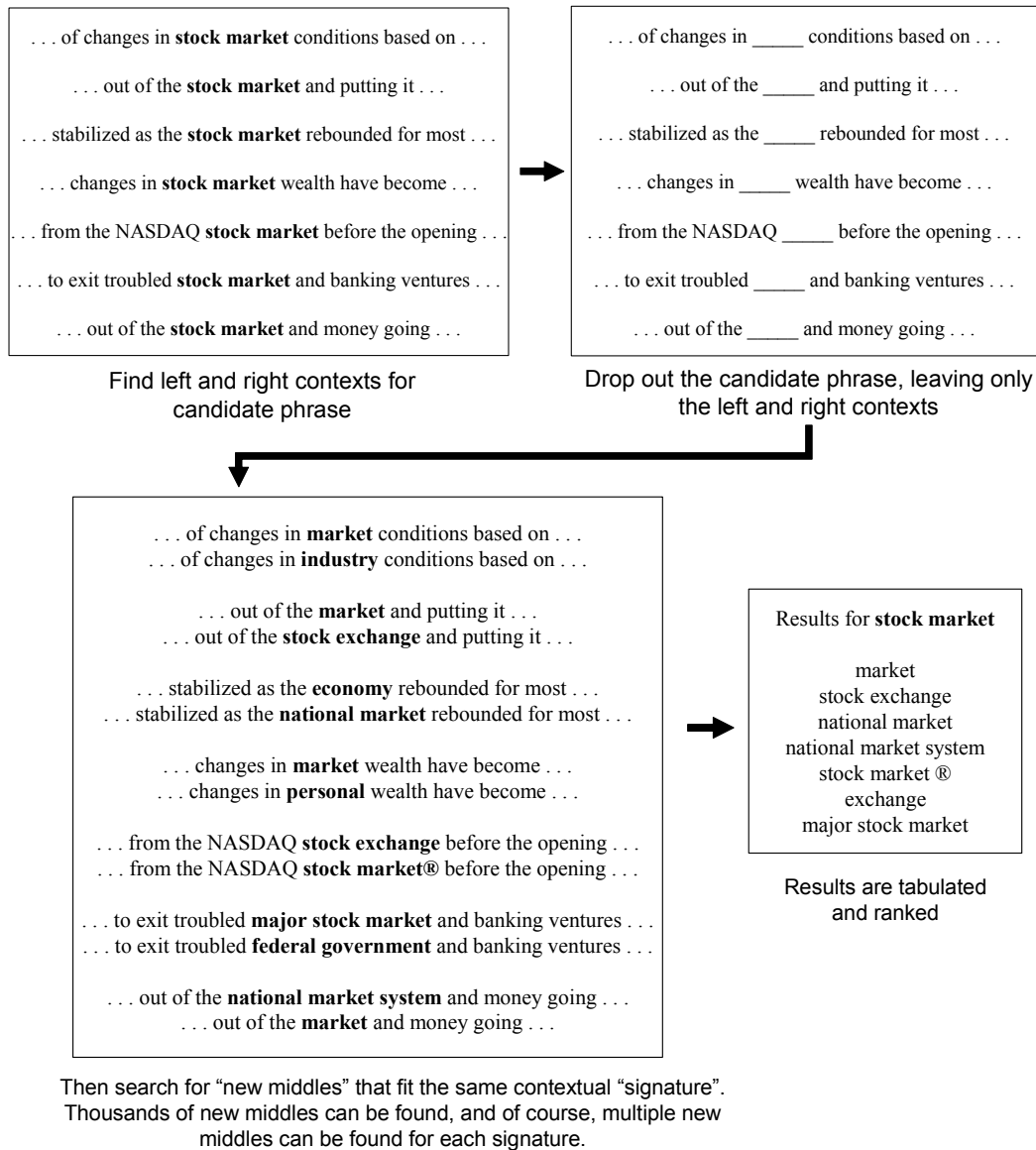**Table 1. Other examples of synonyms and near-synonyms found by this method**

Term	Result
terrorist organization	terrorist network / terrorist group / militant group / terror network
conference	meeting / symposium / convention / briefing / workshop
bin laden	bin ladin / bin-laden / osama bin laden / usama bin laden
nation's largest	country's largest / nation's biggest / nation's leading
watchful eye	direct supervision / close watch / stewardship / able leadership
it is safe to say	it's fair to say / it is important to note / you will find / I can say it is important to recognize / it is well known / it is obvious

Notes: (1) The scores reported in this paper do not reflect the use of the process described above (i.e., supplementing source words and phrases with their synonyms or near-synonyms); (2) other uses of synonyms or near-synonyms in either source or target are beyond the scope of this paper.

3.5   Process 4: Edge Locking

The flooding plus overlap processes are most reliable in the interior of each sentence, as each part of the target translation is confirmed by multiple overlapping n-grams. However, the first word(s) and last word(s) in the sentence are confirmed by one or just a few overlapping n-grams. To obtain better confirmation, we find other source sentences where the starting n-gram from the original source sentence occurs in the interior of the newly found sentences. By performing flooding-style analysis with the original n-gram plus additional context words, we can assess whether the selected translation is consistent with the context, either confirming the top ranking candidate or potentially preferring an otherwise lower-ranking candidate if it satisfies more criteria. The same process is applied to the n-grams at the end of sentences (the right edge) and can be applied to natural breaks within a sentence, i.e., n-grams surrounding commas or other interior punctuation. In this way, we achieve confirmation for "edges" that is close-to-comparable to that of the interior of the sentence.

4   CBMT Results on the BLEU Metric

CBMT was scored on newswire Spanish texts, using the NIST BLEU scoring package, and following the NIST procedure. Since NIST evaluations are for Arabic and Chinese, and the CBMT prototype is for Spanish, we ran our own test with held-out texts (different from development sets) and four reference translations for each sentence. [2] For comparison to other MT systems in Spanish-to-English, we ran the exact same test for the Web-available versions of SDL and SYSTRAN. (Caveat: It is possible that these companies have

somewhat newer internal R&D or other versions of their systems.)

▪ CBMT scored 0.6462 in June 2006 on newswire text on a Spanish-to-English system prototype using still incomplete resources (i.e., v1.0 of the dictionary and only 52GB of indexed target text). On the same test set, SDL scored 0.5610 and SYSTRAN scored 0.5551.

▪ In development testing: CBMT scored 0.6950 on a development set when expanded language resources were simulated: (1) a larger target corpus by online Web crawls seeded by dictionary translations of the source words, and (2) a dictionary that contains all source words (not tailored to the development set, just adding out-of-vocabulary (OOV) words).

▪ The best published results for Arabic-English and Chinese-English MT on the same newswire genre are Google's 2005 NIST scores: 0.5137 (Arabic-English) and 0.3531 (Chinese-English).

Of course, we are well aware that comparing Arabic or Chinese MT to Spanish, even for texts in the same genre, is not an apples-to-apples comparison. Translation from those languages into English is more challenging than from another European language into English. Nevertheless, CBMT, even in its current prototype form, outperformed others: commercial Spanish-to-English, and research Arabic- and Chinese-to-English, which demonstrates the power of the new CBMT technology. An aspect of CBMT is that the more context-dependent and less syntax-dependent the language, the greater advantage CBMT has over other approaches, especially over Rule-Based MT.

An interesting question is how much better can CBMT get? Can it reach human-quality translation, at least with respect to the BLEU score? We can address that question partially by simulating a complete dictionary and a larger target-language corpus. For the former, we simply added OOV words from the test corpus to the bilingual dictionary, without letting the dictionary builders see the test corpus in order to avoid biasing for the right translation in that context. For the latter, we crawled the Web with the possible dictionary

---

[2] The Spanish test set consists of 89 newswire sentences from diverse sources (shortest sentence: 8 words; average: 27 words; longest: 66 words). The Spanish sentences and their four English reference translations may be obtained gratis by e-mailing: mike [-at-] meaningfulmachines.com.

translations of the source words in the text. This simulates having a larger corpus (as the Web is exactly that), though it is not practical to do large numbers of targeted Web crawls for each new text to translate in an operational setting. This resource-enhancement process raised our BLEU score to 0.6950, demonstrating that there is room to improve by resource accretion. Although this "non-blind" result did have the benefit of more complete resources, it did not simulate phrase inclusion in the dictionary, which is anticipated to have a sig-

nificant positive impact. As for other language pair testing, preliminary results on small-scale Arabic and French development systems are positive.

How about algorithmic improvements? We offer the figure below, showing how our BLEU scores have been climbing during our R&D process. Of course, extrapolation is unreliable, but we are striving to reach the lower reaches of the human performance range as measured by BLEU.

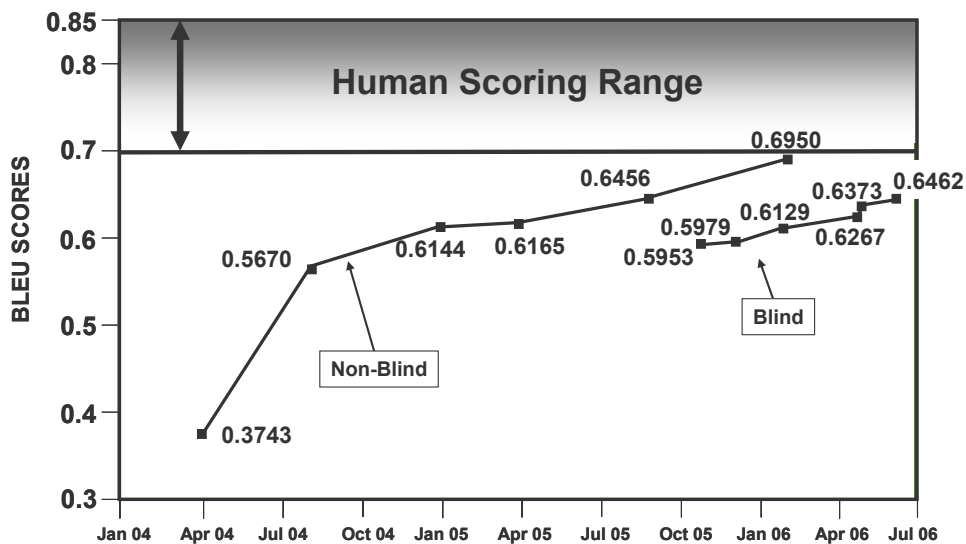**Figure 4. CBMT Scoring Over Time**

Figure 4 Notes: (1) Tests were run on newswire text. (2) Tests used four reference sentences. (3) Non-Blind tests were run on a development set and under conditions as described in section 4. (4) Blind tests were conducted under standard testing conditions. (5) Blind tests were run on an incomplete set of resources, as described above. (6) The Blind test scoring 0.6462 used an indexed target corpus of 52GB; the Blind test scoring 0.6373 used 42GB; other Blind tests used 30GB. The improvements from 0.6267 to 0.6462 in Blind tests were primarily due to corpus size increases, however, bug-fixing was also a factor. The earlier Blind test improvements from 0.5953 to 0.6267 were due to bug fixes/algorithmic improvements. (7) Non-Blind test improvements are largely due to algorithmic development. (8) We scored each human reference set against the other 3 references plus an additional human set so that each was scored against 4 references total. The human scores ranged from 0.7172 to 0.7941.

5 Related Work

Traditional Rule-Based MT systems are comprised of an analysis phase (typically a string-to-tree parser), a transfer phase (typically a rule-based tree-to-tree transformer), and a synthesis phase (typically a tree-to-linear-string generator). Hut-

chins (1986) and Nagao (1989) give excellent overviews. In contrast, interlingua-based systems perform multi-phase analysis and synthesis, but reduce or eliminate the need for a transfer component (Uchida and Zhu, 1993; Carbonell et al., 1994; Mitamura et al., 1994). Both methods rely on extensive human knowledge engineering in all phases. CBMT is radically different.

Corpus-based systems, such as Example-Based MT (Nagao, 1984; Sumita and Iida, 1991; Brown et al., 2003; Kim et al., 2005; Doi et al., 2005) and Statistical MT (Brown et al., 1990; Yamada and Knight, 2002; Och, 2005) are comprised of a translation model and a target language model. The two components are readily evident, for instance, in the original IBM Candide system (Brown et al., 1990), where the central equation is:

$$T_{opt} = Arg \max_{T_i \in Language_T} [P(S \mid T_i) P(T_i)]$$

In other words, the optimal target language sentence is the one that maximizes the product of the translation model (the first probability) and the target language model (the second probability). The decoder is the process that estimates T_{opt}. The translation model is typically trained from a large (e.g., 100MB) sentence-aligned parallel corpus of professionally translated text. The target language model is typically trained from a much larger monolingual corpus.

CBMT is also a corpus-based approach, closest to EBMT, but radically different in terms of requiring no parallel text whatsoever. In a sense, it is reminiscent of the old "Shake and Bake" idea (Whitelock, 1992), the newer EXERGE method in generation-heavy MT (Habash, 2003) and also the METIS work (Dirix et al., 2005) at a very abstract level of letting the target language establish lexical order. It bears some commonality to the work of Brown et al. (Brown et al., 2003) in that it permits combination of lattice entries with overlap (although the relevant CBMT components are covered in patent applications filed in 2001). However, it differs greatly from all previous systems in the maximal overlap principle for decoding and confirmation, in using consistently long n-grams, in near-synonym phrase substitutability, and in requiring no parallel text whatsoever to translate.

Appendix A: Sample CBMT Translations

As is typical of all MT systems, CBMT produces some good translations and some not-so-good. We illustrate a few good examples, which are not atypical (with Web-based SYSTRAN as contrast), drawn from newswire text relating to the Middle East. CBMT does not yet generate true casing.

Example 1 Input: *Un coche bomba estalla junto a una comisaría de policía en Bagdad*
- **CBMT**: a car bomb explodes next to a police station in baghdad
- **SYSTRAN**: A car pump explodes next to a police station of police in Bagdad

Example 2 Input: *Hamas anunció este jueves el fin de su cese del fuego con Israel*
- **CBMT**: hamas announced thursday the end of the cease fire with israel
- **SYSTRAN**: Hamas announced east Thursday the aim of its cease-fire with Israel

Example 3 Input: *Un soldado de Estados Unidos murió y otros dos resultaron heridos este lunes por el estallido de un artefacto explosivo improvisado en el centro de Bagdad, dijeron funcionarios militares estadounidenses*
- **CBMT**: a united states soldier died and two others were injured monday by the explosion of an improvised explosive device in the heart of baghdad, american military officials said
- **SYSTRAN**: A soldier of the wounded United States died and other two were east Monday by the outbreak from an improvised explosive device in the center of Bagdad, said American military civil employees

Acknowledgements

References

Eli Abir, Steve Klein, David Miller, and Michael Steinbaum (2002). Fluent Machines' EliMT System. In *Fifth Conference of the Association for Machine Translation in the Americas*, *AMTA-2002*, Tiburon, California, USA, pp. 216-219.

Regina Barzilay and Kathleen McKeown (2001). Extracting Paraphrases from a Parallel Corpus. In *Proceedings of the 39th Annual Meeting of the*

Association for Computational Linguistics, Toulouse, France, pp. 50-57.

P. Brown, J. Cocke, S. Della Pietra, V. Della Pietra, F. Jelinek, J. Lafferty, R. Mercer, and P. Roossin (1990). A Statistical Approach to Machine Translation. *Computational Linguistics*, 16(2), pp.79-85.

Ralf D. Brown, Rebecca Hutchinson, Paul N. Bennett, Jaime G. Carbonell, and Peter Jansen (2003). Reducing Boundary Friction Using Translation-Fragment Overlap. In *MT Summit IX, Proceedings of the Ninth Machine Translation Summit,* New Orleans, USA, pp.24-31.

Chris Callison-Burch, Philipp Koehn and Miles Osborne (2006). Improved Statistical Machine Translation Using Paraphrases. In *Proceedings NAACL-2006*, New York, USA.

Jaime Carbonell, David Farwell, Robert Frederking, Steven Helmreich, Eduard Hovy, Kevin Knight, Lori Levin, and Sergei Nirenburg (1994). The Pangloss MT System. *Proceedings of the First Conference of the Association for Machine Translation in the Americas,* Columbia, Maryland, USA, pp.240-241.

Peter Dirix, Ineke Schuurman, and Vincent Vandeghinste (2005). METIS-II: Example-Based Machine Translation Using Monolingual Corpora - System Description. In *Proceedings of MT Summit X, Workshop on Example-Based Machine Translation,* Phuket, Thailand, pp.43-50.

Takao Doi, Hirofumi Yamamoto, and Eiichiro Sumita (2005). Graph-Based Retrieval for Example-Based Machine Translation Using Edit Distance. In *Proceedings of MT Summit X, Workshop on Example-Based Machine Translation*, Phuket, Thailand, pp.51-58.

Nizar Habash (2003). Matador: A Large-Scale Spanish-English GHMT System. In *MT Summit IX, Proceedings of the Ninth Machine Translation Summit*, New Orleans, USA, pp.149-156.

W. John Hutchins (1986). *Machine Translation: Past, Present and Future,* Ellis Horwood Ltd., Chichester, U.K., (Halsted Press, New York).

Jae Dong Kim, Ralf D. Brown, Peter J. Jansen, and Jaime G. Carbonell (2005). Symmetric Probabilistic Alignment for Example-Based Translation. In *Proceedings of the 10th Workshop of the European Association for Machine Translation (EAMT-2005),* Budapest, Hungary, pp.153-159.

Dekang Lin, Shaojun Zhao, Lijuan Qin and Ming Zhou (2003). Identifying Synonyms among Distributionally Similar Words. In *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence (IJCAI-03)*, Acapulco, Mexico, pp.1492-1493.

Teruko Mitamura, Eric Nyberg, and Jaime Carbonell (1994). KANT: Knowledge-Based, Accurate, Natural Language Translation. *Proceedings of the First Conference of the Association for Machine Translation in the Americas*, Columbia, Maryland, USA, pp.232-233.

Makoto Nagao (1984). A Framework of a Mechanical Translation between Japanese and English by Analogy Principle. In A. Elithorn and R. Banerji, editors, *Artificial and Human Intelligence*. North-Holland: Elsevier Science Publishers, chap. 11, pp.173-180.

Makoto Nagao (1989). *Machine Translation: How Far Can it Go?* Oxford University Press, Oxford, U.K. Translated by Norman Cook.

Franz Josef Och (2005). Statistical Machine Translation: Foundations and Recent Advances. *MT Summit X*, Phuket, Thailand, Tutorial Note.

E. Sumita and H. Iida (1991). Experiments and Prospects of Example-Based Machine Translation. In *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*. Berkeley, California, USA, pp.185-192.

Hiroshi Uchida and Meiying Zhu (1993). Interlingua for Multilingual Machine Translation. *MT Summit IV*. Kobe, Japan, pp.157-169.

P. Whitelock (1992). Shake-and-Bake Translation. In *Proceedings of the 14th International Conference on Computational Linguistics (COLING 92),* Nantes, France, Volume II, pp.784-791.

Kenji Yamada and Kevin Knight (2002). A Decoder for Syntax-Based Statistical MT. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL-2002)*, Philadelphia, USA, pp.303-310.