A statistical machine translation (MT) system may include a noun phrase/prepositional phrase (NP/PP) translation subsystem to translate NP/PPs as a subtask in an MT operation. The NP/PP translation subsystem may use a model trained on an NP/PP corpus and a decoder to generate an n-best list of candidate translations and a re-ranker to re-rank the candidate translations based on a machine learning method and additional features based on known properties of NP/PPs.
OTHER PUBLICATIONS


Al-Omaizan et al., “Translating withScarce Resources,” 2000, 17th Nat’l Conference of the American Association for Artificial Intelligence, Austin, TX, pp. 672-678.


Yamamoto et al., "A Comparative Study on Translation Units for Bilingual Lexicon Extraction," 2001, Japan Academic Association for Copyright Clearance, Tokyo, Japan.


* cited by examiner
FIG. 2

- 205
- 210

- has
- decided
- to
- renounce
- any involvement in a treaty
- the Bush administration
Train NP/PP translation subsystem on NP/PP parallel corpus

MP system detects NP/PPs in foreign input sentence

Generates an n-best list of possible translations for NP/PPs

Rescore the n-best list using additional features

Provide best translation (or multiple best translations) to the MT system

MT system translates the remaining parts of the sentence and integrates the chosen NP/PP translations

FIG. 3
STATISTICAL NOUN PHRASE TRANSLATION

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims priority to U.S. Provisional Application Ser. No. 60/484,810, filed on Jul. 2, 2003, the disclosure of which is incorporated here by reference in its entirety.

ORIGIN OF INVENTION

The research and development described in this application were supported by DARPA under grant number N66001-00-1-8914. The U.S. Government may have certain rights in the claimed inventions.

BACKGROUND

Machine translation (MT) is the automatic translation from a first language (a “source” language) into another language (a “target” language). Systems that perform an MT process are said to “decode” the source language into the target language.

A statistical MT system that translates foreign language sentences, e.g., French, into English may have the following components: a language model that assigns a probability P(e) to any English string; a translation model that assigns a probability P(le) to any pair of English and French strings; and a decoder. The decoder may take a previously unseen sentence f and try to find the e that maximizes P(ef), or equivalently maximizes P(e)P(lef).

SUMMARY

A statistical machine translation (MT) system may include a noun phrase/prepositional phrase (NP/PP) translation subsystem to translate NP/PPs as a subtask in an MT operation.

The NP/PP translation subsystem may use a model trained on an NP/PP corpus and a decoder to generate an n-best list of candidate translations and a re-ranker to re-rank the candidate translations based on a machine learning method and additional features based on known properties of NP/PPs. In an embodiment, the machine learning method may be a maximum entropy learning method, and the features may include syntactic features and using the web as a language model.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a block diagram of a statistical machine translation (MT) system including a noun phrase/prepositional phrase (NP/PP) translation subsystem.

FIG. 2 illustrates a syntactic parse tree for a sentence including NP/PPs.

FIG. 3 is a flowchart describing a process for training and using an NP/PP translation subsystem.

FIG. 4 is a block diagram of an NP/PP translation subsystem according to an embodiment.

DETAILED DESCRIPTION

FIG. 1 illustrates a statistical machine translation (MT) system according to an embodiment. The MT system 100 may be used to translate from a source language (e.g., French) to a target language (e.g., English). The MT system 100 may include a language model 105, a translation model 110, and a decoder 115.

The MT system 100 may be based on a source-channel model. The language model (the “source”) may assign a probability P(e) to any given English sentence e. The language model 105 may be an n-gram model trained by a large monolingual corpus (or one side of a parallel corpus) to determine the probability of a word sequence.

The translation model 110 may be used to determine the probability of correctness for a translation, e.g., the probability P(lef) of a French string f, given an English string e. The parameter values for computing P(lef) may be learned from a parallel corpus including bilingual sentence pairs. The translation model may be, for example, an IBM translation model 4, described in U.S. Pat. No. 5,477,451.

The decoder 115 may be used to identify the best translation by maximizing the product of P(ef)P(lef). In an embodiment, the decoder may be a greedy decoder, such as that described in co-pending application Ser. No. 09/854,327, filed on May 11, 2001, incorporated herein in its entirety.

In an embodiment, the translation of noun phrases (NPs) and prepositional phrases (PPs) may be performed as a subtask of an MT operation. The MT system 100 may include NP/PP translation subsystem 120. The NP/PP translation subsystem 120 may translate the NP/PPs in a sentence in a source language, e.g., German, that can be translated into NP/PPs in a target language, e.g., English. The NP/PP translation subsystem 120 may then provide the translations to the MT system 100. The MT system 100 may treat the translated NP/PP(s) as fixed translations, i.e., incorporate them into the full translation without changing them.

Certain source language NP/PPs may not translate to NP/PPs in the target language. The NP/PP translation subsystem 120 may treat such NP/PPs as special cases, and handle them separately.

FIG. 2 illustrates a syntactic parse tree 200 for a sentence s that includes NPs 205 and PP 210. In the NP/PP translation subsystem 120, the NP/PPs of the sentence s are the subtrees t, that contain at least one noun and no verb, and which are not part of a larger subtree that contains no verb.

FIG. 3 is a flowchart describing a process for training and using the NP/PP translation subsystem 120 in the MT system 100. The NP/PP translation subsystem may use the same modeling as the overall system (e.g., IBM Model 4), but be trained on only NP/PPs (block 305), e.g., an NP/PP corpus derived from a parallel corpus 125 used to train the MT system 100 (FIG. 1). The MT system 100 may detect NP/PPs in an input foreign language sentence (block 310) and provide them to the NP/PP translation subsystem 120 for translation.

The NP/PPs may be translated in isolation, e.g., without the assistance of sentence context.

FIG. 4 shows an NP/PP translation subsystem according to an embodiment. The NP/PP translation subsystem uses the model 105 and a decoder 107 to generate an n-best list of possible translations (block 315). The NP/PP translation subsystem may include a re-ranker 415 to re-score the n-best list using additional features 420 (block 320). The best translation (or multiple best translations) may then be passed on to the MT system 100 (block 325), which in turn translates the remaining parts of the sentence and integrates the chosen NP/PP translation(s) (block 330).

The NP/PP training corpus 410 may be created by extracting NP/PPs from a parallel corpus, e.g., the parallel corpus 125. The corpus 410 may be word-aligned using, e.g., the toolkit Giza++, which is described in Och and Ney, “Discriminative training and maximum entropy models for statistical machine translation,” Proceedings of ACL (2000).
sides of the parallel corpus may then be parsed with syntactic parsers. This may facilitate detection of the NP/PP's in an input sentence.

In an embodiment, the model 405 may be a phrase translation model that extracts its phrase translation table from word alignments generated by the Giza++ toolkit. Details of this model are described in co-pending application Ser. No. 10/402,350, filed Mar. 27, 2003, which is incorporated herein in its entirety.

The decoder 407 may be a beam search decoder, which employs hypothesis recombination and stores the search states in a search graph which can be mined with standard finite state machine methods for n-best lists. An exemplary beam search decoder is described in Ueffing et al., "Generation of word graphs in statistical machine translation," Proceedings of Empirical Methods in Natural Language Processing (EMNLP) (2002).

The re-ranker 415 may re-rank the n-best list provided by the decoder 407 using a maximum entropy learning method. The decoder 407 may annotate the n-best list of candidate translations with accuracy judgments based on the probability scores that the model assigns to each candidate translation. The initial features are the logarithm of probability scores, i.e., the language model score, the phrase translation score, and the restructuring (distortion) score. The task for the learning method is to find a probability distribution p(e|f) that indicates if the candidate translation e is an accurate translation of the input f. The decision rule to pick the best translation is

$$
e_{best} = \arg \max_{e} p(e|f)$$

Maximum entropy was chosen for its ability to deal with both real-valued and binary features. In alternative embodiments, other machine learning systems, such as support vector machines, may also be used.

The corpus 410 provides the empirical probability distribution by distributing the probability mass of the acceptable translations $e_{best}$ appears as $p(e_{best}|f) | e_{best} = \Sigma_{p(e|f) | e_{best}}$. If none of the candidate translations for a given input f is acceptable, the NP/PP translation subsystem may select the candidates translations that are closest to reference translations measured by minimum edit distance.

The maximum framework parameterizes the probability distribution as $p(e|f) = \exp \sum \lambda_{i} h_{i}(e,f)$ where $h_{i}$'s are the feature values and the $\lambda_{i}$'s are the feature weights. Since the NP/PP translation subsystem may have only a sample of the possible translations e for the given input f, the probability distribution may be normalized so that $\Sigma p(e|f) = 1$ for a sample $e_{i}$ of candidate translations.

Maximum entropy learning finds a set of feature values $\lambda_{i}$ so as to $\max \Sigma_{p(e|f) | e_{i}} \Sigma_{p(e|f) | e_{best}}$. These expectations are computed as sums over all candidate translations e for all inputs f,$\Sigma p(e|f) | e_{i} = \Sigma p(e|f) | e_{best}$

The NP/PP translation subsystem may exploit certain properties of NP/PP translation in order to improve NP/PP translation. The first of these (compounding of words) is addressed by preprocessing, while the others motivate features which are used in n-best list re-ranking.

Compounding of words, especially nouns, is common in a number of languages (German, Dutch, Finnish, Greek), and may pose a problem for machine translation. For example, the German word "Aktionsplan" may not be known to the system, but if the word were broken up into "Aktion" and "Plan", the system could easily translate it into "action plan", or "plan for action". The issues for breaking up compounds include knowing the morphological rules for joining words, resolving ambiguities of breaking up a word (e.g., "Hauptsturm"—"Haupt-Turm" or "Haupt-Sturm"), and finding the right level of splitting granularity (e.g., "Frei"—"Tag" or "Freitag").

In an embodiment, the NP/PP translation subsystem may include a compound splitting module 450, such as that described in application Ser. No. 10/884,174, filed on Jul. 2, 2004, now U.S. Pat. No. 7,711,545, and entitled, "Empirical Methods for Splitting Compound Words With Application to Machine Translation". The compound splitting module may first collect frequency statistics over words in a training corpus. Compounds may be broken up only into known words in the corpus. For each potential compound the compound splitting module may check if morphological splitting rules allow the compound to be broken up into such known words. The compound splitting module may then pick a splitting option S (perhaps not breaking up the compound at all) with highest geometric mean of word frequencies of its n parts $p_i$.

$$S_{best} = \arg \max_{S} \left[ \prod_{i=1}^{n} \text{count}(p_i)^{\frac{1}{n}} \right]$$

The German side of both the training and testing corpus may be broken up in this manner. The model may be trained on a compound-split corpus, and the input broken up before being passed on to the system.

The compound splitting module may work well with a phrase-based translation model, which can recover more easily from too eager or too timid splits than word-based models. One of the features 420 used for re-ranking the n-best list may be web n-grams 452, in which the web may be used as a language model. Preliminary studies indicated that 30% of all 7-grams in new text can be also found on the web, as measured by consulting the search engine Google™, which currently indexes 3 billion web pages. This is only the case for 15% of 7-grams generated by the base translation system.

In an embodiment, the following binary features may be used: Does the candidate translation as a whole occur in the web? Do all n-grams in the candidate translation occur on the web? Do all n-grams in the candidate translation occur at least 10 times on the web? Both positive and negative features may be used for n-grams of the size 2 to 7.

In alternative embodiments the web may be integrated into the system by building a traditional n-gram language model, by using the web frequencies of the n-grams in a candidate translation, or by checking if all n-grams in a candidate translation occur on the web. Smoothing may be used to account for noise.

Syntactic features 454 may also be used for re-ranking. Syntactic features are computed over the syntactic parse trees of both input and candidate translation. The re-ranker may keep the syntactic parse tree inherited from the NP/PP detection process for the input NP/PPs. A part-of-speech tagger and syntactic parser may be employed to annotate the candidate translation with its most likely syntactic parse tree.

In an embodiment, the following syntactic features may be used: preservation of the number of nouns; preservation of prepositions; and with a base NP/PP the determiner generally agree in number with the final noun (e.g., not: "this nice green flowers"). The features may be realized as integers, e.g., how many nouns did not preserve their number during translation? These features encode relevant general syntactic knowledge about the translation of noun phrases, and constitute soft constraints that may be overruled by other components of the system.
In an experiment, the Europarl corpus was used as the training corpus. The Europarl corpus is derived from the European parliament proceedings and consists of 20 million words of German (available at http://www.isi.edu/publications/europarl/). In this corpus, only part of the NP/PPs are translated as such into the foreign language. In addition, the word-alignment and syntactic parses may be faulty. As a consequence, in an experiment, only 43.4% of all NP/PPs could be aligned initially. This number was raised to 67.2% using a number of automatic data cleaning steps. The NP/PPs that partially aligned were broken up, and systematic parse errors were fixed. Certain word types that were inconsistently tagged as nouns in the two languages were harmonized (e.g., the German “wo” and the English “today”). Because adverb+NP/PP constructions (e.g., “specifically this issue”) were inconsistently parsed, the adverb was always stripped from these constructions. In addition, German verbal adjective constructions were broken up if they involved arguments or adjuncts (e.g., “der von mir gegessene Kuchen”–“the by me eaten cake”), because this poses problems more related to verbal clauses. Also, alignment points involving punctuation were stripped from the word alignment, and punctuation was stripped from the edges of NP/PPs.

A total of 737,388 NP/PP pairs were collected from the Europarl corpus as training data. German NP/PPs that did not consistently align to NP/PPs in English were detected at this point. The NP/PP translation subsystem may use the obtained data of unaligned NP/PPs for dealing with these special cases.

To evaluate these methods, all of the 1362 NP/PPs in 534 sentences from parts of the Europarl corpus which are not already used as training data were detected. The evaluation metric was human assessment: Can the translation provided by the system be part of an acceptable translation of the whole sentence? In other words, the noun phrase has to be translated correctly given the sentence context. The NP/PPs were extracted in the same way that NP/PPs in English were initially detected for the acquisition of the NP/PP training corpus. As a result, there were some problems with parse errors, leading to sentence fragments extracted as NP/PPs that cannot be translated correctly. Also, the test corpus contained all detected NP/PP pairs, even untranslatable ones.

The performance of the NP/PP translation subsystem was evaluated on the set of 1362 NP/PPs extracted from 534 sentences. The contributions of different components of the system are displayed in Table 1. Starting from the IBM Model 4 baseline, gains were achieved using a phrase-based translation model (+5.5%), applying compound splitting to training and test data (+2.8%), re-estimating the weights for the system components using the maximum entropy re-ranking framework (+1.5%), adding word count features (+1.7%) and syntactic features (+0.8%). Overall an improvement of 12.3% was achieved over the baseline. Improvements of 2.5% are statistically significant given the size of our test corpus.

<table>
<thead>
<tr>
<th>System</th>
<th>NP/PP Correct</th>
<th>BLEU</th>
</tr>
</thead>
<tbody>
<tr>
<td>IBM Model 4</td>
<td>724</td>
<td>53.2%</td>
</tr>
<tr>
<td>Phrase Model</td>
<td>800</td>
<td>58.7%</td>
</tr>
<tr>
<td>Compound Splitting</td>
<td>838</td>
<td>61.5%</td>
</tr>
<tr>
<td>Re-Estimated Param.</td>
<td>858</td>
<td>63.0%</td>
</tr>
<tr>
<td>Web Count Features</td>
<td>881</td>
<td>64.7%</td>
</tr>
<tr>
<td>Syntactic Features</td>
<td>892</td>
<td>65.5%</td>
</tr>
</tbody>
</table>

Table 1 also provides scores for overall sentence translation quality. The chosen NP/PP translations are integrated into a general IBM Model 4 system that translates whole sentences. Performance is measured by the BLEU score, which measures similarity to a reference translation, and is described in Papineni et al., “BLEU: a method for automatic evaluation of machine translation.” Proceedings of the 40th Annual Meeting of the ACL (2002). As reference translation we used the English side of the parallel corpus. The BLEU scores track the improvements of our components, with an overall gain of 0.027.

According to exemplary embodiments, a machine translation system may include a machine-readable medium having executable instructions embodied thereon. The instructions are executable or otherwise operable to cause the machine translation system to perform a number of tasks. More specifically, the instructions may be executable to cause the machine translation system to detect a noun phrase and/or a prepositional phrase in an input string in a source language, generating a translation in a target language for the noun phrase and/or the prepositional phrase, and generating a translation of the input string in the target language including the translation for the noun phrase and/or the prepositional phrase. The instructions may also be executable to cause the machine translation system to generate a ranked n-best list of candidate translations and re-rank the n-best list using a machine learning method.

A number of embodiments have been described. Nevertheless, it will be understood that various modifications may be made without departing from the spirit and scope of the invention. Accordingly, other embodiments are within the scope of the following claims.

The invention claimed is:

1. A method comprising:
   - receiving by a machine translation system an input string in a source language;
   - detecting by the machine translation system a noun phrase and a prepositional phrase in the input string;
   - generating by the machine translation system possible translations in a target language for the noun phrase and the prepositional phrase;
   - generating a ranked n-best list of candidate translations for the possible translations of the noun phrase and the prepositional phrase;
   - generating by the machine translation system a translation of a remaining portion of the input string in the target language; and
   - integrating the one or more n-best translations for the noun phrase and the prepositional phrase into the translation of the remaining portion of the input string to generate one or more sentence translations.

2. The method of claim 1, wherein said generating the translation for the noun phrase and the prepositional phrase comprises generating the translation using a first method; and wherein said generating the translation of the input string comprises generating the translation using a second model different than the first model.

3. The method of claim 2, wherein the first model comprises a model trained using a corpus consisting of noun phrases and prepositional phrases.

4. The method of claim 2, wherein the first model comprises a phrase-based model.

5. The method of claim 1, wherein said generating possible translations for the noun phrase and the prepositional phrase comprises:
   - re-ranking the n-best list using a machine learning method.

6. The method of claim 5, wherein said machine learning method comprises a maximum entropy learning method.
7. The method of claim 5, wherein said re-ranking comprises re-ranking using one or more additional features based on known properties of noun phrases and prepositional phrases.

8. The method of claim 7, wherein the one or more additional features comprise syntactic features.

9. The method of claim 7, wherein the one or more additional features comprises using the web as a language model.

10. A machine translation system comprising:

a parser stored by a machine-readable medium and executable by a processor and stored in a memory to cause the machine translation system to detect a noun phrase and a prepositional phrase in an input string in a source language;

a translation subsystem including,

a first model stored by a machine-readable medium and executable by a processor and stored in a memory to cause the machine translation system to generate possible translations in a target language for the noun phrase and the prepositional phrase,

a first decoder stored by a machine-readable medium and executable to cause the machine translation system to generate a ranked n-best list including the possible translations, and

a selector to select a translation for the noun phrase and the prepositional phrase;

a second model stored by a machine-readable medium and executable to cause the machine translation system to generate possible translations in the target language for the input string in the source language, each of the possible translations comprising the selected translation for the noun phrase and the prepositional phrase; and

a second decoder to select one of the possible translations for the input string.

11. The machine translation system of claim 10, wherein the first model comprises a model trained using a corpus consisting of noun phrases and prepositional phrases.

12. The machine translation system of claim 10, wherein the first model comprises a phrase-based model.

13. The machine translation system of claim 10, wherein the translation subsystem further comprises:

a re-ranker to re-rank the n-best list using a machine learning method.

14. The machine translation system of claim 13, wherein the machine learning method comprises a maximum entropy learning method.

15. The machine translation system of claim 13, wherein the re-ranker is operative to re-rank the n-best list using one or more additional features based on known properties of noun phrases and prepositional phrases.

16. The machine translation system of claim 15, wherein the one or more additional features comprise syntactic features.

17. The machine translation system of claim 15, wherein the one or more additional features comprises using the web as a language model.

18. An article comprising a non-transitory machine-readable medium including machine-executable instructions, the instruction operative to cause a machine to:

detect a noun phrase and a prepositional phrase in an input string in a source language;
generate possible translations in a target language for the noun phrase and the prepositional phrase;
generate a ranked n-best list of candidate translations for the possible translations of the noun phrase and the prepositional phrase;
generate a translation of a remaining portion of input string in the target language; and
integrate the one or more n-best translations for the noun phrase and the prepositional phrase into the translation of the remaining portion of the input string to generate one or more sentence translations.

19. The article of claim 18, wherein the instructions for generating the translation for the noun phrase and the prepositional phrase comprise instructions operative to cause the machine to generate the translation using the first model; and
wherein the instructions for generating the translation of the input string comprise instructions operative to cause the machine to generate the translation using a second model.

20. The article of claim 19, wherein the first model comprises a model trained using a corpus consisting of noun phrases and prepositional phrases.

21. The article of claim 19, wherein the first model comprises a phrase-based model.

22. The article of claim 18, wherein the instructions for generating possible translations for the noun phrase and the prepositional phrase comprise instructions operative to cause the machine to:

re-rank the n-best list using a machine learning article.

23. The article of claim 22, wherein said machine learning article comprises a maximum entropy learning article.

24. The article of claim 22, wherein the instructions for re-ranking comprise instructions operative to cause the machine to re-rank using one or more additional features based on known properties of noun phrases and prepositional phrases.

25. The article of claim 24, wherein the one or more additional features comprise syntactic features.

26. The article of claim 24, wherein the one or more additional features comprises using the web as a language model.

* * * * *