A Formal Basis for Spoken Language Translation by Analogy

Keiko Horiguchi
D21 Laboratory
6-7-35 Kitashinagawa
Shinagawa-ku, Tokyo 141
Japan
keiko@pdp.crl.sony.co.jp

Alexander Franz
Sony Computer Science Laboratory &
D21 Laboratory
6-7-35 Kitashinagawa
Shinagawa-ku, Tokyo 141
Japan
amf@csl.sony.co.jp

Abstract

Since spoken language is characterized by a number of properties that defy interpretation and translation by purely grammar-based techniques, recent interest has turned to analogical (also known as case-based or example-based) approaches. In this framework, the most important step consists of robustly matching the recognized input expression with the stored examples. This paper presents a probabilistic formalization of analogical matching, and describes how this model is applied to speech translation in the framework of translation by analogy.

Figure 1: Example-based Translation Architecture

1 Introduction

The last decade has seen growing interest in the example-based framework for translation of written and spoken language (Nagao, 1984),(Jones, 1996). This approach, sometimes called analogical, case-based, or memory-based, originated with the following insight. In the course of translating an expression, a skilled human translator often recalls a similar translation that she has performed or studied before, and then carries out the new translation by analogy to the previous case, instead of applying a large number of lexical and grammatical rules in her head. In an example-based translation architecture, pairs of bilingual expressions are stored in the example database. The source language input expression is matched against the source language portion of each example pair, and the best matching example is selected. The system then returns the target language portion of the best example as output. This is illustrated in Figure 1.
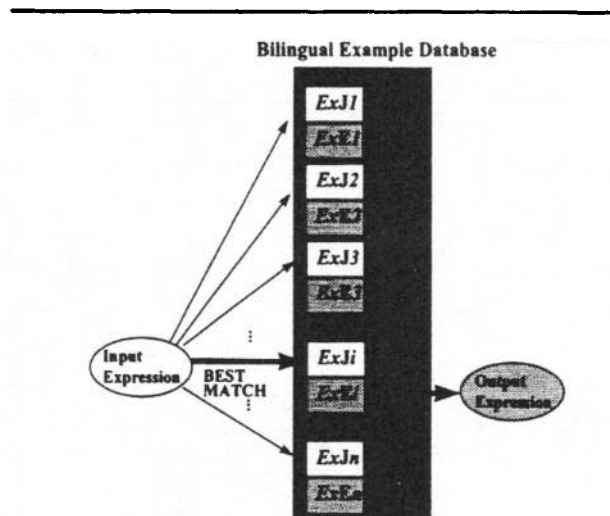
1.1 Pre-translation

The example-based approach has certain advantages over traditional rule-based approaches to translating spoken language. Since an analogical system relies on a database of pre-translated example pairs, it results in high translation quality. High translation quality requires not only that the output be grammatically correct, but also that the output sound natural and idiomatic. Spoken utterances consist of larger portions of fully-lexicalized or semi-lexicalized morpheme sequences, the use of which greatly contributes to sounding natural and native-like, but whose meanings are not totally predicatable from their forms (Pawley and Syder, 1983). An analogical system can generate natural-sounding output more easily than a compositional, rule-based system, because it directly uses the correspondences between source-language and target-language expressions.

32

1.2 Robustness

Another important requirement for spoken language translation is that the system has to be very robust. Spoken utterances contain a lot of disfluencies, such as pronunciation errors, word selection errors, word fragments and repairs. Furthermore, a speech translation module also has to handle the errors introduced by the speech recognition component. In an analogical system, the process that matches the input expression against examples can be very robust, and can always return the best matching output expressions instead of failing completely.

1.3 Improving Translation Quality

An additional requirement of an automatic translation system is that it should be possible to improve the translation quality by expending additional effort. In a traditional rule-based system, as the knowledge sources (such as grammar rules, semantic disambiguation rules, transfer rules, etc.) expand in size, there comes a point at which the complex interrelationships between the different types of information precludes any further improvement. In an analogical system, it is possible to incrementally improve the translation quality by adding more examples to the example database, and by effecting corresponding improvements in the matching function by e.g. refining the thesaurus or re-estimating word similarities from an expanded bilingual corpus.

1.4 The Problem of Scalability

Unfortunately, the pure analogical approach lacks scalability. The effort required to acquire and maintain the example database, the cost of the space required to store the examples, and the cost of the time required to search the database can become prohibitively high, since a pure analogical system requires a separate example for every linguistic variation.

2 A Hybrid Analogical Approach

Since language is productive, a realistic analogical system needs to be able to handle linguistic constructions that do not have an exact match in the example database. Therefore it is important for a system to be able to combine fragments from more than one example expression to cover the input expression.

To meet this requirement, we have designed an architecture for robust, practical translation of spoken language in limited domains that integrates morphological and syntactic linguistic processing with an analogical transfer component. The overall system is described briefly in this section.

2.1 System Architecture

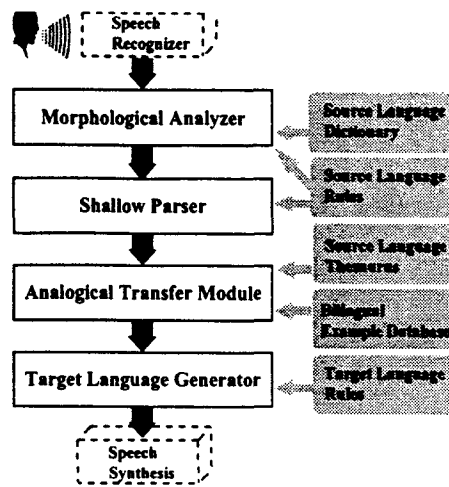The pipelined system architecture, shown in Figure 2, separates speech recognition, morphological

Figure 2: System Architecture

analysis, shallow parsing, and recursive analogical translation into different modules. This separation of general linguistic, domain, and transfer knowledge improves portability and scalability of the system.

2.2 Shallow Source Language Analysis

The purpose of the shallow analysis component is to identify clauses and phrases, to identify modifying relations as long as they are unambiguous (deriving a canonical interpretation in ambiguous cases), and to convert some surface variations into features.

In our prototype implementation, an adapted version of the JUMAN 3.1 Japanese morphological analyzer (Kurohashi et al., 1994) is used for Part-of-Speech disambiguation, and for dictionary and thesaurus look-up.

The second step of source language analysis is carried out using an augmented context-free grammar for the NLYACC parser (Ishii, Ohta, and Saito, 1994), which is an implementation of the Generalized LR parsing parsing algorithm (Tomita, 1985).

The shallow analysis module returns a shallow syntactic parse tree with various lexical and syntactic features. It is robust enough to tolerate extragrammaticalities, disfluencies, and the like in the input.

2.3 Analogical Transfer

The recursive analogical transfer module matches the input shallow syntactic tree against the source language portions of example shallow syntactic trees. The example data is classified into different linguistic constituent levels, such as clause-level examples, phrase-level examples, and word-level examples. The system tries to match the input against examples of the largest unit.

Once the system finds the best matching examples of the largest unit, it checks whether there are portions that differ significantly between the input

and the example. If so, the system performs the analogical matching process again on the identified portion from the input, using examples of the corresponding smaller unit. This recursive process continues until all parts have been matched. Finally, the target language portions of the selected best matches are combined to form the complete target language expression.

The analogical matching step is based on a probabilistic formalization of matching by analogy. The details of this model are described in Section 4.

2.4 Target Language Generation

The target language generation module is designed to perform a number of linguistic operations, such as enforcing subject-verb agreement, ensuring that required definiteness information is present (such as English determiners, quantifiers, or possessives), and generating the appropriate inflectional morphology. In our prototype implementation, we are using the PC-KIMMO system for generating English morphology (Antworth, 1990). After these operations, the shallow syntactic tree is linearized to create an expression in the target language.

2.5 Speech Synthesis

In the final step, spoken output is generated from the target language expression. In our Japanese-English prototype, this step is carried out by the DECTALK system (Hallahan, 1996).

3 Advantages of the Hybrid Analogical Approach

The hybrid approach combines analogical matching and transfer with a rule-based component that accounts for one of the fundamental properties of language: Its productiveness. This section describes what we perceive to be the main advantages of the hybrid analogical approach to speech translation.

3.1 Modular, Natural Knowledge Sources

The system architecture separates general linguistic knowledge, domain knowledge, and transfer knowledge. This means it is easier to port it to different domains, and to apply it to new languages.

We also consider the knowledge sources to be "natural". By this we mean that, from the point of view of knowledge representation, each knowledge source captures certain aspects of the translation process in its most natural form. For example, the example data base captures translation correspondences in a natural way – by means of corresponding natural language expressions in the souce and target language. Other, less natural means of knowledge representation would require significantly more effort to acquire and maintain. As a result, it is easier to improve the translation quality by adding and

modifying examples, and by modifying the thesaurus (if necessary).

3.2 Examples vs. Syntactic or Semantic Grammars

As described above, analogical translation relies on a database of example pairs which can encode idiomatic translation correspondences at the lexical, phrasal, and clausal manner in a natural way. This is an improvement over previous approaches which rely on syntactic or semantic grammars.

For example, the "transfer-driven" approach of (Sobashima et al., 1994) relies on essentially syntactically-based analysis and transfer rules that are manually annotated with examples by providing a sound formal basis for analogical matching. This requires an extensive effort to create a body of rules that covers all possible expressions, and which can handle extragrammatical or disfluent input. As an example of a semantic-grammar based approach, "concept-based" translation (Mayfield et al., 1995) requires an extensive manual knowledge acquisition effort to create detailed, domain and task-specific templates and semantic grammars. In addition, a heavily semantics-based approach such as this work suffers from a lack of generality due to the absence of linguistic processing.

3.3 Examples vs. Interlingua

The framework of Interlingua-based translation rests on the presupposition that there can be a universal, unambiguous, language-neutral, and practically (if not formally) sound knowledge representation formalism to mediate between source and target languages. In practice, defining, maintaining, and extending such a formalism for multiple, not closely related languages has proved to be a major challenge. Analogical speech translation does not rely on this presupposition, and instead seeks to capture intuitive translation correspondences.

3.4 Syntactic and Lexical Distance

In the hybrid analogical approach, the example data is categorized by linguistic constituent. For example, there are translation example pairs at the clause level, phrase level, and word level. This yields a more efficient search procedure during the matching process, while only assuming non-controversial notions of syntactic constituency. By treating syntactically similarity and semantic similarity as two separate aspects of the matching process, we derive an improvement over methods that combine these two aspects. For example, (Sato and Nagao, 1990) combine a measure of structural similarity with a measure of word distance in order to obtain the overall distance measure that is used for matching.

3.5 Computational Efficiency

Analogical translation relies on a large database of example pairs. This incurs a significant computational cost for searching and matching against all the examples, which is proportional to the number of examples multiplied by the average size of the representations of the examples. (In practice, this cost can be mitigated somewhat by clustering and indexing schemes for the example database.) Hybrid analogical translation greatly reduces the number of required examples by relying on the generality of linguistic rules.

Pure statistical machine translation (Brown et al., 1993) must in principle recover the most probable alignment out of all possible alignments between the input and a translation. While this approach is theoretically intriguing, it has yet to be shown to be computationally feasible in practice.

3.6 Linguistic Efficiency

In addition to computational efficiency, we also consider a factor that might be called "linguistic efficiency". We hold that a significant body of systematic linguistic regularities has been identified that must be accounted for somehow during the process of translation. Linguistic efficiency refers to the notion of how efficient the system is with regard to these regularities.

In hybrid analogical translation, the use of a morphological and syntactic module for shallow analysis to derive a linguistic representation with syntactic and lexical features allows us to handle phenomena such as inflections, transformations, and language-specific phenomena (such as the English determiner system and certain Japanese constructions that encode politeness information) in a linguistically efficient manner.

3.7 Translation Adequacy

In order to be able to provide stylistically and pragmatically adequate translations of spoken language, it is not sufficient to merely ignore or tolerate extra-grammaticalities in the input; in many cases, the information carried by such phenomena must be reflected in the target language output. The hybrid analogical approach is able to model such phenomena using probabilistic operators, which are explained in more detail in the next section.

4 A Probabilistic Model for Analogical Matching

When applied to spoken language, the central step in analogical translation is a robust matching step that compares the output of the speech recognition component with the contents of the example database. This section presents the probabilistic model that provides a formal basis for this matching step.
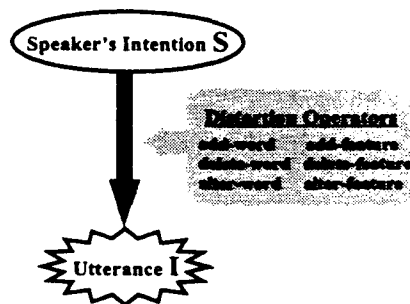

Figure 3: Viewing Input as Distorted Examples

4.1 Notation

Let I denote the input expression, consisting of a sequence of words along with certain features resulting from shallow parsing. Thus, an input expressions I consists of a sequence of words $iw_1, iw_2, \ldots, iw_n$, and a set of features $if_1, if_2, \ldots, if_m$. Similarly, let the source expression E of an example pair consist of $ew_1, ew_2, \ldots, ew_p$ and $ef_1, ef_2, \ldots, ef_q$.

4.2 The Noisy Channel Model

The "noisy-channel" model from information theory has proven highly effective in speech recognition and, more recently, in language understanding (Epstein et al., 1996; Miller et al., 1994). We adopt this model for translation by analogy in the following manner.

Given an input expression, the analogical matching algorithm must determine the example expression that is closest in meaning to the input expression. We denote the probability that an example expression is appropriate for translating some input as the conditional probability of the example, given the input:

(1) $P(\text{Example}|\text{Input})$

Our aim is to find the example that has the highest conditional probability of being appropriate to translate the given input. We denote that example with $E_{\max}$, where the **max** function chooses the example with the maximum conditional probability:

(2) $E_{\max} = \max_{E \in \text{Examples}} [P(E|I)]$

Our approach to determining $E_{\max}$ is as follows. First, we can use Bayes' Law to obtain a re-expression of the conditional probability that needs to be maximized:

(3) $P(\text{E}|\text{I}) = \dfrac{P(\text{E})P(\text{I}|\text{E})}{P(\text{I})}$

Since the input expression, and therefore $P(\text{I})$, remains constant over different examples, we can disregard the term $P(\text{I})$ in the denominator. Thus, we need to determine $E_{\max}$ which can be defined as follows:

35

(4) $E_{\max} = \max_{E \in \text{Examples}} [P(E)P(I|E)]$

The probability distribution over the examples $P(E)$ encodes the prior probability of using the different examples to translate expressions in the domain. It can be used to penalize certain specialized expressions that should be used less frequently. The conditional probability distribution is estimated using a "distortion" model of utterances that is described in the next section.

4.3 Viewing Input as Distorted Examples

The conditional probability distribution $P(I|E)$ is modeled as follows. Consider that the speaker intends to express an underlying message S, but speech errors, certain speech properties, misrecognitions, and other factors interfere, resulting in the actual utterance I, which forms the input to the translation system (Figure 3). This is modeled using a number "distortion" operators:

- **echo-word**(ew_i). This operator simply echoes the ith word, ew_i, from the example to the input.

- **delete-word**(ew_i). This operator deletes the ith word, ew_i, from the example.

- **add-word**(iw_j). This operator adds the jth word, iw_j, to the input.

- **alter-word**(ew_i, iw_j). This operator alters the ith word, ew_i, from the example to the jth word, iw_j, in the input expression. The altered word is different, but usually semantically somewhat similar.

- Corresponding operators for features.

Given these operators, we can view the input I as an example E to which a number of distortion operators have been applied. Thus, we can represent an input expression I as an example plus a set of distortion operators:

(5) $I = \{E, \mathbf{distort}_1, \ldots, \mathbf{distort}_z\}$

This means that we can re-express the conditional probability distribution for an input expression I, given that the meaning expressed by example E is intended, as follows:

(6) $P(I|E) = P(\{\mathbf{distort}_1, \ldots, \mathbf{distort}_z\}|E)$

4.4 Operator Independence Assumption

Two independence assumptions are required to make this model computationally feasible. For the first assumption, we assume that the individual distortion operators are conditionally independent, given the example E:

(7) $P(\mathbf{distort}_k|\mathbf{distort}_1, \ldots, \mathbf{distort}_z; E) \approx$
    $P(\mathbf{distort}_k|E)$

This means that we can make the following simplification:

(8) $P(\{\mathbf{distort}_1, \ldots, \mathbf{distort}_z\}|E) \approx$
    $$\prod_{k=1}^{z} P(\mathbf{distort}_k|E)$$

Thus, we obtain the following:

(9) $P(I|E) \approx \prod_{k=1}^{z} P(\mathbf{distort}_k|E)$

Considering the individual components of the example E, this leads us to the following.

(10) $P(I|E) \approx$
    $$\prod_{k=1}^{z} P(\mathbf{distort}_k|e_1, e_2, \ldots, e_p; ef_1, ef_2, \ldots, ef_q)$$

4.5 Operator Localization Assumption

For the second assumption, we make the assumption that the individual distortion operators only depend on the words and features that they directly involve. In effect, we stipulate that the operators only affect a strictly local portion of the input. For example, we assume that the probability of echoing a word depends only on the word itself, so that the following holds:

(11) $P(\mathbf{echo\text{-}word}(ew_i)|ew_{1,\ldots,p}; ef_{1,\ldots,q}) \approx$
    $P(\mathbf{echo\text{-}word}(ew_i))$

Similarly, we assume that the probability of e.g. deleting a feature depends only on the feature itself, so that the following holds:

(12) $P(\mathbf{delete\text{-}feature}(ef_i)|ew_{1,\ldots,p}; ef_{1,\ldots,q}) \approx$
    $P(\mathbf{delete\text{-}feature}(ef_i))$

This yields the following approximation:

(13) $P(I|E) \approx$
    $$\prod_{k=1}^{z} P(\mathbf{distort\text{-}word}_k(ew_i, iw_j))$$
    $$\prod_{l=1}^{y} P(\mathbf{distort\text{-}feature}_l(ef_i, if_j))$$

4.6 Computing the Match

Given an input I and an expression E. it is straight-forward to determine the probability of the feature distortion. since the features are indexed by name:

$$(14) \quad \prod_{l=1}^{y} P(\text{distort-feature}_l(\text{ef}_i, \text{if}_j))$$

In order to determine the probability of the word distortions. we must find the most probable set of distortion operators. Given an input and an example. there are many different sets of distortion operators that could relate the two. Of course, we are interested in the most straightforward relation between the two, which corresponds to the least cost or highest probability. To further complicate matters, there may not be a single unique set of distortion operators with a unique minimum cost (corresponding to a unique maximum probability); instead, there may be a number of distortion sets that all share the same minimal cost (and maximal probability). In this case, we are content to chose one of the minimal cost sets at random. This set is defined as follows:

$$(15) \quad \text{Distort}_{\text{max}} = \max_{Distort}[P(\text{Distort}|E, I)]$$

We solve this problem with a dynamic programming algorithm that finds a set of distortion operators with maximal probability. First, to obtain a distance measure, we take the negative logarithm of this expression:

$$(16) \quad -logP(\text{Distort}|E, I)$$

Given that we have assumed independence between individual distortion operators above, this can be simplified as follows:

$$(17) \quad -log \prod_{k=1}^{\text{no. of operators}} P(\text{distort}_k|E, I)$$

We have also assumed that the distortion operators are independent of the part of the sentence that does not directly involve them. Thus, we can simplify further as follows:

$$(18) \quad -log \prod_{k=1}^{x} P(\text{distort}_k|\text{ew}_i, \text{iw}_j)$$

This can be further split into the individual distortion operators:

$$(19) \quad \sum_{k=1}^{x} -logP(\text{distort}_k(\text{ew}_i, \text{iw}_j))$$

This corresponds directly to the individual costs that we use for the dynamic programming equation. Let the example expression be $E = e_1, e_2, \ldots, e_p$ and and the input expression be $I = i_1, i_2, \ldots, i_n$. Then, let $D(p, n)$ be the distance between the example and the input. This distance is defined by the following recurrence:

$$D(p,n) = min \begin{cases} D(p-1,n-1) - logP(\text{echo}(\text{ew}_p)) \\ D(p,n-1) \quad -logP(\text{add}(\text{iw}_n)) \\ D(p-1,n) \quad -logP(\text{delete}(\text{ew}_p)) \\ D(p-1,n-1) - logP(\text{alter}(\text{ew}_p, \text{iw}_n)) \end{cases}$$

The result of this is the optimal alignment between the input and the example. as well as the minimum distance between them. The matcher selects the example with the smallest distance to the input, and assembles the target language portions of the selected example pairs to form a complete translation in the target language.

5 Probability Estimation

The method for speech translation by analogy described in this paper was designed to overcome the manual knowledge acquisition bottleneck by relying on techniques from symbolic and statistical machine learning, while still allowing the kind of manual tuning that is necessary to produce high-quality translations.

5.1 Prior Probability Distribution

The prior probability distribution over the example database $P(\text{Examples})$ is used to penalize highly specialized example pairs that should be used less often. After an initial distribution is estimated, these probabilities can be adjusted to solve translation problems due to idiosyncratic examples.

5.2 Alteration Probability Distribution

The two distortion operators **alter-word** and **alter-feature** perform the function of matching semantically similar words or feature values. If a monolingual or bilingual corpus from the application domain is available, these probability distributions can be estimated using iterative methods. If neither type of corpus is available, the probabilities can be estimated with the aid of a manually-constructed thesaurus.

5.3 Thesaurus-based Estimation

A thesaurus is a semantic IA-A hierarchy whose nodes are semantic categories, and whose leaves are words. The traditional method of estimating word similarity, based on counting IS-A links, presupposes that every link encodes equal semantic distance – but in practice, this is never the case (Resnick, 1995). Thus, we adopt a new method for judging semantic distance between two words. If appropriate distributional information for words is available, then the semantic similarity of two words could be estimated from the entropy of their lowest common dominating node *lcdn*:

37

$$(20) \quad \text{Similarity}(ew_i, iw_j) \approx \frac{\text{Entropy}(Root\ Node)}{\text{Entropy}(lcdn)}$$

In the absence of distributional information, the entropy of a node depends only on the number of words that the node dominates.

5.4 Other Distortion Probabilities

The probability distributions for adding and deleting words and features can also be estimated from corpora, if available. Since there are very few Japanese spoken language corpora available, we are currently adopting a word-class based model for the remaining distributions that uses the categories for "strong content words" (nouns and verbs), "light content words" (adjectives and some adverbs), "grammatical function words" (e.g. particles and conjunctions), and "modifiers and adjuncts". In addition, it is possible to assign specific lexical penalties to individual words.

5.5 Learning Example Pairs

Since the contents of the database is central to achieving high-quality translations, it is usually necessary to adjust it manually in response to errors in the translation. At the same time, since the example database must be adapted for every new domain, it is important to minimize the amount of manual effort. For this reason, the example database was designed in such a way that it is possible to acquire new examples by a semi-automatic method consisting of an automatic extraction step from a bilingual corpus (see, for example, (Watanabe, 1993)), followed by a manual filtering and refinement step.

6 Conclusions and Further Work

Based on the probabilistic model of analogical matching, we have implemented a prototype that translates from Japanese to English in a limited domain. The application domain for the prototype are expressions for traveling in a foreign country, such as expressions related to making reservations or dining in a restaurant.

The initial results from our prototype are very promising, but extension of system coverage and subsequent large-scale evaluation is needed. Additional topics for further work include the estimation of lexical probabilities from bilingual corpora; improving the integration between the speech recognition and the translation components in order to increase recognition accuracy and translation robustness; and extending the system to additional languages.

References

Antworth, Evan L. 1990. *PC-KIMMO: a two-level processor for morphological analysis.* Number 16 in Occasional Publications in Academic Computing. Summer Institute of Linguistics, Dallas,TX.

Brown, Peter, Stephen A. Della Pietra, Vincent J. Della Pietra, and Robert L. Mercer. 1993. The mathematics of statistical machine transalation: Parameter estimation. *Computational Linguistics*, 19(2):263-312.

Epstein, M., K. Papieni, S. Roukos, T. Ward, and S. Della Pietra. 1996. Statistical natural language understanding using hidden clumpings. In *ICASSP-96*, pages 176-179, Atlanta, GA.

Furuse, Osame and Hitoshi Iida. 1992. An example-based method for transfer-driven machine translation. In *Proceedings for the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation (TMI-92)*, pages 139-150, Montreal, Canada, June.

Furuse, Osamu and Hitoshi Iida. 1996. Incremental translation utilizing constituent boundary patterns. In *COLING-96*, pages 412-417, Copenhagen, Denmark.

Hallahan, W. 1996. DECtalk software: Text-to-speech technology and implementation. *Digital Technical Journal*, 7(4), March.

Ishii, M., K. Ohta, and H. Saito. 1994. An efficient parser generator for natural language. In *COLING-94*, pages 417-420, Kyoto, Japan.

Jones, Daniel. 1996. *Analogical Natural Language Processing.* UCL Press, London.

Kurohashi, S., T. Nakamura, Y. Matsumoto, and M. Nagao. 1994. Improvements of Japanese morphological analyzer JUMAN. In *Proceedings of the International Workshop on Sharable Natural Language Resources*, pages 417-420, Nara, Japan.

Mayfield, L., M. Gavalda, W. Ward, and A. Waibel. 1995. Concept-based speech translation. In *ICASSP-95*, pages 97-100, Detroit, MI.

McKevitt, Paul, editor. 1994. *Workshop on Integration of Natural Language and Speech Processing.* Seattle, WA. AAAI-94.

Miller, S., R. Bobrow, R. Ingria, and R. Schwartz. 1994. Hidden understanding models of natural language. In *ACL-32*, pages 25-32, Las Cruces, NM.

Morimoto, Tsuyoshi, Masami Suzuki, Toshiyuki Takezawa, Gen'ichiro Kikui, Massaki Nagata, and Mutsuko Tomokiyo. 1992. A spoken language translation system: Sl-trans2. In *Proceedings of the fifteenth International Conference on Computational Linguistics*, volume 2, pages 1048-1052, Nantes.

Nagao, M. 1984. A framework of a Machine Translation between Japanese and English by analogy principle. In A. Elithorn and R. Banerji, editors, *Artificial and Human Intelligence*, pages 173–180. North-Holland.

Pawley, Andrew and Frances Hodgetts Syder. 1983. Two puzzles for linguistic theory: nativelike selection and nativelike fluency. In Jack C. Richards and Richard W. Schmidt, editors, *Language and Communication*, pages 191–227. Longman.

Price, Patti. 1994. Combining linguistic with statistical methods in automatic speech understanding. In *Proceedings of the workshop-The Balancing Act: Combining symbolic and statistical approaches to language*, pages 76–83, Las Cruces, New Mexico.

Rayner, M., D. Carter, P. Price, and B. Lyberg. 1994. Estimating performance of pipelined spoken language translation systems. In *ICSLP-94*, pages 1251–1254, Yokohama, Japan.

Resnick, P. 1995. Using information content to evaluate semantic similarity in a taxonomy. In *IJCAI-95*.

Ringger, Eric K. and James F. Allen. 1996. A fertility channel model for post-correction of continuous speech recognition. In *ICSLP-96*, pages 897–900, Philadelphia, PA.

Sato, S. and M. Nagao. 1990. Towards memory-based translation. In *COLING-90*, pages 247–252, Helsinki, Finland.

Shirotsuka, O. and K. Murakami. 1994. An example-based approach to semantic information extraction from Japanese spontaneous speech. In *ICSLP-94*, pages 91–94, Yokohama, Japan.

Sobashima, Y., O. Furuse, S. Akamine, J. Kawai, and H. Iida. 1994. A bidirectional, transfer-driven machine translation system for spoken dialogues. In *COLING-94*, pages 64–68, Kyoto, Japan.

Stemberger, Joseph P. 1982. Syntactic errors in speech. *Journal of Psycholinguistic Rsearch*, 11(4):313–333.

Tomita, Masaru. 1985. *Efficient Parsing for Natural Language*. Kluwer Academic Publishing, Boston, MA.

Watanabe, H. 1992. Similarity-driven transfer system. In *COLING-92*, pages 770–776.

Watanabe, Hideo. 1993. A method for extracting translation patterns from translation examples. In *TMI-93*, pages 292–301.