Why Corpus-Based Statistics-Oriented Machine Translation

†Keh-Yih Su and ‡Jing-Shin Chang

Department of Electrical Engineering
National Tsing-Hua University
Hsinchu, TAIWAN 30043, R.O.C.
email: †kysu@ee.nthu.edu.tw, ‡shin@ee.nthu.edu.tw

ABSTRACT

Rule-based approaches have been the dominant paradigm in developing MT systems. Such approaches, however, suffer from difficulties in knowledge acquisition to meet the wide variety and time-changing characteristics of the real text. To attack this problem, some statistical translation models and supporting tools had been developed in the last few years. However, a simple statistical model often results in a large parameter space and thus requires a large training corpus. Therefore, it is required to introduce language models that take advantages of well-justified linguistic knowledge to make stochastic MT systems practical. A stochastic model that emphasizes the adoption of well-justified linguistic knowledge in developing the model is called a corpus-based statistics-oriented approach. In this paper, corpus-based statistics-oriented paradigm is proposed, and its characteristics is compared with other methodologies. The recent progress in some corpus-based statistics-oriented models for MT is also reviewed.

1 Introduction

In the past, translation was recognized as a highly intellectual process. Hence, much effort in MT development is placed on the transformation of linguistic knowledge into rules for use in computers. The rules are usually acquired by computer engineers or linguists. Although the acquired rules can describe some general behavior in natural language to some degree, they are unable to cover the wide range of fine-grained knowledge in real applications. The transformation process is especially difficult for large scale systems. With the advance in computer technologies and the availability of large corpora, statistical techniques are proposed in recent years to shift the burden of *knowledge acquisition* from human to computers by inducing linguistic knowledge from large corpora automatically.

Since 1985, different designing methodologies have been explored in the **ArchTran** MT project.[1] The **ArchTran** English-Chinese MT system, like most systems, started with a rule-based approach. As the system scaled up, we found it harder and harder to apply rule-based approaches to acquire the fine-grained knowledge required for the widely-varying text styles from various clients. The research direction was thus shifted toward corpus-based approaches and some encouraging results were observed.

In this paper, we will address some practical problems with the current MT methodologies. In particular, we will discuss (i) some critical topics in designing a practical large-scale MTS (ii) the necessity and feasibility of corpus-based statistics-oriented MT, and (iii) some working examples of the corpus-based approaches, with emphases on their designing philosophy.

2 Designing Goals of a Practical Machine Translation System

Before making comments on any MT methodology, it is necessary to consider the criteria of being a "practical" MT system. This can be evaluated in terms of *quality concerns*, *engineering feasibility* and *cost-effectiveness* for marketing. In a translation center, the domains of the text are widely spread and various clients have their own preference and criteria on translation quality. Under such circumstances, our experiences show that the following factors are very critical to put an MT system into practical use.

Quality for Unseen Text

The translation quality of a system for real applications, rather than its performance in a laboratory environment, is the ultimate performance index of the system. To maintain good translation quality for real applications, some critical factors must be noticed.

- Coverage: Wide-coverage is required for a practical system to cover linguistically uninterested yet practically encountered phenomena. A designing method that can gather the required fine-grained knowledge is desirable for this purpose.
- Robustness: A model that shows good performance in a laboratory environment might perform badly for unseen domain due to the lack of robustness. Therefore, the underlying models of

[1] The ArchTran project started in 1985 as a joint effort between the Department of Electrical Engineering, National Tsing-Hua University, and the Behavior Tech Computer Corporation (BTC). Currently, most research projects are continued in the Behavior Design Corporation. Translation Service Center based on ArchTran had been setup in 1989 both for providing translation services and acquiring feedback to the MT research team [Chen 91]. Most customers are local branches of international computer companies.

a system must not be “over-tuned” to specific phenomena so that it can still maintain good quality even in different application domains.

Meeting Real User Requirements

According to our experiences, many documents, such as technical manuals, are to be published after translation, not just for information retrieval. Therefore, high quality translation, rather than just understandable, is required to meet the customer’s demands. Since the quality should be justified by the customers, not by the designers, the capability to accept user feedback, such as preferred terminologies and style, should be incorporated into a practical system [Su 92]. A practical system thus need to be:

- Adaptable: can be easily adapted to different domains to capture the general preference in the domain.
- Controllable: feedback from customers (or post-editors) can be incorporated into the system in a systematic way so as to meet the specific demands of the users. A feedback-controlled parameterized system [Su 92] could be helpful in this regard.

Capacity of Performance Tuning

- The system performance must be predictable in terms of certain system parameters so that the system can be improved continuously even after the system is scaled up. In other words, changing the system parameters (for example, rules or transition probabilities) should not cause side effects in an unpredictable way. Only when a system can be characterized in this way can it be constantly improved by adjusting the system parameters to maximize the system performance as more data are available.

Cost-Effectiveness

The construction and maintenance of the required knowledge base constitute the major cost of a practical system. Therefore, it is important to be able to acquire the required knowledge in a reasonable time; one should be able to construct the knowledge base systematically and maintain the knowledge base cost-effectively even in front of various domains and styles. Because human involvement is costly, one need a system that can shift the burden of *knowledge acquisition* from human to computers, and a system that can handle messy and uncertain phenomena.

Various methodologies provide various degree of flexibility in meeting these requirements. The characteristics of three different methodologies are discussed in the following sections.

3 Characteristics of Different MT Methodologies

From knowledge acquisition points of view, MT methodologies can be classified into 3 major classes, namely (i) rule-based approaches (ii) purely statistical approaches (iii) corpus-based statistics-oriented (CBSO) approaches. A *rule-based* system is a system that induces its linguistic knowledge from general observation of the behavior of natural languages by human. On the contrary, a *purely statistical* approach induces its knowledge from large corpora; when constructing a stochastic model, only stochastic behavior of the text are assumed without using any linguistic model. A *corpus-based statistics-oriented* approach also induces its knowledge from large corpora. However, the stochastic model is applied on top of well-justified linguistic models.

Rule-Based Approaches

A *rule-based approach* is characterized with several features. (i) It has a strict sense of well-formedness in mind; grammatical errors are explicitly prohibited. (ii) Most rules are based on existing linguistic theories that are linguistically interesting. When the required knowledge does not appear in any literature, *ad hoc* heuristic rules are commonly used. The major advantage of a rule-based system is that existing linguistic knowledge can be incorporated into the system directly. Also, the rules might look more comprehensive to a person.

Purely Statistical Approaches

In contrast, a *purely statistical approach* does not have a strict sense of well-formedness in mind. It does not assume any linguistic models either. Most of the time, the language generation process is simply modeled as a simple stochastic process, such as a Markov chain, and translation is regarded as a decoding process [Brow 90]. Almost no existing linguistic model is used to characterize the highly simplified stochastic process; and most such models acquire the parameters directly from surface strings.

Statistical models, either purely statistical or CBSO, have the following inherent advantages: (i) Uncertainty or preference is interpreted objectively and consistently in statistic sense. (ii) Consistency can be maintained easily even in large scale systems. (iii) The system parameters can be manipulated in a language-independent manner. (iv) Semi-/automatic training is possible with least human intervention. Hence, statistical approaches can be used to remove the burden of rule induction from linguists to machines.

A purely statistical model has the additional advantage that it does not involve any linguistic process, such as lexical analysis or syntax analysis, because no linguistic model is used. The

computation is thus very easy to carry out. One well-known example of such kind is the statistical MT model proposed in [Brow 90]. The underlying "bagged translation" source model, for instance, essentially assumes no structural hierarchy, and the source language is implicitly assumed to be order-free.

Although it is easy to carry out computation in a purely statistical approach, the lack of linguistic model make some useful information unavailable to such a system. For example, due to the lack of high level language model, the possible relationship among various constituents tends to be combinatory explosive. Therefore, an extremely large corpus ([Brow 90]) is required even for a very small test case to cover the large parameter space. Incorporation of well-justified linguistic models in a form that is suitable for computation into a statistical model is thus desirable.

CBSO: Corpus-Based Statistics-Oriented Approaches

A Corpus-Based Statistics-Oriented (CBSO) approach is characterized by an explicit statistical *language model* and a set of *parameters* associated with the model. It has all the advantages of a purely statistical approach described previously. The major difference is the use of a high level language model. In other words, it emphasizes the use of better inductive (statistical) tools without discarding well-justified linguistic models. In a CBSO approach, a language model is established first based on well-justified linguistic knowledge. The model parameters are then extracted from a large corpus to construct a consistent knowledge base.

While a purely statistical approach takes advantages of the stochastic behavior of a language near the surface level, a CBSO approach takes advantages of the stochastic behavior of a language on top of an appropriate language model. By incorporating a high-level language model, well-justified linguistic knowledge can be used directly to resolve interested problems in a large scale system. And all fine-grained knowledge can be acquired statistically from the corpus. Therefore, such an approach is good for integrating linguistic knowledge and statistic behavior.

An Example

To make the differences more clear, consider the problem of lexical ambiguity resolution. In a rule-based system, this problem might be resolved by ruling out all lexical category sequences that violate some constraints such as: a "determiner" can not be followed by a "verb". Such constraints are acquired from general observation of ordinary phrase patterns.

A purely statistical approach might collect all the possible word bigrams in a corpus, estimate the transition probabilities of the word pairs, and conclude that all words belonging to a special class

(the “determiner”) can not be followed by words belonging to another class (the “verb”). Because such an approach does not have the prior linguistic sense of “category”, it has to collect all word patterns to acquire the above conclusion.

For a CBSO approach, firstly, the known lexical knowledge is used to reduce words into equivalent classes such as “determiners”, “nouns”, “verbs”, and so on. The co-occurrence restrictions on the parts of speech are then investigated by estimating the probabilities of the part of speech sequences, instead of word bigrams. By using the linguistic knowledge about category, the parameter space and the required training data can be reduced drastically.

4 Some Observations in MT Development

Given the characteristics of the various methodologies and the requirements for a practical MT, what is a better strategy in developing an MT system? In the course of MT development, several facts about knowledge acquisition that significantly affects the cost-effectiveness of an MT system were observed.

- The required knowledge is huge, messy and fine-grained; there is not a simple set of rules that is capable of covering the demands of various domains.
- Knowledge acquisition is a very expensive and time-consuming process.
- People are competent in abstract language modeling, but are awkward in dealing with large and fine-grained knowledge.
- It is not easy for people to maintain the consistency of a large rule base.
- Tuning the rule base by hand does not guarantee global improvement of the system due to robustness problems.
- Computers are good at processing large corpora.
- Most knowledge required in MT is inductive, not deductive. In fact, linguistics is induced from natural languages, but a natural language is not produced based on any linguistic theory.
- Inductive knowledge can be acquired easily with statistical methods.

In fact, knowledge acquisition is found to be one major bottleneck for a large scale MT [Su 90]. These observations and the issues in the following sections constitute the main reasons for adopting a CBSO approach in MT development.

5 Necessity and Feasibility of Corpus-Based Approaches

Why Not Rule-Based Systems

Although rule-based approaches are widely used, there are several difficulties that make a rule-based approach unfavorable in meeting the designing goals for a large scale system.

- It is hard to handle *uncertainty* due to the lack of an objective preference measure. Therefore, it is awkward in dealing with highly ambiguous constructs. Since real text is not always strictly well-formed, such a system also has little capacity in dealing with *ill-formed* input.
- It is hard to deal with complex and irregular knowledge. When exceptions are encountered, lexicon specific and *ad hoc* procedures are commonly used, which make the system difficult to manage and maintain.
- It is hard to maintain *consistency* of the knowledge bases among different persons at different time. For a practical MT system, which usually takes a long period of development, the inconsistency often introduces diverse effects in the system performance and increases the difficulty and cost in maintenance.
- It is not easy for a rule-based system to attain high *coverage* for real text. First, the knowledge is usually confined to linguistically interested phenomena; linguistically uninterested phenomena are handled specifically as exceptions. Second, the designers tend to incorporate a lot of specific knowledge and increase the complexity of the model to make the system work for some testing cases. However, due to the lack of robustness, the coverage for real text tends to be poor. Consequently, such approaches usually fail to attain the required coverage for unseen text because the required fine-grained knowledge are not included in the *over-tuned* knowledge base.
- A rule usually takes care of only *local* constraints. The effects of adding a rule may cause unpredictable side effects over the global performance of the system, and global improvement is not guaranteed. This prevent a system designer from predicting the system performance in terms of well characterized system parameters; therefore, it is hard to tune such a system.
- Finally, there is usually no systematic and automatic approach to acquire the rules for large-scale applications. The lack of acquisition tools may make a formalism unfavorable in real applications even though it is theoretically interesting.

Accordingly, although rule-based systems could be tuned to show impressive performance in small scale, they may not be satisfactory for a large scale system.

Why Not Purely Statistical Approaches

In general, a statistical model (purely statistical or CBSO) will have poor performance in unseen domain if a large training database is not available. The criteria of being *large* is measured with respect to the complexity of the underlying models or the size of the parameter space. As a rule of thumb, in order to get reasonable estimate of the statistical parameters, the number of training instances must be several times larger than the number of possible outcomes in the parameter space.

Since a purely statistical approach does not use any high level linguistic knowledge in constructing a stochastic model, the parameter space is usually too large to be practical. For instance, to gather the word bigram statistics for the most frequently used 10,000 words, the number of possible bigram patterns is 10^8. The number of training word pairs must be several times larger than this amount. This is one reason why a purely statistical MT, like the one proposed in [Brow 90], requires a large training corpus even for a small vocabulary. Unless the corpus is large enough, the *sparse data* problem is inevitable; reliable statistics will not be available under such circumstances.

Furthermore, to reduce the size of the parameter space, the stochastic model of a purely statistical approach is usually simplified. For instance, because a trigram model requires a training database that is about 10,000 times larger than a bigram model in the last example, the latter is usually adopted at the expense of less contextual information. Hence, a purely statistical model usually cannot deal with long distance dependency. The purely statistical models also suffer from robustness problems due to sparse data when a complex model is used.

Why Corpus-Based Statistics-Oriented MT

With the use of high-level stochastic language models, the sparse data problem in a statistical approach can be relieved significantly. Furthermore, it is possible to make long distance dependency manageable.

Consider the example in the previous section, if the 10,000 words can be reduced to 100 *equivalent classes*, say by using parts of speech bigrams instead of word bigrams, the required training instances will be reduced by a factor of about 10,000. A training corpus of this scale will be much easier to manage. In general, high-level syntactic or semantic knowledge provides the required information to organize linguistic knowledge in terms of equivalent classes that have similar behavior.

As another example, if the co-occurrence preference between constituents is considered in syntactic or semantic level [Su 88, 91, Chan 92a], then the contextual dependency will no longer be confined to a local scope; long distance dependency can thus be explored easily.

These reasons suggest why CBSO approaches are desirable for a practical MT. As described earlier, human are competent in constructing high-level models and computers are good at processing large corpora. The designing philosophy underlying a CBSO approach is, in fact, to combine these capabilities to satisfy the designing goals in an cooperative way [Su 90].

Because large machine readable corpora are becoming available, better statistical language models are being constructed, and cheap computing power can be acquired easily, the adoption of corpus-based statistics-oriented techniques in MT development will be an inevitable new trend in the next MT generation.

6 The CBSO Approaches in the ArchTran MTS

The general tasks involved in a CBSO approach include: (i) language modeling (ii) corpus preparation (iii) parameter estimation, and (iv) adaptive learning (to adjust the estimated parameters in a way that increases the discrimination power and robustness of the model). Although the related techniques are not fully explored in the literature, various techniques have been tried in many aspects with promising or interesting results (e.g., see [Chur 88, Brow 90]).

Since 1987, the **ArchTran** MT project has explored a number of models on corpus-based statistics-oriented approaches for machine translation. These frameworks show encouraging results for adopting the CBSO approaches in a practical MT. The following sections review some of them to show the general designing philosophy. (Refer to the respect papers for details.)

Probabilistic Translation Model

The **ArchTran** approach to the translation problem is to find the most probable target sentence for a source sentence, given the knowledge of the source language and target language, and the transfer knowledge between them. The possibility is measured with the following *translation score* [Chen 91]:

$$
\begin{aligned}
& P(CW_1^{n_C} \mid EW_1^{n_E}, ELM, TM, CLM) \\
=& \sum_{I_E} \sum_{I_C} P(CW_1^{n_C}, I_C, I_E \mid EW_1^{n_E}, ELM, TM, CLM) \\
=& \sum \sum P(CW_1^{n_C} \mid I_C, I_E, CLM, \cdots) \\
& \times P(I_C \mid I_E, TM, EW_1^{n_E}, \cdots) \\
& \times P(I_E \mid EW_1^{n_E}, ELM, TM, CLM) \\
\approx& \sum \sum P(CW_1^{n_C} \mid I_C, CLM) \qquad (generation\ score:\ P_G) \\
& \times P(I_C \mid I_E, TM) \qquad (transfer\ score:\ P_T) \\
& \times P(I_E \mid EW_1^{n_E}, ELM) \qquad (analysis\ score:\ P_A) \\
=& \sum \sum P_G(.) \times P_T(.) \times P_A(.)
\end{aligned}
$$

where $EW_1^{n_E}$ and $CW_1^{n_C}$ are the source (English) and target (Chinese) sentences of length n_E and n_C, respectively; *ELM* and *CLM* are the source and target language models; I_E and I_C are the source and target intermediate representations; and *TM* is the transfer model from source language to target language.

The introduction of the intermediate forms (I_E and I_C) make it possible to resolve the translation problem through three phases, namely, analysis, transfer and generation, with the help of the corresponding *analysis score* $P_A(\cdot)$, *transfer score* $P_T(\cdot)$ and *generation score* $P_G(\cdot)$ in the above formula. The above formulation can thus provide useful scoring information for a transfer-based MT.

The introduction of the intermediate forms is actually a general technique that converts a purely statistical approach to a CBSO approach. By introducing the intermediate representations, different surface forms ($EW_1^{n_E}$ or $CW_1^{n_C}$) that are syntactic or semantic variants of one another can be normalized to the same intermediate form (I_E or I_C). Since transfer is conducted with respect to the intermediate forms, the number of possible mapping between source and target will be greatly reduced in comparison with direct mapping from source to target. In other words, the number of parameters will be reduced from the number of possible ($EW_1^{n_E}$,$CW_1^{n_C}$) pairs to the number of possible ($EW_1^{n_E}$, I_E) pairs, plus (I_E , I_C) pairs, plus (I_C , $CW_1^{n_C}$) pairs.

Score Function: Lexical Score, Syntactic Score and Semantic Score

The *analysis score* in the above translation model is implemented as a *score function* [Su 88, 89, 91, Chen 91, Chan 92a, Chia 92, Lin 92]. It is implemented by decomposing a semantically annotated syntax tree into a set of primitives that describes the lexical, syntactic and semantic features of a particular interpretation. As a result, we have the following *score function*, which can

be decomposed into *lexical score* [Chen 91, Lin 92, Chia 92], *syntactic score* [Su 88, 89, 91, Chen 91, Chia 92] and *semantic score* [Chen 91, Chan 92a] as follows.

$$
\begin{aligned}
& Score\,(Sem, Syn, Lex, Words) \\
& \quad \equiv P\,(Sem, Syn, Lex \mid Words) \qquad (score\ \ function) \\
& \quad = P\,(Sem \mid Syn, Lex, Words) \qquad (semantic\ \ score) \\
& \quad \times P\,(Syn \mid Lex, Words) \qquad (syntactic\ \ score) \\
& \quad \times P\,(Lex \mid Words) \qquad (lexical\ \ score) \\
& \quad \equiv S_{sem} \times S_{syn} \times S_{lex}
\end{aligned}
$$

The lexical feature *Lex* stands for the parts of speech of the input words (*Words*). The syntactic feature *Syn* is a set of sentential forms corresponding to the syntax tree. The semantic feature *Sem* is a set of N-dimensional feature vectors for the tree nodes, which consist of N semantic tags.

The *lexical score* [Chen 91, Lin 92] is implemented as an enhanced version of the lexical tagging model of [Chur 88]. This is done by introducing database smoothing, adaptive learning and discrimination enhancement techniques [Chia 92] to the enhanced baseline model.

The *syntactic score* model [Su 88, 89, 91, Chen 91, Chia 92] is formulated to maximize the transition probabilities between sentential forms; contextual information is considered in evaluating the transitions. As a result, it includes stochastic context-free grammar as a special case, and involves mechanisms for taking context-sensitivity into consideration.

In the *semantic score* [Chen 91, Chan 92a] model, each possible interpretation of a sentence is characterized by an annotated syntax tree. The set of semantic tags that make a syntax tree best annotated is recognized as the best semantic tags and the corresponding annotated syntax tree represent the best interpretation for the input sentence. An effective semantic tagging mechanism is used to attach the semantic tags to each node of the syntax tree.

Probabilistic Transfer and Generation Models

The transfer score and generation score are implemented under the framework of a probabilistic transfer and generation model [Chan 92b]. An automatic algorithm is developed to find the atomic transfer units, called *locally transferable units*, between the source and target languages from bilingual corpora. The intermediate representations of the translation pair are normalized to a form consisting of these units. With these units, the transfer process can be accomplished with only local operations on these units. A mechanism is proposed to train the transfer/generation model bidirectionally from both the source and target sides. In an ordinary transfer approach, the transfer

rules are derived mostly based on the source language. The generated target sentences are therefore unnatural to a native speaker. With such a bidirectional training procedure, the linguistic behavior of the target language can be incorporated into the transfer/generation model. Thus, the generated target sentences will be more close to the real style of the target language.

Feedback-Controlled Model for MT Tuning

To incorporate the preferred style of a customer, a closed loop control can be applied to a parameterized MT system in order to adjust the parameters of the system according to the customer's requirements. A mechanism is proposed in [Su 92] to accept such feedback. The post-edited text, which represents the real translation a customer really want, is compared with the output of the MT system for the corresponding source sentences. The discrepancy between the post-edited text and the current output of the original source sentences, as measured with a distance measure [Su 92], is used to adjust the system parameters. The system parameters are adjusted iteratively until some stopping criteria are met. With such a feedback-controlled mechanism, the parameters could be adjusted to fit particular text style or preference.

Preliminary Evaluation

The general results of the CBSO models in our experiments are encouraging. To give a better sense about their current status, the performance of the lexical score model on the publicly available Brown Corpus is outlined so that it can be directly compared with other frameworks. With the lexical score model, we have been able to reduce the error rate for ambiguous words in the Brown Corpus[2] from 8.56% with [Chur 88] model down to 4.95%; the significant improvement is achieved by using an adaptive learning algorithm and a discrimination enhancement technique in the lexical score model [Lin 92].

In general, we found it profitable to introduce high-level stochastic language models to relieve the burden imposed on computer engineers and linguists in acquiring the find-grained "rules". The introduction of high-level language models also reduces the size of the training corpora substantially. Some of the CBSO approaches had been incorporated into the **ArchTran** MT system. With the encouraging results of these frameworks, it is believed that CBSO approaches to MT are not only important for large scale systems but also technically feasible.

---

[2] Only about 40% of the text in the Brown Corpus have more than one tags. To be precise, only these ambiguous words are used to evaluate the error rate.

7 Concluding Remarks

In this paper, some criteria for evaluating a practical MT are outlined. Rule-based, purely statistical and corpus-based statistics-oriented approaches are compared in terms of their capabilities to meet these designing goals. We found that a rule-based system is not good at acquiring, maintaining and tuning the large, complex and irregular knowledge base, which is required to meet the designing goals of a large scale system. Hence, it is not easy to scale up while retaining good quality and wide coverage in real applications.

Statistical approaches have better capability in acquiring large and consistent knowledge base to fit the demands of real applications. However, purely statistical approaches, which use straightforward statistical models, may have too large a parameter space such that they requires extremely large corpora. Furthermore, long distance dependency might be sacrificed to make the parameter space manageable. These problems suggest that corpus-based statistics-oriented approaches, which emphasize the use of well-justified language models in characterizing the stochastic behavior of languages, should be used to reduce the parameter space and make long distance dependency manageable.

Human are competent in establishing high level linguistic models, but awkward in dealing with a large amount of fine-grained knowledge consistently and cost-effectively. On the other hand, computers are good at inducing knowledge from massive corpora with statistical techniques systematically and cost-effectively. Hence, constructing high level language models based on well-justified linguistic knowledge and estimating the parameters on top of the language models is a promising approach to integrate linguistic and statistic knowledge. It is expected that such corpus-based statistics-oriented approaches will play an important role in practical MT systems of the next generation.

References

[Brow 90] Brown, P. F., et al. 1990. “A Statistical Approach to Machine Translation,” *Computational Linguistics*, vol 16, no. 2, pp. 79–85. 1990.

[Chan 92a] Chang, J.-S., Y.-F. Luo and K.-Y. Su, 1992. “GPSM: A Generalized Probabilistic Semantic Model for Ambiguity Resolution,” to appear in *Proceedings of ACL-92*, 30th Annual Meeting of the Association for Computational Linguistics, University of Delaware, DE, USA, 28 June – 2 July, 1992.

[Chan 92b] Chang, J.-S. and K.-Y. Su, 1992. “A Corpus-Based Statistics-Oriented Transfer and Generation Model for Machine Translation,” to submit.

[Chen 91] Chen, S.-C., J.-S. Chang, J.-N. Wang, and K.-Y. Su, 1991. "ArchTran: A Corpus-Based Statistics-Oriented English-Chinese Machine Translation System," *Proceedings of Machine Translation Summit III*, pp. 33–40, Washington, D.C., USA, July 1–4, 1991.

[Chia 92] Chiang, T.-H., Y.-C. Lin and K.-Y. Su, 1992. "Syntactic Ambiguity Resolution Using A Discrimination and Robustness Oriented Adaptive Learning Algorithm", to appear in *Proceedings of COLING-92*, 14th Int. Conference on Computational Linguistics, Nantes, France, 20–28 July, 1992.

[Chur 88] Church, K., "A Stochastic Parts Program and Noun Phrase Parser for Unrestricted Text," *ACL Proc. 2nd Conf. on Applied Natural Language Processing*, pp. 136-143, Austin, Texas, USA, 9-12 Feb. 1988.

[Lin 92] Lin, Y.-C., T.-H. Chiang and K.-Y. Su, 1992. "Discrimination Oriented Probabilistic Tagging", to submit.

[Su 88] Su, K.-Y. and J.-S. Chang, 1988. "Semantic and Syntactic Aspects of Score Function," *Proc. of COLING-88,* vol. 2, pp. 642–644, 12th Int. Conf. on Computational Linguistics, Budapest, Hungary, August 22–27, 1988.

[Su 89] Su, K.-Y., J.-N., Wang, M.-H. Su, and J.-S. Chang, 1989. "A Sequential Truncation Parsing Algorithm Based on the Score Function," *Proceedings of International Workshop on Parsing Technologies (IWPT-89)*, pp. 95–104, Pittsburgh, PA, USA.

[Su 90] Su, K.-Y., and J.-S Chang, 1990. "Some Key Issues in Designing MT Systems," *Machine Translation*, vol. 5, no. 4, pp. 265-300, 1990.

[Su 91] Su, K.-Y., J.-N. Wang, M.-H. Su, and J.-S. Chang, 1991. "GLR Parsing with Scoring". In M. Tomita (ed.), *Generalized LR Parsing*, Chapter 7, pp. 93-112, Kluwer Academic Publishers.

[Su 92] Su, K.-Y., M.-W. Wu and J.-S. Chang, 1992. "A New Quantitative Quality Measure for Machine Translation Systems," to appear in *Proceedings of COLING-92*, 14th Int. Conference on Computational Linguistics, Nantes, France, 20–28 July, 1992.