DOCUMENT RESUME

ED 380 987 FL 022 682

AUTHOR Juola, Patrick
TITLE Self-Organizing Example-Based Machine Translation, A Prototype.
PUB DATE [95]
NOTE 8p.
PUB TYPE Reports - Descriptive (141)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Algorithms; *Computational Linguistics; *English; Evaluation Methods; *French; Grammar; *Machine Translation; Uncommonly Taught Languages; *Urdu

ABSTRACT

This paper describes an attempt to combine the advantages of both example-based translation and stochastic translation methods in an attempt to develop a method for inferring symbolic transfer functions from a bilingual corpus. By formalizing the translation process and applying standard optimization techniques, a system can be developed that will identify grammatical categories and produce coherent transfer functions between languages. The validity of this approach is demonstrated in a prototype system that can learn transfer functions between English, French, and Urdu. Contains 13 references. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 380 987

Self-Organizing Example-Based Machine Translation, A Prototype

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Patrick
Juola
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)"

Patrick JUOLA
Institute for Cognitive Science and
Department of Computer Science
University of Colorado at Boulder
Boulder, Colorado USA 80309
juola@cs.colorado.edu

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it
Minor changes have been made to improve reproduction quality
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

Abstract

This paper describes an attempt to combine the advantages of both example-based translation and stochastic translation methods in an attempt to develop a method for inferring symbolic transfer functions from a bilingual corpus. By formalizing the translation process and applying standard optimization techniques, a system can be developed that will identify grammatical categories and produce coherent transfer functions between languages. The validity of this approach is demonstrated in a prototype system that can learn transfer functions between English, French, and Urdu.

Keywords: Hybrid approaches, corpus-based NLP, machine translation

Introduction

The development of symbolic machine translation systems is difficult and expensive, but the development of non-symbolic MT systems is typically not much easier or less expensive. An ideal MT system would be able to to identify the structure of the source and target languages without the assistance of human engineers, but at the same time be easily understood, corrected, and updated to reflect new information and situations. For the second, a symbolic approach is almost a necessity, as subsymbolic or completely automatic inference systems tend to produce nearly opaque sets of numbers as their "outputs." However, the process of describing human languages in symbolic terms is difficult and knowledge-intensive. I describe here a prototype system for deriving linguistically plausible and understandable transfer functions from paired corpora without the necessity of human intervention or preanalysis.

Background

The major engineering bottleneck to machine translation, in general, is the development of the knowledge base, such as the linguistic analysis tools and the bilingual dictionary. The costs of developing dictionaries or linguistically sophisticated parsers is comparable to the cost of developing an expert system. Many researchers have looked for tools that could be used to automatically develop an analyzer for the source language (or a generator for the target language) from samples, or for that matter, a method of developing a bilingual dictionary from samples.

Nagao(Nagao, 1984) proposed just such an analytical method with his proposal of what would become known as "example-based translation." Instead of encoding explicit transfer rules, translation systems would store a collection of translation examples that would provide coverage *in context* for the input. A novel sentence would then be compared against the sentences most similar to it in the database, and the similar components would be combined into a translation. Although the notions of "similar" and "component" still require a fair amount of human engineering, systems have been built(Sumita et al., 1990; Sato, 1991) which show good results without an explicit dictionary.

Another approach(Brown et al., 1990) that avoids the knowledge acquisition bottleneck uses a huge bilingual corpus to attempt to solve the translation problem as a mapping between Markov chains. The researchers treat every sentence as generated by a Markov process and then consider every sentence to be a possible translation of every other sentence. The "translation" is simply the most probable translation of a given source sentence, Although this approach avoids the knowledge acquisition problem and can

BEST COPY AVAILABLE

achieve good results, the grammar and analysis techniques are psychologically implausible.

Self-Organizing Automatic Translation

The following work rests largely on the work of Green and others(Green, 1979; Morgan et al., 1989; Mori and Moeser, 1983) and their results on the psycholinguistics of language learning and the inherent structure of natural language. In particular, their research indicates that grammatical constructions are typically marked at surface-level by the appearance of a closed-class word or morpheme. A transparent example of this can be seen in Japanese. The basic sentence structure of Japanese is subject–object–verb. However, the noun phrase that fills the subject role is specifically *marked* by the appearance of a particle such as 'wa' or 'ga' immediately following. Similarly, the object noun phrase is marked by the particle 'o.' Similarly, in English, a noun phrase ("a noun phrase") is typically structured as a determiner or quantifier, a possible set of modifiers, and then the head noun itself. The determiner strongly signals the beginning of a complete noun phrase. If a prepositional phrase follows, it is marked by the appearance of a preposition, and then another noun phrase with its attendant structure. Such markers have not only been attested cross-linguistically, but have been shown to greatly help language learning in psychological experiments. This set of observations together constitute the Marker Hypothesis.

This hypothesis may provide a way to cast translation problems into a form solvable by standard multivariate optimization techniques. If languages are assumed to be largely compositional and grammatically structured, most of the preliminary work in translation is to identify the grammatical structures and corresponding structures in the target language as well as to find the appropriate expression of those structures. These marker constructions can provide powerful cues to these structures. Using a rewrite-rule formalism defined below, one can define the translation problem as the problem of simultaneous identification of :

- the grammatical organization of the source language
- the changes necessary to convert the grammar of the source language into that of the target language (which in a CFL can be done by permutation operations)
- the "dictionary" : a table lookup expressing the most appropriate translation for each grammatical element in the source language.

The input to this system is simply a set of paired sentences in the source and target languages. The only preprocessing done is the compilation of a list of the tokens appearing the the input data, to simplify the construction of a source-target dictionary. Unlike (Sumita et al., 1990) et al., there is no extensive semantic processing, so the development of a new data set for a new language pair requires only data entry from an existing body of translated text. The output of this learning, of course, is a description of the translation algorithm as defined above.

The Formalism

(Juola, 1994) demonstrates the existence of a normal form for context-free grammars (CFGs) which incorporates some of the properties of the Marker Hypothesis as described above. Specifically, grammars in *marker-normal form* have all rules in one of the following forms :

$A \rightarrow \epsilon$

$A \rightarrow a$

$A \rightarrow bBcCdD\cdots$

and therefore all non-terminal constituents are marked by a terminal symbol. This template thus gives a skeletal grammar-scheme that can be used to describe large classes of grammars in terms of a relatively small number of parameters (how many rules are there, what terminal symbol appears at position i of rule j, what non-terminal symbol appears at position $i+1$ of rule j, and so forth).

Given a filled set of these parameters, a closely-related grammar can be used to partition sentences into their constituents. In the third example above, the initial A-phrase would be partitioned into the "phrases" (bB) (cC) (dD)$\cdots$ and each phrase would be recursively parsed using additional rules or sets of rules.

Any phrase (typically a single-word phrase) which cannot be further broken down by this procedure will be treated as a unit for purposed of translation. Such lexicalized units can simply be looked up in a dictionary and replaced by a corresponding target word, most often a direct translation. These translations are used as the basic blocks for constructing the target sentence. The entire process can be learned and run without human intervention, but is understandable enough to permit human analysis and correction if necessary.

The Learning Algorithm

If there is such a thing as a "most general" problem in computer science, one candidate is certainly non-linear multivariate optimization. Almost any task of interest can be described as an

ERIC

equation or a set of equations over some complex mathematical space. The reduction of boolean logic to algebra is well-known, and symbolic (discrete) tasks can be easily converted to continuous search spaces by incorporation of an additional term that adds large amounts of error to a system when the parameters are "distant from" the desired discrete numbers. Because of this simple reduction, any semi-mathematical learning situation can be easily cast as a task of simultaneous parameter estimation.

Within this learning framework, it is easy to turn the individual components of the translation system described into semi-numerical optimization tasks. For example, most of a dictionary is a simple word-to-word mapping. With some gross simplifying assumptions, one can simply guess the entire dictionary mapping and then tune the guess, one entry at a time, to reflect the actual contents of the dictionary. Words with multiple definitions (for example, 'you' can be translated as 'tu' or 'toi' depending upon its grammatical role) can be handled by simply learning a large number of mappings to be used in the "appropriate" contexts. One-to-many and many-to-one translations (for example, 'not' to 'ne/pas') can be handled by accepting the null string as a potential word, and thus insertions and deletions can be done in certain specific contexts.

As has been stated above, the task of restructuring the source parse tree to reflect target structure can be performed by a simple permutation of the daughters at a given node. The task of learning a specific minimal-cost permutation has been well-studied in optimization theory as the "Traveling Salesman Problem." This can also be solved by a guess-and-tune algorithm where two cities (or constituents) are swapped at each iteration and compared against the desired results. Although in theory each rule of the source grammar will have a permutation to learn, there are few constituents per rule and the problem should still be tractable.

The hardest subproblem to cast into this optimization framework is the acquisition of the grammar. As has been discussed above, the existence of marker words makes the task much easier. If every constituent is marked by the appearance of a closed-class word, then by identifying the relevant marker words, one can segment the sentence into its constituents and parse recursively. Some work(Smith and Witten, 1993) has been done on sophisticated methods for identifying such words, but for this prototype system, the input vocabulary was small enough that the system could simply acquire the marker word set and all words are therefore potential marker words. The acquisition proceeds as : for each rule in the grammar, one must identify the subset of marker words which are relevant to that rule (for example, determiners are relevant to noun phrases) and the classes of the constituents themselves (for example, $S \rightarrow \bar{\mathrm{N}}\ \bar{\mathrm{V}}$). Guess-and-tune will obviously serve for subset selection (each marker word is either in or out of the relevant set and can be changed individually), and can be made to serve for class-identification by enumerating all potential non-terminal symbols and allowing the system to select among them for every symbol on the right hand side of a skeletal grammar. Given this grammar, the system can use it to parse and translate selected sentences from the training set.

Once the numerical framework is in place, one has a free hand in selecting the learning algorithm itself; the work described below uses a simple variant of simulated annealing. Originally a model of crystal growth, simulated annealing has the advantage of being well-known, well-studied, and relatively uncontroversial as an optimization technique. This technique is a modified version of a simple random walk through the dataspace of interest. The system starts out at some location in the dataspace (the guess) and makes modifications, at random, to the parameter set. At each stage, the new parameter set is measured against the old parameter set, and if it results in improved performance, the old parameter set is discarded and replaced (the tuning). Even if the new set results in reduced performance, the old set may be discarded and replaced if the reduction isn't too bad. As the tuning progresses, the notion of "too bad" is gradually tightened until the system accepts only improving changes and will eventually find the global optimum. Despite the simplicity of this algorithm ("It guesses some random grammar and then looks to see if it works well enough. If not, it randomly changes something to see if that makes it better."[1]), simulated annealing is actually a relatively powerful and efficient technique for nonlinear optimizations. Other experiments are in progress using other optimization techniques such as genetic algorithms, but the results have been largely similar to the work presented here and so are omitted for brevity. For interim evaluation as required for error measurement, the system uses an edit-graph (*diff*, to UNIX programmers) formalism(Myers, 1986) as an approximate measure of the amount of work that would be necessary to convert the results of the translation into the actual target sentences from the example database.

[1] James Martin, University of Colorado, personal communication of July 28, 1994

BEST COPY AVAILABLE

The Translation Algorithm

Once a skeletal transfer function has been generated, and particularly at the conclusion of the learning phase, the system should have a list of numerical parameters that describe the actual transfer process. A sentence to be translated is first parsed by the marker-normal form grammar generated above. All leaf nodes are replaced by their translations as defined by the context-sensitive dictionary representing their parent class. The system then proceeds to restructure the parse tree in strict bottom-up fashion. The children of a particular parent are reordered in keeping with the permutation function, and then the children's words are concatenated into a constituent phrase which replaces the (sub)tree of the parent.

The parent, then, is now a leaf and a child in its own right for another node, and its phrase will be incorporated into a larger phrase until the entire tree has been reordered and flattened into a translation of the entire source sentence. This sentence can either be presented to the user or compared against the target sentences in the example base for evaluation.

Experiments

The above formalism has been tested in a series of experiments designed to determine the strengths and limitations of the approach as well as the most efficient algorithms to use for error metrics and optimization techniques. I report here on the two most interesting from a linguistic/theoretical point of view. In the first experiment, we created artificial context-free grammars covering interesting and relatively complex subsets of French and English (including sentential complements, relative clauses, and gender distinctions), and attempted to learn the appropriate transfer functions from English to French based on a twenty-nine sentence artificial corpus developed and tested against the judgements of a native speaker of French. I then tested the system both against the training corpus and against novel sentences that had not been seen in the training phase.

The other experiment was designed as an attempt to model a "situated cognition" task, and to compare the performance of the system against a comparable task for humans, that of learning a second language (Urdu) by exposure to instructional documents. In particular, the system was given the vocabulary and grammar examples from a particular sample of instructional text (lesson two from (ur Rahman, 1958)) and told to learn appropriate transfer functions. As above, the resulting system was tested against both the training corpus and novel sentences (the exercises from the same lesson).

Evaluation

The problem of evaluating translations, whether human or machine, is difficult indeed. The following sections provide examples of two sorts of evaluations done on the system. The first, a black box examination, is a simple comparison of the system's results with the correct translation on novel material. For most stochastic MT systems, this is the only type of analysis that can be done, because the tables of probabilities are, to all practical purposes, opaque. This level of analysis typically will show the presence of errors, but not necessarily the source, the method of correction, or even what sorts of training data should be included to reduce them. However, because the system uses the symbolic information incorporated into the marker hypothesis (as well as a more linguistically plausible description in the form of a CFG), one can also do a glass box examination of the details of the translation functions and use this additional information to identify the precise source (in many cases a single decision) of errors and correct them.

Black Box Evaluation

In the English→Urdu experiment, the training data consisted of a vocabulary list for both languages and a set of seven sentences providing coverage for copula-locatives and imperative sentences. The training set was learned without errors, and the system was then tested on the exercises incorporated into lesson two from (ur Rahman, 1958). Except in cases where an individual lexical item had not been included in the training data, the test set was also translated perfectly. Expansion of the training data to include the complete system vocabulary in context resulted in a perfect transfer function for the constructions studied.

As is to be expected, the higher syntactic complexity of the English→French data reduced system performance considerably. I performed this experiment six times, using different random seeds and annealing schedules. Over several trials, the system was usually able to reproduce between 30% and 70% of the sentences correctly. Typical error patterns were instructive, however. The system usually identified the verb correctly, and the most common errors were context sensitive deletion errors, for instance of a pronomial direct object, a reflexive particle or a complementizer. Over the repeated experiments, only 5% of the sentences were translated as gibberish.

ERIC

Trials on novels sentences seem to bear this out. The experiment with the longest amount of learning time was selected for further analysis, and presented with novel sentences developed by native English speakers from the system's vocabulary without knowledge of the experimental results. These sentences were post-edited to limit the sentences to syntactic forms present in the training data (for example, the system only saw "glass" as a noun, not an adjective) and presented to the system without further modification. The system correctly translated 46% of the novel sentences, and produced the minor deletion errors or simple lexical production errors on another 8%. (These numbers can be compared, respectively, with the 51% perfect and 44% minor errors that this particular trial achieved on the training data.) In particular, the system had the most difficulty with the areas where the French structure is the most different from English; in the preverbal (pronomial) direct and indirect objects, the non-optional sentence complementizers, and reflexive particles.

Glass Box Evaluation

A major advantage of this approach is that it is possible to directly read a grammar and transfer functions from the output of the simulated annealing. This allows us to explain the behavior of the system at a conceptual level appropriate to the task at hand. For example, by examining the set(s) of words used as grammatical markers, one can identify which words the system felt were used in similar contexts and identify mistakes in class-identification. This allows us (or any user) to provide human guidance at any point in the learning cycle.

As an example of this approach, I will conduct a partial evaluation of the English→French transfer function discovered by the system. The fundamental unit for analysis should be a sentence, e.g.:

the man washes the car

The system attempts to translate this sentence by noting that words like 'washes', 'falls', 'creates', etc. comprise a class, and that everything before the appearance of the first word of such type should be translated as a unit. A similar process identifies the words 'a', 'the', 'this', and 'that' as another class which divide the third utterance of the sentence from the second.

(the man) (washes) (the car)

These utterances are themselves translated

(le homme)[2] *(lave) (la voiture)*

and then permuted (via the identity permutation) and concatenated to form the French translation.

Upon this level of analysis, the reasons for certain types of errors are clearly shown. For example, the identity permutation learned above is applicable in some instances but not in others (pronomial objects, for instance), merely reflecting the most common rule in the training set. A more sophisticated system would have the capability of applying different permutations depending upon the relevant utterances. It may be impractical on a large scale to list the verbs in a language explicitly, and some form of on-line tagging(Cutting et al., 1992) may be useful. The most serious flaw is that the classes learned are imperfect (for example, 'woman' is mis-classified as a verb, resulting in a serious mistranslation of the sentence "**(the) (woman washes) (the car)*"), but these are easily identifiable and correctable by more sophisticated search algorithms or by human intervention as necessary. The error here is more sophisticated than merely garbage in the partition. Because the system is allowed to reuse the partition classes to allow for recursive productions, occasional local maxima are found where two classes are parsed and translated by the same function—in the case of this particular error, the set described above as "verbs" is also used to distinguish between masculine and feminine nouns so that articles can be correctly translated.

Some translations as performed by the METLA system are attached as tables 1 and 2. These tables show, for each language pair, sample source sentences, their initial division into components, their translated components, and the final sentence as translated. As can be seen, the Urdu sentences are not only translated perfectly but also broken down into logical and linguistically cogent categories such as subjects, objects, and prepositional phrases.

The results for the French are not as uniformly positive, but for each of the incorrect sentences, it is possible to identify and explain the source of the errors. For example, the division of the third sentence is incorrect—"the man that touches the car" is an entire component and the main verb of the sentence is the *second* token of 'touches.' This is an artifact of the admittedly broken METLA-1 parsing algorithm, which divides at the first appearance of a given token. That this sentence is correctly translated at all is a tribute to the remarkable structural similarity between this sentence and its French translation. The fifth sentence is an example of a so-called "reflexive" verb;

[2]The process that converts *le homme* to *l'homme* is almost purely phonetic, and was ignored for purposes of this experiment.

BEST COPY AVAILABLE

the man is in the shop ((the man) ((is) (in the shop))) ((admi) ((hai) (dukan men))) admi dukan men hai
bring the letter from the shop (bring) ((the letter) (from the shop)) (lao) ((chitthi) (dukan se)) chitthi dukan se lao
wait in the office (wait) (in the office) (thairo) (daftar men) daftar men thairo
put the box on the table (put) ((the box) (on the table)) (rakho) ((sanduq) (mez par)) sanduq mez par rakho

Table 1: Sample English→Urdu translations with partial analysis

the glass touches a car (the glass) (touches) (a car) (le verre) (touche) (une voiture) le verre touche une voiture
she washes a cat (she) (washes) (a cat) (elle) (lave) (un chat) elle lave un chat
the man that touches a car touches a glass (the man that) (touches) (a car touches a glass) (le homme qui) (touche) (une voiture touche un verre) le homme qui touche une voiture touche un verre
that man washes a car that she creates (that man) (washes) (a car that she creates) (ce homme) (lave) *(une voiture qui elle creee) *ce homme lave une voiture qui elle creee
this cat washes (this cat) (washes) () (ce chat) (lave) () *ce chat lave

Table 2: Sample English→French translations with partial analysis

the proper translation should be "ce chat se lave," where 'se' is a general pronoun meaning 'self.' In English, certain verbs can be intransitive when the subject and object of the verb are the same—for example, "I shave (myself) every morning," "I wash," and so forth. Some of these verbs, in turn, *must* be expressed with the reflexive particle in French but with an ordinary direct object otherwise. This leads, in turn, to another example of the multiple-necessary-permutation problem discussed above.

A similar analysis can be done for the early experiments in English→Urdu translation. In this case, the errors can be tied directly to the fact that the word "knife", although learned as a single lexical item, was never presented *in context*, and the system had no way of identifying its grammatical class. When presented as part of the test data, the system determined (randomly) that it was a determiner, received no evidence to disprove this during training, and mistranslated accordingly. When the training data was extended to cover this case, the percentage correct rose to 100%.

It is important to not let the errors made by a prototype system overshadow the larger results. Despite having no a priori knowledge of semantic/syntactic categories, the system correctly identified the important grammatical constructs most of the time and built appropriate transfer functions. The errors were easily identifiable by examining the set of "marker words" used in parsing. Adding a new language to the system required only typing in the training data; no system modifications were required. If this approach scales well to larger corpora and vocabularies, this should cut the time necessary to develop and maintain large MT databases by a large fraction.

Obviously, many more experiments are required. Although the Urdu experiments demonstrated the ability to identify and permute high-order structures (such as basic Greenbergian word order), further testing is indicated. In addition, the small artificial corpora used should clearly be replaced by aligned real-world corpora as used by (Brown et al., 1990). (Wu, 1994) describes an experiment in aligning parallel Chinese-English texts that would provide data for a non-Indo-European language in large enough quantities to provide a significant test. Using this or a similar corpus should provide enough information to allow METLA to be improved considerably in future versions.

Conclusions

It may, at this point, be useful to revisit some of the differences between this work and some of the major projects in EBMT and statistical approaches. This work does not exclusively focus on grammatical induction. Although grammatical induction is an important part of the task, neither the problem (translation) nor the approach guarantees that the system will learn anything usable for grammaticality judgements. At the same time, the system would presumably be robust enough to translate malformed phone numbers without causing system errors. This is clearly an advantage in dealing with real-world input, where typographical errors and misphrasings are not uncommon. At the same time, this system includes grammatical structure which should result in more robust, understandable, and linguistically plausible translation functions than the Markov chains developed by (Brown et al., 1990).

Finally, although the system uses examples to develop its translation functions, there are several crucial differences between the proposed work and the more mainstream EBMT paradigm. First, other than the notion of paired sentences, there is no preanalysis of the translation database, which greatly reduces the load on the developers of the system. This system also produces a reduced database, explicitly extracting patterns rather than finding them as needed in on-line examples.

These results suggest that induction of transfer functions from untagged, unanalyzed corpora is both theoretically and computationally viable as an approach for the development of machine translation systems. In particular, of computer time can be substituted for human time if the appropriate bilingual corpus is available, as it is for most major languages in instructional text. The above formalism appears to work well and produces results which can be impressive in their own right as well as easily modified and improved.

References

Brown, P. F., Cocke, J., Della Pietra, S. A., Della Pietra, V. J., Jelinek, F., Lafferty, J. D., Mercer, R. L., and Roossin, P. S. (1990). A statistical approach to machine translation. *Computational Linguistics*, 16(2):79–85.

Cutting, D., Kupiec, J., Pedersen, J., and Sibun, P. (1992). A practical part-of-speech tagger. In *Proceedings of the Third Conference on Applied Natural Lanugage Processing*, Trento, Italy. Association for Computational Linguistics. Also available as Xerox PARC technical report SSL-92-01.

Green, T. R. G. (1979). The necessity of syntax markers: Two experiments with artificial languages. *Journal of Verbal Learning and Verbal Behavior*, 18:481–96.

Juola, P. (1994). Self-organizing machine translation : Example-driven induction of transfer functions. Technical Report TR-722-94, Computer Science Department, University of Colorado. Also available as cmp-lg/9406012.

Morgan, J. L., Meier, R. P., and Newport, E. L. (1989). Facilitating the acquisition of syntax with cross-sentential cues to phrase structure. *Journal of Memory and Language*, 28:360–74.

Mori, K. and Moeser, S. D. (1983). The role of syntax markers and semantic referents in learning an artificial language. *Journal of Verbal Learning and Verbal Behavior*, 22:701–18.

Myers, E. W. (1986). An O(ND) difference algorithm and its variations. *Algorithmica*, 1:251–56.

Nagao, M. (1984). A framework of a mechanical translation between Japanese and English by analogy principle. In Elithorn, A. and Barnerji, R., editors, *Artificial and Human Intelligence*, pages 173–80. North-Holland.

Sato, S. (1991). Example-based translation approach. In *International Workshop on Fundamental Research for the Future Generation of Natural Language Processing*, Kansai Science City, Japan. ATR Interpreting Telephony Research Laboratories.

Smith, T. C. and Witten, I. H. (1993). Language inference from function words. Technical Report 1993/3, University of Waikato, New Zealand.

Sumita, E., Iida, H., and Kohyama, H. (1990). Translating with examples : A new approach to machine translation. In *The Third International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Language*.

ur Rahman, A. (1958). *Teach Yourself Urdu in Two Months*. Azizi's Oriental Book Depot, II, K, 14/4, Nazimabad, Karachi–18, Pakistan, 22nd edition.

Wu, D. (1994). Aligning a parallel English-Chinese corpus statistically with lexical criteria. In *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics (ACL-94)*.