Virtual Machine Translation

Daniel Jones
Centre for Computational Linguistics
UMIST
Manchester, UK
danny@uk.ac.umist.ccl

Abstract

This paper proposes a new method of analysing and translating text based on example-based language processing techniques. Experiments have been conducted where statistical techniques utilising linguistic descriptions of language examples have been used to demonstrate the feasibility of removing conventional rule-based processing algorithms from the system architecture. Initial experiments have been conducted in order to illustrate the salient features of the approach as well as to demonstrate its feasibility.

Keywords: Analogy, Translation, Example-based, Cloning.

Introduction

There has been a recent increase in interest in the area of example or memory-based language processing, particularly in the field of machine translation. Several researchers including Nagao (1984), Sadler (1989), Sumita et al. (1990), and Sato & Nagao (1990), Brown et al. (1991) have all contributed to the area of example-based translation by proposing a variety of statistically-based techniques with contributions from linguistic description playing their part to some degree or other.

On one hand, for example, Brown et al. describe the effectiveness of using parallel text alignment techniques without the use of linguistic description at all in order to translate between languages. By way of contrast, Nagao (1984) introduces the idea of using case frames as a necessary representative framework in which to represent and match examples with source texts to be translated.

This paper reports research which has attempted to use statistical processing and linguistic descriptions together as fundamental features of an example-based natural language processing engine but without rule-based processing. However, although such an approach does utilise linguistic information it does not necessarily assume an *a priori* need for explicit linguistic rules which predominate in conventional second generation MT architectures.

Representations

One of the most important features of example-based language processing is the ability of a system to be able to match a given portion of input text with the examples stored in the dataset and to decide which is the "best" matching example. As Nagao has said:

> The most important function in the utilisation of example sentences ...is how to find out the similarity of the given input sentence and an example sentence. [1984, p. 178].

The concept of *similarity* is obviously very important, but what exactly is it? There are several axes of similarity which can be measured i.e. lexical; syntactic; semantic and pragmatic. Measurement of similarity can be equated with measuring the distance between an input and example text, but what do the units of distance have to be for successful matching to occur?

Measuring Distance

The concept of *distance* or similarity is of prime importance in example-based MT. The type of information which is to be measured, is, of course, task-related, as not all applications may need to determine semantic or pragmatic distance if the scope of an application is limited. However, it is suggested that being able to take into account a wide range of different types of information contained in language examples in order to correctly measure the distance between an input and and an example (or how far apart an input and a particular example are along any axis of similarity whether it be lexical, semantic, or pragmatic) is advantageous as it can assist in the matching process.

As a consequence, the research undertaken has attempted to represent examples at a number of different levels thereby making available a range of linguistic and non-linguistic information to the matching procedures.

The specific purpose of the experiments was to determine what information in the initial analysis phase

could be usefully employed by a particular statistical process with the aim of determining, firstly, the most closely matching examples to a given input text and, secondly, to translate the input with reference to the translation of the matched example.

A Functional Language Perspective and Virtual Translation

An interesting feature of example-based language processing is its ability to represent language in a non-atomistic manner and thereby associate with examples "pre-packaged" information about the way fragments of language interact with each other as well as how they relate to external objects, i.e., the physical world, as well as with the mental intentional structure of language users in communicative acts. For example, a language fragment can be represented along a number of axes of description in parallel e.g.

Figure 1: A Generalised Example

$$\textit{example}\left[\begin{array}{l} \textit{syntax}\left[\begin{array}{l} < \text{morpho-lexical} > \\ < \text{phrase structure} > \end{array}\right] \\ \textit{semantics}\,[< \text{case roles} >] \\ \textit{pragmatics}\,[< \text{rhetorical structure} >] \end{array}\right]$$

where an input text can be measured against an example along all or one of these parameters.

Conventional rule-based language processing has a general tendency to emphasize morpho-lexical and syntactic information as the initial source of analytic information with semantic and pragmatic processing taking second place. This is not necessary when using data shown in Figure 1 as semantic and pragmatic information is available concurrently. For the purposes of translation, Figure 1 can be augmented so as to include a translated equivalent of a particular source text example thereby yielding Figure 2.

Notwithstanding this, the need to match the source input against the source examples remains the same. However, once this matching process has succeeded, the generation of the translation is relatively straightforward as the translation, in some sense, is already known by the system. It is in this sense that example-based translation can be considered as *virtual* translation. As Somers et al. (1990) note:

> It could be argued that ...generation of a target text on the basis of something other than a source text is not 'real' translation. Such an argument might derive from an overly traditional view of translation ...where a translator ...translates 'blind' i.e. with no actual knowledge of the writer's intentions, goals, etc. [Vol. 3, p. 272]

Figure 2: A Generalised Example Translation Fragment

$$\left(\begin{array}{l} \textit{example}\ (\text{source})\left[\begin{array}{l} \textit{syntax}\left[\begin{array}{l} < \text{morpho-lexical} > \\ < \text{phrase structure} > \end{array}\right] \\ \textit{semantics}\,[< \text{case roles} >] \\ \textit{pragmatics}\left[< \begin{array}{l}\text{rhetorical}\\ \text{structure}\end{array} >\right] \end{array}\right] \\ \textit{example}\ (\text{target})\left[\begin{array}{l} \textit{syntax}\left[\begin{array}{l} < \text{morphology} \\ < \text{phrase structure} > \end{array}\right] \\ \textit{semantics}\,[< \text{case roles} >] \\ \textit{pragmatics}\left[< \begin{array}{l}\text{rhetorical}\\ \text{structure}\end{array} >\right] \end{array}\right] \end{array}\right)$$

The method of expressing the translation is already 'known' by the translation system so the emphasis now lies on the system being able to match *what* the user wants to say with the *way* it can be expressed. Such an approach allows the computational linguist to break away from the conventional rule-based bias toward syntax. As the example text fragments can be regarded as pre-bundled syntactic packages, semantic information automatically gets promoted in importance during the matching process. It is this Functional view of language use (Dik, 1978) which has been adopted in the research reported here.

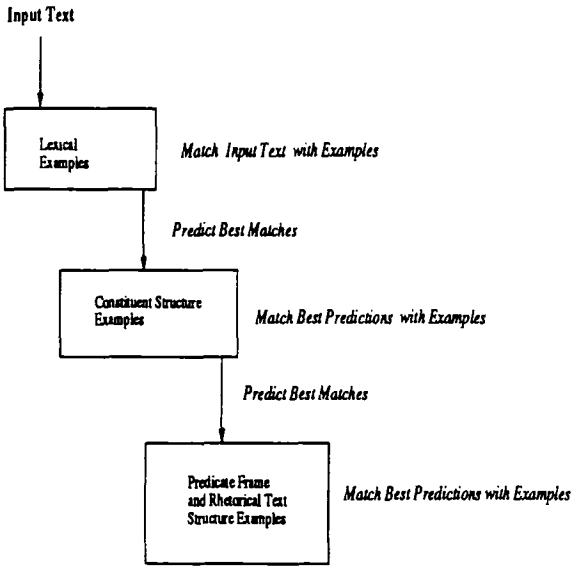
Experiments

Some experiments have been conducted with the aim of demonstrating the feasibility of using examples represented at different levels as in Figure 1 in order to accurately measure the distance between an input text and an example. However, only the first two levels of representation were used i.e. syntax and semantics.

These examples were represented as vectors of feature variables. It is envisaged that the semantic level would be represented as predicate frames in the style of Dik's Functional Grammar (ibid.), and rhetorical structure text profiles would be represented in a manner akin to representations suggested by Mann and Thompson (1988).

The driving force behind the measuring of syntactic distance was the statistical prediction algorithm proposed by Skousen (1989) called Analogical Modeling (AM) which allows for the construction of datasets of examples represented as arrays of features which are present *in* a fragment of text, e.g. determiners, verbs, etc., or *outside* of the text fragment i.e. what context the fragment occurs in e.g. the last three phonemes of the preceding word. A fragment of text can be any portion of text taken from a corpus. It can range from an individual word or morpheme through to groups

Figure 3: A Cascaded Dataset Architecture



of words representing syntactic structure. It can be thought of as a window of a variable size. AM uses a network of examples in order to calculate the analogical effect an entire dataset (set of examples) has on a particular input. It is this probability which is used as a distance measure in the reported experiments.

Cascaded Dataset Models

Figure 3 shows a simplified form of the kind of architecture created in the experiments. Initially the input text is matched against the lexical examples to predict the lexical categories of the input morphemes. These lexical category predictions are then used in the word group or constituent assignment predictions. The aim of the word grouping module is to predict the most likely constituent structuring of the input as well as to assign the most likely case roles to the predicted constituents.

Once this has been done, the predicate frame dataset can be entered where the syntactic and semantic (case role) predictions are combined to locate the most likely functional match or matches across this dataset. In terms of translation, it is at this point that the virtual generation of the target text can occur as, once this abstract level of representation has been achieved, by using examples as in Figure 2, all the information (both semantic and syntactic) is available to "produce" the translation [see Van der Korst (1989)]. Pragmatic and rhetorical text structure information is used for checking the appropriateness of an example for a given context. If, for example, a text fragment example is described as performing a text function such as BACKGROUND i.e. it is essentially introducing some information related to the *purpose* of communication, it would score less well as a candidate if the current rhetorical point in the "generated" text was, for example, post-nuclear. As Hatim and Mason (1990) point out:

> Although different languages may prefer different structural formats, ultimately, the limits on structural modification in translation are reached when the rhetorical purpose of the ST [source text] begins to be compromised. In such cases, the SL [source language] format must be considered the overriding factor. [p. 173]

The process of matching input and example texts can be thought of as a *cloning* procedure – cloning in the literal sense of making a near replica or look-alike of the input. The assumption made in example-based translation is that examples are not fixed and unchangeable but malleable to some extent with respect to the demands of the input. The underlying concern of the system is to clone examples in relation to any given input. The onus is on the examples to attempt to clone themselves as they are the repositories of all linguistic and non-linguistic information. The input is a *tabula rasa* in this respect. Exact clones, however, will generally not be possible as examples will tend to have to undergo some form of change or modification during the cloning process.

Creation and Use Examples

For the lexical data, a small corpus of business letter texts was automatically segmented to provide contextual clues for lexical categories. A simple heuristic was tested where word triples were extracted (w1, w2, w3). The last three and first three letters of w1 and w3 were retained and the last four of w2 were retained. These information was stored in a vector with one addition, i.e., the outcome which was the actual lexical category of w2. This information was added by hand. A vector produced from the words "for example the" would be of the form:

[f, o, r, m, p, l, e, t, h, e] ==outcome==> NOUN

Once an input string of words was automatically segmented in the same way, all the lexical categories for w2s were predicted.

The same text data was used here as in the lexical dataset except that sentences were represented as variable vectors where each variable represented possible predictions from the previous level in the cascade i.e. lexical categories. A typical "empty" input vector would be of the form:

[DETERMINER, NOUN, PRONOUN, PREPOSITION, ADJECTIVE, CONJUNCTION, NEGATIVE, ADVERB, QUANTIFIER, COMPLEMENTIZER, CONDITIONAL] ==outcome==> ?

with a possible instantiated form being:

Table 1: Lexical Predictions

INPUT WORD	PROBABLE CATEGORY	PROBABILITY
you	Pronoun	100%
can	Aux	95%
import	Verb	43%
widgets[1]	Noun	35%
from	Prep	95%
America	Aux	50%
"	Adj	14%
"	Noun	11%

Table 2: Word Group Predictions

WORD GROUP	PROBABLE CASE	PROBABILITY	CLONING PREDICATE
"you"	Agent	20%	import
"	Goal	20%	import
"	Agent	20%	supply
"	Recipient	20%	supply
"import"	Goal	100%	import
"from America"	Source	100%	import

[*determiner*, *noun*, PRONOUN, PREPOSITION, *adjective*, CONJUNCTION, NEGATIVE, ADVERB, QUANTIFIER, COMPLEMENTIZER, CONDITIONAL] ==outcome==> noun phrase

The vector entries labelled *category* are activated. Features only become active when a proposed word group's categories contain those lexical categories. In the experiment described here, lexical items whose categories were predicted to be AUX or VERB were not used to constitute word groups; therefore, with the input sentence "You can import widgets from America" only *You*, *widgets*, *from*, and *America* were considered to be important as viable translation units. Word groups were iteratively enlarged by adding the next input word to the word group vector. Once the probability for any particular outcome fell after previously rising, a word group boundary was hypothesised at that point.

The lexical examples were generated semi-automatically. The simple representation of characters meant that groups of words could be edited fully automatically. The labelling of the outcomes, however, had to be done manually. The whole process could be automated if the corpus was already tagged with lexical categories.

The creation of the word group examples based on predicate frame analysis of sentences is a much harder job to do as it involves a much greater amount of knowledge. Consequently, the entire task was done manually. Following the Functional Grammar convention, text fragment examples were represented as nuclear case frames with associated surface expression functions for generating surface texts (although these were not invoked at all in the current experiments). Predicate frames were used as a representational medium in order to capture the situational states with their participants e.g. Agent, Instrument, etc. By mapping the probabilistic results of the previous syntactic components onto this level, knowledge of actions, events and their participants can potentially be revealed in a completely analogical and non-rule-based way. The derivation of the predicate frames was achieved by human analysis and description of the same material from which the syntactic example information was derived.

Results

Although the syntactic level of representation might seem to be too simplistic, the predictions made by the system were encouraging even though the number and scale of the experiments were limited. At the lexical level, the predictions were 80-90% accurate. A typical result is shown in Table 1.

The results of the word grouping predictions and associated case role probabilities was not as successful as more anomalies were produced. This was probably due to the over simplistic representation of clause structure in the example vector as well as the compounding of earlier analogical errors from the previous lexical predictions – a typical situation to be found in this kind of cascaded architecture. However, the results were still reasonably good. Table 2 shows the result of segmenting the input text of Table 1.

The information gained from the word group level

[1] This category was predicted correctly even though the word "widgets" was unknown to the system.

predictions is to associate the slots of the cloning predicate i.e. the parent predicate which the analysis was based on, with the words in the current input. This then links the case role information associated with the cloning predicate with the pragmatic information in the same example fragment (see Figure 1) which can be used as a further measure of the appropriateness of this example in the wider pragmatic context. Associating input word groupings with example predicate frame slots can be thought of as akin to locating translation units – a necessary step before attempting any structural modifications.

Conclusions

The predictive power of the cloning process was surprisingly good given the under-specification of examples at the syntactic level. However, full integration with the semantic frame representations as well as pragmatic information still needs to be further tested and evaluated.

Encouragingly, even though the representations used were limited, they showed the feasibility of doing without any explicit rule-based processing during the analysis phase. The implications of this in terms of generation in a machine translation system, however, still need to be explored more fully.

The type of human interface envisaged in such a system is user-driven with interactions with the user occurring when the system believes that there is insufficient information available in the way of examples to translate some current input. This approach to translation really is best-suited to environments where authors of texts require their source language input to be translated while it is being created. Such a scenario naturally lends itself to a co-operative dialogue interface where, as the author types in their text, the system, by analogical means, attempts to translate in real time and only prompts the user when there are no, or insufficient examples on which to base a translation of the input. In such circumstances, the user may be prompted with alternative ways to express themselves to facilitate translation or to choose from a number of alternatives the system feels would be appropriate given the current syntactic/pragmatic context. Such possibilities remain to be explored.

References

Brown, P., Della Pietra, S., Della Pietra, V., Mercer, R. (1991). *Word Sense Disambiguation Using Statistical Methods.*

In Proceedings of 19th ACL, Berkeley. pp. 264-270.

Dik, S. (1978). *Functional Grammar.* North-Holland Linguistic Series.

Hatim, B. & Mason, I. (1990) *Discourse and the Translator.* Longman.

Mann, W.C. & Thompson, S.A. (1988). *Rhetorical Structure Theory: Toward a Functional Theory of Text Organization.* Text 8(3). pp. 243-281. Mouton.

Nagao, M. (1984) *A Framework of a Mechanical Translation between Japanese and English by Analogy Principle.* In Elithorn, A. Artificial and Human Intelligence. Elsevier.

Sadler, V. (1989). *Working with Analogical Semantics: Disambiguation Techniques in DLT.* Foris Dordrecht.

Sato, S. & Nagao, M. (1990) *Towards Memory-based Translation.* In Proceedings of COLING 90 (Helsinki), Vol. 3. p. 247. Ed. Hans Karlgren. University of Helsinki.

Skousen, R. (1989). *Analogical Modeling of Language.* Kluwer Academic Publishers.

Somers, H., Tsujii, J., Jones, D. (1990). *Machine Translation without a Source Text.* In Proceedings of COLING 90 (Helsinki), Vol. 3. p. 271. Ed. Hans Karlgren. University of Helsinki.

Sumita, E., Iida, H., Kohyama, H. (1990). *Translating with Examples: A New Approach to Machine Translation.* In Proceedings of the Third International Conference on Theoretical and Methodological in Machine Translation of Natural Languages. LRE, University of Texas at Austin.

Van der Korst, B. (1989). *Functional Grammar and Machine Translation.* In Connolly & Dik (Eds.). Foris.