Towards Machine Translation Using Contextual Information

Tim Cornish, Kimikazu Fujita, Ryochi Sugimura
{tim,fujita,sugimura}@isl.mei.co.jp

Matsushita Electric Industrial Co Ltd
Information and Communications Technology Laboratory

1 Abstract

A proposal is made for the use of contextual information in the machine translation of Japanese and English. This paper describes the use of a Context Monitor to maintain contextual information dynamically and the augmentation of appropriate features to a semantic network to enable simple inference. The approach taken is that of "*best guess*" processing with the contextual information being handled with semantic information on a shallow level.

2 Introduction

Current Machine Translation (MT) systems process input sentence by sentence. However, experience with English and Japanese has shown that some languages differ to such a degree that sentential translation yields poor results. Let us first compare the results of a conventional MT system with those we expect to get for MT with context:

1. 去年開発主任が新しいテレビ二機種と新しいビデオ四機種を開発した。
2. ビデオは販売部に出荷した。
3. 二機種はすぐに発売した。
4. とても良く売れた。

This might be translated by a current machine translation system as shown in Figure 1[1]:

It can clearly be seen that meaning in many sentences is obscured. Let us compare this with the results of a system using simple contextual information as shown in Figure 2:

This second translation is much more coherent and better preserves the meaning of the original sentence.

An attempt has therefore been made to solve some of the problems of translating languages such as Japanese and English using contextual information. Due to the considerations of wanting to produce a high quality small-sized MT system, the approach taken is to use the resources available in an existing MT system and to process the contextual information

1. The Chief Development Engineer developed two new TV models and four new video models last year.
2. a) A video was shipped to the Sales Section.
   b) We/Someone shipped a video to the Sales Section.
3. a) Two models were released straight away.
   b) We/Someone released two models straight away.
4. It sold very well.

Figure 1: Conventional MT Results

1. The Chief Development Engineer developed two new TV models and four new video models last year.
2. He shipped the videos to the Sales Section.
3. They released two models straight away.
4. They sold very well.

Figure 2: Contextual MT Results

on a shallow level only, using the information gained to guide the translation on a "*best guess*" basis. This kind of feature with rather light processing for the production of a higher quality translation is desirable in a practical MT system because the advantages of large-scale processing for deep contextual information are likely to be limited in this application.

3 The MT System

The translation system presented here is a model system which is being used to investigate the techniques proposed. The translation part is carried out in PROLOG using an LFG-like grammatical formalism [2]. The current dictionaries contain information to translate about 300 words. There are 350 grammar rules which cover a wide range of sentence patterns.

The context monitor operates using information retrieved from the *f*-structure of a sentence after analysis. This information is then used during the transfer

[1] There is obviously a great difference in results between systems, but these translations represent typical (unedited) results from a number of systems. a) and b) options depend on the default settings of individual systems

[2] The original program for English-Spanish translation developed by G. Amores [Amores'89] has been widely adapted and enlarged to J-E & E-J translation.

of the source *f*-structure to the target *f*-structure. As context processing is carried out on only a shallow level, only information for lexical item, number, person, gender, case role etc is used in the context system, along with semantic information from the semantic network. The way that this information is used will be explained below in regard to the specific problems that the use of context is intended to resolve.

4 The Context Monitor

The context monitor proposed in this paper uses a standard focussing theory as a basis ([Sidner 81]),[Sidner 86]), although somewhat simplified according to the *best guess* approach that we are adopting. It is planned to increase the complexity of this initial algorithm to reflect more current versions of the theory as the system is developed.

The context monitor has a number of basic data structures: Current Focus, Actor Focus, Potential Focus List, Potential Actor Focus List, Discourse Segment Stack and Actor Focus Stack. There is also a Current State List that maintains a record of all the semantic items currently held in any of the other data structures and the semantic features to which they are linked. This list is updated (entries added and removed) after every sentence.

In order to limit the scope of the context information required in the context monitor, an analysis was made of the main differences between Japanese and English that provide problems for MT systems. The basis of the analysis was to find what information can be gained from context to solve these problems.

4.1 Plural Forms

Japanese is (in general[3]) unmarked for number. English, however, differentiates between singular and plural. This fact causes problems when translating from Japanese to English as the number information required for the inflection and declension of English is not available from the analysis of the Japanese. For example:

少年は犬が好きだ。
(boy(s)$_{subj}$ dog(s)$_{obj}$ like)

The boy likes the dog.
The boys like the dog.
The boy likes the dogs.
The boys like the dogs.
Boys like dogs.

In the current system an initial sentence analyzed by the system is processed to find possible foci. Items which are in the plural or are in conjunction are stored as a set. The set as a whole is given plural number, but consists of individual items or, as in the case of "*two new TV models and four new video models*", as subsets. Subsets or individual items within the sets are available as antecedents to subsequent referring expressions.

Thus, in the example text in Section 1, after the initial sentence is analysed, the proposed focus is *two new TV models and four new video models*, the structure of which is shown in Figure 3 below:[4]

```
[set1: [set2:
          [pred:terebi(TV),
           num:plur,
           mod:[pred:kishu(model),
                spec:[pred:ni(two),
                       ...]]],
           ref:set2],
        [set3:
          [pred:bideo(video),
           num:plur,
           mod:[pred:kishu(model),
                spec:[pred:yon(four),
                       ...]]],
           ref:set3],
        ref:set1]
```

Figure 3: PROLOG Structure

Sentence 2 (S2) is analysed and a test is made to see if any items in that sentence confirm or reject the proposed focus. The structure for the item ビデオ ("a video/videos") is matched by unification with the structure for the proposed focus and can be matched with a subset of it, namely 新しいビデオ四機種 ("4 new video models"). That item is therefore taken to confirm the proposed focus.

That proposed focus is, however, immediately PUSHed onto the focus stack because the subset of 4 videos is taken as the current focus[5]. The item ビデオ of S2 inherits the features of the set of videos from S1 and is therefore expressed in the English with a plural form: "*videos*". It is hoped that in this way the context monitor will be able to distinguish between singular and plural in at least some cases.

In some cases there is no way of distinguishing between singular and plural reference in Japanese as in the case of the sentence below:

太郎と花子はケーキを買った。そして øで公園で ø食べた。

Taro and Hanako bought a cake. *They ate* it *in the park.*
Taro and Hanako bought (some) cakes. *They ate* them *in the park.*

[3]Note the use of 達 ('*tachi*') with mainly people and animals, and some pronouns: 彼 ('*kare*'-he) vs 彼ら ('*karera*'-*they*)

[4]in a simplified form, showing relevant detail only. Italics are translations for explanation only and do not appear in the structure proper

[5]The system currently deals only with local focus - there is no account of global focus

In such cases the context monitor cannot resolve singular or plural and so the MT system default will be relied on. However, the context monitor at least allows for coherence with subsequent pronouns.

4.2 Translation of Pronouns

Japanese makes much use of the zero pronoun (marked here by "ø"), especially in the subject position, but equally for other roles. For example:

ø	ø	もう	食べて	しまった。
(ø	ø	mou	tabete	shimatta)
($ø_{subj}$	$ø_{obj}$	*already*	*eat*	*AUX*]

{ *(I) have already eaten (it)*
(They) have already eaten (them)
(He) has already eaten (them)
...

This means that there is no information available from the single sentence to aid the choice of equivalent English pronoun (which must normally be expressed). As shown in 2 and 3 of the example text in Figure 1, MT systems use a number of methods to add an overt pronoun, often involving the user in the final choice.

It is claimed that if there is a pronoun in a sentence, it must refer to the focus of that text segment (in order to continue the current segment) and if there are more than one pronouns, at least one of them must refer to the focus. By tracking the focus of a text segment, ø pronouns in Japanese should be able to be resolved so that an appropriate overt pronoun in English can be selected for the translation.

When a zero pronoun is detected in a sentence, if an antecedent can be found for it, and that antecedent is a set of items, the overt pronoun inserted in the English will be plural.

Thus in 4 of the example from Figure 2, we see that the zero pronoun in "$ø_{subj}$ とても良く売れた" is analysed as referring to the two video models released and is therefore translated with a plural pronoun: "*They sold very well*".

Note, however, that there is ambiguity in Sentence 3 between whether the zero pronoun in ø二機種をすぐに発売した *(ø released two models straight away)* refers to the Chief Development Engineer or the Sales Section.

When faced with ambiguity such as this, large-scale attempts at context understanding might use inference plans to solve the ambiguity. However, because of the limitations of a small size MT system and the fact that even large scale deep level semantic processing has not been satisfactorily realised for unlimited domains (with which our MT system is intended to work), we decided to attempt limited inferencing by the addition of some features and links to the semantic network of the MT system. The inferencing able to be performed by such a method is quite simple, but is hoped to be sufficient for our needs in accordance with the *best guess* policy.

4.3 Semantic Networks

Semantic networks are basically a hierarchy of concepts which are linked to one another in a network type structure. Semantic networks were introduced by Quillian in 1968 [Quillian 68] and were widely used in attempts at Knowledge Based Systems, particularly during the late 1980s.

As an example of such a system let us briefly consider the system for Japanese-English translation using contextual information proposed by H. Isahara and S. Ishizaki ([Isahara 86], [Isahara 87], [Ishizaki 89] and [Isahara 90]) as one Knowledge Based approach and compare it to the techniques used in the system proposed in the current paper.

The translation system CONTRAST translated Japanese newspaper articles into English. However, a major difference regarding our system is that context understanding involved analysing a sentence A (eg with an overt subject) and a sentence B (eg with a covert subject) and then matching these sentences against a number of sentence patterns. If a match was found these would form a text pattern C, with A and B as subparts. The subject of A would be used (if suitable) to provide the subject for B. By adding further sentences and text patterns, a representation of the entire text would be formed and this text representation translated into an English equivalent text.

However, this technique relies on the fact that you can predict all the types of sentence that will occur and how they combine to form an entire text (perhaps possible for the types of newspaper articles the CONTRAST system aimed to translate). However, if a sentence cannot fit into one of the preprepared patterns, the system will fail. Our system is intended for more general language and as we cannot predict the length of a text or what kinds of sentence will occur within that text, the Context Monitor provides ongoing contextual analysis without prepresuming the length or nature of the text.

CONTRAST also relies on making a representation of the entire text. In our system there is no understanding of the overall text structure (according to our shallow level approach). Instead, the objects and events referred to in the text are analysed and made available to resolve subsequent analysis problems. The translation remains sentence by sentence, although the general context of the text is monitored.

Finally, Semantic Networks, such as that proposed by Isahara et al., are static networks. The links do not change between nodes. The possible paths that are available through the network may change but the links themselves do not change. In our system, the basic semantic network is static, defining irrefutable relations between the concepts in the hierarchy, but on top of this, other links are augmented onto the network and these links can change dynamically in respect to the specific objects and concepts referred to in the text. This provides a powerful augmentation to the basic network.

4.4 The Augmented Semantic Network

The semantic network in this system is basically a hierarchy of Objects, States and Events. The addition of features to the semantic network in effect adds links to the network. Two kinds of link are proposed: permanent links and temporary links. Permanent links are conditions that must be true for a certain action or state-of-affairs to hold. The other, temporary, links are used to create a default state for the objects mentioned in the text. As the text is processed, these links may change, so that the information available to the system will differ from one sentence to the next.

4.4.1 The Links

The division between Objects, States and Events is reflected in the type of feature given to semantic items. For example, Events typically contain features about the sort of things that are affected by that event; States contain information the types of objects that may be in that state; Objects contain information about any subparts or if they themselves are typically part of another (larger) object and what type of Event they are typically involved in.

On this basis, the following types of link are proposed:

⋆ Condition (=c): (permanent) a condition that must hold for a State or Event to come about.
⋆ Before Condition (BC::): (permanent) a condition that must be true before an Event or State comes about.
⋆ After Condition (AC::): (permanent) a condition that becomes true after an Event or State comes about.
⋆ Has Subpart (has): (temporary) an Object has related subparts or is a subpart of another Object.
⋆ Characteristic (has Semantic_Label): (temporary) an Object has the characteristic of Semantic_Label (usually an Abstract_Relation: Size, Shape Colour etc.). This takes the form of: "Item has Semantic_Label", such as "*Peter* has Existence Lifespan". This states that an item with the semantic item *Peter* has an existence of some kind and further locates that item on a path of the network to the abstract relation of Lifespan. In this way, nodes between these two points are all available for reference by the inference system.
⋆ Ability (able_to): (temporary) This is not fully defined in the current system but represents characteristic features of items e.g. "*door*" often appears in the theme position of the Events *Open* and *Shut*.

These links are considered sufficient for the current capabilities of the system. Links may be deleted or others added as the range of the sytem widens, if this is thought necessary or desirable.

4.4.2 Permanent & Temporary Links

The difference between permanent and temporary links is in the nature of the information that they convey. Permanent links are those that are augmented to the network and connect nodes one to the other in accordance with the features found in those nodes. '*Before Condition* and *After Condition* links are permanent, although the information contained in the nodes that they connect to will only become available to the context monitor in accordance with the tense and aspect of the verb (i.e. an *After Condition* is obviously only valid *after* the completion of the (for example) action denoted by a verb has finished. Temporary links are those that supply default information to the context monitor concerning nodes that it is concerned with. Thus, for example, an entry for a bird might state that it is Able_To Fly. However, if the input text were to state that a particular bird is unable to fly, that Able_To link would be cancelled. Thus temporary links provide the information that the context monitor uses, using the temporary links to spread throughout the network (within set search constraints) and gathering information that can be used for inferencing.

4.4.3 Example of the Augmented Features

An example of the features used to augment the semantic network can be given using the example:

> Peter heard that John had died. He was very sad.

Given the dictionary entry shown (here simplified) below, "*Peter*" will be analysed as a male proper noun.

```
dic(n,'Peter',[semfeat:[human:yes],
              proper:yes,
              gender:masc,
              pred:peter]).
```

When the embedded clause is analysed, "*John*" will be analysed in a similar way. The semantic feature `human:yes` locates these two lexical items as subsumed by the semantic feature "Living" in the network. Augmented features for a male human such as the objects referred to by the names *Peter* and *John* are shown below in Figure 4 below along with possible entries for the Event *die* and the State *be sad*.

It can be seen that one of the Before Conditions of the Event *Die* is that the actor role is filled by an item that has the semantic feature "Living". The default assumption for "*John*" is that he is Human

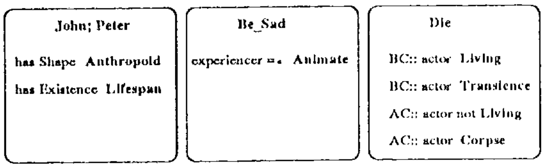

Figure 4: Augmented Features

and therefore Living. However, the After Conditions of the Event *Die* cancel the feature Living in connection with "*John*" (*'not'* means that a feature and all the other features underneath it in the tree should not be reachable by that item), and state that the item should be associated with the feature "Corpse" (a semantic label in the system for something that was living but is no longer). Thus the semantic item "*John*" is first linked with the semantic feature "Human" and all the other features inherited from that feature. However, the features associated with the semantic item "*Die*" cause the links associated with "*John*" to change. This means that when the second sentence is analysed, the possible candidates for the experiencer role of the semantic item "*Be_Sad*" are analysed, an item with the semantic feature "Animate" will be sought, and so the item "*John*" will not be considered in the search as it is no longer on a path reachable by "Animate". "*Peter*" is therefore the only possible antecedent.

4.5 Articles

Japanese does not use definite and indefinite articles and so when there is no overt determiner in the Japanese, one must be supplied for the English translation. For example, Sentence 2 of our example text:

ビデオは販売部に出荷した。
They passed *the videos* to the Sales Section.

Where a simple default rule is used for articles, this could equally be machine translated as: *they passed videos to the Sales Division*[6], where it can be considered that some of the sense of the original sentence is lost.

While the use of contextual information cannot solve all of the problems of articles, it is hoped that at least in some cases incorrect possibilities can be eliminated (following the "best guess" policy). In the cases that the context monitor cannot decide an article, the MT system default will be relied upon.

To decide between a definite and indefinite article in English, a simple rule of thumb in the present system is that once an object has been specified in a context, all subsequent references to that particular object in the same context will be definite[7].

In the method proposed here, as objects are analysed, they are given a unique reference number (`ref`) that separates them from all other objects of the same type. Thus, the first time that an object is analysed, it will be made indefinite, unless the reference can be analysed as being a generic one (e.g. *The lion is a dangerous animal* etc).

From then on, if an item in the text can be linked to an item which is the current focus, a potential focus or an item on the focus stack, it will be made definite in the English translation. Therefore, the two video models of Sentence 3 are recognised as a subset of the four videos that form the focus and are given the definite article.

Note also that as subparts of objects are included in the features attached to semantic items using the **has** feature, objects related to an item already mentioned can also be treated to some extent and translated with definite articles:

> Hanako bought a new video. She took it back to the shop as *the tape head* was damaged.

This, however, a very simple approach and cannot account for all possible uses of the definite/ indefinite articles. However, the approach outlined above also follows the "*best guess*" strategy; where this strategy fails the normal default rules of the translation system take over.

4.6 Restrictions on the Repetition of Pronouns

In English, overt pronouns are repeatable and in some cases obligatory in a sentence to preserve meaning. In Japanese, however, overt pronouns are not repeatable as shown in the below[8].

He does *his* work when *he* wants to.

* 彼は 彼が したい ときに 彼の 仕事を する
(he $_{topic}$ he $_{subj}$ wants to when his work $_{obj}$ does)
彼は φ したい ときに φ 仕事を する
彼は 自分が したい ときに φ 仕事を する
himself $_{subj}$
彼は φ したい ときに 自分の 仕事を する
his own

It is therefore desirable to have a routine in an MT system to replace overt pronouns in English with ø or 自分 (*'jibun'* oneself) in Japanese. In this case, the use of the pronoun *he* in English will be analysed and recognised as referring to the same person using

[6] assuming that the noun is defined as plural by some other process, otherwise *a video* is also a possibility

[7] This basic principle is supplemented by rules based on syntactic constructions etc

[8] This example taken from [Wada 90]

the processes outlined above. Separate rules concerning co-occurrence of pronouns can then be used to substitute ø or 自分 (*'jibun'* himself) in the Japanese translation.

5 Limitations & Problems

As shown above, the inferencing carried out is very simple. It depends entirely on the links between nodes of the network and there is an obvious limit as to how complicated those links may become before the processing required to search all the nodes linked to a particular item becomes prohibitive. At the current stage of planning, a structure (a semantic item) may be linked to another via one node (constrained to be an Abstract Relation). There are no current plans to increase the number of such linking nodes.

The inference mechanism is also expected to perform poorly where actions denoted by a verb are complex. This is due to the very simple feature descriptions that we use in the system. It might therefore be desirable that, if the processing is not completed within, for example, a constrained time, the process be terminated and the context monitor left to rely on semantic feature matching alone.

Another major problem is writing the features for the links in the network. At the moment, all features are written by hand, but it is hoped that similar information might be extracted from semantic and caseframe dictionaries.

The context monitor is currently written in PROLOG[9]. The program currently consists of several hundred lines of PROLOG.

6 Final Remarks

The idea of using contextual information in Machine Translation has been proposed before (for example [Wada 90],[Eberle 92],[Haenelt 92]), however, there seems to be little research carried out in the field. MT research still seems to take the sentence as the basic unit of translation and the quality of their raw output suffers as a result. We have proposed how some of the errors of J-E & E-J translation can be solved and have outlined a Context Monitor with simple inferencing.

The ***best guess*** approach tries to define a problem and specify the information needed to solve that problem. The context monitor system searches for specific information from the input sentence and if it cannot find it, it simply does nothing, allowing the defaults of the translation system to supply the necessary information. The search routines of the context monitor look for that specific information at as earlier a stage as possible in the process and so if that information is not found, the next routine is tried as quickly as possible in order not to decrease the overall translation speed by a significant amount.

Even when the context monitor fails and the MT system defaults are relied upon, the context monitor ensures consistency with subsequent sentences.

Complicated texts are likely to lead to the Context Monitor failing often but it is still felt that the better translation produced in many more cases and the fact that interference with the speed of the translation is negligible mean that the prospects for a compact-sized personal MT system producing better quality translations are very promising.

[9] Not all of the features mentioned in this paper are currently implemented

References

[Amores 89] G. Amores, MSc. Dissertation, UMIST, 1989.

[Eberle 92] K. Eberle, W. Kasper, C. Rohrer, *Contextual Constraints for MT*, TMI-92 Proceedings, pp213-224.

[Haenelt 92] K. Haenelt, *Towards a Quality Improvement in Machine Translation: Modelling Discourse Structure and Including Discourse Development in the Determination of Translation Equivalents*, TMI-92 Proceedings.

[Isahara 86] H. Isahara & S. Ishizaki, *Context Analysis System for Japanese Text*, COLING86, pp244-246.

[Isahara 87] H. Isahara, K. Hashida and S. Ishizaki, *Constraint Based Inference for Concept Unification in Contextual Analysis* (in Japanese), Proc of 35th Meeting of the Information Processing Society of Japan, 1987, pp1240-1249.

[Ishizaki 89] S. Ishizaki & H. Isahara, *Contextual Information Translation System CONTRAST* (in Japanese), Journal of the Information Processing Society of Japan, vol.30, no.10, pp1405-6, 1989.

[Isahara 90] H. Isahara & S. Ishizaki, *Natural Language Understanding System with Concept Hierarchy*, Proc. of Pacific Rim International Conference on AI '90, pp322-27.

[Quillian 68] M. R. Quillian, *Semantic Memory*, in Semantic Information Processing, ed M. Minsky, 1968, pp227-270.

[Sidner 81] C. L. Sidner, *Focusing for Interpretation of Pronouns*, American Journal of Computational Linguistics, Vol.7, 217-213.

[Sidner 86] C. L. Sidner, *Focusing in the Comprehension of Definite Anaphora* in Readings in Natural Language Processing, eds B.J. Grosz, K.S. Jones, B.L. Webber, 1986.

[Wada 90] H. Wada, *Discourse Processing in MT: Problems in Pronominal Translation*, COLING-90, pp73-75.