Machine Translation: Productivity and Conventionality of Language

JUN'ICHI TSUJII
UMIST

Abstract

The linguistics-based machine translation (LBMT) has been a dominant framework in MT research since the beginning of the eighties. However, I argue that several assumptions on which the research in LBMT has been based do not hold in translation of actual texts. In particular, I discuss why the notions of possible translation and compositionality of translation, both of which have their roots in monolingual studies of syntax and semantics, have been wrongly promoted by theoretical linguists and how these notions have (mis)lead the researchers in MT in wrong directions. Then, I discuss how we should proceed in the future and what types of research should be pursued. In conclusion, I illustrate what an ideal architecture for MT systems should look like.

1 Introduction

There had been strong interest and high expectations in Machine Translation throughout the 80s in the research community, commercial industry and among potential users of MT systems. However, the interest and high expectations seem to be waning somewhat in the 90s. This is partly because quite a few commercial products have been brought onto the market and MT systems have become common to the general public as well as these communities. This is also because people's expectations become more modest and reasonable. People are beginning to realise now that MT systems are not very special but are simply ordinary information processing tools.

It is generally a good thing that people have a clear picture and reasonable expectation of this new technology. However, it is also the case that the current MT technology does not meet the initial expectations that people had and that the current MT systems do not cover the demands of a potentially very large translation market.

There is also acute frustration about the fact that theoretical research in MT has not contributed to the development of MT systems at all. Though it generally takes some time for the results of research to be reflected in commercial products, it is, nonetheless, frustrating.

In this paper, I will discuss what went wrong with theoretical research in MT and what is lacking in the current MT technology to bridge the gap between research and development.

2 Disappointment

While there are many successful applications of MT systems, there is also disappointment about the current state of MT systems among people who have either invested in the field or been involved in research and development.

Investors are disappointed because the market for MT systems as they are now is much smaller than they thought. Users are disappointed because the quality of translation produced by systems does not meet their standards. Researchers are disappointed because their methodologies failed to deliver what they thought they could.

Part of the disappointment is due to the fact that their expectations at the beginning of the 80s were unrealistically high. However, even those who claim that current MT systems are quite successful (actually, I am one of them) may admit that the success of the current MT is a fairly restricted one. In order to widen the range in which MT systems can be used and in order not to repeat the same mistakes in the future, we have to learn lessons from past experience, in particular the experiences of these last 10 years. The following are the lessons I think we can draw from past experience.

1. Linguistics in a narrow sense is not as useful as we expected.
2. MT systems as we conceived of at the beginning of the 80s do not meet actual market demands.
3. There is no such thing as a universal MT system.

These lessons may sound all too familiar to those who were engaged in MT in the 50s and 60s (unfortunately, I was not). The first one, for example, simply says that linguistics alone is not able to solve most of the problems MT systems encounter. As Bar-Hillel claimed long time ago, translation requires understanding, which in turn requires real-world knowledge.

However, I do not claim that *Therefore, in order to improve quality of MT we have to integrate understanding or processing based on real world knowledge with MT*. This has been claimed since a long time ago and serious attempts have been made during the 80s, which were equally unsatisfactory. We cannot be so naive now.

I start with two myths in theoretical research on MT, which have influenced the way of thinking in the whole research community and which I believe has lead the research field in a wrong direction.

3 Myth-1: Compositionality of translation

Language is infinite. The infiniteness of language is the main cause of difficulties in NLP applications including MT. The linguists who have been involved in MT since the beginning of the 80s have emphasised the importance of how to cope with the infinite nature of language.

The solution they propose is *compositionality* of translation. Like compositionality of meaning in mono-lingual theories, it associates translation with linguistic structures of some sorts. That is, translations (or meanings) of complex expressions are determined by their parts, and the relationship between the complex expressions and their parts are, for example, determined by their phrase structures. Though most of MT systems use more *abstract* levels of representation than phrase structures, the basic scheme remains the same.

The strict form of compositionality of translation seems to be based on the following two assumptions.

[ASP 1] Translation equivalence by identity of meaning: Assuming the existence of meanings which are independent of context, translation equivalent expressions in different languages have the same meanings.

[ASP 2] Independent status of structure equivalence: Assuming that a complex expression in one language can be decomposed into its sub-expressions with a constructor[1], the translation can be constructed from translation equivalences of the sub-expressions, by using the constructor of the target language which is translationally equivalent to the constructor of the source. Like Montague's semantic theory, they assume that translation equivalence of constructors in two languages can be established, regardless of sub-expressions which are combined by them.

Most of the transfer-based MT systems assume, to varying degrees, that **[ASP 2]** is the case. The transfer phase descends down the structural

[1] Constructors can be syntactic, semantic, etc. In the case of syntax, constructors can be individual phrase structure rules or grammatical functions such as SUBJ, OBJ, etc. In semantics, they can be deep cases or thematic relations such as Agent, etc.

description of a source sentence from top to bottom and at each level, it decomposes a complex expression into its sub-expressions and constructor. Then, it ascends from bottom to top to construct a target sentence, at each level of which it composes a complex expression by using a translation equivalent constructor.

If **[ASP 2]** is really the case, the transfer phase is a simple recursive process as described above. However, developers of MT systems which are being used for actual translation know, through their experience, that the transfer phase cannot be so neat as described above.

There are many cases where the independent status of constructor equivalence is challenged and equivalences of constructors are affected by sub-expressions to be combined. Such cases are abundantly observed in such examples as terminological expressions, lexical gaps, idioms, pseudo-idioms, speech patterns (Alshawi 1991), etc.

While we know that the naming or labeling of real world entities by words are arbitrary and simply the conventions of individual languages, conventionality of language use is much more pervasive than we thought. In other words, conventionality permeates the other aspect of language use, i.e., productivity, which *compositionality of translation* emhasises. These conventionalised expressions cause difficulties for compositional theories in general, but they are more serious in translation, because two languages have their own different conventions.

There are basically two alternative ways of treating problems. One is to admit empirical facts and demote the status of constructors. MT systems based on lexicon-oriented views, perhaps inadvertently, took this path. They use mono-lingual constructors only as descriptors to define a translation equivalent pair which contains a specific word or words. In their frameworks, there is no such thing as a constructor equivalence (or structure equivalence). As a result, structure transfer is performed as part of lexical transfer.

The other alternative is to ignore empirical facts and push **[ASP 2]** to the extreme. If **[ASP 2]** is the case, one of its logical consequences is the possibility of discovering a set of universal constructors. While to establish a set of universal lexical items is hard simply due to the sheer size of vocabulary, the number of constructors seems fairly small, whether they are syntactic, semantic or pragmatically motivated ones. EUROTRA seems to have taken this line of reasoning and reached the idea of *simple transfer*. As in lexicon-oriented MT systems, structure transfer is eradicated, but for a very different reason. In this framework, the status of constructor

equivalence obtains supreme independence and is actually represented as universal constructors. Though lexical transfer can change the structure, this is treated as an exception.

ROSETTA also maintained **[ASP 2]** and tries to co-ordinate the constructors of two languages (Appelo 1987; Landsbergen 1989). The result seems to be proliferation of constructors which cannot be justified monolingually.

The results of the two attempts, EUROTRA and ROSETTA, show that **[ASP 2]** is empirically wrong and that frameworks based on this assumption do not work.

[ASP 1] is explicit in a naive interlingual approach, in the sense that the meaning which guarantees translation equivalence of expressions in various languages is explicitly represented at the level of interlingua. However, all sentence-based MT systems implicitly share this assumption. As the ROSETTA group rightly claims, the meaning which guarantees the translation equivalence need not be explicitly represented, but when one defines two expressions (structures and/or words) as translation equivalents, one assumes that the meanings of the two are the same.

More precisely, when *compositionality of translation* decomposes translation equivalents of complex expressions into translation equivalents of sub-expressions, it implicitly assumes that the meanings of these expressions (the sub-expressions as well as the complex expressions) are the same as those of their corresponding target expressions and that the meanings which matter can be established independently from the contexts. Thus, translation equivalence of a larger expression can be reduced to a collection of translation equivalences of smaller expressions, the equivalences of which are established regardless of the larger units of expressions.

However, **[ASP 1]** is very doubtful from an empirical point of view. It is rather normal than exceptional in human translation that extra phrases or words are added or that phrases or words in the source disappear in the target. Obviously, in human translation, two translation equivalent sentences do not have the same meanings which are established independently of context.

More seriously, as we will see in Section 5, the context independent meaning postulated in **[ASP 1]** is not so crucial in translation, but context dependent interpretation plays the decisive role in translation. As we discuss later, this context dependent interpretation, together with conventionality of language use, affects translation in subtle ways and makes *compositionality of translation* an irrelevant straightjacket.

4 Myth-2: Possible translation

Since the early 80s, research in MT has been getting more and more similar to research in theoretical linguistics of a certain type. Both heavily rely on human intuition. As linguists (of a certain type) do, researchers in MT tend to ignore phenomena occurring in real translation by human translators, pick up artificial examples they produce and thus put disproportional emphases on certain specific problems.

Though it is problematic in some cases, grammatical judgement by intuition works and has played a major role in monolingual research on syntax. However, the same methodology has not worked for MT research so well. To judge *correctness* of translation and thus define translation equivalence of a single sentence without considering context has turned out to be much more problematic. The same sentence can and should be translated differently, depending on the *context* in which it appears[2]

Because it is generally difficult to circumscribe the context in which a sentence appears, the judgement become subjective, i.e., the judgement depends on the context which a reader of that sentence happens to come up with. Researchers in the linguistics-based MT paradigm tried to dissociate translation from context by introducing the distinction of *possible* translation and *good* or *correct* translation, the distinction which reminds us of the distinction of *competence* vs. *performance*. They argue that, given a context, only a subset of the possible translations are *correct* ones, and that one has to concentrate on *possible* translation in theoretical research.

However, the distinction seems more problematic and fragile than the distinction of *competence* and *performance*. Firstly, as S.Nirenburg rightly pointed out, it avoids problems related with ambiguities, which pose real difficulties in actual MT systems.

More seriously, unlike grammatical judgement (in which native speakers have to say only *yes* or *no*), one has actually to generate all instances of possible translations of a given expression in every conceivable context. Without serious empirical investigation, it is very difficult, if not impossible, to generate all possible translations for a given sentence in a contextual vacuum. As a result, while *ideally* a set of *possible* translations has to be determined independently of a theory or a particular system, those who are engaged in MT development or MT theory, not translators, determine such a set by themselves. The consequence is that the definition of *possible*

[2] I may sound that I emphasise the context dependency of translation. But this is not my intention in this paper. See Section 5.

translation becomes a theory internal concept. In short, a set of possible translations is defined as a set of translations which a given system (or theory) produces but from which a system (or theory) cannot choose *correct* ones (due to the lack of context, etc.). A set of *possible* translations as such have nothing to do with a set of translations which actually appear as translations in real texts (see Section 5).

In short, the concept of *possible* translation provides a convenient excuse for researchers to play with toys, and contributes to cutting theoretical MT research off from its empirical basis. It has lead the whole research in a wrong direction.

It is also obvious that, due to this excuse, researchers have been able to ignore the obvious fault of **[ASP 1]**. On the other hand, **[ASP 1]** gives the illusion that translation problems can be discussed without referring to context (because translation equivalent relationships can be defined in terms of context independent meanings), and reinforces the myth of *possible translation*.

5 Examples: Metonymic nature of language and translation

Let us see several examples to illustrate the points I have made.

[Fact 1] (Kitamura & Matsumoto 1995) reported that only 236 Japanese-English word pairs are registered in one of the most comprehensive bilingual dictionaries for human use, out of 948 word-pairs which their alignment program discovered from real texts (24%).

[Fact 2] We examined the manual of UNIX to find that, among 15 Japanese equivalents for the English verb *to match* listed in an English-Japanese dictionary, only two Japanese equivalents (*taiousuru* and *icchisuru*), appear as translations for 125 occurrences of the word.

These facts show that even *possible* translations given by lexicographers, who are more empirical than linguists or computer engineers, do not reflect actual translations produced by translators. To see why the discrepancies like **[Fact 1]** arise, let us consider the following simple example.

[Example 1] (by Jiping Sun, UMIST)
English : I will go to see my GP tomorrow.
Japanese : `Watashi(-I)-wa Asu(tomorrow) Isha(-GP)-ni Mite(-check) Morau(beneficiary causative).`

Literal Translation of Japanese: I will ask my GP to check me tomorrow (and I will benefit from the action)

While a compositional translation of English into Japanese such as:

`Watashi`(-I)-`wa` `Asu`(-tomorrow) `Isha`-(GP)-`ni` `ai`(-meet)-`ni` `iku`(-go)

is possible, this translation implies that the speaker will meet his/her GP to discuss about something unusual (like mis-diagnosis, fees etc).

What is happening in this example is that, although the two languages, English and Japanese, describe the same situation (an aggregation of actions), they verbalise different actions in the aggregation. Which aspect of a complex reality a language verbalises is somewhat fixed, and when one does not follow the convention, additional meanings are conveyed.

The process of human translation of the above example is roughly described by using the KBMT framework and a Schankian type of representation, as follows.

[Step-1: Understanding] The first phase is to understand what situation is described. The result of understanding would be, though naive, represented by something like:

AGGR-1 `[the speaker GOES to some place like a hospital],`
`[s/he and her/his GP MEET]`
`[the GP CHECKs him/her]`, etc.

This step uses knowledge about conventions in English that the expression "go to see one's GP" is used to describe a situation, which can be described by AGGR-1.

[Step2: Paraphrase] This phase is to choose which part of AGGR-1 is to be verbalised in Japanese, following the convention of Japanese. That is, a human translator knows that in Japanese a situation like AGGR-1 is described by verbalising the part of it, i.e., `[the GP CHECKs the speaker]`, and using a beneficiary causative to express the speaker's initiative (in English, this part is expressed implicitly by [I go] and [I see my GP]).

The actual process would be more complicated. The understanding step in human translation is a more dynamic and flexible process which associates the compositional meaning of *to go to see one's GP* with a typical situation (like AGGR-1) of someone visiting his/her GP. This interpretation process definitely uses general knowledge and context, either by inference or by association. Because of the dynamic nature of the interpretation phase,

if the context indicates that an unusual incident happened between the speaker and his/her GP, human recognises it and the same compositional meaning of the sentence would be linked with another aggregation. Because the GP may not check him/her up medically in such circumstances, Japanese translation would be different.

Or, the paraphrasing phase would be equally more dynamic. That is, if the context suggests that the speaker's action of *going* is crucial, then the second phase has to choose a different construction in which *go* is verbalised as the main verb and *ask GP to check him/her* is realised as a subordinate clause.

Though it may sound trivial, **[Example 1]** illustrates how human translation is performed through understanding of what is actually described. However, it is not my intention to emphasise the context dependency (or dynamic) nature of human interpretation or paraphrasing. My point here is the interaction of metonymic nature of language and conventionality of language use, which makes compositionality of translation irrelevant to actual translation and which is revealed even in the translation example given in **[Example 1]**.

That is, a sentence in one language describes metonymically part of the complex reality which it intends to describe, and which part of the same reality is explicitly expressed depends on individual languages. It is obvious that understanding results such as AGGR-1 are completely different from the meanings intended in **[ASP 1]**, which are established independently of context and which can be computed from context-independent meanings of individual words like *go*, *see* and *GP*. As a result, we have rather strange translation pairs, strange from the compositional view of translation, such as a pair of **[X sees Y]** and **[Y check X]** even in a *normal* situation. Such a correspondence can hardly be imagined when one tries to enumerate possible translations of *to see*. [3]

However, as **[Fact 1]** indicates, it seems that such ad-hoc correspondences are rather the norm than exceptions. In reality, it does not make any sense to discuss correspondences in terms of linguistic structures of two sentences, because the two sentences describe different parts of the reality and

[3] Though this example is somewhat similar to the well-known example (miss - manquer in English and Frech) – and I feel some continuum –, this correspondence is very specific unlike (miss, manquer). If *doctor* is replaced by *my lawyer*, then we have to use *discuss* or *consult* instead of *check* and the causative construction is no longer approriate.

their compositional meanings are consequently very different. The basic assumption of *compositional* translation does not hold.

The example given by M.Kay, *validate a ticket* and *invalidate a ticket* in French and German, illustrates a similar point that different conventions of verbalisation shared by the two speech communities result in un-conceivable translation pairs like ([**X validate Y**], [**X invalidate Y**]). The same state of affairs is expressed metonymically in Japanese by focussing on a particular action like *punch a ticket*, and thus we end up with equally un-conceivable pair like ([**X validate Y**], [**X punch Y**]), which no one expects to be in a bi-lingual dictionary.

So far, we have discussed rather general examples. [**Fact 2**], however, indicates a different aspect to problems of *possible* translations. That is, given the fixed context of Unix[4], a set of possible translation which lexicographers enumerate without context is simply too large and thus makes the problem of disambiguation unnecessarily difficult.

Furthermore, if a specific context such as Unix manuals is fixed, we have much more conventions which texts have to conform to. Here again, different languages follow different conventions.

[**Example 2**] Maruyama (1992) observed drastic structure changes are often required in a production manual of mechanical devices such as follows.
Japanese: `Buhin(-parts)-no Iro(-colour)-ha Hyo(-Table)-2-ni yoru-mono-to suru.`
English : See Table-2 for the colours of the parts.

While the Japanese sentence literally means *As for the colour of parts, one is supposed to follow Table-2*, there are no corresponding expressions *follow* and *be supposed to* in the English translation. On the other hand, *see* appears in English.

This is because the manual in the two languages follow different conventions to express the same information, i.e., *the colour of the parts are listed in Table 2*. Maruyama said that there are many such significant structure changes in the manual he examined. Again, it seems ridiculous to claim that the two sentences have the same compositional meanings.

[4] I use the term *context* in a broader sense, which includes the communicative environment where a text is prepared.

6 Conceptual design of a simple MT system

In the previous section, I claim that human translation is based on understanding of what is described or what is intended. However, this does not imply, for example, that an MT system as an engineering system has to represent understanding results such as AGGR-1 explicitly and simulate the human process of *interpretation* and *paraphrasing*.

First of all, though I used an informal Schankian script to represent understanding results, it is not at all clear how actually we can represent them in a computationally sound way. As Kay's example illustrates, the symbolism at this level is not trivial at all (how can we represent *a ticket is valid* or *a ticket is not valid* without referring to the whole social system associated with *tickets* ?).

Secondly, even if one could represent them, how can one relate them with the target-oriented paraphrasing ? The target language may have some general principles that determine which part of a complex reality should be expressed explicitly. Unfortunately, we know almost nothing about this process. Unless the process which manipulates objects like AGGR-1 and which can change dynamically the interpretation and the paraphrase is realised computationally, to represent them explicitly does not contribute to MT.

Thirdly, there are many cases like **[Example 2]** and Kay's example, where conventions about how to verbalise are really specific to individual events or the information to be described. In other words, they have characteristics similar to terminological expressions, and we have to treat them in the same way as we treat terminological expressions. Neither structure of expressions nor the internal structure of understanding results are not crucial for translation.

Let us assume here the extreme. That is, conventionality, not productivity of language plays a dominant role in translation and that the role of translation is to transfer conventions of one language to the corresponding conventions of another language. And assume that almost all linguistic expressions have characteristics similar to terminological expressions. In other words, complex expressions in one language are related with expressions in another language, regardless of their internal linguistic structures. As correspondences between terminological expressions (terms) are often expressed through language independent concepts, let us use objects like AGGR-1 as links without analysing their internal structures.

Then, the correspondences in the two examples would be represented as follows:

ENG:[e(X) go to see e(X's) GP]
↔ **AGGR-1[X]**↔
JPN:[j(X)-wa j(X)-no isha-ni mite morau]

ENG:[See e(X) for e(Y)]
↔**AGGR-2[X, Y]**↔
JPN:[j(Y)-ha j(X)-ni yoru mono-to suru]

I use an informal notation in the above such as e(X) and j(X) which means translation of X in English and Japanese, respectively. X is a variable in AGGR-1.

Unlike pure terminological terms, the expressions to be related in the above examples contain variables like X and Y. In order to make the correspondence of the first one more general, one can introduce another variable with its own restriction such as:

ENG:[e(X) go to see e(X's) e(Y)]
↔ **AGGR-1[X, Y]**↔
condition[Y is a medical-profession]
JPN:[j(X)-wa j(X)-no j(Y)-ni mite morau]

This rule can be compared with another rule, which is concerned with a similar but different situation like *to go to see one's lawyer* and which leads to a different translation in Japanese, like:

ENG:[e(X) go to see e(X's) e(Y)]
↔ **AGGR-3[X, Y]**↔
condition[Y is a legal-profession]
JPN:[j(X)-wa j(X)-no j(Y)-ni soudan-ni iku]

Here we have a word *soudan* (*consult* in English) in Japanese and the beneficiary causative is no longer used.

As you easily see, these correspondence rules look like transfer rules which are used in actual commercial MT systems. They speculate correspondences which can hardly be justified on a purely linguistic basis. They are also very specific in the sense that they contains many individual words like *see*, *go miru*, *soudan* and *morau*. Furthermore, they introduce rather ad-hoc classification of nouns like *medical-profession* and *legal-profession*.

While researchers with theoretical orientations considered rules such as these awkward and ad-hoc and started to tide them up, it seems to me that they actually reflect certain essential aspects of translation. What was

essentially wrong with their attitude is that they took these exceptions to general rules or simply ignored empirical facts alltogether in translation.

The framework is obviously naive in many points. In particular, it may need to have structural annotations in ENG and JPN in order to use these correspondences for constructing larger expressions which contain them as parts. It may also be desirable for the structural annotations to be done at a certain abstract level, in order to allow some freedom in the generation phase, etc. However, structural annotations as such play only the roles of descriptors for defining correspondences between complex expressions, and do not have the independent status on which translation equivalence is defined.

7 Other frameworks and future directions

I have illustrated how rules in an MT system I have in mind look like, and indicated that they have more similarities with rules in traditional transfer-based systems than with those in *linguistically motivated* proto-type MT systems. The reason for this is that linguistically motivated research has made several wrong assumptions on translation, notably those related with *compositionality* of translation and *possible* translation, which miss the true nature of translation.

However, several other research frameworks like EBMT (Example-based MT) (Nagao 1984; Nagao 1992; Sumita 1991; Furuse 1992; Jones 1992) SBMT (Statistics-based MT) (Brown 1992) and KBMT (Knowledge-based MT)(Nirenburg 1989) have been proposed and have attracted more and more interest of the research community. These frameworks have taken orientations distinctly different from linguistically motivated MT (LBMT) and do not have the defects I discussed in this paper. EBMT and SBMT which use translation produced by human translators as a major source of knowledge, for example, will never be separated from empirical facts in translation.

In the following, I will summarise my points by referring to these frameworks as well as the traditional transfer paradigm.

[1] If the nature of transfer rules are like the ones I described, then recursive transfer which heavily relies on the structure of a source sentence may not work well. The straightjacket imposed by the recursive transfer has to be relaxed. The transfer process is more like the process of solving jigsaw puzzles.

[2] The framework in Section 6 and EBMT share many things in common. The major difference is whether one introduces variables and their conditions explicitly or not. The framework in Section 6 assumes a phase of Knowledge Preparation in which individual correspondence rules are identified (e.g., how many distinct AGGRs have to be established) and parts to be represented by variables are identified. In this phase, one has to examine translations given by human translators.

The obvious advantage of EBMT is to avoid this Knowledge Preparation phase and just use examples as *transfer rules*. However, as several groups in the EBMT camp admit, the careful scrutiny of examples is vital for success of EBMT, which implies that EBMT also has to have a knowledge preparation phase of some sort.

[3] In Section 6, I argued as if disambiguation between AGGR-1 and AGGR-3 are to be made simply by referring to the properties of Y. However, there is no guarantee that disambiguation is possible only by examining the internal structures of expressions to be transferred, although most traditional transfer-based systems assume that this is the case.

As we saw in Section 5, the same expression *go to see one's GP* has to be related with different AGGRs, depending on the context in which the expression appears. Whether disambiguation has to be performed by a rationalistic way (through explicit understanding of what a text describes, like in KBMT) or otherwise remains to be seen. The key problem here is how to characterise context which affects selection. EBMT, for example, provides an empiricists' alternative to the problem. It is also plausible to apply methods, statistical or connectionist ones, which have been proven to be effective for sense disambiguation of lexical items.

[4] I treat AGGRs as un-analysable wholes, which directly connect expressions of two languages. It might be possible, as the KBMT camp and (Dorr 1994) have done to a certain extent, to analyse the internal structures of AGGRs and discuss (and implement) the processes of *interpretation* and *paraphrase*, which dynamically associate linguistic forms with AGGRs. However, considering our current understanding of these processes, it would be too ambitious to map a source text to AGGRs and then generate a target text.

[5] Although I have not discussed in this paper, I share with many people the belief that there is no such thing as a universal MT system. The belief has two consequences.

One consequence is that every MT system has to be tuned towards specific subject domains and text types. In my model, for example, a set of correspondence rules have to be prepared for every different sublanguages (Ananiadou 1990). Consequently, the knowledge preparation phase plays a crucial role, and I believe that the technology of automatising this phase will be vital for broadening the range of MT application (Tsujii 1992; Tsujii 1993).

The other consequence of this belief is that the architecture of MT systems has to be diversified. Which architecture, Transfer-based MT, EBMT, KBMT, LBMT or the framework illustrated here, is most appropriate is highly dependent on the complexities of translation required by a given sublanguage and the functionality of a system required in a specific application.

REFERENCES

Alshawi, Hiyan, D. J. Arnold, R. Backofen, D. M. Carter, J. Lindop, K. Netter, S.G. Pulman, J. Tsujii & H. Uszkoreit. 1991. EUROTRA-*6/1 Final Report*. Technical Report produced for the Commission of the European Communities, Luxembourg. Cambridge: S.R.I. International, Cambridge Computer Science Research Centre.

Ananiadou, Sofia. 1990. "Sublanguage Studies as the Basis for Computer Support for Multilingual Communication". *Proceedings of Terminology and Planning (Termplan-90)*, 10-13. Kuala Lumpur.

Appelo, Lisette, C. Fellinger & J. Landsbergen. 1987. "Subgrammars, Rule Classes and Control in the ROSETTA Translation System". *Proceedings of 3rd Conference of the European Chapter of the Association for Computational Linguistics (EACL'87)*, 118-133. Copenhagen, Denmark.

Brown, Peter F., S. A. Della Pietra, V. J. Della Pietra, J. D. Lafferty & R. L. Mercer. 1992. "Analysis, Statistical Transfer, and Synthesis in Machine Translation". *Proceedings of the 4th International Conference on Theoretical and Methodological Issues in Machine Translation (TMI-92)*, 83-100. Montreal, Canada.

Furuse, Osamu & H. Iida. 1992. "Transfer-Driven Machine Translation". *Proceedings of the 2nd International Workshop on Fundamental Research for the Future Generation of Natural Language Processing (FGNLP)*. Manchester: Centre for Computational Linguistics, UMIST.

Maruyama, Hiroshi & H. Watanabe. 1992. "Tree Cover Search Algorithm for Example-Based Translation". *Proceedings of the 4th International Conference on Theoretical and Methodological Issues in Machine Translation (TMI-92)*, 173-184. Montreal, Canada.

Jones, Daniel. 1992. "Non-Hybrid Example-Based Machine Translation Architectures". *Proceedings of the 4th International Conference on Theoretical and Methodological Issues in Machine Translation (TMI-92)*, 35-43. Montreal, Canada.

Landsbergen, J. 1989. *The Power of Compositional MT*. Eindhoven, The Netherlands: Philips Research Laboratories.

Nirenburg, Sergei. 1989. "Knowledge-based Machine Translation". *Machine Translation* 4:1.5-24.

Nirenburg, Sergei, J. Carbonell, M. Tomita & K. Goodman. 1992. "Machine Translation: A Knowledge-based Approach". San Mateo, Calif.: Morgan Kaufmann.

Nagao, Makoto. 1984. "A Framework of a Mechanical Translation between Japanese and English by Analogy Principle". *Artificial and Human Intelligence* ed. by A. Elithorn & R.Banerji, 173-180. Amsterdam: North-Holland Elsevier.

Nagao, Makoto. 1992. "Some Rationales and Methodologies for Example-Based Approach". *Proceedings of the 2nd International Workshop on Fundamental Research for the Future Generation of Natural Language Processing (FGNLP)*. 82-94. Manchester: Centre for Computational Linguistics, UMIST.

Sumita, Eiichiro & H. Iida. 1991. "Experiments and Prospects of Example-Based Machine Translation". *Proceedings of 29th Meeting of the Association for Computational Linguistics (ACL'91)*, 185-192. Berkeley, California.

Tsujii, Jun'ichi, S. Ananiadou, I. Arad & S. Sekine. 1992. "Linguistic Knowledge Acquisition from Corpora". *Proceedings of the 2nd International Workshop on Fundamental Research for the Future Generation of NLP (FGNLP)*, 61-81. Manchester: Centre for Computational Linguistics, UMIST.

Tsujii, Jun'ichi & S. Ananiadou. 1993. "Knowledge-based Processing in MT". *Knowledge Building and Knowledge Sharing* ed. by K. Fuchi & T. Yokoi, 68-85. Amsterdam, The Netherlands: IOS Press.