

## ON EVALUATING MT SYSTEMS

*Jacques Durand*

*University of Salford*

### INTRODUCTION

The more I think about the title I selected for this paper, the more I am struck by its arrogance given, not only the time available, but also the limited competence I have in a number of areas which are directly relevant to the assessment of MT systems. Having recently perused a good sample of the forthcoming book *An Introduction to Machine Translation* by Hutchins and Somers did not help matters. Their chapter 9 'Evaluation of MT systems' offers an excellent overview of this specialised domain and all I can do is refer you to this chapter and the references therein for a thorough discussion.

What I shall be concerned with is a rather limited aspect of the assessment of MT systems: I will look at this question from the point of view of a potential user and I will mainly concentrate on this problem from a linguistic standpoint. The kind of evaluation I will be dealing with is 'black box' evaluation rather than 'glass box' evaluation. That is, I will assume that the examination of a given system is not done by developers/researchers but by users who have access only to inputs and outputs and who are particularly interested in the linguistic aspect of a system's performance.

One warning is in order from the outset: my potential user is in a sense a 'mythical beast'. Real users are not disembodied entities but businessmen, translators, teachers, students, MT specialists, computational linguists or whatever. Each type of background will trigger a different range of expectations. Moreover, evaluations are constrained by the type of system one is faced with: Are we dealing with a research system? Or with a near commercial system? Or is the system part of the translation work of a whole business team? How much post-editing is acceptable to these translators? These questions and others are highly relevant to the practical assessment of real world systems. The recently published paper by Isabella Moore 'Criteria for selection MT systems' (1990) provides a good example of the factors governing the choice of an MT system within a small business environment. By comparison, much of what I will say further on abstracts away from these problems. It is nevertheless my belief that the majority of the issues I broach in this paper have to be addressed if we do not want our evaluation to be random or 'dilettante' if I may be allowed a little foray into the French language.

### QUALITY

Whatever system one is concerned with, the question of 'quality' of translation arises. The assessment of quality is usually broken down into broad categories of assessment. Traditionally, the translation of sentences would be split into 'form' and 'content' but this division is too inclusive to be of much use. In Hutchins and Somers (1992), it is suggested, along with other work in MT, that three criteria should be taken into account: (i) fidelity, (ii) intelligibility, (iii) style.

'Fidelity' requires that the same 'information' be conveyed by the target text and the source text. Various tests can be put forward. For instance, within the ALPAC investigation, people were asked to read the output and compare them with the original in terms of 'information'. Nowadays, in assessing the instruction manuals which accompany various products (say a coffee machine), the translation can be assessed in terms of a reader being able to carry out the tasks described in the translation. 'Intelligibility' typically involves a ranking of the text from perfectly intelligible to unintelligible. Various techniques can be used to refine measures of assessment: for instance the Cloze technique which involves the masking of a number of words, then asking readers to fill in the blanks and finally comparing the guessed words with the original. The relationship between fidelity and accuracy is an interesting one to study and it was the conclusion of Carroll's experiments, as part of the ALPAC report, that averaged over sentences and raters, fidelity and intelligibility scores were highly correlated. I can only refer you here to the paper by Lee Humphreys in this volume and the report he gives on evaluation criteria devised by the Essex MT group. The third criterion 'style', according to Hutchins and Somers, is "as subjective as the global ranking of intelligibility" but they do point out that stylistic factors are important. Thus, in instructions manuals, very often, the translation from English into French of an imperative should be an infinitive and not an imperative, e.g.:

- (1) (a) Depressurize hydraulic system
- (b) Dépressuriser le système hydraulique.

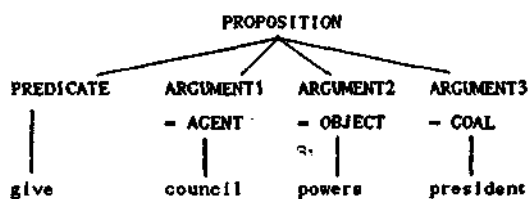
The overview by Hutchins and Somers is a very useful one. From a linguistic point of view, I will now present an alternative scheme which should cover all the points they make in a way which complements their approach rather than compete with it. In so far as possible, I wish to argue that good quality translation involves the following components in relation to sentences within a text. It should preserve: a) propositional content, b) thematic information, c) pragmatic interpretation, d) stylistic value. I will not introduce a separate category of 'textual information' as the role of co-text and context is relevant for each area and induces in each case a partition in terms of higher vs. lower quality. A recent book which discusses the various parameters of human translation in an interesting way from a linguistic point of view is Basil Hatim and Ian Mason's *Discourse and the Translator* (1990).

### **Propositional Content**

The propositional content of a sentence is what is sometimes called the 'message' or the 'gist' of the sentence in everyday speech. In more technical parlance, specialists would speak of preserving the 'cognitive content' of a sentence, its truth-value, or its 'ideational' meaning in the linguistic theory of M.A.K. Halliday (see Halliday, 1985 – but for our purposes we can neglect the differences between various accounts of meaning). To take a simple example, consider a family of sentences such as:

- (2) (a) Council gave these powers to the President
- (b) Council gave the President these powers
- (c) These powers were given to the President by Council
- (d) The president was given these powers by Council

We expect an adequate MT system to give us a translation based on the information that there was an act of giving involving three participants (a source or agent *Council*, an object *these powers* and a goal *the President*). Put semi-formally, a basic type of information to be conveyed is a schematic predicative structure of the form:

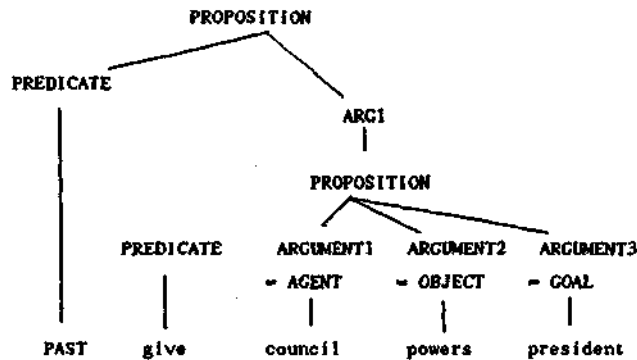


**Figure 1**

In addition, the information must be conveyed that the action of giving took place in the past, which we could express semi-formally as in Figure 2.

These trees are not given here as model representations but as a way of coding the information in question. In so far as the cognitive content of sentences is preserved – a mixture of fidelity and accuracy – a system can be said to be successful. Whether to achieve this it has resource to statistical techniques or linguistic techniques or a mixture of the two is not pertinent to the point at issue here.

It must be stressed that recovering the propositional content from surface strings is by no means a trivial matter. It requires among other things that a vast amount of disambiguation, both lexical and grammatical, has taken place. Leech in his introduction to *The Computational Analysis of English: a Corpus-based Approach* (edited by Garside, Leech and Sampson, 1987) notes that the grammatical analysis of the Brown corpus (officially known as the Brown University Standard Corpus of Present-day American English), which was based on a system called TAGGIT, had a success-rate of 77-78%. He adds: "If this strikes the reader as an unremarkable achievement, it is worth bearing in mind that English is a notoriously ambiguous language with respect to homography – for example in the eight-word sentence *Norman forced her to cut down on smoking* each word is grammatically ambiguous, so that about 3500 different word-class labellings of this sentence (in terms of the labels provided by the LOB tagset) are in theory possible" (p. 9).



**Figure 2**

From the point of view of translation, ambiguity resolution is obviously fundamental for the preservation of propositional content. Consider, for instance, the following sentence from an article in *Le Monde* on the Greenpeace affair (part of a passage quoted in Hatim and Mason, 1990):

- (3) A ce stade, il est impossible de savoir si ces trois personnalités sont directement impliquées, ou simplement concernées en raison de malentendus et de non-dits lors des discussions sur Greenpeace (Le Monde, 18.9.1985).

which was translated in the Guardian as follows:

- (3') At this stage, it is impossible to know if these three were directly implicated or simply involved through misunderstandings and incomplete information during discussions on Greenpeace (The Guardian, 18.9.1985).

We need to know that FR *stade* is not used in the sense of "stadium" (a case known as 'homograph resolution'). We also need to be able to determine that the preposition *sur* in *des discussions sur Greenpeace* is triggered by *discussions* and not to be attached to *Greenpeace* and therefore could also be translated by *about*. There is a small syntactic clue showing that *sur* does not form a modifier group with *Greenpeace*, namely the absence of an article. If the discussions had taken place on the boat called Greenpeace, the French would have to be *des discussions sur le Greenpeace*. Such examples could be multiplied ad infinitum and I will spare you a long list of translation equivalents in various languages. What should be realised,

however, is that fully determining the propositional content of an utterance must involve some strategy for accessing discourse and background information.

First of all, many ambiguities can only be resolved if the co-text and the context are taken into account. Secondly, when we talk informally of the message conveyed, we assume that the reference of pronouns has been fixed. But, in very many cases, the internal structure of the sentence does not give sufficient clues to resolve the matter. Much work in modern syntax helps to establish the sentence-internal conditions under which co-reference is allowed or forbidden but it does not offer a strategy for actually establishing co-reference in real texts. The work done by syntacticians is indispensable as shown by a brief consideration of (the simple examples in (4):

- (4) (a) Mr Delors spoke to him  
 (b) He spoke to Mr Delors  
 (c) Mr Delors wrote the report after he left  
 (d) After he left, Mr Delors wrote the report

We can see that in (a), *him* cannot refer to Mr Delors, nor *he* in (b). On the other hand, in (c) and (d), co-reference between *he* and *Mr Delors* is possible. Professional translators can effortlessly take such facts in their stride but building all the appropriate conditions into an MT system is no mean achievement. And, in addition, as pointed out earlier, we need to actually fix the reference of pronouns to translate adequately between languages. 'They' is neutral between male/female and human/non-human in English but French, like many other languages, forces us to state whether it is 'ils' or 'elles'.

Of course, it may be quite easy in post-editing to settle such matters but the more ambiguity is left on various fronts the less acceptable the output will be. On the basis of the points I have just sketched, we can therefore divide systems between those that can capture the translation of propositional content on a sentence internal basis only and those which, at least in a number of areas, can access discourse information to determine the meaning of various lexical and phrasal units.

### Thematic information

By thematic information here I shall be referring to the way that each utterance is organized as a message in terms of topic-comment or theme-rheme in Hallidayan terminology (hence the term 'thematic' in the use adopted here). Thematic structure is closely related to the organisation of each message as a piece of information transfer (sometimes referred to as given vs. new information) and, more generally, it is part of the textual function of language. A passive, to take the most salient example of thematic information, is not simply equivalent to an active although in many cases active and passive sentences have the same truth-value. In attempting to characterize thematic information, we would work from families of sentences such as the one below which involve actives, passives, cleft sentences, topicalized sentences, etc.:

- (5) (a) The Government rejected the proposal  
 (b) The proposal was rejected by the Government  
 (c) It was the Government that rejected the proposal  
 (d) It was the proposal that the Government rejected  
 (e) It was the proposal that was rejected by the Government

A number of translation routines can indubitably be based on sentence information alone with reasonable accuracy. But, here again, there is no doubt that much work remains to be done in exploring equivalences between corresponding constructions within various languages: e.g. short passives, impersonal passives, 'reflexive' constructions, 'one' constructions and the like. As for the contrastive study of the highlighting of various elements according to the co-text and the context, it is still in its infancy (but see M.-P. Woodley, 1991 as an interesting study of the differences between passives in English and French, including a discussion of Fr, 'on' vs. the English passive).

### Pragmatic interpretation

The theory of pragmatics variously defined as 'the relation of signs to their interpreters' (Morris, 1938) or 'the study of linguistic acts and the contexts in which they are performed' (Stalnaker, 1972) is a vast and growing area of research. Here, I shall only point out that reasonably quality translation has often been argued to require the calculation of what speech act is being performed. Are we dealing with a statement, a command, a request, a promise, a warning, etc.? Earlier on it was pointed out that for instance in translation from English to French it is often more appropriate to use an infinitive instead of an imperative:

- (6) (a) Depressurize hydraulic system  
 (b) Dépressuriser le système hydraulique.

Although this was presented as an example of stylistic equivalence earlier, it is more accurate to describe it as a case of speech act equivalence. Clearly, the study of pragmatics, which has been on the centre stage of linguistic theorising in recent years, does not limit oneself to speech acts. It often encompasses presuppositions, implicatures, aspects of discourse structure, or even the study of all the effects which can be derived from utterances in context. Much of this is relevant to translation but unfortunately the discussion is usually restricted to general principles and does not pretend to offer detailed strategies for dealing with individual speech act. This is an area much trodden by Artificial Intelligence specialists and no ambitious MT system could ignore the pioneering work of AI specialists such as Wilks and Schank in this domain.

### Style

The last area to be covered is style. All authors who write on style stress the diversity of usage. The following quote is not untypical:

"Style is one of the thorniest concepts to be dealt with in this encyclopedia. To Samuel Wesley, it was 'the dress of thought'; to Jonathan Swift, it was 'proper words in proper places'; to W. B. Yeats, it was 'high breeding in words and in argument'. And so we could continue, through several hundred definitions and characterizations. It is a remarkable career for a word that originally meant no more than a 'writing-implement' – a pointed object, or *stilus*, for inscribing wax" (Crystal, 1988: 66)

It is not the case that all aspects of style are irretrievably subjective and teachers of translation are able to guide students in establishing stylistic equivalences between languages (see Vinay

and Darbelnet, 1958: 201-219 *et passim* for a particularly interesting example). Nevertheless it has to be admitted that in many areas we are dealing with simple preferences and that it is not clear how their combination can be formalised. What is perhaps worth pointing out in this context is that if we are able to recover the propositional content of an utterance and to determine various aspects of its thematic orientation and its pragmatic value, we will go a long way towards capturing important aspects of style.

## ERROR ANALYSIS

Discussions of quality frequently engender the impression that one is dealing with rather subjective matters and the above paragraph on style may have convinced the reader that quality just lies in the eyes of the beholder. A solution to the problem of evaluation is often argued to lie in the analysis of errors. Error counting looks like an objective way of evaluating a system. The problem here is that unless we are provided with a classification of errors – i.e. a classification related to what should be achieved – we are likely to be faced with a rather atomistic, piecemeal account of the performance of a translation system. For instance, we are often told when a new MT system is presented to us that problems of various types could be solved by just altering the dictionary entries and making full use of the features that the dictionary incorporates. On closer inspection, however, such claims often turn out to be empty. Suppose for example that we feed into an MT system a compound such as:

(7) smoke-detector

and that the system, translating into French either prints out a sequence such as *\*\*smoke-detector* (or *@@smoke-detector*), or some message warning us that this sequence has not been recognized, or just offers a literal string such as *'fumée-détecteur'* in French. This may seem a sensible strategy if the forms are indeed not part of the dictionary as we would all accept that dictionaries have to be constantly up-dated. But what strikes me as a linguist interested in the evaluation the system – on the assumption that the same type of response is provided for a number of other compounds – is that the system in question appears to have no strategy for the translation of compounds. Compounding, in Germanic languages, is highly productive. A simple test of this productivity is the following observation: if we list all the noun-noun compounds in a given text and reverse them (e.g. *water-pressure* → *pressure-water*), it is always possible to find an interpretation for each new compound given appropriate contextual conditions. What has to be appreciated is that compounding in languages like English allows the generation of an infinite set of new words (cf. *manager, floor manager, factory floor manager, steel factory floor manager*, etc.). It is not an accident that the famous Canadian project Taum-Aviation, after the success of its immediate predecessor Taum-meteo, failed in large part because of an inability to handle the long nominal compounds characteristic of aviation vocabulary.

What I am therefore arguing is that, in assessing an MT system and trying to analyze errors, we need a great deal of appreciation of the linguistic potentialities of the languages under consideration. Broadly speaking, we can divide the structures we come across into a two by two matrix of the following kind:

(8)

|         | WORD LEVEL        | SENTENCE LEVEL        |
|---------|-------------------|-----------------------|
| FORM    | morphology        | syntax                |
| MEANING | lexical semantics | grammatical semantics |

From the point of view of morphology, systems to be tested need to be able to cope with inflection, derivation and compounding. Inflection has to do with grammatically determined variants of lexical units: e.g. plural (*cats*), tense endings (*worked*), etc. By and large, all modern MT systems incorporate subroutines for recognizing inflections. Where systems are far less adequate is in dealing with derivation (the production of new lexical units from bases, cf. *regular*, *regularize*, *regularization*, *de-regularization* etc.) and with compounding (the production of new lexical units from words, cf. *machine code*, *machine code analysis*, *machine code computer analysis*, etc.). It is important to realize that word-formation is truly dynamic and open-ended. Not all words can be listed and coping with new formations is not a simple matter of updating dictionaries.

At sentence level, from a formal syntactic point of view, there are a number of dimensions that need examining in testing any system. For instance, we expect phrases (noun phrases, verb phrases, adjectival phrases, etc.) to be correctly sequenced. It is true that there are disagreements among linguists about what is the best theoretical description of this or that domain. Nevertheless, there are many areas of sentence structure which are reasonably well understood. To take a simple example the structure of the so-called Auxiliary in English is well encapsulated in formulae such as the following (see Chomsky, 1957, 1965; and Gazdar et al. 1985, for a more detailed formalisation):

- (9) Auxiliary → Tense (Modal)(Perfect)(Progressive)  
 Tense → {Present, Past}  
 Modal → {may, will, ... , can}  
 Perfect → have + past participle  
 Progressive → be + present participle

The rules of (9) account for verbal structure ranging from *The computer exploded* to *The computer may have been working*. Given that corresponding formulae for verbal structures are available for other languages, it is surprising that many commercial systems still fail to translate sentences which incorporate fairly straightforward verbal complexes.

But correct sequencing is far too general a requirement. The kind of areas we should pay attention to are all the processes which allow structures to be mapped into other structures: control (*The Government expects to revoke the treaty*), raising (*The Government seems to be in trouble*), relativisation (*The committee which supports this proposal*), subordination (*They realized that it was impossible*), etc. In other words, making sure that the recursive backbone of sentence structure is handled by an MT system is crucial for an understanding of its potentialities.

If we now turn to meaning, we can distinguish lexical semantics from grammatical semantics. As mentioned earlier, one of the major difficulties in MT is lexical disambiguation (cf. the old chestnut of *bank1* financial institution vs. *bank2* side of a river). A typical strategy for lexical disambiguation involves the use of lexical semantic features in defining words (e.g.  $\pm$ artefact,  $\pm$ human,  $\pm$ animate, etc.). What many commercial systems mean when they advertise



the inclusion of Artificial Intelligence in their programmes is simply the incorporation of a number of lexical semantic features in the specification of content words. In assessing a system it is important to know what kind of feature grid is used, how the features are defined, what tools are available for linking features to the co-text, etc.

Finally, we turn to grammatical semantics. Here we have in mind areas such as determination and quantification (e.g. *the/all* reports), mood (e.g. *Open the meeting* vs. *They opened the meeting*), modality (e.g. *They might/could/should* open the meeting) or tense and aspect (*They had opened* the meeting). There seems little doubt that a proper translation of these areas cannot really be morphosyntactic (or, surfacy, if you prefer). Unless the MT system integrates semantic refinements, translations are likely to be of extremely poor quality.

The above paragraphs are in a sense no more than headings pointing us in a variety of directions (see Allegranza et al., 1991, and Durand et al., 1991, for more extensive description of the areas sketched here within a transfer approach). My aim though was to emphasize that much is known about linguistic structure which can be cast in a fairly neutral way and can be used to evaluate a system. In addition to monolingual adequacy we need of course some account of equivalence between various areas of linguistic structure. But it does not seem so far-fetched to expect of the documentation accompanying an MT system that it should specify what range of structures are covered. In the same way as manuals for various types of software often offer a description for the layman followed by a technical description for the programmer/developer, it would be nice to be told what kinds of structures are covered in an MT system. To my knowledge, the leaflets and manuals on the market are sadly deficient in this respect.

## SUBLANGUAGES

I am all too aware that the linguistic dimensions I have outlined as relevant to an evaluation of MT systems have on the whole been exemplified here with ordinary language data. Many specialists claim, with some plausibility, that by devising systems based on 'sublanguages' or 'restricted' languages, the task of MT becomes tractable and has a fair chance of success. By sublanguages is usually meant sets of texts which deal with one field only, and hence have a specialised vocabulary or terminology, and a syntax exhibiting a much narrower range of constructions than general purpose language. Examples taken from Kittredge (1988: 60-62) are weather bulletins or aircraft maintenance manuals as respectively typified below:

### (10) Weather bulletin

FORECASTS FOR YUKON AND NORTHWESTERN BC  
 ISSUED BY ENVIRONMENT CANADA AT 5:30 AM PDT  
 FRIDAY JULY 11 1980 FOR TODAY AND SATURDAY  
 KLONDIKE  
 BEAVER CREEK  
 STEWART RIVER  
 RAIN OCCASIONALLY MIXED WITH SLEET TODAY  
 CHANGING TO SNOW THIS EVENING. HIGHS 2 TO 4.  
 WINDS INCREASING TO STRONG NORTHWESTERLY THIS  
 AFTERNOON. CLOUDY WITH A FEW SHOWERS  
 SATURDAY. HIGHS NEAR 6.

(11) Aircraft maintenance manual  
 REMOVAL AND INSTALLATION OF PRESSURE SWITCH - NO. 1 SYSTEM

Removal procedure:

- (a) Depressurize hydraulic system
- (b) Disconnect electrical connector on pressure switch
- (c) Disconnect line at pressure port
- (d) Disconnect line at drain port elbow
- (e) Loosen the two mounting bolts and remove switch.

In the same article, Kittredge (1988: 59) defines a sublanguage as any subsystem of a language with the following properties:

- (i) the language subsystem is used in reference to a particular domain of discourse, or family of related domains,
- (ii) the sets of sentences and texts in the language subsystem reflects the usage of some 'community' of speakers, who are normally linked by some common knowledge about the domain (facts, assumptions, etc.) which goes beyond the common knowledge of speakers of the standard language,
- (iii) the subsystem has all the 'essential' properties of a linguistic system, such as 'consistency', 'completeness', 'economy of expression', and so forth,
- (iv) the language subsystem is maximal with respect to the domain, in the mathematical sense that no larger domain has the same properties.

The first two clauses of this definition do seem to me to correspond to our intuitive, pre-theoretical definition of what a sublanguage is. The last two clauses, however, move towards a formal characterization of a sublanguage treating it as closed under a certain set of operations. It is reminiscent of the definition of simple formal structures such as groups in mathematics illustrated by clock arithmetic. But do sublanguages behave in line with clauses (iii) and (iv)? One general point about formal systems is that the notions of *completeness* (derive all and only the correct sentences or formulae) and *consistency* (do not derive a contradiction) are very hard to demonstrate. Thus, in the classical work of the logician and mathematician Kurt Gödel in the 1930's, there are claims of incompleteness even with respect to well-understood systems such as number theory or propositional logic (cf. Gödel, 1931; and for more readable accounts see Nagel and Newman, 1958; Regis, 1987: ch. 3). All this is to say, that I find the analogy between sublanguages and closed, formal languages (which are, by definition, regimented) rather misleading.

My own experience of working with terminologists within the Eurotra project has convinced me that dealing with texts in restricted domains (e.g. satellite telecommunications) reduces the problem of ambiguity but by no means solves it. As Jennifer Pearson of Dublin City University recently reminded me, many texts which would unhesitatingly be described as prototypical examples of sublanguages show that ordinary language grammar and vocabulary is intermixed with the more restricted syntax and terminology characterizing the sublanguage in question. Nor is it the case that the vocabulary of sublanguages can generally be treated as closed, and hence amenable to simple listing. In many languages, compounding and derivations processes, as pointed out earlier, allow the production of an indefinitely large set of lexical units. And Taum-aviation is reported, among other things, to have foundered on

the problem of complex compounding in the field of aerospace. Be that as it may, the extensibility of systems based on highly regimented sublanguages must be in doubt.

## CONCLUSION

Let me move on to a few concluding remarks. I have emphasized in this paper the need for more sophisticated testing of the linguistic aspects of MT systems which are on the market. It could be objected that commercially available systems are not intended to deal with natural language in its full richness and complexity but are devised for restricted languages which the user must program in. Whether sublanguages can be sharply isolated from ordinary language is at least questionable. And, in any case, it will be noted that much of the publicity for commercial systems seems to be based on quite complex texts such as business letters or semi-scientific literature. In so far as little information is available on the syntactic and semantic aspects of business MT systems, it is practically impossible to evaluate claims regarding their potential if properly used.

It is not my aim to knock commercial systems down. It is undeniably easier to talk about systems than to devise them and I have the greatest admiration for system-designers and developers. What I have tried to argue is that the evaluation of the linguistic capabilities of MT systems should not be done in a random and piecemeal way, as I often feel is the case in reading reviews of MT systems in either journals or magazines. Like scientific theories, all classifications and linguistic test suites are deficient in some respect or other. I have nevertheless tried to demonstrate that linguistic structures can be looked at from a variety of separable dimensions relevant to translation and that more awareness of such structures leads to a better understanding of what MT systems can and cannot do.

## REFERENCES

- Allegranza, V., Bennett, P., Durand, J., van Eynde, F., Humphreys, L., Schmidt, P. and Steiner, E. (1991) 'Linguistics for Machine Translation', *Studies in Machine Translation and Natural Language Processing*, Vol. 1, 15-123, Luxembourg: Office for Official Publications of the European Communities.
- Chomsky, N. (1957) *Syntactic Structures*, The Hague: Mouton.
- Chomsky, N. (1965) *Aspects of the Theory of Syntax*, Cambridge, Mass.: MIT Press.
- Crystal, D. (1987) *The Cambridge Encyclopedia of Language*. Cambridge: Cambridge University Press.
- Durand, J., Bennett, P., Allegranza, V., van Eynde, F., Humphreys, L., Schmidt, P. and Steiner, E. (1991) 'The Eurotra Linguistic Specifications: An Overview', *Machine Translation* 6: 103-147.
- Halliday, M.A.K (1985). *An Introduction to Functional Grammar*. London: Edward Arnold.
- Hatim, B. and I. Mason (1990). *Discourse and the Translator*. London: Longman.

- Hutchins, W.J. and H.L. Somers (1992). *An Introduction to Machine Translation*. London: Academic Press.
- Garside, R., Leech, G. and G. Sampson (1987). *The Computational Analysis of English: a Corpus-based Approach*. London: Longman.
- Gazdar, G., E. Klein, G. Pullum and I. Sag (1985) *Generalized Phrase Structure Grammar*. London: Basil Blackwell.
- Gödel, K. (1931) 'Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I', *Monatshefte für Mathematik und Physik*, 38: 173-198.
- Humphreys, L. (this vol.) 'Evaluation. Who should do it and how?'
- Kittredge, R.I. (1988) 'The significance of sublanguage for automatic translation'. In S. Nirenburg (ed.) *Machine Translation. Theoretical and Methodological Issues*. Cambridge: Cambridge University Press.
- Moore, I. (1990). 'Criteria for selection MT systems'. *Translating and the Computer* 10.
- Morris, C.W. (1938) *Foundations of the Theory of Signs*. Chicago: Chicago University Press.
- Nagel, E. and J.R. Newman (1958) *Gödel's Proof*. London: Routledge and Kegan Paul.
- Regis, E. (1987) *Who Got Einstein's Office?* Harrnondworth: Penguin Books.
- Stalnaker, R. (1972) 'Pragmatics'. In D. Davidson and G. Harman (eds.) *Semantics of Natural Language*. Cambridge: Cambridge University Press.
- Vinay, J.-P. and J. Darbelnet (1958) *Stylistique comparée du français et de l'anglais*. Paris: Didier.
- Péry-Woodley, M.-P. (1991) 'French and English passives in the construction of text'. *Journal of French Language Studies* 1-1: 55-70.

#### **AUTHOR**

Jacques Durand, Professor of French and Linguistics, Department of Modern Languages,  
University of Salford, The Crescent, Salford, M5 4WT, UK