

# Teaching Contrastive Linguistics for MT

Paul Bennett

Centre for Computational Linguistics  
UMIST  
P O Box 88  
Manchester M60 1QD  
Great Britain  
Paul.Bennett@umist.ac.uk

## Abstract

We discuss a framework for teaching contrastive translation-oriented linguistics to students of machine translation. The framework embodies a classification of cross-linguistic differences and a demonstration of how the neutralization of such differences in abstract representations can make transfer modules simpler. We illustrate by means of examples from a range of languages, and then discuss a proposal for linguistic typology to which our system can be fruitfully related.<sup>1</sup>

## 1 Introduction

Leaving aside computational aspects, what kind of linguistics should be taught to students of machine translation (MT)? To some extent, it should be the kind of linguistics taught as part of any introductory linguistics course, concentrating on morphology, lexis, syntax and semantics, though perhaps with a special emphasis on ambiguity and its types. But, in addition, there needs to be a great deal of attention paid to the ways in which languages differ from each other and how this creates problems for MT. How, though, are these linguistic differences to be approached?<sup>2</sup>

One possible answer would be the kind of course or module already taught in many linguistics departments, viz. on linguistic typology or universal grammar, concentrating on the ways languages do and do not differ and perhaps on classifying languages into different types. This is the kind of approach found in textbooks such as Comrie (1981) and Croft (1990). For a number of reasons, however, such a course is not suitable.

Firstly, typological work sometimes deals with relatively superficial aspects of language (such as case-marking and agreement) and often refers to ‘exotic’ languages with relatively few speakers which are of little relevance for MT. Secondly, it generally fails to provide the kind of explicit bilingual knowledge that is needed. Let’s just take one example:

If a language uses a nonzero case marking for a direct object on the animacy/definiteness hierarchies, then it uses a nonzero case marking for direct objects higher on the hierarchies. (Croft 1990, p. 128)

1. Thanks to Miriam Butt, Federico Gaspari and Piklu Gupta for comments on an earlier version.

2. See Bennett (to appear) for some more general considerations on the relevance of linguistics for MT, in a volume aimed at translators; with a few exceptions, points made in that paper are not repeated here.

(The general idea here is that some combination of animacy and definiteness triggers overt case-marking on an object – cf. Spanish *a.*) This simply does not help in MT: the rules for case-marking in a particular language need to be stated irrespective of how they relate to these hierarchies, and in any case such a ‘surfacy’ feature is largely irrelevant to translation.

Instead, contrastive work which emphasizes the main languages used in MT R&D (and those most often spoken by the students themselves) is preferable. It should be oriented towards bilingual issues (though not concerned with just two languages), and should pay attention not only to morphosyntax but to semantics and the links between meaning and structure. Some typological work does achieve this last point, e.g. work on the links between semantic notions like action and syntactic categories like verb (Croft 1990, pp. 139–143). Indeed, we shall look at some other relevant typological work in Section 5. In addition, MT-oriented linguistics must not be tied to any particular linguistic theory, though that does not prevent reference to theoretical constructs. Moreover, it needs to be rather wider in scope and depth than presentations of linguistic differences and transfer problems in MT textbooks (e.g. Hutchins & Somers (1992, ch. 6), Arnold, Balkan, Humphries, Meijer & Sadler (1994, ch. 6) and Trujillo (1999, pp. 124–128)).

## 2 Canonical Form

What I have done is to develop a framework for teaching contrastive linguistics that is specially designed for MT. It embodies a classification of cross-linguistic differences, illustration of them from a range of languages, and a demonstration of how the neutralization of such differences in abstract representations can make transfer modules

simpler. I make no great claims for its originality, and it is deliberately eclectic. While not based on any particular MT system or linguistic theory, the framework is indebted (inter alia) to research on transfer-based MT (especially the Eurotra system – Allegranza, Bennett, Durand, Van Eynde, Humphreys, Schmidt & Steiner (1991)), the classical deep structure of standard theory transformational grammar (Chomsky 1965), relational grammar (Blake 1990) and some work in translation theory (Nida & Taber 1969, Malone 1988). The framework has been used on the MSc in Machine Translation at UMIST.

The system is based on the notion of a canonical form (CF), an active declarative sentence form which involves the most straightforward representation of predicates together with their arguments and modifiers. Comparable reconstructions of predicate–argument structure are the foundation of many implementations of transfer-based MT.<sup>3</sup> CFs are generally given in an informal way as English sentences, since it is better not to bias analyses towards particular formalisms. It is, however, also perfectly possible to represent them more abstractly using notions such as predicator and logical subject (I do not on the whole make use of case labels, as these are hard to define in general and to apply in specific instances).

Sentences and constructions are classified as to the ways in which they differ from CF, as languages vary in the kinds of divergence from CF that they permit. As an initial example, the sentence (1a) would be assigned the CF shown in (1b):

- (1) a. Mary is believed by John to  
prefer red wine

3. This is not the place to compare our approach with other more formalised work on MT, so a reference to the flat semantics of Verbmobil (Dorna & Emele 1996) will have to suffice.

- b. John believes that Mary prefers red wine

Example (1a) differs from the CF in that *Mary* has been raised into the main clause, which has then been passivized. *Mary* in (1a) is thus some distance from the verb of which it is a semantic argument, *prefer*, unlike in (1b).

### 3 A Contrastive Framework

Divergences from CF are classified into three types: movement, insertion and omission. That is to say, items can either be moved around (e.g. from one clause to another), be inserted or be omitted. We now go on to illustrate these three types and their various subtypes, including discussion of how CF helps to simplify transfer and also how languages differ from each other in terms of divergence from CF. Discussion and exemplification in this section is massively reduced, for reasons of space. See Bennett (1997) for more detail.

#### 3.1 Movement

Movement is divided into three kinds, depending on how far items have moved. Movement usually involves a concomitant change in grammatical relation, rather than just a change in linear order (as in Scrambling). No commitment to transformational grammar is implied here, just the idea that logical and actual positions of constituents may vary. The general ideas can be captured in, for instance, Lexical-Functional Grammar-based accounts that undo some kinds of movement via structure-sharing (Dorna, Frank, van Genabith & Emele 1998).

**Local Movement** refers to movement within a single clause. A simple example is the DATIVE alternation:

- (2) a. The man gave a book to the girl

- b. The man gave the girl a book

Example (2a) is the CF, with the logical indirect object having its function explicitly marked by the preposition *to*. In (2b), the *to* has disappeared, and the logical indirect object now appears immediately following the verb. A similar alternation occurs in Dutch, for instance, but not in French, where the shifted equivalent is ill-formed. It is not just that the verb *donner* ‘give’ disallows Dative in French; the language simply has no such alternation at all.

The classic example of local movement, however, is the PASSIVE. Besides straightforward examples such as *The dog was rescued by Jim*, English also allows prepositional passives (*This bed has been slept in by someone*) and recipient passives where the logical indirect object is subject (*The girl was given a book by the man*). Passive can be seen as normally involving promotion of a direct object to subject; we have just seen that French lacks Dative, so it naturally lacks recipient passives too. Prepositional passives are fairly rare cross-linguistically, while some languages allow an adversative passive not found in English, as in this Japanese example:

- (3) Taroo ga sensei ni musuko  
Taroo NOM teacher by son  
o sikarareta  
ACC scold-PASS-PAST  
‘Taroo was adversely affected by the teacher’s scolding his son’

I use the opportunity to discuss the impersonal passives of Dutch and German.

**Bi-Clausal Movement** involves movement from one clause to the next higher clause; it is not usually distinguished in theoretical linguistics as a separate category, but it does seem useful for our purposes.

SUBJECT-TO-SUBJECT RAISING involves the subject of the subordinate clause being moved up to be subject of the matrix

clause, as in *Peter seems to like coffee*. SUBJECT-TO-OBJECT RAISING is where a subject moves up to object position (*I believe Peter to be ill*) – this is a controversial construction, but I adopt the raising analysis because of its interaction with Passive (see (1a)). It does not occur in French or German, and is cross-linguistically rather rare. Then there is OBJECT-TO-SUBJECT RAISING (also known as Tough-Movement), where the embedded object is made into the matrix subject, as in *John is easy to convince*. A different kind of bi-clausal movement is NEG-RAISING, as in *I don't think United will ever win the title again*, where the negative belongs semantically in the subordinate clause and licenses the negative polarity item *ever*. CLITIC CLIMBING, as in Italian and Spanish, is another example of bi-clausal movement.

**Long-Distance Movement** (or, unbounded movement) sees a phrase being moved potentially any distance. The classic example is WH-MOVEMENT, as in *wh*-questions: in *Who did Bill say that Jim believes that Fred claimed that Martha loves?*, *who* is logically the object of *loves*. In contrast, some languages regularly have the questioned item occurring in its ordinary position (*wh-in-situ*), as in this Chinese example:

- (4) Ni he shenmo?  
 you drink what  
 'what are you drinking?'

And English allows adposition stranding in *wh*-questions (*Who did you lend it to?*), which most languages do not. These distinctions, and discussion of them, relate to both direct and indirect *wh*-questions.

Other than questions, Long-Distance Movement also occurs in TOPICALIZATION; and since an unbounded dependency is involved, I use this opportunity to discuss various strategies for RELATIVE CLAUSE FORMATION, such as leaving a gap or use

of a resumptive pronoun.

### 3.2 Insertion

Insertion describes constructions where some empty or dummy element which has no place in canonical form occurs, as *it* and *there* in the following:

- (5) a. It is unlikely that he will win  
 b. It is easy to convince him  
 c. There is a man waiting outside

The CF for (5a), for instance, is *That he will win is unlikely*. In the case of insertion structures, there is a discussion of the circumstances in which the different dummies can be used, and also of cross-linguistic contrasts. For instance, English *it* in (5a, b) is a subject, as it can be raised (*I consider it to be unlikely that he will win*), whereas German *es* is a mere place-holder, rather than an actual subject.

### 3.3 Omission

Omission covers a variety of examples where some item present in CF is absent from the surface form. One clear instance is the PRO-DROP phenomenon, whereby an argument of the verb is omitted, sometimes still coded in verbal agreement, but sometimes not, as in these Italian and Chinese examples:

- (6) a. parlo  
 speak-1SG  
 'I speak'  
 b. kan shu  
 read book  
 '() is reading a book'<sup>4</sup>

4. *Parlo* in (6a) also codes tense and mood (present indicative). The translation of (6b) is intended to show that the subject is completely undetermined: it could be *I*, *They*, etc. So translating this example may be compared to the task of identifying the antecedent of a pronoun, as in *It is not working*.

Translating these into English, and many another language, requires restoring the subject at CF (obviously this is easier in Italian than in Chinese).

SHORT PASSIVES, with no *by*-phrase, are another example of omission, and provide an opportunity to discuss the linguistic and translational differences between them and ergatives (e.g. *The ship was sunk* vs. *The ship sank*), and also between eventive and stative passives. SMALL CLAUSES are considered as omission of a form of *to be*, as in *I consider Andrew a fool*. GAPPING is found when an identical verb is omitted in a coordination, as in *John likes coffee, and Bill tea*. Gapping is simply not possible in Chinese, while in Japanese it is the second of two identical verbs that is omitted:

- (7) Taroo wa enpitu o to Ziroo  
 Taroo TOP pencil ACC and Ziroo  
 wa kami o katta  
 TOP paper ACC bought  
 ‘Taroo bought pencils and Ziroo paper’

Restoring the omitted item will help transfer from English to both Chinese and Japanese.

Turning now to CONTROL, we encounter a complex linguistic phenomenon, with a number of translational implications. ARBITRARY CONTROL concerns examples where a ‘missing’ subject is interpreted in a non-specific way, as in *To convince him is easy* (who does the convincing?). In contrast, NON-ARBITRARY CONTROL relates to cases where there is a clearly-identifiable controller in the sentence:

- (8) a. Jack intends to write a novel  
 b. Fred wishes to leave early  
 c. I persuaded John to leave  
 d. He urged me to think again

It is important to explain the properties of these control constructions and how they differ from examples of raising (cf. remarks

on bi-clausal movement in §3.1). From a translational point of view, a number of considerations arise. For instance, English *believe* is not a subject control verb (*\*Peter believes to be intelligent*), but its French equivalent is:

- (9) Pierre croit être intelligent  
 Pierre believes to-be intelligent  
 ‘Pierre believes himself to be intelligent’

Additionally, languages which lack infinitives naturally lack control-type structures. In Greek, for instance, all verb forms are obligatorily marked for the person and number of the subject, so translation of an English control structure will require identifying the controller:

- (10) O janis prospathise na fiji  
 the John tried NA he-go  
 ‘John tried to leave (lit. ‘John tried that he leave’)’

The exact representation of the omitted element in control structures varies from one theory to another, so I am deliberately non-committal on this point.

## 4 Simplifying Transfer

I have already made the point that the kinds of neutralized canonical forms we adopt are useful in simplifying transfer in MT, and I now discuss this point at greater length. Below, and in teaching, I generally rely on English as either source or target language, as it is the only language that every student will know, but in many cases similar comments could be made for a number of other languages.

We can begin with a simple example involving Dative (see (2)). The French translation of (2b) is:

- (11) L'homme a donné un livre à la  
the-man has given a book to the  
fille  
girl  
'The man gave a book to the girl'

It is straightforward to undo Dative during the English analysis phase and map it to the CF (2a). Transfer then becomes a purely lexical matter, with no need for it to perform any change of structure. This is the general line of argument we shall take here: that neutralizing surface differences by means of CF simplifies transfer, at least in the cases where source and target language differ in that one permits a 'deviation' from CF that the other does not (e.g. English has the Dative alternation, while French lacks it). Languages differ less at CF than at the surface.

Similarly in the case of *wh*-questions. With the Chinese example (4) as the source sentence, translating it into *What are you drinking?* is far easier if the CF of the latter is (ignoring aspect) *You drink what?*. Similar remarks hold for the various raising constructions (see under Bi-Clausal Movement above). Undoing the raising may make transfer simpler, e.g. *I believe Peter to be ill* should be rendered as *I believe that Peter is ill*, to facilitate transfer into (say) German *Ich glaube, dass Peter krank ist* (cf. (12) below).

The examples of omission that we have already discussed in §3.3 offer a further illustration of the general point. Translating an English example involving Gapping into Chinese or Japanese is far easier if the gapped item is restored. Equally, the translation of *John tried to leave* into Greek (10) is easier if the 'missing' subject of *leave* is restored in some way.

Some particularly convincing illustrations can be found in the interaction of phenomena, so let us look at one instance of this. I mentioned that Subject-to-Object Raising does not occur in, among other

languages, German. To translate an English sentence involving Raising plus Passive, such as *Peter is believed by everyone to be rich*, one has to undo two processes and produce the CF *Everyone believes that Peter is rich*, which can then be translated word-for-word:

- (12) Jeder glaubt, dass Peter reich ist  
each believes that Peter rich is  
'Everyone believes that Peter is rich'

The benefits derived from CF as a form of abstract representation do not just relate to the elimination or reduction of structural transfer. They also relate to issues such as selectional restrictions and word-sense disambiguation. In the example just discussed, for instance, it may well be helpful in transfer to know that *rich* is being predicated of a person, rather than (say) a cake. Or take the case of German *tragen*: 'wear' when the object is material or fabric, otherwise 'carry' (Niven 1997). The correct translation of *tragen*, then, requires identifying its canonical object, however and wherever it is actually realised in the sentence.

The notion of CF naturally does not enable all cross-linguistic differences to be neutralized, but its limitations are themselves instructive as to other elements of transfer. It needs to be supplemented by discussing, inter alia, support verbs (Damos 1992), morphology-syntax interactions (Bennett 1993), and the general cases of translation divergence (Vandooren 1993), as in this French example:

- (13) Elle a traversé la Manche en  
She crossed the Channel by  
avion  
plane  
'She flew across the Channel'

And Germanic languages allow resultatives, as in *John beat the metal flat*, while Romance and Slavic languages and Greek do

not permit this (having to express this meaning by something like ‘John flattened the metal by beating it’). Beyond this are all the problems relating to tense, aspect, anaphora, and many other areas, about which my framework has nothing to say.

It might also be argued that the use of CFs is illegitimate, as neutralization sometimes leads to loss of information. The Dative alternation is a case in point:

- (14) a. John taught the students French  
b. John taught French to the students

For many speakers at least, (14a) entails that the students learned some French, while (14b) carries no such implication. But even if one accepts that this distinction is valid and translationally-relevant, it is very much a special case. Goldberg (1995), while arguing that distinct constructions cannot be both semantically and pragmatically synonymous, accepts that the two Dative alternants are in general semantically equivalent, whatever their pragmatic or discourse differences might be.

The study of alternations and their semantic basis has recently received much attention (Dixon 1991, Levin 1993), including some work on cross-linguistic aspects (Frense & Bennett 1996). The usefulness of this research for MT (Baldwin, Bond & Hutchinson 1999) and the teaching of MT seems fairly clear, though much remains to be done.

## 5 Typology Revisited

A possible comment on the approach dealt with above is that, while it provides a useful perspective on cross-linguistic differences, it does so in a largely arbitrary manner, simply recording linguistic contrasts but not really doing anything more. However, I shall now argue that this method can be inter-

preted in a more general and typologically-oriented way, in terms of a typology of language that looks at syntax–semantics mappings.

Hawkins (1986) examines contrasts between English and German. Among his findings are the results summarised in Table 1: raising is far more widespread in English than in German. More generally, English permits far more argument-trespassing structures (where arguments occur in syntactic positions where they do not belong semantically) than German does. In our terms, English tolerates more deviations from CF. Equally, grammatical relations are far more diverse semantically in English than in German. Subjects, for instance, can cover far more semantic cases in English; e.g. German has no literal translation of *This tent sleeps four*, and instead one has to say:

- (15) In diesem Zelt können vier  
in this tent can four  
Personen schlafen  
people sleep  
‘Four people can sleep in this tent’

There would therefore be fewer ‘oblique’ subject alternations (Levin 1993, pp. 79–83) in German than English.

Hawkins goes on to propose:

a typological continuum whereby languages vary according to the degree to which surface forms and semantic representations correspond, with English being nearer the negative end, and German nearer the positive end of this continuum. (Hawkins 1986, p. 123)

In a comparable vein, Kakouriotis (1995) argues that the relations between argument structure and syntax are more transparent in Greek than in English, thus placing Greek towards the positive end of Hawkins’ continuum.

Müller-Gotama (1992) takes Hawkins’ work considerably further, with the ideal

Table 1: Raising in German and English (Hawkins 1986, p. 97)

| Raisings           | German       | English |
|--------------------|--------------|---------|
| Subject-to-Subject | Basically No | Yes     |
| Subject-to-Object  | No           | Yes     |
| Object-to-Subject  | Limited      | Yes     |

types on the continuum labelled as grammaticizing and transparent. Movement processes are seen as affecting transparency to different degrees:

- (16) Scrambling > Across VP > Extraction (Wh-Movement) > Raising > Preposition-stranding

Scrambling has least effect, as items stay within the same maximal projection and retain their coding properties, whereas raising (say) is more disruptive, as constituents cross a clause boundary. In my terms, local movement has less impact than bi-clausal movement, which in turn is less disruptive than long-distance movement. Müller-Gotama claims that Korean is a highly transparent language, far more so than German: promotion of non-agents to subjecthood is very restricted, for instance, and there is no Subject-to-Subject Raising. Bahasa Indonesia, in contrast, is highly grammaticizing (far closer to the ideal type than English is), except that only subjects can be moved or extracted, which makes it of the transparent type in some respects.

In subsequent work, Müller-Gotama (1994) extends this research, and proposes the ideal types seen in Table 2; he also relates the typology to right- versus left-branching structures, which I ignore here. Of course, most languages are somewhere in between the extreme types, with Dutch, for instance, falling between English and German, which latter language is itself less transparent than Russian. Not all logically-possible types occur, though: languages with overt nominal case-marking are always

highly semantically transparent overall.

My point here is not so much the typological validity of such a classification as its pedagogic potential from an MT angle. Differences between languages in the semantic range of their grammatical relations will often give rise to translational problems (see (15) above). Equally, the status of relation-changing rules such as Passive and Raising (which may vary from absent through restricted to common) can create sizeable differences between surface forms, which the notion of CF helps to explicate, as witness (12) earlier.

I would like to suggest that – other things being equal – strongly grammaticizing languages may be harder for MT (analysis components in particular) to cope with than the strongly transparent type. This is because the former require more processing and manipulation to reconstruct the predicate-argument structure of CF (undoing passive and extraction, for instance). On the other hand, scrambling may raise its own problems for a parser, and stranding of an adposition may be helpful as it provides an indication of the extraction site. It would thus not be relevant that stranding an adposition ‘removes the only element which indicates the role of the noun phrase it governs’ (Müller-Gotama 1994, p. 27). Rather, guides to where a displaced item has been ‘moved from’ can be very useful. Ambiguous examples such as *Where did you say you met Mary?* illustrate this point. The fact that most languages fall between the extremes may reduce the impact of these considera-



Table 2: Ideal types in the Semantic Typology (Müller-Gotama 1994, p. 28) – GRs = grammatical relations

| Strongly transparent             | Strongly grammaticizing        |
|----------------------------------|--------------------------------|
| Little semantic diversity of GRs | Much semantic diversity of GRs |
| GRs overtly marked               | GRs not overtly marked         |
| Free scrambling                  | Fixed order in clause          |
| No GR-changing rules             | Many GR-changing rules         |
| No extraction                    | Frequent extraction            |

tions, but does not eliminate it altogether.<sup>5</sup>

## 6 Conclusion

I have presented a framework for teaching contrastive linguistics to students of MT. My experience has been that this framework is useful in teaching students about (a) monolingual issues relevant to NLP in general, (b) contrastive problems which may create difficulties for MT, (c) possible solutions in terms of a CF-like interface representation, (d) phenomena requiring more abstract interface representations or change of structure in transfer, and (e) the role of linguistics in contributing to the identification and resolution of MT problems. Finally, I explored some typological work from the linguistics literature which may provide a broader foundation for the contrastive framework. This has not been exploited in teaching, but I think it may turn out to be helpful.

## References

- Allegranza, V., Bennett, P., Durand, J., Van Eynde, F., Humphreys, L., Schmidt, P. & Steiner, E. (1991), Linguistics for machine translation, in C. Copeland, J. Durand, S. Krauwer & B. Maegaard, eds, *The Eurotra Linguistic Specifications*, Office for Official Publications of the European Communities, Luxembourg, pp. 15–123.
- Arnold, D., Balkan, L., Humphries, L., Meijer, S. & Sadler, L. (1994), *Machine Translation: an Introductory Guide*, NCC Blackwell, Manchester and Oxford.
- Baldwin, T., Bond, F. & Hutchinson, B. (1999), A valency dictionary architecture for machine translation, in *Proceedings of the 8th International Conference on Theoretical and Methodological Issues in Machine Translation [TMI-99]*, Chester, pp. 207–217.
- Bennett, P. (1993), The interaction of syntax and morphology in machine translation, in F. Van Eynde, ed., *Linguistic Issues in Machine Translation*, Pinter, London and New York, pp. 72–104.
- Bennett, P. (1997), *Canonical Form, Linguistic Diversity and Machine Translation*. CCL/UMIST.
- Bennett, P. (to appear), The relevance of linguistics for machine translation, in H. Somers, ed., *Computers and Translation*, Benjamins, Amsterdam.
- Blake, B. (1990), *Relational Grammar*, Routledge, London.

5. It may be worth adding that Hawkins (2001) argues that languages like English are made easier for people to parse by the tendency for syntactically-related items to be adjacent. This would apply, *inter alia*, to *This tent sleeps four* (cf. (15)) and the assignment of a locative role to its subject once the adjacent verb is encountered.

- Chomsky, N. (1965), *Aspects of the Theory of Syntax*, MIT Press, Cambridge, Mass.
- Comrie, B. (1981), *Language Universals and Linguistic Typology*, Basil Blackwell, Oxford.
- Croft, W. (1990), *Typology and Universals*, Cambridge University Press, Cambridge.
- Danlos, L. (1992), Support verb constructions: linguistic properties, representation, translation, *Journal of French Language Studies* **2**, 1–32.
- Dixon, R. (1991), *A New Approach to English Grammar, on Semantic Principles*, Clarendon Press, Oxford.
- Dorna, M. & Emele, M. (1996), Semantic-based transfer, in *Proceedings of 16th International Conference on Computational Linguistics*, Copenhagen, pp. 316–321.
- Dorna, M., Frank, A., van Genabith, J. & Emele, M. (1998), Syntactic and semantic transfer with f-structures, in *Proceedings of 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics*, Montreal, pp. 341–347.
- Frñense, J. & Bennett, P. (1996), Verb alternations and semantic classes in English and German, in K. Jaszczolt & K. Turner, eds, *Contrastive Semantics and Pragmatics*, Pergamon, Oxford, pp. 305–317. Also available as *Language Sciences* **18**.
- Goldberg, A. L. (1995), *Constructions: a Construction Grammar Approach to Argument Structure*, University of Chicago Press, Chicago and London.
- Hawkins, J. (1986), *A Comparative Typology of English and German*, Croom Helm, London and Sydney.
- Hawkins, J. (2001), Why are categories adjacent?, *Journal of Linguistics* **37**, 1–34.
- Hutchins, W. J. & Somers, H. L. (1992), *An Introduction to Machine Translation*, Academic Press, London.
- Kakouriotis, A. (1995), On the double object construction in English and Modern Greek, *Studies in Language* **19**, 1–35.
- Levin, B. (1993), *English Verb Classes and Alternations*, University of Chicago Press, Chicago and London.
- Malone, J. (1988), *The Science of Linguistics in the Art of Translation*, State University of New York Press, Albany.
- Müller-Gotama, F. (1992), Towards a semantic typology of language, in M. Kefer & J. van der Auwera, eds, *Meaning and Grammar*, Mouton de Gruyter, Berlin and New York, pp. 137–178.
- Müller-Gotama, F. (1994), *Grammatical Relations: a Cross-Linguistic Perspective on their Syntax and Semantics*, Mouton de Gruyter, Berlin and New York.
- Nida, E. & Taber, C. (1969), *The Theory and Practice of Translation*, Brill, Leiden.
- Niven, B. (1997), Verb-noun combinations in machine translation, *Web Journal of Modern Language Linguistics* **2**. <<http://wjml1.ncl.ac.uk/issue02/niventrans.htm>>.
- Trujillo, A. (1999), *Translation Engines: Techniques for Machine Translation*, Springer, Berlin, Heidelberg and New York.
- Vandooren, F. (1993), Divergences de traduction et architectures de transfert [Translation divergences and transfer architectures], in P. Bouillon & P. Clas, eds, *La traductique*, Presses de l'Université de Montréal, Montréal, pp. 77–90.