STRUCTURAL NON-CORRESPONDENCE IN TRANSLATION

Louisa Sadler,
Dept. of Language and Linguistics,
University of Essex,
Wivenhoe Park,
Colchester, CO4 3SQ, Essex, UK.
louisa@uk.ac.essex

Henry S. Thompson,
Human Communication Research Centre,
University of Edinburgh,
2 Buccleuch Place,
Edinburgh, EH8 9LW, UK.
ht@uk.ac.ed.cogsci

ABSTRACT

Kaplan et al (1989) present an approach to machine translation based on co-description. In this paper we show that the notation is not as natural and expressive as it appears. We first show that the most natural analysis proposed in Kaplan et al (1989) cannot in fact cover the range of data for the important translational phenomenon in question. This contribution extends the work reported on in Sadler et al (1989) and Sadler et al (1990). We then go on to discuss alternatives which depart from or extend the formalism proposed in Kaplan et al (1989) in various respects, pointing out some directions for further research. The strategies discussed have been implemented.

0. Introduction

Recent work in LFG uses the notion of projection to refer to linguistically relevant mappings or correspondences between levels, whether these mappings are direct or involve function composition (Halvorsen & Kaplan (1988), Kaplan (1987), Kaplan et al (1989)). Mapping functions such as ψ (from c to f structure) and σ (from c to semantic structure) are familiar from the LFG literature. Kaplan et al (1989) extend this approach to Machine Translation by defining two translation functions τ (between f-structures) and τ' (between semantic structures). By means of these functions, one can 'co-describe' elements of source and target f-structures and s-structures respectively. Achieving translation can be thought of in terms of specifying and resolving a set of constraints on target structures, constraints which are expressed by means of the τ and τ' functions.

The formalism permits a wide variety of source-target correspondences to be expressed: τ and ψ can be composed, as can τ' and σ. The approach allows for equations specifying translations to be added to lexical entries and (source language) c-structure rules. For example:

(1) $(\tau\ (\uparrow\ \text{SUBJ})) = ((\tau\ \uparrow)\ \text{SUBJ})$

composes τ and ψ, equating the τ of the SUBJ f-structure with the SUBJ attribute of the τ of the mother's f-structure. Thus (1) says that the translation of the value of the SUBJ slot in a source f-structure fills the SUBJ slot in the f-structure which is the translation of that source f-structure. What results is an interesting, and apparently extremely attractive approach to MT, which finds echoes in a good deal of recent work in MT and Computational Linguistics generally (van Noord et al (1990), Zajac (1990)). Among the apparent advantages are:

(i) that it avoids the problems that arise in traditional stratificational transfer systems where a variety of (often incompatible) kinds of information must be expressed in a single structure.

(ii) that, because it uses the formal apparatus of LFG, it is at least compatible with a large body of well worked out linguistic analyses.

Perhaps most important, however, the examples that Kaplan et al give suggest that the notation is both natural and expressive: natural, in the sense that adequate τ relations can be stated on the basis of reasonable intuitive and well-motivated linguistic analyses; expressive, in the sense that it is powerful enough to describe some difficult translation problems in a straightforward way.

In this paper we show that the notation is not as natural and expressive as it at first seems. We first show that the most natural analysis proposed in Kaplan et al (1989) cannot in fact cover the range of data for the translational phenomenon in question. These cases are important because these constructions represent a pervasive structural difference between languages which one wants to be able to deal with in a natural way. We then go on to discuss alternatives which depart from or extend the formalism

proposed in Kaplan et al (1989) in various respects, pointing out some directions for further research.

Section 1

The discussion here concerns the translational data of head-switching or splitting/fusing discussed in Kaplan et al (1989) as cases of differences of embedding:

(2a) John just arrived.
Jean vient d'arriver.
(b) Jan zwemt graag/toevallig.
John likes to/happens to swim.
(c) I think that John just arrived.
Je pense que Jean vient d'arriver.
(d) Ik denk dat Jan graag zwemt.
I think that John likes to swim.
(e) Ich glaube dass Peter gern schwimmt.
I believe that Peter likes to swim.

Kaplan et al (1989) sketch two alternative approaches to head-switching. The first assumes that the adverb is essentially an f-structure head subcategorising for a sentential ARG. This has the effect of making the source and target f-structures rather similar to each other but faces a number of serious problems. Firstly, it violates the LFG assumption that the f-structure of the highest c-structure node contains the f-structures of all other nodes. Secondly, while such an analysis might well be correct for semantic structure, it cannot be justified on monolingual grounds as a surface syntactic feature structure (an f-structure). For example, it is the verb, not the adverb, which is the syntactic head of the construction, carrying tense and participating in agreement phenomena. This proposal therefore lacks monolingual motivation and thus is not attractive as an approach to translation. Thirdly, the approach cannot be extended to the case discussed here.

In this paper we focus on the alternative proposal, in which head-switching (or alternatively, splitting) is a τ operation.[1] Adverbs are taken to be f-structure SADJs. The τ annotation to ADVP states that the τ of the mother f-structure is the XCOMP of the τ of the SADJ slot (Kaplan et al's fig. 26):[2]

(3)

$$\begin{array}{lccc} S \rightarrow & NP & ADVP & VP \\ & (\uparrow SUBJ) = \downarrow & (\uparrow SADJ) = \downarrow & \\ & & (\tau\ (\uparrow SADJ)\ XCOMP) = (\tau\ \uparrow) & \end{array}$$

This has the effect of subordinating the translation of the f-structure which contains the SADJ to the translation of the SADJ itself. The lexical entries themselves contribute further τ equations (following Kaplan et al's fig. 21):

(4)
arrive: V $(\uparrow PRED) = arrive\langle SUBJ\rangle$
$((\tau\ \uparrow) PRED\ FN) = arriver$
$(\tau\ (\uparrow SUBJ)) = ((\tau\ \uparrow) SUBJ)$

just: ADV
$(\uparrow PRED) = just$
$((\tau\ \uparrow) PRED\ FN) = venir$

john: N
$(\uparrow PRED) = john$
$((\tau\ \uparrow) PRED\ FN) = jean$

This is unproblematic for examples such as (2a,b). The τ equations and further information from the target monolingual lexicon collaborate to relate (5) to (6):

(5)
$$\begin{bmatrix} PRED & arrive\langle SUBJ\rangle \\ SUBJ & \left[PRED\ john\right]_{f4} \\ SADJ & \left[PRED\ just\right]_{f5} \end{bmatrix}_{f3}$$

(6)
$$\begin{bmatrix} PRED & venir\langle SUBJ,XCOMP\rangle & \\ SUBJ & \left[\ \right]_{\tau f4} & \\ XCOMP & \begin{bmatrix} PRED & arriver\langle SUBJ\rangle \\ SUBJ & \left[PRED\ jean\right]_{\tau f4} \end{bmatrix}_{\tau f3} & \end{bmatrix}_{\tau f5}$$

However a problem with the τ equations arises in translating these sentences in an embedded context (2c,d,e). The English structure, and lex-

[1] Notice that a third possibility, not discussed in Kaplan et al (1989), involving a flat f-structure but treating the adverb as a semantic head at s-structure, simply means that the mapping problem we describe below arises in the monolingual mapping between f-structure and s-structure.

[2] Note that here and subsequently, following Kaplan et al, we ignore the monolingual potential for more than one ADVP with its attendant problems for translation.

ical entries are:

(7)

$$\begin{bmatrix} PRED\ think\langle SUBJ,COMP\rangle \\ SUBJ\ \left[PRED\ I\right]_{f2} \\ COMP\ \left[\cdots\right]_{f3} \end{bmatrix}_{f1}$$

(8)
think: V (↑PRED) = think<SUBJ,COMP>
((τ ↑) PRED FN) = penser
(τ (↑ SUBJ)) = ((τ ↑) SUBJ)
(τ (↑ COMP)) = ((τ ↑) COMP)

I: N (↑ PRED) = I
((τ ↑) PRED FN) = je

The equations in (3) and (8) require the translation of the f-structure immediately containing the SADJ attribute (τ (f3)) to be both a COMP and an XCOMP in the target f-structure:

(9)
((τ f1) PRED FN) = penser
(τ (f1 SUBJ)) = ((τ f1) SUBJ)
(τ (f1 COMP)) = ((τ f1) COMP)
((τ f2) PRED FN) = je
((τ f4) PRED FN) = jean
(τ f3) = (τ (f3 SADJ) XCOMP)
((τ f3) PRED FN) = arriver
((τ f3) SUBJ) = (τ (f3 SUBJ))
((τ f5) PRED FN) = venir

Notice that since (f1 COMP) = f3, and (f3 SADJ) = f5), we have the following equations from the emphasised lines:

(τ (f3) = ((τ f1) COMP)
(τ (f3)) = ((τ f5) XCOMP)

This results in a doubly-rooted DAG (10).

(10)

$$\begin{bmatrix} PRED\ venir\langle SUBJ,XCOMP\rangle \\ SUBJ\ \left[\ \right]_{\tau f4} \\ XCOMP\ \begin{bmatrix} PRED\ arriver\langle SUBJ\rangle \\ SUBJ\ \left[PRED\ jean\right]_{\tau f4} \end{bmatrix}_{\tau f3} \end{bmatrix}_{\tau f5}$$

$$\begin{bmatrix} PRED\ penser\langle SUBJ,COMP\rangle \\ COMP\ \left[\cdots\right]_{\tau f3} \end{bmatrix}_{\tau f1}$$

This is clearly not what is required and on standard linguistic assumptions, will not be accepted by the target generator. It does not give a correct translation of the source string.

In this section we have shown that the proposal as outlined in Kaplan et al (1989) does not produce an adequate analysis of these cases. The problem, which is not at first apparent, arises from the combination of the regular and irregular equations from the emphasised lines. Note that there is no problem stating this correspondence in the French -> English direction (see below).

Section 2.0.

In this section we will briefly consider a number of alternatives.[3] To facilitate discussion, it is worth noting that the proposal in Kaplan et al (1989) involves basically three elements:

(11a) a set of (regular) equations constraining both source and target:

(τ (↑ SUBJ)) = ((τ ↑) SUBJ)

(11b) a set of (regular) equations assigning target PRED values:

((τ ↑) PRED FN) = *sem form*

(11c) a (special) equation on ADVP constraining both source and target:

((τ (↑ SADJ)) XCOMP) = (τ ↑)

The problem noted above arises from the combination of an equation from (a) with the equation (c).

Section 2.1.

The first alternative we considered involves maintaining equations of type (a) (so that (τ (f3) is indeed ((τ f1) COMP)), and then switching heads only (rather than whole constructions). The basic idea is that the τ annotations to ADVP provide a PRED value for τf3 and specify that the τ of the PRED of f3 is the PRED in (τf3 XCOMP). To do this, the PRED value must be made into a complex feature, and heavy use is made of τ equations on the c-structure rules, so that the mapping is essentially structurally determined.

[3] These alternatives have been explored by using at Essex an implementation of PATR due to Bob Carpenter, and at Edinburgh a version of MicroPATR.

Intuitively, the approach works by building target constructions without assigning them PRED values directly, then specifying the target PRED values in such a way that it is possible to switch the heads for the cases in question. .LP In fact, though this works for cases such as (2c,d,e), it is limited to cases in which it is correct to raise all the dependents of a predicate to the same slot in the construction headed by the translation of the adverb. It thus fails with (12a) in which *races* must remain a dependent of the embedded construction, and (12b) in which the same is true of *Jean*:

(12a) Peter zwemt graag wedstrijden.
Peter likes to swim races.

(12b) I said that John will probably come.
J'ai dit qu'il est probable que Jean viendra.

This is of course an immediate consequence of the fact that the proposal works by switching not constructions but heads. .SH Section 2.2.

It is clear from the above that any solution must achieve constructional integrity in translation. This idea can be achieved in a number of (slightly different) ways. In the following we exploit the path equation variables available in LFG, which permit one to use a value assigned elsewhere as an attribute (that is, our proposal here is modelled on the use of (↑ (↓ PCASE) = ↓) in LFG.

We alter (11a) and (11b) so that the paths that they constrain are sensitive to the value of an attribute (which we call CTYPE):

(13)
((τ ↑) (↑ CTYPE) PRED FN) = *sem form*
((τ ↑) (↑ CTYPE) SUBJ) = (τ (↑ SUBJ))
((τ ↑) (↑ CTYPE) OBJ) = (τ (↑ OBJ))[4]

The value of CTYPE is given by the adverbial annotations:

(14a)
on ADVP:
(τ ↑) = (τ ↓)
(↑ CTYPE) = (↓ TYPE)

(14b)
on the adverb *just*:
(↑ TYPE = XCOMP)

Notice that the τ annotation to ADVP (which states that the translation of the containing f-structure is the translation of the f-structure associated with the ADVP (i.e. the SADJ slot)) simply equates the τ of two f-structures and avoids the problem which beset the proposal in Kaplan et al (1989). This can be seen from the equation set for (2c) in (15). Note that when there is no adverb, the value of (↑ CTYPE) must be ε (since paths are regular expressions ((τ ↑) (ε) GF) = ((τ ↑) GF)).

(15)
((τ f1) PRED FN) = penser
(τ (f1 SUBJ)) = ((τ f1) SUBJ)
(τ (f1 COMP)) = ((τ f1) COMP)
((τ f2) PRED FN) = je
((τ f3) XCOMP PRED FN) = arriver
((τ f3) XCOMP SUBJ) = (τ (f3 SUBJ))
((τ f4) PRED FN) = jean
(τ f3) = (τ f5)
((τ f5) PRED FN) = venir

What is the cost of this proposal? The translational correspondences in all lexical entries will be sensitive in this way to the value of the CTYPE feature. We must guarantee that when a value is not contributed by the type feature on the adverb, the value of (↑ CTYPE) is ε, either by some priority union operation to initialise it to ε, or by some other convention with this effect, or by assuming different versions of the c-structure rules (with VP contributing (↑ CTYPE) = ε) as appropriate.

Variants of this solution which do not exploit the path equation variable apparatus of LFG are also possible, though at the cost of massively increasing the size of the lexicon. For example, lexical translation correspondences could be disjunctions as in (16).

(16)

$$\left\{\begin{array}{l} [\ \uparrow SADJ\ TYPE =_c predic \\ \quad ((\tau\ \uparrow) XCOMP\ PRED) = swim \\ \quad ((\tau\ \uparrow) XCOMP\ SUBJ) = (\tau\ (\uparrow SUBJ)) \\ \quad ((\tau\ \uparrow) XCOMP\ OBJ) = (\tau\ (\uparrow OBJ))\ ] \\ \\ [\ ((\tau\ \uparrow) PRED) = swim \\ \quad ((\tau\ \uparrow) SUBJ) = (\tau\ (\uparrow SUBJ)) \\ \quad ((\tau\ \uparrow) OBJ) = (\tau\ (\uparrow OBJ))\ ] \end{array}\right\}$$

[4] Notice that *all* dependents of a head must be made sensitive to the value of the CTYPE attribute (to maintain constructional integrity). To deal with non-subcategorised constituents such as SADJs (whose τ equations are given by c-structure rule) we must annotate ADVP with: τ (↑ SADJ) = ((τ ↑) (↑ CTYPE) SADJ)

It remains to be seen whether one needs the constraining condition to rule out unwanted 'partial' or 'extra' translations, or whether one can rely on completeness and coherence checks on the target side.

Section 2.3.

Our third alternative involves giving the path equations some sort of functional uncertainty interpretation. Our starting point is the problematic pair of equations repeated in (17) and the observation that the required target structure embeds the XCOMP within the COMP.

(17)
(τ (↑ COMP)) = ((τ ↑)COMP) ▪
(τ (f3)) = ((τ f1) COMP)

(τ ↑) = (τ (↑ SADJ) XCOMP) ▪
(τ (f3)) = ((τ f5) XCOMP)

The interpretation of the (τ (↑ COMP)) = ((τ ↑) COMP) could be loosened on the source side, as in (18):

(18)
(τ (↑ COMP GF)) = ((τ ↑) COMP) ▪
(τ (f3 GF)) = ((τ f1) COMP)

which specifies that the translation of some f-structure on a path from the source COMP (e.g. the COMP SADJ) fills the COMP slot in the translation. This avoids the problem in (17). Equally, the interpretation could be loosened on the target side:

(19)
(τ (↑ COMP)) = ((τ ↑) COMP GF) ▪
(τ(f3)) = ((τ f1) COMP GF)

which says that the translation of the COMP fills some path from the COMP slot in the translation (e.g. the COMP XCOMP). This proposal raises a number of interesting questions for further research about whether functional uncertainty can be used here while still guaranteeing some determinate set of output structures to be validated (or not) by the target grammar. Notice however that for the case in hand, the uncertainty equation can be quite specific - all that is required is the source functional uncertainty:

(τ (↑ COMP SADJ)) = ((τ ↑) COMP)

3. Conclusion.

Our starting point in this paper was the observation that a treatment proposed for cases such as (2) in Kaplan et al (1989) is unworkable. We have then discussed alternative approaches available within the general model assumed by Kaplan et al (1989). We have shown that the problem is to achieve simple general statements of the correspondence mapping which cover exceptional cases without spreading the effect of exceptionality throughout the grammar. The discussion in section 2 raises intricate technical issues about the formalism itself, but also relates to wider issues concerning the modularity of the approach to translation proposed in Kaplan et al (1989) as well as the suitability and expressivity of the formalism, raising serious questions about the feasibility of a large MT system along these lines.

We also noted that these cases are unproblematic in the "fusing" direction, for then we do not run into problems with the functionality of the τ correspondence. In this direction, the 'special' equations are within the lexical entry for *venir*:

(20)
((τ ↑) SADJ PRED FN) = just
(τ ↑) = (τ (↑ XCOMP))

Substituting variables for clarity, combining these equations with the regular equation from the embedding verb (*penser*) produces no inconsistency, since the path specifications τf1 COMP and τf5 can be equated:

(21)
(τ (f1 COMP) = ((τ f1) COMP) ▪
(τ f3) = ((τ f1) COMP)

(τ f3) = (τ (f3 XCOMP)) ▪
(τ f3) = (τ (f5))

((τ f3) SADJ PRED FN) = just

This observation raises interesting questions concerning the directionality assumed in Kaplan et al (1989). It seems that the correct way to view all this is that we have a system of correspondences relating 4 structures (Source and Target c and f structures). For a given set of correspondences and a partially determined set of structures, three possibilities exist:

- no solution can be found;

- a finite number of solutions can be found;
- an indeterminate and/or infinite number of solutions can be found.

We might expect therefore that a solution may be found even if we state correspondences in the French -> English direction but supply the partial determination from the English side (that is, when English is source). The system for translating in either direction would then be a pair of monolingual grammars with a set of τ equations stated in the "fusing" direction (i.e. in the French grammar). This is currently under investigation.

Preliminary results suggest that this approach will in fact cleanly overcome the specific problem at hand. It has proved possible to translate sentences (2b) and (2d) above from Dutch to English using grammars and lexicons in which τ only appears in the English rules and entries. But this work has in turned raised a number of fundamental issues, some of which apply not only to LFG but to any other attempt at theory-based translation:

- Exactly what does the formal definition of the 'translates' relation look like, in LFG or any other theory-based approach to translation?
- Can this formal definition actually be implemented? Existing approaches to generation from f/s-structure in LFG are too restrictive (Wedekind 1988), and our current implementation over-compensates.
- Is the functional nature of correspondences appropriate to the τ family, or would a relation be more appropriate? If so, what would the theoretical and practical consequences be?
- What is the relation between strict theory-based 'translation' and translation in the ordinary sense of the word? Is it not likely that its applicability will in practice be limited to closely related language pairs?
- Is there a substantive difference between the structures - and - correspondences approach of LFG and the single - structured - sign approach of HPSG or UCG? Translation seems a strenuous test.

ACKNOWLEDGEMENTS

We thank an anonymous EACL reviewer for helpful comments and constructive criticisms, and Doug Arnold and Pete Whitelock for useful discussion. All remaining errors are, of course, our own.

REFERENCES

Halvorsen, P-K and R.M. Kaplan (1988) "Projections and Semantic Description in Lexical-Functional Grammar" presented at the *International Conference on Fifth Generation Computer Systems*, Tokyo, Japan.

Kaplan, R. (1987) "Three seductions of computational psycholinguistics", in P.Whitelock, M.M.Wood, H.L.Somers, R.Johnson and P.Bennett (eds) *Linguistic Theory and Computer Applications*, Academic Press, London, pp 149-88.

Kaplan, R., K. Netter, J. Wedekind and A. Zaenan (1989) "Translation by Structural Correspondences", *Proceedings of 4th EACL*, UMIST, Manchester, pp 272-81.

Sadler, L., I. Crookston and A. Way (1989) "Co-description, projection and 'difficult' translation", *Working Papers in Language Processing No. 8*, Dept. of Language and Linguistics, University of Essex.

Sadler, L., I. Crookston, D. Arnold and A. Way (1990) "LFG and Translation", in *Third International Conference on Theoretical and Methodological Issues in Machine Translation*, Linguistics Research Center, Austin, Texas, (Pages not numbered).

van Noord, G., J.Dorrepaal, P. van der Eijk, M. Florenza, and L. des Tombe (1990) "The MiMo2 Research System" in *Third International Conference on Theoretical and Methodological Issues in Machine Translation*, Linguistics Research Center, Austin, Texas, (Pages not numbered).

Wedekind, J. (1988) "Generation as structure driven derivation" *Proceedings of COLING-88*, vol 2, pp. 732-737, Budapest.

Zajac, R. (1990) "A Relational Approach to Translation" in *Third International Conference on Theoretical and Methodological Issues in Machine Translation*, Linguistics Research Center, Austin, Texas, (Pages not numbered).