GENERATION AND TRANSLATION— TOWARDS A FORMALISM-INDEPENDENT CHARACTERISATION

Henry S. Thompson
Human Communication Research Centre
University of Edinburgh
2 Buccleuch Place
Edinburgh EH8 9LW
SCOTLAND

ABSTRACT

This paper explores the options available in the formal definitions of generation and, parasitically, translation, with respect to the assumed necessity for using a single grammar for analysis and synthesis. This leads to the consideration of different adequacy conditions relating the input to the generation process and the products of analysis of its output.

I. A SCHEMATIC DEFINITION OF 'GENERATES'

We start from the assumption of a constraint-based theory of linguistic description, which supports at least the notions of derivation and underlying form, in that the definition of grammaticality appeals to a relation between surface strings and some formal structure. We will attempt to remain agnostic about the shape of this formal structure, its precise semantics and the mechanisms by which a grammar and lexicon constrain its nature in any particular case. In particular, we take no stand on whether it is uniform and monolithic, as in the attribute-value matrices (hereafter AVMs) of PATR-II or HPSG, or varied and partitioned, as in the trees, AVMs and logical formulae of LFG. We will use the phrase *products of analysis* to refer to the set of underlying structures associated by a grammar and lexicon with a surface string, viz

> for a grammar G and sentence s $\in L_G$, we refer to the set of all products of analysis
> $X_s \equiv \{\chi \mid \Delta_G(s,\chi)\}$,
> where we use Δ for the 'derives' relation.[1]
> We will also use χ_s to refer to an arbitrary member of X_s.

We will also assume that the formal structures involved support the notions of subsumption and its inverse, extension, as well as unification and generalisation. Whether this is accomplished via appeal to a lattice, or in terms of simulations, will only become relevant in section IV.

We can now provide schematic definitions of generation and, with a few further assumptions, translation. We say

> Definition 1.
> $\Gamma^{\gamma}_{G}(\chi,s)$ (a structure χ generates$_\gamma$ a string s for grammar G)
> iff $\exists\ \chi_s \ni \gamma(\chi_s,\chi)$[2]

Most work to date on building generators from underlying forms (e.g.

[1] In this we follow Wedekind (1988), where we use X/χ for an arbitrary underlying form, as he uses Φ/ϕ for f-structure and Σ/σ for s-structure.

[2] Again our γ is similar to Wedekind (1988)'s adequacy condition C.

Wedekind 1988, Momma and Dörre 1987, Shieber, van Noord, Pereira and Moore 1990, Estival 1990, Gardent & Plainfossé 1990) have taken the adequacy condition γ to be strict isomorphism, possibly of some formalism-specific sub-part of the structures χ_s and χ, e.g. the f-structure part in the case of Wedekind (1988) and Momma and Dörre (1987). In the balance of this paper I want to explore alternative adequacy conditions which may serve better for certain purposes. Although some progress has been made towards implementation of generators which embody these alternatives, that is not the focus of this paper. As far as I know, aside from a few parenthetical remarks by various authors, only van Noord (1990) addresses the issue of alternative adequacy conditions—I will place his suggestion in its relevant context below.

II. WEAKER FORMULATIONS

Work on translation (Sadler and Thompson 1991) suggests that a less strict definition of γ is required. Consider the following AVM, from which features irrelevant to our concerns have been eliminated:

$$\begin{bmatrix} \text{cat} & \text{s} \\ \text{pred} & \text{like} \\ \text{comp} & \begin{bmatrix} \text{subj} & \begin{bmatrix} \text{cat} & \text{np} \\ \text{pred} & \text{Robin} \end{bmatrix} \\ \text{pred} & \text{swim} \end{bmatrix} \end{bmatrix}$$

Figure 1. Exemplary 'underspecified' χ

Under the γ is identity approach, this structure will not generate the sentence *Robin likes to swim*, even though one might expect it to. For although we suppose that somewhere in the grammar and lexicon there will be a constraint of identity between the subject of *like* and the subject of *swim*, which should be sufficient to as it were 'fill in' the missing subject, the strict isomorphism definition of γ will not allow this.

II.1 Subsumption and extension

If γ were loosened to extension, the inverse of subsumption, this would then work straightforwardly (i.e. $\gamma(\chi_s,\chi)$ iff $\chi_s \sqsupseteq \chi$, that is, χ_s extends χ, χ subsumes χ_s). It is just this sort of thing which seems to be required for translation, see for example Sadler and Thompson (1991) and the discussion therein of Kaplan et al. (1989), where χ for the desired target arises as a side effect of the analysis of the source, and χ_s is additionally constrained by the target language grammar[3].

Note that for Wedekind (1988) this move amounts to removing the coherence requirement, which prevents the addition of additional information during generation. Not surprisingly, therefore, implementation of a generator for γ as subsumption is in some cases straight-forward—for the generator of Momma and Dörre, for example, it amounts to *removing* the constraints they call COH_A and COH_B, which are designed to implement Wedekind's coherence requirement.

van Noord (1990) discusses allowing a limited form of extension, essentially to fill in atomic-valued features. This avoids a problem with the unconstrained approach, namely that it has the potential to overgenerate seriously.

[3] Note that appealing to subsumption assumes that both the inputs to generation (χ) and the results of analysis (χ_s) are fully instantiated.

For the above example, for instance, the sentence *Robin likes to swim on Saturdays* could also be generated, on the assumption that temporal phrases are not subcategorised for, as χ_S in this case clearly also extends χ. Rather than van Noord's approach, which is still too strong to handle e.g. the example in Figure 1 above, some requirement of minimality is perhaps a better alternative.

II.2 Minimal extension

What we want is that not only should χ_S extend χ, but it should do so minimally, that is, there is no other string whose analysis extends χ and is in turn properly extended by χ_S. Formally, we want γ defined as[4]

Definition 2.

$\gamma(\chi_S,\chi)$ iff

$\chi_S \sqsupseteq \chi$ and

$\nexists s' \ni \chi_{S'} \sqsupseteq \chi \wedge \chi_S \sqsupset \chi_{S'}$

This rules out the over-generation of *Robin likes to swim on Saturdays* precisely because χ_S for this properly extends χ_S for the correct answer *Robin likes to swim*, which in turn extends the input χ, as given above in Figure 1.

II.3 Maximal Overlap

Unfortunately, the requirement of any kind of extension is arguably too strong. We can easily imagine situations where the input to the generation process is over-specific. This might arise in generation from content systems, and in any case is sure to arise in certain approaches to translation (see section III below). By way of a trivial example, consider the input given below in Figure 2.

$$\begin{bmatrix} \text{cat} & \text{s} \\ \text{pred} & \text{swim} \\ \text{subj} & \begin{bmatrix} \text{cat} & \text{np} \\ \text{gender} & \text{masc} \\ \text{pred} & \text{Robin} \end{bmatrix} \end{bmatrix}$$

Figure 2. Exemplary 'overspecified' χ

In the case where nouns in the lexicon are not marked for gender, as they might well not be for English, according to Definition 2 no sentence can be generated from this input, as χ_S for the obvious candidate, namely *Robin swims*, will not extend χ as it would lack the gender feature. But it seems unreasonable to rule this out, and indeed in our approach to translation to enforce the extension definition as above would be more than an inconvenience, but would rather make translation virtually unachievable. What seems to be required is a notion of maximal overlap, to go along with minimal extension, since obviously the structures in Figures 1 and 2 could be combined. What we want, then, is to define γ in terms of minimal extensions to maximal overlaps:

[4] Hereafter I will use the 'intensional' notation for extension, subsumption, unification and generalisation, using square-cornered set operators, as follows:

$ss \sqsubseteq ls$	ss subsumes ls; ls extends ss
$ss \sqsubset ls$	ss properly subsumes ls; ls properly extends ss
$ss_1 \sqcup ss_2 = ls$	ss_1 and ss_2 unify to ls
$ls_1 \sqcap ls_2 = ss$	ls_1 and ls_2 generalise to ss

The intuition appealed to is that of the set operators applying to sets of facts (ss—smaller set; ls—larger set).

Definition 3.

$\gamma(\chi_s,\chi)$ iff

χ_s and χ are compatible, that is, $\chi_s \sqcup \chi \neq \bot$ and

they are maximally overlapped, that is, $\nexists s' \ni \chi_s \sqcap \chi \sqsubset \chi_{s'} \sqcap \chi$ and

χ_s minimally extends its overlap with χ,

that is,

$\nexists s'' \ni \chi_{s''} \sqcap \chi = \chi_s \sqcap \chi$
$\wedge \chi_s \sqsupset \chi_{s''}$

Roughly speaking, χ_s must cover as much as possible of χ with as little left over as possible. Note that we have chosen to give priority to maximal overlap at the potential expense of minimal extension. For example, supposing all proper nouns are marked in the lexicon for person and number, and further that commitative phrases are not sub-categorised for, then given the input

$$\left[\begin{array}{ll}
\text{cat} & \text{s} \\
\text{pred} & \text{swim} \\
\text{subj} & \left[\begin{array}{lr}\text{cat} & \text{np} \\ \text{person} & 3 \\ \text{number} & \text{sg} \\ \text{pred} & \text{Robin}\end{array}\right] \\
\text{comm} & \left[\begin{array}{lr}\text{cat} & \text{pp} \\ \text{pcase} & \text{comm} \\ \text{pred} & \text{Kim}\end{array}\right]
\end{array}\right]$$

Figure 3. Exemplary χ for overlap/extension conflict

we will prefer *Robin swims with Kim*, with its extensions for the person and number features, as opposed to the non-extending *Robin swims*, because the latter overlaps less. Note that in the case of two alternatives with non-compatible overlaps, two alternative results are allowed by the above definition.

Note that this approach is quite weak, in that it contains nothing like Wedekind's completeness condition—if the grammar allows it, output may be produced which does not overlap large portions of the input structure, regardless of its status. For example structures which may be felt to be ungrammatical, as in Figure 4 below, may successfully generate surface strings on this account, i.e. *Hours elapsed*, despite 'leaving out' as 'important' a part of the underlying form as the direct object.

$$\left[\begin{array}{ll}
\text{cat} & \text{s} \\
\text{pred} & \text{elapse} \\
\text{subj} & \left[\begin{array}{lr}\text{cat} & \text{np} \\ \text{person} & 3 \\ \text{number} & \text{pl} \\ \text{pred} & \text{hour}\end{array}\right] \\
\text{obj} & \left[\begin{array}{lr}\text{cat} & \text{np} \\ \text{person} & 3 \\ \text{number} & \text{sg} \\ \text{pred} & \text{Kim}\end{array}\right]
\end{array}\right]$$

Figure 4. Exemplary 'ungrammatical' χ

If it is felt that generating anything at all from such an input is inappropriate, then some sort of completeness-with-respect-to-subcategorised-for-functions condition could be added, but my feeling is that although this might be wanted for grammar debugging, in principle it is neither necessary nor appropriate.

Alternatively one could attempt to constrain not only the relationship between χ_s and χ, but also the nature of χ itself. In the example at hand, this would mean for instance requiring some form of LFG's coherence restriction for subcategorisation frames. In general I think this approach would be overly restrictive (imposing completeness in addition would, for example, rule out the χ of Figure 1 above as well), and will not pursue it further here.

It is interesting to note the consequences for generation under this defintion of input at the extremes. For $\chi = \top$ (or any structure with no grammatical subset), the result will be the empty string, if the language includes that, failing which, interestingly, it will be the set of minimal sentences(-types) of the language, e.g. probably just intransitive imperative and indicative in all tenses for English.

The case of $\chi = \bot$ is trickier. If $\bot$ is defined such that it extends everything, or alternatively that the generalisation of anything with $\bot$ is the thing itself, then 1) $\bot$ is infinite so 2) no finite structure can satisfy the maximal overlap requirement; but in any case $\bot$ fails to satisfy the first clause of 3, namely the unification of χ_S and χ must not be $\bot$, since if χ is $\bot$ then χ_S and χ unify to $\bot$ for any χ_S.

Finally note that in cases where substantial material has to be supplied, as it were, by the target grammar (e.g. if a transitive verb is supplied but no object), then Definition 3 would allow arbitrary lexicalisations, giving rise to a very large number of permissible outputs. If this is felt to be problem, then restricting (in the sense of (Shieber 1985)) the subsumption test in the second half of Definition 3 to ignore the values of certain features, i.e. pred, would be straight-forward. This would have the effect of producing a single, exemplary lexicalisation for each significantly different (i.e. different ignoring differences under pred) structure which satisfies the minimaximal requirements.

II.4 A Problem with the Mini-maximal Approach

One potential problem clearly arises with this approach. It stems from its dependence on subsumption and its friends. Since subsumption, in at least some standard formulations (e.g. Definite Clause Grammars) fails to distinguish between contingently and necessarily equivalent sub-structures, we will overgenerate in cases where this is the only difference between two analyses, e.g. for *Kim expects to go* and *Kim expects Kim to go* on a straight-forward account of Equi. One can respond to this either by saying that this is actually correct, that Equi is optional anyway (wishful thinking, I guess), or by adding side conditions to Definition 3 which amount to strengthening subsumption etc. to differentiate between e.g. the two graphs in Figure 5. As I do not at the moment see any way of expressing these side conditions formally without making more assumptions about the nature of underlying forms than I have so far had to (c.f. for example (Shieber 1986) where subsumption is defined in terms of a simulation plus an explicit requirement on the preservation of token identity), I will leave this point unresolved.

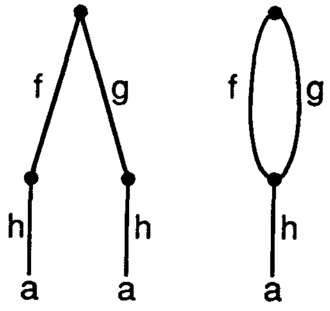

Figure 5. Two structures not distinguished by subsumption

III. Theory-Based Translation

As mentioned above, the need to consider more carefully the nature of the adequacy conditions for the generation relation has arisen from developments in theory-based translation (Kaplan et al. 1989, Sadler and Thompson 1991, van Noord 1990). Although a range of different approaches fall uńder this description, they all share some amount of grammaticalisation of translation regularities. Furthermore, they all appeal to some form of reversibility or bi-directionality. Figure 6 below provides a schematic characterisation of all these approaches, where Δ and Γ are as before, and T is for an optional transfer component.

$$\begin{tikzcd}
s_{source} \arrow[r, "\Delta_{G_{source}}"'] & \chi \\
(\chi \arrow[r, "T_{T_{source/target}}"'] & \chi') \\
\chi(') \arrow[r, "\Gamma^{\gamma}_{G_{target}}"'] & s_{target}
\end{tikzcd}$$

Figure 6. Schematic characterisation of translation

The important point about these approaches is that the output of the analysis process is the input to the generation process. This is in contrast to previous transfer approaches, in which transfer produces some distinct new structure for input to generation. If a transfer component is included in the approaches I'm concerned with, as in van Noord (1990), its rules function to elaborate the product of analysis, not replace it, and they could without loss of generality be incorporated into the source and/or target grammars.

Now we can formalise the picture in Figure 6 as follows:

Definition 4.

$TP^{\gamma}_{G_s,G_t}(s,t)$ (a string s translates$_{\gamma}$ to a string t for grammars Gs,Gt)
iff $\exists\, \chi_s \ni \Delta_G(s,\chi_s)$ and $\Gamma^{\gamma}_{G}(\chi_s,t)$

The goal of this enterprise has been to provide a version of γ which makes this a practical definition of theory-based translation, and it should be clear how all the phenomena which were used in section II to motivate the Definition 3 version of γ are likely to arise in translation. In particular, the necessity for allowing the overlap between χ_s and χ_t to be less than total arises from the obvious asymmetry which will exist between the syntactic contents of the two—in whatever form is appropriate to the grammatical theory involved, χ_s will contain a full syntactic analysis in the source domain, and possibly only a root S node for the target, while for χ_t the situation will be reversed. The mini-maximal approach given above covers this case straight-forwardly.

IV. Beyond Subsumption

The use of subsumption as the basis for my explorations of γ has another problem, in that typically definitions of subsumption require that the structures to be compared share a common root. For reasons which would take too long to set out, this constraint too may prove over-strong in certain translation cases. By way of illustration, consider translating into a language in which overt performatives are required for all grammatical utterances. We would then find that the translation into this language of e.g. *Robin swims* would involve a higher predicate, so for various parts of the product of analysis, the appropriate relationship would hold not between root and root, but between root and sub-part. This suggests that a weaker relationship, perhaps the existence of a homomorphism, should replace subsumption in the definition of γ.

V. IMPLEMENTATION

I have made some progress towards implementing a generator based on Definition 2 of section II. I believe it will be possible to provide an implementation which is guaranteed to provide all and only the correct outputs if any exist, but may fail to terminate if no output is possible. The basic idea is to constrain the generator to produce results in node-cardinality order, that is, smallest first. In fact, there is some slop in the most straightforward way of implementing this, in that it is fairly simple to limit the number of nodes allocated, but more difficult to constrain the number eventually used. What is guaranteed, however, is that structures are produced in an order which respects subsumption, in that if χ_S subsumes χ_S', then it will be generated first. This in turn means that one can enforce the minimality constraint of Definition 2.

The problem arises with certain classes of recursive definition, both the simple left recursion cases of more traditional grammars, and the more complex ones of categorial-style ones. My best guess for these is to anticipate that it would be possible to (semi-)automatically prove that any such rule produced via recursion a structure which was 'subsumed' (as per section IV above) by one with less recursion. This in turn would mean that provided some result had been found, the recursion could be terminated, since any further downstream result would fail the minimality constraint. If however no result could be found, there would be no basis for stopping the recursion other than a very ad-hoc shaper test (Kuno 1965), based on some more or less arbitrary (depending on the application) limit on the size of the expected output.

At the moment I have no ideas on how to implement a generator which respects Definition 3.

ACKNOWLEDGEMENTS

The work reported here grew out of work carried out while the author was a visitor to the Embedded Computation and Natural Language Theory and Technology groups of the Systems Science Laboratory at the Xerox Palo Alto Research Center. These groups provided both the intellectual and material resources required to support that earlier work, for which thanks. Many of the ideas presented here were first articulated in discussions with Jo Calder and Mike Reape, and many of the HCRC coffee room regulars also contributed patience and suggestions—my thanks to them all. Thanks also to Klaus Netter, in particular for first calling to my attention the behaviour of Definition 3 with respect to the extreme cases.

REFERENCES

Estival, D. [1990] Generating French with a Reversible Unification Grammar. In Karlgren, H. (ed.) *COLING90*, 1990, pp106-111.

Gardent, C. and Plainfossé, A. [1990] Generating from a Deep Structure. In Karlgren, H. (ed.) *COLING90*, 1990, pp127-132.

Kaplan, R., Netter, K., Wedekind, J. and Zaenen, A. [1989] Translation by structural correspondences. In *Proceedings of the Fourth Conference of the European Chapter of the Association for Computational Linguistics*, University of Manchester Institute of Science and Technology, Manchester, UK, 10-12 April, 1989, pp272-281.

Kuno, S. [1965] "The predictive analyzer and a path elimination technique", *Communications of the ACM*, **8**, 687-698.

Momma, S. and Dörre, J. [1987] Generation from f- Structures. In Klein, E. and van Benthem, J. (eds.) *Categories, Polymorphism and Unification*, pp148-167. Edinburgh and Amsterdam: University of Edinburgh, Centre for Cognitive Science and Institute for Language, Logic and Information, University of Amsterdam.

Sadler, L. and Thompson, H. S. [1991] Structural Non- Correspondence in Translation. In Kunze, J. and Reimann, D. (eds.) *Proceedings of the Fifth European Association for Computational Linguistics*, Berlin, April, 1991, pp293-298.

Shieber, S. M. [1985] Using restriction to extend parsing algorithms for complex-feature-based formalisms. In *Proceedings of the 23rd Annual Meeting of the Association for Computational Linguistics*, pp145-152.

Shieber, S. M. [1986] *An Introduction to Unification- based Approaches to Grammar*. Chicago, Illinois: The University of Chicago Press.

Shieber, S. M., van Noord, G., Pereira, F. C. N. and Moore, R. C. [1990] Semantic-Head-Driven Generation. *Computational Linguistics*, 16, 30-42.

van Noord, G. [1990] Reversible Unification Based Machine Translation. In Karlgren, H. (ed.) *COLING90*, 1990, pp299- 304.

Wedekind, J. [1988] Generation as structure driven derivation. In *COLING88*, Budapest, Hungary, 22-27 August, 1988, pp732-737.