MEASURING COMPOSITIONALITY IN TRANSFER-BASED MACHINE TRANSLATION SYSTEMS*

Björn Gambäck
Swedish Institute of Computer Science
Box 1263, S - 164 28 KISTA, Stockholm
gam@sics.se

Hiyan Alshawi, David Carter and Manny Rayner†
SRI International
Cambridge Computer Science Research Centre
23 Millers Yard, Cambridge CB2 1RQ, U.K.
hiyan@cam.sri.com, dmc@cam.sri.com, manny@cam.sri.com

Abstract

The paper describes a simple method for objectively evaluating the compositionality of a transfer-based Machine Translation system. The question is the extent to which rule interaction gives rise to (unwanted) side-effects. An example is given of the use of the method in the context of the BCI (Bilingual Conversation Interpreter), an interactive transfer-based bidirectional Machine Translation system.

Introduction

When trying to evaluate a Machine Translation system, two different approaches are possible: either the system's behaviour in its proposed environment is assessed, or the theoretical coverage and worth of the transfer formalism is evaluated. The first type of evaluation concentrates on translation quality and effectivess, while the latter seeks to specify which linguistic constructions the system can handle. Most work in the field have been concerned with system behaviour; here, we will concentrate on linguistic coverage.

In the literature on Machine Translation, a number of criteria are mentioned as significant when evaluating the worth of a transfer formalism; among these are *expressiveness*, *simplicity*, *generality*, *reversibility*, *language-independence*, *monotonicity* and *compositionality*. Unfortunately, when trying to convince others of the worth of one's own approach, it soon becomes evident that most of these are not easy to measure objectively, if they are not absolute properties of the formalism. (In particular, a pure unification-based formalism is guaranteed to be monotonic). To say, for example, that a formalism is "good" from the point of view of expressiveness, and then back this up with five carefully-chosen examples, is not really to say very much.

Compositionality, however, can be measured objectively. Here, we will describe a simple method for evaluating the compositionality of a transfer-based MT system, and give an example of its use in the context of the BCI (Bilingual Conversation Interpreter) (Alshawi *et al* 1991), an interactive transfer-based bidirectional system currently being developed in a co-operation between SICS and SRI Cambridge. The main components of the BCI are English (Alshawi ed. 1991) and Swedish (Gambäck, Lövgren & Rayner 1991) versions of the SRI Core Language Engine, transfer taking place at the level of Quasi Logical Form (QLF) (Alshawi & van Eijck 1989); the transfer formalism is unification-based and bidirectional. Our approach to Machine Translation is aimed at keeping the transfer component as simple as possible, while depending on fully constrained reversible monolingual grammars for correct analysis and synthesis.

*Part of the research described in this paper was also reported on at the *Meeting of the International Working Group on Evaluation of Machine Translation Systems*, Les Rasse, Switzerland, April 1991.

†The work reported here was funded by the Swedish Institute of Computer Science, and the greater part of it was carried out while the fourth author was employed there.

Measuring compositionality

Perhaps the most important factor in keeping transfer simple is the degree to which the transfer relation is a homomorphism, i.e. the degree to which transfer rules are compositional.

For compositionality to be a meaningful notion in the first place, it must be possible for transfer rules to apply to partial structures. These structures can consequently occur in different contexts; other transfer rules will apply to the contexts as such. The question is the extent to which particular combinations of rules and contexts give rise to special problems. In a perfectly compositional system, this will never happen, although it seems a safe bet that no such system exists today. What we want is a method which objectively measures how closely we approach the compositional ideal.

Our first step in this direction has been the construction of *compositionality tables*, in which a set of rules and a set of contexts are systematically combined in all possible meaningful combinations. This is done in order to figure out the extent to which the complex transfer rules continue to function in the different contexts.

In the following three diagrams, we give an example of such a table for the current version of the BCI. Table 1 gives a set of rules, which exemplify six common types of complex transfer. Table 2 gives a set of twelve common types of context in which the constructions referred to by the rules can occur. Finally, Table 3 on the next page summarizes the results of testing the various possible combinations.

To test transfer compositionality properly, it is not sufficient simply to note which rule/context combinations are handled correctly; after all, it is always possible to create a completely *ad hoc* solution by simply adding one transfer rule for each combination. The problem must rather be posed in the following terms: if there is a single rule for each complex transfer type, and a number of rules for each context, how many *extra* rules must be added to cover special combinations? It is this issue we will address.

Table 1: Types of complex transfer used

Type	Example
Different particles	John likes Mary John tycker om Mary
Passive to active	Insurance is included Försäkring ingår
Verb to adjective	John owes Mary $20 John är skyldig Mary $20
Support verb to normal verb	John had an accident John råkade ut för en olycka
Single verb to phrase	John wants a car John vill ha en bil (lit.: "wants to have")
Idiomatic use of PP	John is in a hurry John har bråttom (lit.: "has hurry")

Table 2: Transfer contexts used

Context	Example
Perfect tense	John has liked Mary John har tyckt om Mary
Negated	John doesn't like Mary John tycker inte om Mary
YN-question	Does John like Mary? Tycker John om Mary?
WH-question	Who does John like? Vem tycker John om?
Passive	Mary was liked by John Mary blev omtyckt av John
Relative clause	The woman that John likes Kvinnan som John tycker om
Sentential complement	I think John likes Mary Jag tror John tycker om Mary
Embedded question	I know who John likes Jag vet vem John tycker om
VP modifier	John likes Mary today John tycker om Mary idag
Object raising	I want John to like Mary Jag vill att John ska tycka om Mary ("I want that J. shall like M.")
Change of aspect	John stopped liking Mary John slutade tycka om Mary ("J. stopped like-INF M.")

Table 9: Compositionality Table
(Swedish-English shown above English-Swedish)

Transfer context	Different particles	Active to passive	Verb to adjective	Support verb to normal verb	Single verb to phrase	Idiomatic use of PP
Present tense	OK OK	OK OK	OK OK	OK OK	OK OK	OK OK
Perfect tense	OK OK	generator OK	OK OK	OK OK	OK OK	OK OK
Negated	pres-not pres-not	pres-not pres-not	pres-not pres-not	past-not past-not	pres-not pres-not	transfer transfer
YN-question	OK OK	OK OK	OK OK	OK OK	OK OK	OK OK
WH-question	OK OK	OK OK	OK OK	OK OK	OK OK	OK OK
Passive	OK OK	- -	- OK	- -	- OK	- -
Relative clause	OK OK	OK OK	OK OK	OK OK	OK OK	OK OK
Sentential complement	OK OK	OK OK	OK OK	OK OK	OK OK	OK OK
Embedded question	OK OK	OK OK	OK OK	OK OK	OK OK	OK OK
VP modifier	OK OK	transfer transfer	OK OK	OK OK	OK OK	OK OK
Change of aspect	OK OK	OK OK	OK OK	OK OK	OK OK	OK OK
Object raising	transfer OK	transfer OK	transfer OK	transfer OK	transfer OK	transfer OK

Each square in Table 3 consists of two entries, the first for the Swedish-English, and the second for the English-Swedish direction. The entries are to be interpreted as follows:

- **-** means that the combination was not applicable, i.e. that the construction referred to by the rule cannot occur in this context.
- **OK** means that analysis, transfer and generation all functioned correctly, without any extra rule being necessary to deal with the particular context.
- **generator** means that the generator component was unable to generate the correct target language sentence.
- **transfer** means that the transfer component was unable to make a correct transfer.
- All other entries are names of rules needed to deal with special combinations of rule and context. For this table, only two extra rules were needed: **pres-not**, which reverses the relative scope of the operators for negation and the present tense and **past-not**, which performs a similar function for the past tense

The actual results of the tests were as follows. There were 136 meaningful combinations (some constructions could not be passivized); in 115 of these, transfer was perfectly compositional, and no extra rule was needed.

Of the remaining 21 rule/context/direction triples, seven failed for basically uninteresting reasons: the combination "Perfect tense + Passive-to-active" did not generate in English, and the six sentences with the object-raising rule all failed in the Swedish-English direction, since that rule is currently uni-directional. The final fourteen failures are significant from our point of view, and it is interesting to note that all of them resulted from mismatches in the scope of tense and negation operators.

The question now becomes that of ascertaining the generality of the extra rules that need to be added to solve these fourteen unwanted interactions. To reorder the scopes of tense, negation and modifiers, and account for the scope differences between the English and Swedish QLFs arising from the general divergences in word-order and negation of main verbs relevant here, two rules involving general transformations of the QLF structure were added. These solved ten of the outstanding cases.

The four bad interactions left all involved the English verb *to be*; these were the combinations "Passive to active + VP modifier" and "Idiomatic use of PP + negation", which failed to transfer in either direction. Here, there is no general solution involving the addition of a small number of extra rules, since the problem is caused by an occurrence of *to be* on the English side that is not matched by an occurrence of the corresponding Swedish word on the other. The solution must rather be to add an extra rule *for each complex transfer rule in the relevant class* to cover the bad interaction.

Summarizing the picture, to solve the specific examples in the test set, two extra rules were thus required. The tests revealed that all bad interactions between the transfer rules and contexts shown here could be removed by adding four extra rules to cover the 124 possible interactions.

Extending the framework

It should be pointed out that the compositionality table presented here is still too small to detect more than a fraction of the bad rule interactions that may occur in the current system. Most important is to extend systematically the set of contexts, taking note of the fact that many of the features they are intended to represent are in fact orthogonal to each other.

A full set of contexts would include at a minimum all legal combinations of independent choices along the following dimensions:

- *Tense*: Present, past or future.
- *Mood*: Active or passive.
- *Negation*: Positive or negative.
- *Modification*: Unmodified, PP modification, ADVP modification, modified by fronted constituent.
- *Clause-type*: Declarative sentence, Y-N question, WH-question, relative clause, sentential complement, embedded question, progressive VP complement, object raising.

Multiplying out all the choices gives a total of 384 distinct contexts; this must then be multiplied by the number of transfer rule types to be tested, and doubled to get both directions of transfer. With the figures given above, 4608 sentences would have to be tested. In practice, of course, not all combinations are possible. Specifically, passives don't interact well with other rule-contexts, leading to a total size of the test set of 3082 sentences.

Developing the software support needed to be able to run tests of this size regularly is clearly not a trivial task, but our opinion is that being able to do so greatly contributes to maintaining the system's reliability and integrity. We are thus giving high priority to constructing the necessary tools in the current phase of the project.

Also worth noting is that the tests described above are exclusively at the sentence level. For complete tests of the compositionality of transfer, one would have to construct test schemes for at least the noun phrase level, as well. The compositionality tables for NPs should account for the interactions (in various positions) of different NP-modifiers. Thus, the transfer contexts should be something like the ones suggested in Table 4 and the transfer types should include the ones given in Table 5. This will be further studied in the next phase of the project.

Conclusions

We have described a straight forward way of measuring the compositionality of transfer-based MT systems by the use of "compositionality tables" We claim this to be a good method for the objective evaluation of one aspect of MT systems, even though the tables given in this paper should be further extended to capture more transfer contexts and types of transfer rules, as well as NP-structures

Table 4 NP transfer contexts

Transfer context	Example
Plural	car parks parkeringsplatser
Definite	the car park parkeringsplatsen
Genitive	car park's parkeringsplatsens
Pre-modified by Adjective	big car park stor parkeringsplats
Pre-modified by Genitive	his car park hans parkeringsplats
Post-modified by PP	car park here parkeringsplats här
Post-modified by Relative clause	car park which I use parkeringsplats som jag använder

Table 5: Complex NP transfer types

Transfer type	Example
Adjective Noun to Noun	bad luck otur
Noun PP to Noun	chairman of the board styrelseordförande
Noun Noun to Noun	car park parkeringsplats
Past Participle to Adjective	The broken cup Den trasiga koppen
Adjective to Relative clause	The uninsurable car Bilen som inte kan försäkras
PP to Genitive	The end of the story Sagans slut

References

Alshawi, H. and J. van Eijck. 1989. "Logical Forms in the Core Language Engine". *27th Annual Meeting of the Association for Computational Linguistics*, Vancouver, British Columbia, 25-32.

Alshawi, H . ed. 1991 (to appear). *The Core Language Engine*. Cambridge, Massachusetts: The MIT Press.

Alshawi, H., D. Carter, B. Gambäck and M. Rayner. 1991. "Translation by Quasi-Logical Form Transfer". *29th Annual Meeting of the Association for Computational Linguistics*, Berkeley, California.

Alshawi, H., D. Carter, B. Gambäck and M. Rayner. 1991 (to appear). "Swedish-English QLF Translation". In H. Alshawi (ed.), op. cit.

Falkedal, K. Forthcoming. "Evaluation Methods for Machine Translation Systems: An Historical Overview and a Critical Account". *Technical Report*, ISSCO, Geneva.

Gambäck, B., A. Lövgren and M. Rayner. 1991 (to appear). "The Swedish Core Language Engine". *SICS Research Report*, Swedish Institute of Computer Science, Stockholm.

Lehrberger, J. and L. Bourbeau. 1988. *Machine Translation: Linguistic characteristics of MT systems and general methodology of evaluation*. Amsterdam: John Benjamins Publishing.

Way, A. 1991. "Developer Oriented Evaluation of MT Systems". *Technical Report*, The University of Essex.