

Aspects of the Translation of English Subordinate Clauses into German

Published in: Problems and Potential of English-to-German MT Systems. Workshop at the 8th International Conference on Theoretical and Methodological Issues in Machine Translation. TMI-99. Chester, England. August 1999.

Stephan Mehl

Gerhard-Mercator-Univ. - GH Duisburg
Computational Linguistics
D-47048 Duisburg
he234me@unidui.uni-duisburg.de
<http://verdi.uni-duisburg.de/Englisch/mehl.html>

Martin Volk

Universität Zürich
Department of Computer Science
Winterthurer Str. 190
CH-8057 Zürich
volk@ifi.unizh.ch
<http://www.ifi.unizh.ch/CL/volk/>

Abstract. While many errors in machine translated texts are due to word sense ambiguity, syntactical transfer still poses problems, too. For a particular type of syntactic structure, namely, subordinate clauses, this paper distinguishes five kinds of translation difficulties: cross-language differences, syntactic ambiguity, functional ambiguity, stylistic preferences, and grammatical coverage.

1. Introduction

In recent years, commercial MT systems running on PCs have attracted much attention. Although the translations still contain many errors, they are usually comprehensible, which makes them useful for people with a poor knowledge of the source language. A closer look at machine made translations reveals that the majority of mistakes are due to deficiencies in lexical choice since the system often has to select randomly among multiple readings within the same part-of-speech (e.g. *bank*: *Bank* or *Ufer*). While such semantic mistakes can easily be explained or even predicted, syntactic errors are much harder to understand: Often a particular construction is translated well, while a similar-looking sentence causes surprising problems.

This paper gives some reasons for the diverging results in syntactic processing and estimates the chances of improvement. This is done for one particular syntactic domain, namely, the translation of subordinate clauses. The following five problem classes will be distinguished:

1. Syntactic analysis can be hampered by structural ambiguities such as the attachment of a prepositional phrase or a relative clause, or the detection of clause boundaries.

2. A related problem is functional ambiguity. There are cases in which a subordinate clause might function either as a relative or a complement clause, due to the ambiguity of *that* as a pronoun or a complementizer.
3. While the structure of English and German subordinate clauses are similar in most cases, some constructions cause difficult transfer problems.
4. A special case of transfer problem arises if the transfer component has to choose between alternative German translations of an English structure. Stylistic preferences come into play here.
5. There are still some (combinations of) grammatical phenomena not covered by the grammar of some systems.

These five types of problems will be discussed in the following sections and illustrated with translations from three commercial MT systems: IBM's Personal Translator plus 98 (PT+98), Langenscheidt's T1 Professional 3.0 (T1), and Globalink's PowerTranslator Pro 6.5 (PwTr). This paper does not aim at comparing the translation quality of these systems. We use multiple translations solely to avoid a bias towards coincidental correct translations by any one system.

Furthermore, to avoid a concentration on obsolete constructions, example sentences are taken from the WWW page of the City of Chester as an arbitrary sample of a natural text.¹

2. Structural Ambiguity

Many types of English subordinate clauses can be identified easily, their form and function being indicated by punctuation marks, complementizers, relative pronouns, etc. The structural transfer of these clauses into German does not pose any problems. These comprise all kinds of relative and complement clauses, including reduced forms. Of course, individual clauses are difficult to translate because of additional structural ambiguities that interfere with the correct interpretation of the subordinate clause. Some examples:

- PP attachment ambiguity: Adjunct prepositional phrases can either be attached to the adjacent phrase or to its matrix clause. PT+98 assigns adjuncts at the end of a sentence to the main clause, which may sometimes be right, as in (1), but is generally the wrong heuristic, as illustrated in (2), especially with the preposition *of* that has a strong tendency towards attachment to the adjacent phrase. Recent studies (e.g. Mehl, Langer, Volk 1998) show that statistical methods can help in resolving such PP attachment ambiguities.

¹ (Mehl, Heidemann, Volk 1998) compiled a systematic list of clause types, together with a synopsis of the results of translating these types by the above-mentioned MT systems. The data gathered in this study are accessible from our homepages.

(1) *You'll find everything you need to book your accommodation in Chester Visitor Guide.*
 PT+98: *Sie finden alles, das Sie brauchen, um Ihre Unterkunft zu buchen, in Chester Besucherführung.*

(2) *It will be of interest to all who are concerned with the history of the north-west region.*
 PT+98: *Es wird von Interesse zu allen der Nordwestregion sein, die um die Geschichte besorgt sind.*

- **Coordination:** Of course, coordination as one of the most serious problems in syntactic analysis plays an important role in subordinate constructions, too. (3) is a coordination of gerunds, one of which was misinterpreted as a participle forming a relative clause; in (4) two coordinated verbs wrongly share an object:

(3) *Are you thinking of booking a meeting, planning a conference, association gathering?*
 PT+98: *Denken Sie daran, eine Besprechung zu buchen, die eine Tagung, Verbandsversammlung plant?*

(4) *Everyone has to use a pencil, as pens can leak and permanently damage documents.*
 PT+98: *Jeder muß einen Bleistift verwenden, da Stifte Dokumente durchsickern lassen und dauernd beschädigen können.*

As global structural ambiguities, these cases can be solved by semantic analysis only. In (3) the system would need to know that *booking* and *planning* are related concepts, and in (4) selectional restrictions must specify that the object of "leak" needs to be a fluid or gas.

3. Functional Ambiguity

Even if their attachment is clear, certain clause types will usually fail to be translated properly for lack of semantic knowledge.

- **Adverbial clauses** are a typical representative of this category: Since many English conjunctions have various meanings (e.g. *when: wenn, als, während; since: seit, da; if: falls, wenn, ob*), it is rarely possible for an MT system to find out the correct semantic relation between main and subordinate clause, and to select an appropriate German construction. (5) illustrates this for *as* which can be interpreted as a causal conjunction (done correctly by PT+98 in this example), as modal conjunction (chosen by T1) or as temporal conjunction (chosen by PwTr):

(5) *Everyone has to use a pencil, as pens can leak and permanently damage documents.*
 PT+98: *Jeder muß einen Bleistift verwenden, da Stifte Dokumente durchsickern lassen und dauernd beschädigen können.*

T1: *Jedes muss einen Bleistift benutzen, wie Schreibstifte Dokumente dauerhaft durchlassen und beschädigen können.*

PwTr: *Jeder muß einen Bleistift benutzen, als Kugelschreiber lecken können und können permanent Dokumente beschädigen.*

- A similar problem arises with the interpretation of infinitives. As complements, they have to be translated according to the respective German subcategorization requirements; as adjuncts in a purpose clause they are rendered as an infinitive with *um zu*. (6) and (7) are examples of these two cases; PT+98 chose the wrong interpretation in both cases, perhaps because it selected the wrong verb reading.

(6) *At Chester Archives we ask students working in groups of 3 or more to make a prior appointment.*

PT+98: *An Chester Archiven fragen wir Studenten, die in Gruppen von 3 oder mehr arbeiten, um einen vorherigen Termin zu machen.*

(7) *It's always worth ringing us to see how we can help you.*

PT+98: *Es ist immer wert, uns zu läuten, zu sehen, wie wir Ihnen helfen können.*

This problem arises in part because the correct subcategorization frame for *ask* is missing. But it is nevertheless unclear why the system treats both cases in a different way. In addition to that, there may be cases where the adjunct reading as a purpose clause (as chosen by PT+98) is correct (cf. *He asked me again to make me angry.*) We admit that this is a rare construction, however. A manual inspection of some dozen sentences with *ask* + infinitive in the COBUILD Word Bank did not reveal any occurrence of a purpose clause.

- Detached participle clauses are another case where only semantics can help.² The literal translations in (8) and (9) are comprehensible, but unidiomatic; however, alternative solutions such as *Auf einem Gelände von ... / Seit unserer Gründung 1973 ...* are impossible to achieve for any MT system, if it does not know these phrases as complex patterns. If the order of participle and complement is reversed (as done by PwTr), the otherwise unusual participle construction in German leads to a better translation.

(8) *Set in 110 acres of landscaped gardens, it houses many rare and endangered creatures in its wide collection of mammals, birds, reptiles and fish.*

PT+98: *Eingesetzt in 110 Morgen landschaftlich gestaltete Gärten, bringt sie viele seltene und gefährdete Kreaturen in ihrer breiten Sammlung von Säugetieren, Vögeln, Reptilien und Fisch unter.*

(9) *Founded in 1973 [...], we undertake the full range of archaeological responsibilities in Chester District.*

² Source sentence (9) sounded strange to one of the reviewers, but querying the COBUILD Word Bank we found many similarly structured sentences, that started with *Founded* and a reduced participle clause.

PT+98: *Gegründet in 1973 [...] übernehmen wir den vollen Bereich von archäologischen Verantwortungen in Chester Bezirk.*

PwTr: *In 1973 [...] gegründet, unternehmen wir die volle Auswahl archäologischer Verantwortungen in Chester Gebiet.*

4. Transfer Problems

While English and German have corresponding structures for many clause types, some others need a structural transfer. Typical examples are gerund and participle clauses as well as some types of infinitives. In contrast to the cases listed in sections 2 and 3, however, there are syntactic indications that support a correct analysis. This section illustrates that some MT systems nevertheless fail to translate these structures properly. In section 5 we show that in case a correct translation is found, it does not always obey stylistic constraints.

4.1 Infinitives with a subject of their own

A *for-PP* preceding an infinitive usually represents the infinitive's subject. In (10), all three MT systems interpret the PP as complementing the adjective instead.³

(10) *Material is available for the public to consult.*

PT+98: *Material ist für die Öffentlichkeit verfügbar zu beraten.*

T1: *Material ist für die Öffentlichkeit vorhanden beraten zu.*

PwTr: *Material ist für die Öffentlichkeit verfügbar, zu konsultieren.*

Unfortunately, not every *for-PP* represents the subject of a subsequent infinitive; hence we have a structural ambiguity again:

(11) *All Chester residents over the age of 18 are entitled to vote for City Councillors to represent them at elections.*

PT+98: *Alle Chester Bewohner über dem Alter von 18 sind berechtigt, zu wählen, daß Stadtratsmitglieder sie an Wahlen darstellen.*

But a small corpus survey showed that by far in the majority of occurrences of a *for-PP + infinitive* the PP contains the infinitive's subject. We manually inspected 75 sentences with a *for-PP + infinitive* and we found only 12 negative examples where the PP did not contain the subject. This mostly happened with verbs taking a *for-PP* as a prepositional object and the PP containing a noun that takes an infinitival complement.

(12) *However, an unregistered body [...] had applied for permission to stage a public show on 1 May.*

³ As pointed out by one of the reviewers (10) is particularly hard to translate since it lacks the overt direct object of the verb *consult*.

(13) *They are being tried for conspiracy to commit rebellion or insurrection ...*

Thus, a good heuristic would ask for all *for*-PPs + infinitives to be translated as the infinitive's subject unless the *for*-PP is subcategorized for by a preceding verb and its noun takes an infinitival complement.

4.2 Distinction of gerunds and participles

Both gerund and participle constructions may be translated as subordinate clauses, but while gerunds correspond to complement clauses, and may alternatively be rendered as a noun phrase, participles correspond to relative clauses, and may also be translated as adjective phrases. Since gerund clauses often start with a preposition, which is not possible for participle clauses, there are few ambiguous cases such as the famous *They discussed walking on the beach* (see also (3) above).

However, in (14) PT+98 misinterprets the preposition as a particle, and the gerund as a participle introducing a relative clause. The other systems fare better. T1 inserts a pronominal adverb (*dabei*) while PwTr omits the preposition. All three systems fail to correctly attach the *by*-PP as object of the passive verb. This is surprising as one would expect that such an attachment would constitute a good heuristic in combination with passive verbs.

(14) *The Lord Mayor is assisted in carrying out civic and ceremonial duties by the Sheriff and the Deputy Lord Mayor.*

PT+98: *Dem Oberbürgermeister wird ein geholfen, bürgerliche und zeremonielle Pflichten durch den Sheriff und den stellvertretenden Oberbürgermeister ausführend.*

T1: *Dem Oberbürgermeister wird dabei geholfen, Bürger und Zeremonie Pflichten vom Sheriff und dem Stellvertreteroberbürgermeister durchzuführen.*

PwTr: *Dem Herrn Mayor wird geholfen, staatsbürgerliche und zeremonielle Pflichten vom Sheriff und dem Abgeordnete-Herr-Bürgermeister auszuführen.*

The cases mentioned in this section have in common that many MT systems fail to process them correctly despite the fact that they can in general be identified with the help of syntactic clues, i.e. mostly verbs' subcategorization requirements. In (14) PT+98 could have avoided the mistranslation by relying on the subcategorization of the verb *assist* taking the preposition *in* + gerund. Obviously, this leaves those cases unresolved where a verb has multiple subcategorization requirements and more than one of them is applicable to a given sentence.

5. Stylistic Preferences

In section 4 we showed that clauses such as gerunds, participles or infinitives may be hard to translate properly. An additional aspect is that there are different stylistic variants for translating them, so that an MT system should ideally apply stylistic constraints to find the best solution. For example, a successful strategy might aim at a mix of verbal and nominal elements in order to avoid multiple embeddings of one or the other type of structure. We will illustrate this with few examples here; cf. Mehl 1996 or Gdaniec 1998 for further discussion.

5.1 Gerunds

Most of the time the realisation of an English gerund as a German verb gives the best results. In (15) the verbal translations by PT+98 and T1 sound more natural than the nominalization chosen by PwTr.

(15) *We know that much of Chester's success so far is based on using its great history to attract visitors, over 9 million last year.*

PT+98: *Wir wissen, daß so viel von Chesters Erfolg so weit auf Grundlage davon ist, seine große Geschichte zu verwenden, um über 9 Millionen letztes Jahr Besucher anzuziehen.*

T1: *Wir wissen so viel von dem Erfolg von Chester bisher gründet darauf seine große Geschichte zu benutzen um Besucher anzuziehen daß Ende 9 Millionen halten werde Jahr.*

PwTr: *Wir wissen, daß viel von Chester 's Erfolg so wird weit auf dem Benutzen seiner großer Geschichte, um Besucher anzuziehen, über 9 Million letztes Jahr, gegründet.*

A nominal realization may lead to long distances between subject and predicate (as with PwTr in 15 and 16) which are difficult to process for a human reader. A verbal realization lends itself much easier to a natural sequence and embedding of clauses.

(16) *[...] continue working together to develop the cultural life of the district despite the news that there is unlikely to be lottery money available for the project*

T1: *[...] daß man fortfährt, zusammen zu arbeiten, um so dort das kulturelle Leben des Bezirks trotz der Nachricht zu entwickeln, daß es unwahrscheinlich ist, ein für das Projekt vorhandenes Lotteriegeld zu sein.*

PwTr:*[...] das Zusammenarbeiten, um das kulturelle Leben des Gebietes trotz der Nachrichten zu entwickeln, die es unwahrscheinlich geben, Lotterie-Geld für das Projekt verfügbar zu sein, fortzusetzen.*

Although a nominalization is always possible, it is especially inappropriate in case of reflexives (as in 17) or tenses other than present:

(17) *Loans must be booked out in advance by contacting Mike Hardman.*

PT+98: *Darlehen müssen gebucht werden aus im voraus durch Sich Wenden an Mike Hardman.*

5.2 Infinitives

Similarly, infinitives can be translated as nouns or verbs. In (18), PT+98 chose to translate both infinitives as subordinate clauses. Although this is not incorrect, a combination of nouns and verbs (*um Ihnen bei der Ausarbeitung Ihrer Reise zu helfen*) will be stylistically preferred.

- (18) *The Tourist Office has an AA Route Planner facility to help you work out your journey.*
 PT+98: *Das Reisebüro hat eine AA Streckenplanereinrichtung, um Ihnen zu helfen, Ihre Reise auszuarbeiten.*

5.3 Participial Clauses

Participles may be translated either as a subordinate clause, as in (19), or an adjective phrase embedded in a noun phrase, as in (20):

- (19) *Henry, Prince of Wales, issues an order expelling the Welsh from Chester.*
 PT+98: *Henry, Prinz von Wales erteilt eine Ordnung, die das Walisische aus Chester ausweist.*
- (20) *The Roman Amphitheatre is the largest arena ever uncovered in Britain.*
 PT+98: *Das römische Amphitheater ist die jemals in Großbritannien aufgedeckte größte Arena.*

However, the latter solution is stylistically unacceptable in case of multiple embeddings. PwTr fares much better in (21) than PT+98 since it preserves the order of subordinate embeddings:

- (21) *This publication of papers given at the highly successful one-day conference held at Chester in November 1995 summarises the latest research on the subject from the Iron Age to the twentieth century.*
 PT+98: *Diese Veröffentlichung von an der an Chester im November 1995 gehaltenen hoch erfolgreichen eintägigen Tagung gegebenen Papieren faßt die letzte Forschung über das Subjekt vom Eisen Alter zur zwanzigsten Jahrhundert zusammen.*
 PwTr: *Diese Veröffentlichung von Dokumenten, die bei der sehr erfolgreichen eintägigen Konferenz gegeben wird, die in November 1995 bei Chester gehalten wird, faßt die spätesteste Forschung auf dem Thema vom Eisernen Alter zum zwanzigsten Jahrhundert zusammen.*

6. Problems in Grammatical and Lexical Coverage

Sometimes we encounter mistranslations that can only be explained by a lack of grammatical coverage. (22) contains a conjunction of an auxiliary-adjective-particle and a verb plus infinitive predicate. The sentence becomes complex because the object of these two predicates is fronted. While PT+98 produces a decent translation, T1 fails totally, presumably because its grammar does not cover a coordination of predicates with double fronting.

(22) *The Environment is something we should all be aware of and be prepared to protect and safeguard for the sake of our children and future generations.*

PT+98: *Die Umgebung ist etwas, das wir ganz wissen und bereit werden sollten, zu schützen und zu schützen, wegen unserer Kinder und zukünftiger Generationen.*

T1: *Die Umgebung ist etwas wir sollte alles Be bewusst von und vorbereiten um zu schützen und sichern wegen unserer Kinder und der künftigen Generationen.*

PwTr: *Die Umgebung ist etwas wir, wenn alle von bewußt sein sollten und bereit sein sollten, zu schützen und wegen unserer Kinder und künftiger Generationen zu schützen.*

As a general tendency we find that systems performing a detailed syntax analysis (such as T1) fare better if they succeed in finding a complete parse. But they fare worse if they fail to find a parse. Systems performing only shallow analysis show opposite results. In (23), PwTr takes a reduced complement clause to be a relative clause although there is no indication at all for this:

(23) *I am sure you won't be disappointed.*

Pt+98/T1: *Ich bin sicher, daß Sie nicht enttäuscht sein werden.*

PwTr: *Ich bin sicher der Sie wird nicht enttäuscht werden.*

An additional factor is lexical coverage. In (24), PwTr misinterprets the gerund as a noun; as soon as lexical information is added that the verb *include* may subcategorize for a gerund, the result becomes slightly better:

(24) *These include drawing up policies for the preservation and investigation of archaeological remains.*

PT+98: *Dies umfassen, Politiken für die Konservierung und Untersuchung von archäologischen Überresten hinauf zu zeichnen.*

T1: *Diese umfassen, Policen für die Erhaltung und Untersuchung der archäologischen Überreste anzufertigen.*

PwTr (with *include* subcategorizing for direct objects only): *Diese schließen Zeichnung auf Politik für die Bewahrung und die Untersuchung archäologischer Überreste ein.*

PwTr (with *include* subcategorizing for gerunds): *Diese schließen das Vorfahren bei Politik für die Bewahrung und die Untersuchung archäologischer Überreste ein.*

7. Summary

While some improvements can be made by adding stylistic constraints and adding rules to the grammar (like special coordinations and absolute participle constructions) many of the translation errors reported in this paper are due to missing semantic information and are therefore unavoidable with the current state of the art.

Some improvements could be made by using statistical preferences. We showed for the *for*-NP-infinitive constructions that corpus evidence can lead to clear preferences for the most likely reading.

It is striking that the three MT systems under investigation often show complementary results. It would therefore be worthwhile investigating whether one could use them in parallel and have the user decide which translation she would like to use. It is also conceivable that a statistic module ranks the translations according to a measure that favors the most likely sequence of words (similar to part of speech tagging).

Bibliography

Gdaniec, Claudia 1998: Lexical Choice and Syntactic Generation in a Transfer System: Transformations in the New LMT English-German System. In: "Machine Translation and the Information Soup", eds. David Farwell, Laurie Gerber, Eduard Hovy, 1998.

Mehl, Stephan 1996: Systematic Alternatives in Lexicalization: The Case of Gerund Translation. *Machine Translation* 11: 185-216.

Mehl, Stephan; Heidemann, Britta; Volk, Martin 1998: Zur Problematik der maschinellen Übersetzung von Nebensätzen zwischen den Sprachen Englisch und Deutsch. In: Rita Nübel; Uta Seewald-Heeg (eds.): Evaluation of the Linguistic Performance of Machine Translation Systems. Proceedings of the Workshop at the KONVENS-98. Bonn: 1998. 75-92.

Mehl, Stephan; Langer, Hagen; Volk, Martin 1998: Statistische Verfahren zur Zuordnung von Präpositionalphrasen. In: Proceedings of KONVENS-98. Bonn: 1998. 97-110.