

CONSTITUENT SHIFTS IN THE LOGOS ENGLISH-GERMAN SYSTEM

Claudia Gdaniec and Patricia Schmid
Logos Corporation
{cgdaniec, pschmid}@logos-usa.com

ABSTRACT

In this paper we will be concerned with certain patterns where semantically equivalent English and German sentences differ in terms of constituents. We will present these patterns, select three for more detailed analysis, and then show how transformation from the English source to the German target is implemented in the Logos MT system.

1 INTRODUCTION

In his *Linguistic Theory of Translation* (1965), Catford describes **category shifts** as "departures from formal correspondence in translation." (76) In this paper we will concentrate on one type of category shift: when the SL and TL equivalents are located at different syntactic levels, we say that a **constituent shift** has occurred during translation.

2 SOME CONTRASTS BETWEEN ENGLISH AND GERMAN

Constituent shift is a widespread practice of translators, for whom the strategy of structure-preserving translation is "a last resort" (Somers et al. 1988:5). We have categorized examples of English-German translation equivalents in order to illustrate various forms of constituent shift. The translations were provided by native German speakers and are not the only ones possible. All of them can be produced by Logos' English-German development system.

2.1 Finite Verb → Adverb

- (1) She **likes** watching horror movies.—> Sie sieht **gerne** Horrorfilme.
- (2) Prices **continue** to rise.—> Preise steigen weiter.
- (3) Cast iron **is apt** to break.—> Gußeisen bricht **leicht**.

In examples (1) through (3), the verb (or copula and predicate adjective) in the main clause in English is changed to an adverb and the two clauses are collapsed into a single clause in the German translation. What was the verbal complement in the English sentence is a conjugated main verb in the German sentence.

2.2 Noun Phrase → Verb Phrase

- (4) When a **drop in temperature occurs**, the experiment should be stopped.—> Wenn **die Temperatur sinkt**, sollte das Experiment abgebrochen werden.
- (5) **There was an increase in prices.** —> **Die Preise stiegen**

In these sentences, the subject of an English clause is rendered as a verb, the object of the prepositional phrase is made the subject, and the semantically schematic¹ verb is dropped altogether.

2.3 Prepositional Phrase → Subordinate Clause

- (6) (a) **On their return**, they talked about their adventures. → **Als sie zurückkehrten**, sprachen sie über ihre Abenteuer.
(b) **On their return**, they will talk about their adventures. → **Wenn sie zurückkehren**, sprechen sie über ihre Abenteuer.

In these sentences, a prepositional adverbial phrase in English is translated as an adverbial clause in German. The preposition is changed into a conjunction, the possessive adjective becomes the subject pronoun, and the object of the preposition is translated as the verb. Both conjunction and verb pick up the tense from the main clause: in the conjunction it is lexicalized, and in the verb it is marked morphologically.²

2.4 Adverbial Clause → Prepositional Phrase

- (7) **When merging a file**, you can eliminate blank lines. → **Beim Mischen einer Datei** können Sie leere Zeilen entfernen.
(8) **When you start your computer**, the Board initializes itself. → **Beim Anschalten Ihres Computers** initialisiert sich die Platine von selbst.
(9) The code must not be lost **when the battery is changed**. → Der Code darf **beim Batteriewechsel** nicht verlorengehen.
(10) **When there is a power or machine failure**, disk space is lost. → **Im Falle eines Strom- oder Maschinenausfalls** geht Plattenspeicher verloren.
(11) Do not try to operate the washing machine **when the door is open**. → Versuchen Sie nicht, die Waschmaschine **bei offener Tür** zu bedienen.

In sentences (7) through (11), the adverbial clause in English is translated as an adverbial prepositional phrase in German. This involves eliminating redundant or dummy subjects (such as *you* and *there*) and schematic verbs (such as *is*).

2.5 Elimination of Infinitive Clause

- (12) She was annoyed/pleased **to hear** that Clinton had won the election. → Sie ärgerte/freute sich (darüber), daß Clinton die Wahl gewonnen hatte.

In sentence (12) the infinitive is dropped in the German translation and the English passive verb is translated as a reflexive active verb in German.

3 ANALYSIS OF CONSTITUENT SHIFTS

¹ The term is from Langacker (1991:205).

² Somers et al. (1988) discuss the problem that MT is confronted with when dealing with nominal constructions that have to be denominalized in the target language.

3.1 Finite Verb → Adverb

Translations (1-3) have the pattern V1 + V2(non-finite) → V2(finite) + ADVERB, where V1 shifts from being a verb constituent in the main clause to being an adverbial constituent in a clause that does not embed another clause. This shift can be characterized as a **demotion**³ in the sense that the embedding clause is eliminated in the German translation and its constituents are moved down into the originally lower clause. V1 in the English main clause loses its status as matrix verb. Looked at from a different perspective, we have a **promotion**. V2--the verb in the embedded clause in English--is promoted into a non-embedded clause. This shift is obligatory in order to produce a grammatically acceptable translation of sentences (1-3).

3.2 Adverbial Clause → Prepositional Phrase

Translation (9) has the pattern WHEN+SUBJECT+PASSIVE VERB → BEI+MODIFIER+NOUN. This shift can be considered a **promotion**, since the adverbial clause is changed into a prepositional phrase and moved up into the main clause. In the German translation, it becomes an integrated constituent of the higher clause, positioned in the middle of the sentence. The arguments from the original English clause, however, are **demoted** in German: the subject of the passive verb becomes the modifier of the nominalized verb. The unmarked form would be a postnominal modifier in the genitive case or a prepositional phrase (with *von*), e.g., *when the battery is changed* → *bei Wechsel der Batterie*. An additional demotional shift occurs if the demoted subject is compounded with its head, e.g., → *bei Batteriewechsel*.

3.3 Elimination of Infinitive Clause

Translation (12) has the pattern HUMAN SUBJECT+PASSIVE VERB+INFINITIVE+COMPLEMENT → SUBJECT+REFLEXIVE VERB+COMPLEMENT. Since the infinitive is eliminated in the translation from English to German, the embedded complement is **promoted** to direct complement of the main verb. There is a whole semantic class of verbs (*upset, surprised, concerned, pleased, mortified, etc.*) that seem to render their infinitive complements semantically redundant if they are of a certain type. Accordingly, these complements need not be translated. The verb *hear* expresses the transmission of the information contained in its complement, while the main verb expresses its subject's psychological state in reaction to this information. The main verb presupposes knowledge of the that-clause; therefore the omission of the infinitive in the German translation does not reduce the relevant informational content.

4 CONSTITUENT SHIFTS IN LOGOS

4.1 The Logos System

³ We've borrowed the terms "demotion" and "promotion" from Bonnie Dorr (1990).

Before demonstrating how the Logos system produces the source-to-target constituent shifts, we will briefly describe the system itself. We will begin with a description of SAL (Semantico-syntactic⁴ Abstraction Language), which provides both the ontological and structural basis of the system. We will then give an overview of the system's architecture, and describe in more detail the structure and functioning of the Logos grammar rules.

SAL is a numeric classification scheme which subdivides all of the traditional parts of speech into semantico-syntactic categories. The categories are arranged in a four-level hierarchy so that any word is classified, with increasing specificity, according to **wordclass**, **superset**, **set**, and **subset**. For example *battery* is a member of wordclass 1, **noun**, superset 3, **concrete**, set 35, **agents**, and subset 750, **machines or systems**. There are approximately 1000 semantico-syntactic primitives in SAL. In addition, a word is classified according to **form**, i.e., morphological and syntactic information. SAL is not a set of properties or feature/value pairs. It is an actual language into which natural language is mapped. Thus the initial SAL representation of *battery* is **1 3 35 750 1** (noun, concrete, agent, machine, singular).

The Logos system consists of a pipeline of modules to gradually "disambiguate and decomplexify" (Scott 1989) a source sentence and construct corresponding sentences in various targets. Each module in the pipeline performs a specialized task. The first and last modules handle formatting. Linguistic processing begins with the dictionary lookup module. During lookup, each word of the input text is replaced by one or more strings of numbers which symbolize its SAL classifications. Each SAL string is called an **s-work**. Lookup is followed by two RES ("Resolution") modules, which disambiguate homographs. The output of RES is exactly one s-work per word. RES also builds a representation of the sentence's clausal constituents. Then, four TRAN modules work together to produce a bottom-up parse. When a syntactic constituent is identified, its component s-works are "concatenated" under the s-work of the constituent's head. Thus the lower-level TRANs are specialized for lower-level constituents while the higher-level TRANs are specialized for higher-level "concatenated" constituents. Many TRAN rules also call target language-specific components which perform transfer operations. Any number of targets may operate off the source analysis. Roughly, the functions of the TRAN modules are as follows: TRAN1 identifies and analyzes noun phrases. TRAN2 identifies right-attached modifiers of noun phrases, including prepositional attachments, relative clauses, and adjective and participial clauses. TRAN2 also identifies verb-particle constructions. TRAN3 identifies coordinated noun phrases, analyzes verb-argument structure, and determines final prepositional attachment. TRAN4 looks at the sentence as a whole, making final identification of clauses, subjects, objects, and inter-clausal relations. All four TRAN modules obtain parsing and transfer information from SEMTAB, the "semantic table", which contains collocational information about specific words, differentiates specific senses of words, and specifies translations overriding those from the dictionary.

⁴ defined as: semantic properties that correlate with syntactic behavior

Each of the rulebases in the Logos grammar contains several thousand rules, each consisting of an SAL pattern and a corresponding procedural routine. The organization of the rulebases is based on SAL. A sort algorithm uses the numeric symbols of the SAL patterns to index and order the rules according to their specificity and length. Any new rule that is added fits automatically into the appropriate place in the rulebase organization. SAL also controls the rule-matching process in all modules. In accessing a rulebase, the elements of the input SAL string are the search arguments, much as a word is the search argument in accessing a dictionary. Because of the way the rulebase is sorted, the most specific, longest rule appropriate to the input string unifies with the input string, and the procedure it contains is executed. In each RES and TRAN module, the input string is processed from left to right: its SAL elements are changed and eventually discarded, and an output SAL string is created. When the input string is empty, processing is complete. The output string is sent on to the next module.

4.2 Constituent Shifts in Logos

The three types of constituent shift described in Section 3 are handled in different places in the Logos system.

4.2.1 Finite Verb—> Adverb

(2) *Prices continue to rise.* —> *Preise steigen weiter.*

This shift is achieved through interaction of the TRAN and SEMTAB modules. The SAL string coming in to TRAN3 consists of five s-works:

```
20 900 2   1 462   2 748 1   2 579 28   20 10 1
BOS5     PRICES  CONTINUE  RISE      EOS
```

Each s-work consists of three numbers: wordclass, type, and form. The type field contains the superset/set/subset information: only the most specific of these is shown here. Processing of the sentence proceeds from left to right. At the third element, *continue*, there is a match on the following rule:

```
COMMENT: V   .S.   TO_V
SOURCE PATTERN: 02-01-1 52-02-1 02-0258
                means: (verb) (look ahead/conditions) (infinitive verb)
CONDITIONS:
                element 2: skip adverbs, time nouns, and prepositional phrases not
                attached to the verb
SOURCE ACTION:
                1. Send all s-works, plus the saved subject of element 1 to SEMTAB.
```

⁵BOS stands for Beginning of Sentence; EOS stands for End of Sentence; PUNC stands for Punctuation or Conjunction; .S. stands for STRETCH, which specifies elements that can be skipped in matching.

1. Check all rules indexed for element 1.
2. Change the wordclass of element 1 to 17 to show that it has been to SEMTAB.
3. Continue processing at first element.

TRANSFER (GERMAN): none

The rule above is a very general rule, matched by any verb followed by an infinitive. Step 1 under SOURCE ACTION sends all the matching input s-works, augmented by the saved subject s-work, to SEMTAB. The s-work string in this case matches a SEMTAB rule written specifically for *continue* followed by an infinitive.⁶

INDEX: CONTINUE

COMMENT: CONTINUE TO_INFINITIVE = WEITER(ADV) V

SOURCE PATTERN: 0274800 0400058

means: (continue) (infinitive verb)

CONDITIONS: none

SOURCE ACTION:

1. Change form field of element 2 to indicate it is attached to a verb.

TRANSFER:

1. Mark element 1 for demotion from verb to adverb.
2. Store *weiter* as the adverb transfer.

The source part of the SEMTAB rule marks the infinitive as linked to the first verb, while the transfer part stores information about the demotion of *continue* in the German translation. This target information is read in TRAN4, where it causes several things to happen:

1. The target person, number, and tense, as well as the source form field are copied from *continue* to *rise*, effectively promoting it to main clause verb.
2. The adverb *weiter* (the demoted verb) is introduced into the output array.

The Generation module reads these signals and generates *Preise steigen weiter*.

4.2.2 Adverbial Clause → Prepositional Phrase

(9) *The code must not be lost when the battery is changed.* →
Der Code darf beim Batteriewechsel nicht verlorengehen.

This shift occurs in TRAN4, the clause-level module. The SAL string coming in to TRAN4 consists of nine s-works:

⁶In matching, a TRAN rule must correspond sequentially to the s-works of the input string, but s-works may match SEMTAB rules in any order. It is not necessary for all s-works sent to SEMTAB to match in order for a rule to match.

20 900 2 1 986 17 12 897 6 2 588 60 20 966 19 1 750 17
 BOS CODE MUST LOST WHEN BATTERY

12 886 3 2 203 80 20 10 1
 IS CHANGED EOS

At the fifth element there is a match on the following rule:

COMMENT: WHEN N BE V'D PUNC
 SOURCE PATTERN: 2096619 01-0191 12886-1 02-0180 20-02-1
 means: (*when*) (all possible subject nouns) (*be* as verb) (past participle)
 (conjunction or punctuation)

CONDITIONS:
 element 4 not marked for passive-active transformation
 element 5 not a coordinating conjunction

SOURCE ACTION:
 1. Label noun as part of higher-level clause.
 2. Continue processing at second element.

TRANSFER (GERMAN):
 1. Save verb information in global storage area.
 2. Translate element 1 as *bei*, with dative case governance.
 3. Load element 1 into output array.
 4. Mark element 4 for nominalization.
 5. Call routine to decide if element 2 should become pre- or postmodifier
 of nominalized verb.
 6. If punctuation is comma, null it.
 7. Change adverb modifiers of verb to adjectives.

The information is used by subsequent TRAN4 rules as well as by the Generation module to produce the German prepositional phrase in place of the English subordinate clause. It is relatively simple for this module to produce the nominalized German form, since in the dictionary a verb is linked with its nominalization.

4.2.3 Elimination of Infinitive Clause

(12) *She was annoyed/pleased to hear that Clinton had won the election.* →
Sie ärgerte/freute sich, daß Clinton die Wahl gewonnen hatte.

This shift is also achieved in TRAN4. There is a match on the following rule, which is based on the SAL code of the class of verbs of psychological state that can embed another clause:

COMMENT: V(set 77-PASSIVE) .S. TO_V COMPLEMENTIZER
 SOURCE PATTERN: 0307765 52-01-1 02-0138 20-02-1
 means: (passive verb in main clause of set 77=psychological state verbs
 that can embed complements)(look ahead)(infinitive verb linked with
 main verb)(complementizers *how, that, which, what*)

CONDITIONS: none

SOURCE ACTION:

1. Mark infinitive for promotion into main clause, i.e., collapse the two clauses.
2. Continue processing at first element.

TRANSFER (GERMAN):

If infinitive denotes perception/incoming information, such as *hear*, *receive*, *learn*, *discover*, *find out*, mark infinitive for no transfer.

The information is used by subsequent TRAN4 rules and the Generation module, which does not generate a translation for the infinitive.

5 CONCLUSION

Constituent shifts occur at varying levels of generality. In the Logos rulebases, contrastive patterns ranging from very specific to very general can be reproduced as rules at different syntactic levels.

Some constituent shifts are necessary to produce a grammatically correct translation while others are optional and stylistically motivated. For example, the compression of information which occurs when a clause is reduced to a phrase is appropriate for some texts and not for others. The development version of the Logos English-German system offers a rather systematic way of making stylistic choices between a more condensed, nominalized style and one that is more verbal and expanded. We are developing a mechanism by which the user of this system will be able to choose the style that is better suited to the texts being translated.

REFERENCES

- Catford, J.C. (1965), *A Linguistic Theory of Translation*. Oxford: Oxford University Press
- Dorr, Bonnie J. (1990), *Lexical Conceptual Structure and Machine Translation*. Ph.D. Dissertation at M.I.T.
- Langacker, Ronald W. (1991), *Foundations of Cognitive Grammar*, Vol. II. Stanford: Stanford University Press
- Scott, Bernard E. (1989), "The Logos System," Paper Delivered at the MT SUMMIT II Conference, Munich
- Somers, Harold, Hideki Hiraoka, Seiji Miike, Shinya Amano (1988), "The Treatment of Complex English Nominalizations in Machine Translation," *Computers and Translation*, Vol. 3, Number 1:3-21