

*Helle V. Dam*

## **Text condensation in consecutive interpreting - summary of a Ph.D. dissertation**

On April 10th 1996 I defended a Ph.D. dissertation on the above topic at the Aarhus School of Business, Faculty of Modern Languages (Dam 1995). The following is a presentation of the main aspects of the dissertation.

As opposed to the bulk of previous research on interpreting, which has mainly been dedicated to general theory formation, the main purpose of the present dissertation is to provide an empirical description of a specific interpreting phenomenon, namely *text condensation*, ie. target text manifested length reductions of a source text, as it occurs in the context of consecutive conference interpreting. An additional aim of the dissertation is to propose a framework for an empirical investigation of this phenomenon.

The approach to the investigation is interdisciplinary insofar as the analyses are conducted by applying concepts and methods from disciplines other than interpreting. As the focus of the investigation is on the textual manifestation of the phenomenon under study, rather than on eg. process-related aspects, concepts and methods applied in the analyses mainly stem from the field of text linguistics. A particular source of inspiration was found in text linguistic studies dedicated to summarization and related phenomena, especially in Teun A. van Dijk's theory of Macrostructures in discourse (eg. van Dijk 1980), which served as a theoretical frame of reference in some of the analyses, and Kjersti Fløttum's doctoral dissertation on French school summaries (Fløttum 1989), which was a main source of methodological inspiration.

### **1. Data**

The investigation is based on a text corpus consisting of recordings of, on the one hand, a Spanish speech on drug policy delivered in an experimental setting on the basis of a manuscript in 7.5 minutes (the source text) and, on the other hand, five consecutive interpretations of this speech into Danish performed by five professional conference interpreters with Danish as their A-language (the target texts).

For the purpose of the analyses, the recordings of both source and target texts were transcribed, basically in accordance with orthographic standards. Furthermore, in order to neutralize the language differences between the two sets of texts for the purpose of making them comparable in terms of length, the Spanish source text was translated into Danish as literally as possible, but without violating the target language rules. The translated source text serves as

a *tertium comparationis* between source and target texts in the analyses. Finally, the translated source text and the target texts were divided into smaller units, here referred to as *segments*. With few exceptions, a segment consists of a series of words grouped around a verb, which may be finite or infinite. The further delimitation of the segment is generally based on valency grammatical principles as described in detail in the dissertation (pp. 114-123). The segments constitute the units of analysis in the subsequent, comparative analyses of source and target texts.

A first, quantitative comparison of the two sets of texts reveals that all the target texts are shorter than their source text, the average length reduction constituting 25,9% measured in characters on the basis of the transcripts. All the target texts are, in other words, condensed versions of the source text.

## 2. Analyses

The model of analysis designed for the investigation consists of two major modules: (1) a so-called *formal* analysis, and (2) a so-called *linguistic* analysis. The formal analysis aims at describing how the condensed target texts of the corpus are formed on the basis of their source text with the specific purpose of identifying the underlying condensation operations or forms. The linguistic analysis aims at describing the linguistic elements to which the condensation operations or forms identified in the formal analysis are typically applied. An additional aim of the second module of analysis is to explain - by means of the linguistic description - why those particular elements have been subjected to condensation.

### 2.1. Formal analysis

In the first stage of the formal analysis, a series of categories are established in order to describe how the target text segments are constructed on the basis of the source text segments. The categories are established on the basis of the corpus, but are inspired by the above mentioned studies by van Dijk (1980) and Fløttum (1989).

The following formal categories of target text segments are found:

(1) **Parallel segments**, ie. target text segments that are practically identical - semantically as well as formally - to those source text segments on the basis of which they are constructed. Thus, parallel segments are constructed by transferring each of the elements contained in the corresponding source text segments as directly as possible to the target text.

(2) **Selective segments**, ie. target text segments that contain or consist of selected elements from one or several source text segments. Thus, selective segments are constructed by transferring some of the elements contained in one or

several source text segments to the target text, whereas the rest of the elements of the source text segment(s) in question are deleted.

(3) **Substituting segments**, ie. target text segments that contain or consist of elements that are constructed by replacing some or all the elements of one or several source text segments by different elements. The substituting segments are subcategorized according to the semantic or pragmatic relation that obtains between the substituting target text elements and the substituted source text elements: (3.1) *formally substituting segments*, (3.2) *generalizing segments*, and (3.3) *integrating segments*. The semantic relation between (3.1) and the substituted source text segments is one of equivalence. In the case of (3.2) and (3.3), the relation is described in terms of degree of globality, and segments ascribed to category (3.2) - generalizing segments - are characterized by being semantically more global than the generalized segments, whereas segments ascribed to category (3.3) - integrating segments - are conceptually (or pragmatically) more global than the integrated source text segments. A common denominator of the three subcategories of substituting segments is that they are shorter than the source text segments on the basis of which they are constructed.

(4) **Additional segments**, ie. target text segments that contain or consist of elements that have no counterpart in the source text.

Apart from the above four categories, the corpus contains a small number of target text segments (approximately 3%) which cannot be categorized as any of these main types and which are judged to be irrelevant in the present study, since they are generally longer than the source text segments on the basis of which they appear to be constructed. These target text segments are ascribed to a category referred to as *other segments*.

It should be noted that not all the target text segments of the corpus are 'pure' in the sense that they are ascribable to one of the above categories only. Rather, there are also some complex type segments, ie. segments that can be - and are - ascribed to two or more categories at the same time.

In the second stage of the formal analysis, a synthesis of the formal source-to-target text transfer is performed, partly by juxtaposing each of the source text segments with its five target text representations tagged according to the above categories, partly by quantifying the findings. The juxtaposition of source and target text segments allows for a few additional observations concerning the representation of the source text in the target texts, the most important one in this context being the registration of 0-representation, ie. deletion of entire source text segments. The quantification of the findings of the formal analysis is summarized in the table below. The left column lists the percentages of

source text (ST) segments which are represented in the target texts (TTs) in the manner indicated in the right-hand column:

<b>ST segments<sup>1</sup></b>	<b>Representation in the TTs</b>
14%	Parallel segments
53%	Selective segments
18%	Formally substituting segments
3%	Generalizing segments
1%	Integrating segments
3%	Other segments
17%	0-representation (full deletion) <sup>2</sup>

The above results call for a few comments. First and foremost, the frequency of source text representation in the target texts by the category referred to as 'selective' is striking (53%). As will be recalled, selective segments are basically constructed by transferring selected elements from the corresponding source text segments directly to the target text, while deleting the rest. In other words, the procedure of source-to-target text transfer underlying the construction of selective segments may be described as one of 'copy-delete-copy-etc.'. It is worth noting that precisely the same principle of transfer is reflected in the parallel segments (14%), which are constructed by transferring an entire source text segment directly to the target text, as well as in the deletion of entire source text segments (17%). This means that the so-called 'copy-delete-copy procedure', which apparently operates at the surface of the source text, is clearly the dominant principle of transfer. On the other hand, the categories reflecting a source-to-target text transfer that involves a synthesis, ie. the generation of completely different linguistic elements, are considerably less frequent (formally substituting segments: 18%; generalizing segments: 3%; integrating segments: 1%). Incidentally, these observations seem to be in contradiction with one of the most influential interpreting theories - namely a theory originally proposed by the French interpreting scholar Danica Seleskovitch (eg. Seleskovitch 1975) and sometimes referred to as 'la théorie du sens'. The main tenet of this theory is that interpreting does *not* proceed directly from source to target text, but through a language-independent ('deverbalized'), mental representation of the meaning of the source text (the so-called 'sens' or

<sup>1</sup> The source text consists of 124 segments, which means that the total number of possible representations of source text segments in the five target texts amounts to 620.

<sup>2</sup> Apart from the indicated source text representations, there is quite an impressive number of additional target text segments (ie. additions) in the corpus, namely a total of 241. However, by far the largest number of these additions are segment-internal and constitute single words or particles (eg. 'and', 'well', 'so'), rather than entire passages or segments.

the 'meaning meant' by the source text speaker). The existence of the stage of deverbalization is supposed to result in the production of a target text whose individual linguistic elements differ substantially from those of its source text - evidently a claim that is not confirmed by the findings of the formal analysis of the present study.

The main contribution of the formal analysis to the present study, however, is to have identified the relevant condensation operations or forms as *deletion* and *substitution*. As we have seen, deletion is reflected partly in the segments categorized as selective, partly in the segments represented by full deletion in the target texts. Substitution is reflected in the three subcategories of substituting segments, ie. in the formally substituting, the generalizing and the integrating segments. Obviously, source text segments represented in the target texts by one of these condensation forms are candidates to be carried on to the subsequent module of analysis, ie. to the linguistic analysis. However, given the low frequency of both generalizing and integrating segments, source text segments ascribed to these categories are not passed on to the linguistic analysis, which is therefore only based on those source text segments that are represented in the target texts by selective segments, full deletion or formally substituting segments.

## 2.2. Linguistic analysis

As already explained, the purpose of the linguistic analysis is to describe the linguistic elements to which the principal condensation operations or forms - now defined as deletion and formal substitution - are typically applied in the corpus of the study. Consequently, the source text segments represented by deletion or substitution in the target texts by *most* of the subjects, ie. by a minimum of three of the five interpreters, were selected for this analysis.

In the analysis of **deletion**, fully deleted segments and deleted elements from the selective segments are described at two different levels: (1) a syntactic level, and (2) a semantic-pragmatic level. This dual analysis is motivated by two opposing hypotheses formulated in previous studies dealing with the phenomenon of deletion in interpreting. The first hypothesis is proposed by Andrzej Kopczynski (Kopczynski 1980), who suggests that deletion (or omission) is syntactically determined, a suggestion supported by the corpus-based observation that source text elements in so-called modifying positions (ie. syntactically subordinate elements) tend to be deleted, whereas constitutive parts of sentences or phrases tend to be transmitted to the target text. The 'syntactic hypothesis' would imply that deletion is basically a mechanical procedure governed by more or less random factors. The second hypothesis is present as an underlying assumption both in a study by Bistra Alexieva (Alexieva 1983) and in a study previously conducted by myself (Dam 1993), in which deleted

elements are described in terms of their semantic or pragmatic position in the source text. Thus, in these studies it is assumed that deletion is semantically and/or pragmatically determined, ie. determined by the meaning or information value of the deleted elements. The 'semantic-pragmatic hypothesis' would imply a less random selection on the part of the interpreters than would the syntactic one. In view of these two opposing hypotheses, the main purpose of the analysis of deletion of the present study is to establish which one is empirically more correct, ie. to find out whether the phenomenon of deletion is dominantly governed by syntactic (random) or by semantic-pragmatic (non-random) factors. Accordingly, the analysis is conducted at both these levels.

The *syntactic* analysis of deletion confirms the tendency registered by Kopczynski to delete modifying elements and, conversely, to transmit constitutive ones, insofar as approximately 62% of the deleted elements appear in modifying positions in the source text. Still, though deletion of syntactic modifiers is dominant, the syntactic hypothesis is not completely verified, as approximately 38% of the deleted elements are constitutive parts of sentences or phrases.

In the *semantic-pragmatic* analysis of deletion, a series of categories are established in order to describe the semantic-pragmatic position of the deleted source text elements in relation to the surrounding source text elements which have been transmitted to the target texts. As in the formal analysis, the categories are established on the basis of the corpus of the investigation, but, again, some of them are inspired by similar studies, especially by Fløttum's study on summaries (Fløttum 1989).

The following ten semantic-pragmatic categories of deleted source text elements are found:

- (1) **Repetitions**, ie. expressions containing information which has already been given in an identical or a different wording.
- (2) **Reformulations**, ie. expressions by which a given piece of information is rephrased so as to make its implications explicit. Deletion may either affect the reformulated information or the reformulation itself.
- (3) **Explicitations**, ie. expressions containing information which is inferable from the linguistic or encyclopedic context by means of a so-called default-interpretation on part of the listener.
- (4) **Expressions of contrast to the main theme**, ie. expressions introducing a theme which is different from the main theme of the text or the text part in question. This different theme is compared - or contrasted - with the main theme, in relation to which it is subordinate (eg. '*as opposed to a legal market, an illegal market is characterized by ..*').

(5) **Metatextual expressions**, ie. expressions that do not contain information about the world outside the text, but whose only function is to comment on the structure, contents etc. of the text or part of the text itself (eg. 'as already mentioned').

(6) **Circumstantial expressions**, ie. expressions indicating the place, time, manner, cause, consequence etc. of principal actions or states described in expressions which *are* transmitted to the target texts. As opposed to the transmitted expressions, circumstantial expressions are characterized by containing only contextually unbound information.

(7) **Categorizing expressions**, ie. expressions which are semantically very general and which normally indicate the category to which a given - transmitted - element belongs (eg. 'the *disease* of AIDS').

(8) **Modal expressions**, which are a broad category of semantic modifiers, including expressions of epistemic modality and/or certain evidentials. Deletion of modal expressions generally has the effect that the affected target text utterances become more categorical than their source text counterparts.

(9) **Evaluative expressions**, ie. expressions by which the speaker indicates his attitude towards the information he is giving. Deletion of evaluative expressions is restricted to non-asserted, ie. non-focalized, speaker evaluations, whereas asserted evaluations are typically reproduced in the target texts. As opposed to the non-asserted evaluative expressions, the surrounding - asserted - information is generally transmitted to the target texts.

(10) **Expressions contained in enumerations**, ie. expressions listed as part of a series of two or more items. In enumerations in which the individual expressions together designate a progression or a paradigm, the deleted expressions normally designate objects or concepts at a lower step or with a lesser range, whereas the transmitted expressions typically designate objects or concepts at a higher step or with a wider range.

Having established the above semantic-pragmatic categories of deleted elements, each of them is evaluated in terms of the degree of importance of its elements in relation to the non-deleted elements and the text as a whole. It is argued that, except for the category of modal expressions (8), all the categories contain information that is either semantically and/or pragmatically *redundant* in its context, insofar as it is either given elsewhere or can be inferred by the listener (this is the case of categories (1), (2) and (3), and normally also (5) and (7)), or *less important* than the surrounding, non-deleted information (categories (4) and (6), sometimes (5) and (7), and normally also (9) and (10)).

On the basis of these observations, it is concluded that the phenomenon of deletion in consecutive interpreting is more likely to be governed by semantic-

pragmatic than by syntactic factors. Thus, deletion is likely to involve a non-random - be it conscious or unconscious - process of selection guided by the information value of the individual parts of the source text. On the other hand, the dominance of deleted elements in so-called modifying (ie. syntactically subordinate) positions in the corpus cannot be neglected. Therefore, it is hypothesized that while the phenomenon of deletion is probably mainly governed by the semantic-pragmatic position of source text elements, this position is frequently (but not always) syntactically marked, in the sense that elements which are less important in terms of information value tend to appear as syntactic modifiers, whereas more important elements often appear as constitutive parts of sentences or phrases.

The linguistic analysis is concluded by a description of **formal substitution**. Given the fact that substitution occurs far less frequently than deletion, this part of the analysis is correspondingly less comprehensive than the analysis of deletion.

The source text elements that are represented by formal substitution in most of the target texts may be ascribed to one of the following three categories:

(1) **Anaphors**, ie. text items whose interpretation proceeds via a previous text item (the antecedent). As is well-known, anaphors may have different linguistic representations with varying degrees of explicitness - and length, one may add. Thus, they may for instance occur as definite noun phrases or as pronouns. In the corpus of the study, a considerable number of anaphors represented in the source text as relatively long, definite noun phrases are substituted by pronouns or by shorter - and less explicit - noun phrases.

(2) **Structural markers**, which constitute a heterogeneous category of elements (other than anaphors) by means of which the structure of the text is explicated. The interpreters tend to substitute such explicit structural markers by considerably less explicit - and shorter - ones (eg. 'one implication is ..., a second implication is ..., a third implication is ...' is substituted by 'firstly ..., furthermore ..., and ...').

(3) **Markers of written discourse**, ie. certain expressions (other than explicit anaphors and explicit structural markers) which are characteristic of the written - or literate - register. Markers of written discourse are rather frequent in the source text of the corpus, because it is manuscript-based. Not surprisingly, many of these markers are substituted by expressions more typical of the spoken - or oral - register in the target texts, and some of these target text expressions are, in fact, shorter than the corresponding source text expressions. This is for instance the case of most source text nominalizations, which are typically verbalized in target texts (eg. 'drugs are subjected to a prohibition ..' is substituted by 'drugs are forbidden ..').



To a large extent, substitution of linguistic phenomena like the ones described above depends on their paradigmatic nature, which means that they may occur in different linguistic representations. Whereas the source text representations are relatively long, the interpreters seem to prefer the shortest possible ones - which is not surprising in view of the pressure inherent in their task.

### 3. Concluding remarks

The results yielded by the analyses of the present dissertation - however interesting some of them may be - evidently need to be treated with caution. First and foremost, the amount of data analyzed is far too restricted for the results to be anywhere near conclusive. Furthermore, if valid at all, the findings may only be true for the specific set of parameters (ie. language pair, type and direction of interpreting, setting, text type and genre, etc.) selected for this particular study. Therefore, the results obviously need to be further tested and documented before they can be regarded as valid - or even universally valid. Still, such testing may be conducted within the theoretical and methodological framework proposed in the context of the present dissertation, which may therefore be a valid contribution to interpreting research in this respect.

### References

- Alexieva, Bistra (1983): Compression as a means of realisation of the communicative act in simultaneous interpreting. In: *Fremdsprachen*, 4, pp. 232-238.
- Dam, Helle Vrønning (1993): Text condensing in consecutive interpreting. In: Gambier, Y. & Tømmola, J. (eds.): *Translation and Knowledge*. University of Turku.
- Dam, Helle Vrønning (1995): *Tekstkondensering i foredragstolkning: Formel og sproglig analyse på grundlag af spansk-danske tolkninger*. Det Erhvervsproglige Fakultet, Handelshøjskolen i Århus. Ph.d. afhandling. (*Text condensation in consecutive interpreting: Formal and linguistic analyses based on Spanish-Danish interpretations*. The Faculty of Modern Languages, Aarhus School of Business, Denmark. Ph.D. dissertation.)
- van Dijk, Teun A. (1980): *Macrostructures. An interdisciplinary study of global structures in discourse, interaction, and cognition*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Fløttum, Kjersti (1989): *La nature du résumé scolaire. Analyse formelle et informative*. University of Trondheim.
- Kopczynski, Andrzej (1980): *Conference interpreting. Some linguistic and communicative problems*. Adam Mickiewicz University of Poznan.
- Seleskovitch, Danica (1975): *Langage, langues et mémoire. Études de la prise de notes en interprétation consécutive*. Paris: Minard.



