

## Textual Cohesion and Translation

Balkrishan Kachroo

---

Volume 29, Number 2, juin 1984

URI: <https://id.erudit.org/iderudit/003279ar>

DOI: <https://doi.org/10.7202/003279ar>

[See table of contents](#)

---

Publisher(s)

Les Presses de l'Université de Montréal

ISSN

0026-0452 (print)

1492-1421 (digital)

[Explore this journal](#)

---

Cite this article

Kachroo, B. (1984). Textual Cohesion and Translation. *Meta*, 29(2), 128–134.  
<https://doi.org/10.7202/003279ar>

# TEXTUAL COHESION AND TRANSLATION\*

BALKRISHAN KACHROO

An authentic translation involves more than translating sentences, however accurate grammatically. One has to bear in mind the interaction between the sentences and the semantic and stylistic implications of this interaction. Thus an authentic translation has to consider factors which are in some way beyond the sentence. We propose that one of these factors is 'textual cohesion'.

## TEXTUAL COHESION

"Cohesion is defined as the property that distinguishes a sequence of sentences that form a discourse from a random sequence of sentences" (Singh 1979). Singh suggests that in studying cohesion a distinction should be made between 'linguistic cohesion' and 'pragmatic cohesion' which parallels the notions of 'linguistic presupposition' and 'pragmatic presupposition' (Keenan 1970). Consider the following discourses :

- (A) 1. John likes Helen.  
2. She, however, hates him.
- (B) 1. Do you have coffee to go ?  
2. Cream and sugar ?

In (A) the link between 1 and 2 is provided by pronominalization which is a purely linguistic link ; in (B), on the other hand, the semantic link is contingent on knowledge of the real world. In this study we are concerned with the links of the former type (A) i.e. with linguistic cohesion. We will ignore pragmatic cohesion, exemplified by type (B) links. Linguistic presupposition and pragmatic presupposition differ in a similar manner. Whereas in linguistic presupposition the information can be extracted from the linguistic context, the information is deduced from outside the linguistic context in the case of pragmatic presupposition. Example :

- (C) John gave his brother two books.

*Linguistic Presupposition* : John has a brother.

*A possible Pragmatic Presupposition* : John's brother likes books.

The 'pragmatic presupposition' *John's brother likes books* can be deduced from a possible real world belief that you give people what they like to receive, which even if true is not contained in (C). In other words, while linguistic presupposition depends entirely on the linguistic context, the domain of pragmatic presupposition is largely outside the linguistic context.

Hallyday and Hasan have identified five kinds of cohesion in English : reference, conjunction, substitution, ellipsis and lexical cohesion. The following pairs of sentences serve as examples of the above-mentioned cohesion devices :

---

\* This work was in part supported by an SSHRC grant to R. Kittredge and R. Singh (grant #410-79-0070). I am thankful to R. Kittredge, Jim Lees, R. Singh and I. Spilka for helpful comments on an earlier version of this paper.

**Reference** : Reference is defined by Halliday as a semantic relation that "ensures the continuity of meaning in a text". It involves items that cannot be meaningful on their own, but refer to something else. Consider the following :

1. John is very athletic.
2. *He* loves to ski.

In 2 *he* is meaningful only because it refers to *John* of 1.

**Conjunction** : The link provided by words like *nevertheless*, *and*, *furthermore*, *however*, etc. Example :

1. He was very tired.
2. Nevertheless he kept on working.

**Substitution** : A grammatical relation where one linguistic item substitutes for another.

Example :

1. My *car* is running badly.
2. I should buy a new *one*.

In the example above, *one* of 2 substitutes for the word *car* of 1. Consider another example :

1. Helen *runs six miles every day*.
2. I should *do the same*.

In the example above the string *run(s) six miles every day* is replaced by *do the same*.

**Ellipsis** : Ellipsis involves the deletion of either the noun or the verb. In English, ellipsis occurs most frequently in question-answer pairs of the following type :

1. Which car do you like ?
2. This is the nicest.

Here the word *car* of 1 is elided in 2 (in other words replaced by a zero).

**Lexical** : Lexical cohesion involves the use of same or similar or even related words in successive sentences. Consider the following :

1. John *likes* apples.
2. Helen *likes* oranges.
3. Robert *hates* oranges and apples.

The link is provided by *likes* in the pair 1 — 2 and by the lexical contrast *likes* — *hates* in the pair 2 — 3.

The categories identified by Halliday and Hasan are, however, much too general. They group a number of cohesive relations under the same category. Consider the category lexical cohesion : it covers relation such as word repetition, synonymy, hyponymy, lexical contrast, etc. We think that textual cohesion can be studied better with a broader and more refined taxonomy, especially when one is dealing with more than one language. Consider two languages that behave in the same way when we consider the broad category 'lexical cohesion'. It is entirely possible, in fact quite likely, that these two languages differ considerably when different relation (cohesive) such as synonymy, hyponymy, lexical repetition are considered individually. The following is a concise version of the taxonomy that we have constructed for the present study. Each category in this taxonomy represents a cohesion device, a means for linking two adjacent sentences. A pair of sentences is referred to as an adjacency pair.

**Anaphora** : This major category describes referential relations binding one element to an element mentioned in a preceding sentence<sup>1</sup>. The following are two possible kinds of anaphoric relations :

1. Restricted to adjacency pairs. Undoubtedly, notions such as anaphora and cataphora have to be expanded when larger sequences of sentences constitute the domain of investigation. However, the operant principle which we discuss here should obtain.

(a) *Pronominalization* : Any instances where a noun phrase is replaced by a pronoun. Example :

1. *La dissertation est défendue publiquement par le récipiendaire.*
2. *Elle est transmise au jury au moins un mois avant la date fixée pour l'épreuve.*

(b) *Definitivization* : In these cases the second element is definite whereas the first element is indefinite. Example :

1. *En deuxième épreuve, l'étudiant choisit, outre des cours obligatoires, un ensemble de cours à option pour une durée de 105 heures au moins.*
2. *Parmi ces cours à option, l'étudiant peut choisir 60 heures au maximum...*

*Cataphora* : It is similar to anaphora in that it involves the same kind of referential relation, and yet different in that it operates in the opposite direction, i.e. the first element refers to an element which follows. Theoretically the same kinds of relations should be found under both cataphora and anaphora. However the usage of cataphora is much more restricted.

(a) *Cataphoric pronominalization* : Any instances where a pronoun makes reference to a noun phrase in the immediately following sentence. Example :

1. He ran towards the river.
2. John dived headlong only to appear at the other bank.

This type of referential relationship is not possible in English within the boundaries of the sentence. Consider the sentence *she thinks that Mary likes apples*. In this sentence *Mary* and *she* cannot be coreferential, whereas in the sentence pair 1 and 2 *he* and *John* are coreferential. Clearly, the properties governing discourse are different from those governing syntax.

*Derivational morphology* : This device covers the use of a lexical item and its derived form. Changes of tense, aspect, gender and case are not included. Thus, while the relationship expressed over an adjacency pair, between *realize* and *realization* falls under the category derivational morphology, the relationship between *ask* and *asked* does not (which is included under inflectional morphology). Example :

1. Most households in North America use four to five *electric* gadgets.
2. The average *electricity* bill ranges from fifteen to twenty dollars.

*Inflectional morphology* : This category accounts for verb forms i.e. changes in tense, aspect, etc. Singular/plural morphology is not included here. (See lexical repetition.) Example :

1. *En les coiffant Cendrillon songeait...*
2. *Une autre que Cendrillon les aurait coiffées de travers...*

*Lexical repetition* : This category covers lexical repetition over two sentences (i.e. when the same word is repeated over an adjacency pair). Singular/plural are not differentiated and instances like *dog-dogs* are counted as lexical repetition. Example :

1. This cat is not mine.
2. My cat is black.

The cohesion in the above pair is achieved by the repetition of the word *cat*.

*Synonymy* : This semantic relation describes terms which are similar in meaning. Example :

1. John's explanation of the facts was very *lucid*.
2. He made it abundantly *clear* that...

*Hyponymy* : This category covers the relationship expressed by a pair of the following type : *animal* — *cat*, where one word represents a class and the other is a member of that class. Example :

1. I like *flowers*.
2. And especially *roses*.

In the pair *rose* — *flower* the word *flower* represents the class and *rose* is a member of that class.

*Synechdoche* : This category refers to a semantic relationship where an item signifying a part is used to signify the whole. Example :

1. John, the farmer has only ten *men* working for him this year.
2. He could easily do with a few more *hands*.

In the pair above, *men* of 1 is replaced by *hands* in 2.

*Conjunction* : This category covers the link provided by words like *and*, *nevertheless*, *furthermore*, etc. Example :

1. John was tired.
2. Nevertheless he kept on working.

The word *nevertheless* connects 1 and 2. In addition cohesion is also ensured by anaphoric pronominalization.

#### TEXTUAL COHESION AND TRANSLATION

The purpose of this paper is to provide evidence for the hypothesis that the 'extra-sentential' factors that influence the quality of a translation are to a great degree in the domain of textual cohesion ; that the overall semantic accuracy of a translation is directly proportional to the similarity that the 'target-text' has with the 'source text' in terms of the distribution of cohesion devices ; but more importantly that an authentic translation of an  $L_1$  text (Let's say Hindi children's literature) into  $L_2$  (English) will mean that the 'target-text' will exhibit a strong resemblance to the sublanguage<sup>2</sup> (English children's literature) in terms of the distribution of cohesion devices.

The hypothesis may be stated more precisely and formally as follows : if  $X_1$  is the distribution of cohesion devices for the source text (we assume that  $X_1$  is representative of the sublanguage under consideration),  $X_2$  the distribution of cohesion devices in the target text and  $X_3$  the distribution of cohesion devices in the target sublanguage...then : an authentic translation will strive to match  $X_2$  to  $X_1$  but more importantly  $X_2$  to  $X_3$ , and that the preceding will be truer in the case of a more authentic translation than a less authentic one.

Thus, a good translation attempts to do two things simultaneously : make the target text as authentic as possible in the target sublanguage, and retain as much of the source sublanguage as can be accommodated, given the interlanguage differences.

#### METHODOLOGY

To test the validity of the hypothesis we analysed one Hindi and one English text (children's literature) of fifty adjacency pairs each. Counts of the distribution of cohesion devices were made in each case. Five native speakers of Hindi were then asked to translate the Hindi texts into English. These translations were then evaluated for their authenticity and a mark was assigned to each. These texts (translations) were later analysed for the distribution of cohesion devices. Evaluation of the translations was done by two different judges and there was no significant disagreement between the two evaluations (the differential being only 4%). The mean result was assigned to each translation as a final mark.

2. A language may be analysed as a 'mixture' of sublanguages. English, for example, may be analysed as a 'mixture' of 'scientific English', 'children's literature', 'stock-market reports' and so on. The differentiation is made because each 'sublanguage' exhibits certain peculiarities in terms of its syntax (Kittridge 1980) and distribution and use of cohesion devices (Singh and Kachroo 1980).

## RESULTS

It was found that the 'best' (most authentic) translation resembled the 'actual' English text<sup>3</sup> in terms of the distribution of cohesion devices more than did all the other translations. In the case of pronominalization (anaphoric), the Hindi text had seven instances, the 'actual' English text eight instances, and the translations as follows: T<sub>1</sub> (the best) eight instances, T<sub>2</sub> (the second best) six instances, and the less authentic ones i.e. T<sub>4</sub> and T<sub>5</sub> two and three instances respectively.

Distribution of cohesion devices (summarized in Table 1) in the better translations matched the Hindi text and especially the English text more closely than the less authentic translations.

TABLE - I

	Hindi	English	T <sub>1</sub>	T <sub>2</sub>	T <sub>3</sub>	T <sub>4</sub>	T <sub>5</sub> *
Anaphora:							
Pronominalization	7	8	8	6	3	2	3
Definitivization	1	0	0	1	0	0	0
Cataphora:							
Pronominalization	1	1	0	0	0	0	0
Derivational morphology	1	0	0	0	0	0	0
Inflectional morphology	5	2	3	5	6	7	6
Lexical repetition	18	19	20	24	21	26	20
Synonymy	4	3	3	2	2	2	3
Hyponymy	0	2	1	1	0	0	0
Synechdoche	0	1	0	0	1	0	0
Conjunction	2	3	2	6	4	8	5

\* (T<sub>1</sub>.....T<sub>5</sub> are the translations of the Hindi texts in order: T<sub>1</sub> scoring 80%, T<sub>2</sub> and T<sub>3</sub> 60% and T<sub>4</sub> and T<sub>5</sub> 45 - 40%).

The category inflectional morphology seems to be an exception. The Hindi text has 5 instances, the 'actual' English text 2, in the translations they are distributed as follows: T<sub>1</sub> 3, T<sub>2</sub> 5, T<sub>3</sub> 6, T<sub>4</sub> 7, T<sub>5</sub> 6 (T<sub>5</sub> being the least authentic). However, a closer examination of the figures reveals that the exception is only an apparent one, for the most authentic translation alone was able to reduce the interlanguage differences between Hindi and English in terms of the general use and distribution of inflectional variants.

3. The 'actual' English text refers to the sample text of children's literature in English which was analysed for cohesion device distribution.

In other words  $T_1$  manages, to the greatest extent, to look like an English text and at the same time retain the semantic integrity of the source text ; while the less authentic translations were merely sentential translations demonstrating more affinity to Hindi in terms of cohesive behaviour. Perhaps, some kind of differentiation should be made among cohesion devices as to which ones can be kept in parallel and which ones cannot, given a pair of languages.

In general, all translations seem to confirm the hypothesis. However, the confirmation is not complete. Take  $T_5$  for example : it scored the least (40%) for semantic stylistic authenticity, but matches more closely with the 'actual' English text in terms of the distribution of cohesion devices than does  $T_2$ , which scored 60%.

Assuming that the criteria used for evaluation were uniform, the distribution of cohesion devices in  $T_5$  undermines our hypothesis, at least marginally. The problem is that the distribution of cohesion devices in  $T_5$  resembles more the 'actual' English text than does  $T_4$  or  $T_3$ , which goes against our hypothesis. It is, however, possible to disregard the 'counter-evidence' provided by  $T_5$  on the basis that it is in fact 'bad' at the sentential level, and to argue that comparison should be made only amongst translations which are 'good' at this level.

Consider the category conjunction :  $T_4$  has eight instances whereas the Hindi text has only two instances. Given the 'usual' distribution of cohesion devices (we assume that the distribution of cohesion devices in the actual English text is representative of that sublanguage i.e. children's literature) in the sublanguage under consideration, the frequency is unusually high. Perhaps, the translator, in some way, 'over-translated' the text in an attempt to make it more coherent.

Though coherence is a requirement in discourse, one cannot translate coherence by merely 'connecting' sentences. The point is that taking one coherent text and translating it into another coherent text is not enough. The distribution patterns of cohesion devices have to be followed, whether consciously or intuitively.

The problem becomes clearer when we consider that each language is going to display variant properties both at the sentential and discourse level. For example : in Cree-Montagnais the equivalent of the sentence *She thinks that Mary likes apples* (*Mary* and *she* being coreferential) is acceptable (Jim Lees, personal communication) as opposed to English in which it is not. Note that it is, however, possible in English to have : *Mary thinks that she likes apples* (*Mary* and *she* being coreferential). The condition which rules out the sentence *She thinks that Mary likes apples* (*Mary* and *she* being coreferential) in English is that anaphora may not precede its antecedent within the sentence unless it is in a subordinate clause. In English this condition also seems to operate on the discourse level, though not as strongly, making the use of cataphoric pronominalization rather rare. It is interesting to note that in Cree-Montagnais the use of cataphoric pronominalization is not as rare as in English. Hindi also seems to make use of cataphoric pronominalization more frequently than English.

On textual (discourse) level it is also clear that English differs from Hindi and Cree-Montagnais with respect to direct discourse. In fact there is no evidence that Hindi (Singh 1979) and Cree-Montagnais make the distinction between direct discourse and indirect discourse. The usage of cataphora may be related to the presence or absence of rule(s) of indirect-discourse formation.

#### CONCLUSION

We do not claim the preceding to be a definitive study. Not so, for two main reasons : first, the data at our disposal is not substantial, consequently, we have had to assume that the distribution of cohesion devices in the 'actual' English text is representative of that sublanguage (children's literature). Also, as was correctly pointed out by K.

Connors (personal communication) "'good' translations should, perhaps, be defined by a set of explicit criteria which are independent of the analysis of cohesion".

Although we do not have a set of explicit criteria for defining a 'good' translation, an evaluation of the translations under consideration did precede the count of cohesion devices, thus making the evaluation independent of the analysis of cohesion.

In conclusion, we believe that there is enough evidence to indicate that the use of cohesion devices plays a crucial role in determining the accuracy of a translation, and that a study on a much larger scale should be undertaken to supplement this preliminary inquiry.

#### REFERENCES

- HALLIDAY, M.A.K. and R. HASAN (1976) : *Cohesion in English*, London, Longman Group.
- KEENAN, E. (1970) : "Two Kinds of Presupposition in Natural Language", in Charles Fillmore and D. Langendoen (edit.), *Studies in Linguistic Semantics*, New York, Holt, Rinehart and Winston.
- KITTREDGE, R. (1980) : "Variation and Homogeneity of Sublanguages", to appear in Kittredge and Lehrberger (edit.), *Sublanguage : Studies of Language in Restricted Domains*, de Gruyter.
- SINGH, R. (1979) : "Contrastive Textual Cohesion", Montréal, Université de Montréal. Unpublished.
- SINGH, R. and B. KACHROO (1981) : "Cohesion in Hindi", Paper presented at the Conference of South-Asian languages, at the State University of New York, Stony Brook, May 1981.
- SINGH, R. (1979) : "Modern Hindi and the Direct Discourse Hypothesis", *Gaveshna*, Agra, India, Kendriya Hindi Sansthan.