TEXT COHERENCE IN TRANSLATION

Distributed Language Translation

The goal of these series is to publish texts which are related to computational linguistics and machine translation in general, and the DLT (Distributed Language Translation) research project in particular.

Series editor
Toon Witkam
B.S.O./Research
P.O. Box 8348, NL-3503 RH Utrecht
The Netherlands

Other books in this series:

1. B.C. Papegaaij, V. Sadler and A.P.M. Witkam (eds.)
   Word Expert Semantics
2. Klaus Schubert
   Metataxis

Bart Papegaaij and Klaus Schubert

TEXT COHERENCE IN TRANSLATION

1988
FORIS PUBLICATIONS
Dordrecht - Holland/Providence RI - U.S.A.

Published by:
Foris Publications Holland
P.O. Box 509
3300 AM Dordrecht, The Netherlands

Sole distributor for the U.S.A. and Canada:
Foris Publications U.S.A., Inc.
P.O. Box 5904
Providence RI 02903
U.S.A.

Sole distributor for Japan:
Toppan Company, Ltd.
Sufunotomo Bldg.
1-6, Kanda Surugadai
Chiyoda-ku
Tokyo 101, Japan

CIP-DATA

In co-operation with BSO, Utrecht,
The Netherlands

ISBN 90 6765 360 8 (Bound)
ISBN 90 6765 361 6 (Paper)

© 1988 Foris Publications - Dordrecht

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission from the copyright owner.

Printed in the Netherlands by ICG Printing, Dordrecht.

Contents

Foreword

The present study was written with the aim of discussing an attempt to render the coherence of texts in a machine translation process with a broader public in theoretical and computational linguistics. It is based on work done for the Distributed Language Translation system, a long-term research and development project of the BSO software house in Utrecht for advanced semi-automatic multilingual machine translation.

We wrote this book during the winter of 1987/88, a period of intense implementation work for the first prototype of the DLT system, demonstrated in December 1987, and a second version shown in March 1988. We should like to express our gratitude to all our colleagues in the DLT team for making it possible during these very labour-intensive months for us to work in peace and to think beyond the limits of the first (sentence-based) prototypes.

The practical body of this study, chapter 2, is by Papegaaij, and chapters 1 and 3 are by Schubert, but of course we have discussed manuscripts and contributed to one another's work throughout. A draft version of the second chapter was studied very carefully by Ian Fantom of Newbury, Great Britain (previously connected with the DLT project). Toon Witkam, research director of BSO/Research, also took a critical interest in the manuscript. An inspiring forum was provided by those of our colleagues who took part in our course on text grammar, held in BSO/Research in the winter of 1986/87. We thank them all for many helpful remarks and suggestions. We are also grateful to Dan Maxwell for improving our English and to Ronald Vendelmans, Maiumi Sadler and Hanni Schroten for their contribution to the graphical and typographical form of this book.

Utrecht, April 1988

B. C. Papegaaij
Klaus Schubert

Text coherence in machine translation

An appraisal of the problem and some prerequisites for its solution

The objects of translation are **texts**. A translation process starts from a given text, it analyses the text, and its final product is a new text in another language. One way that **Machine translation** is different from a human translator's work is that the rules and the knowledge that guide the process have to be made explicit and fixed with the degree of precision required for computer programs.

The prime and major source of rules and knowledge in this respect is **grammar**. But unfortunately grammatical research has for many decades and even centuries focused much more on clauses and sentences than on entire texts. That grammatical investigations with a broader domain are a desideratum was noted even before machine translation entered the scene, and the last sixty years or so have seen a slowly but steadily increasing interest in **text linguistics, discourse analysis** or **text grammar**. These fields of study have come into blossom and have brought about a sizeable collection of studies, models and theories. However, the available grammatical rules that might be applied for the translation of texts still cover considerably less ground than sentence-grammatical ones, are much less reliable, and are in no way ready to use for machine translation. The higher awareness of text phenomena among linguists has also led to the insight that the results of the endeavour have remained relatively modest (compared to sentence level grammar) – not so much due to a lack of attention to the problem, but rather because the regularity in text structure is of a considerably less obvious nature.

The present study is an attempt to estimate the impact of text phenomena on a machine translation process, to assess the feasibility of a number of suggested models for this particular application and to propose solutions. In view of the size and complexity of the problem, we do not aim to resolve the entire set of text-grammatical tasks in machine translation in any exhaustive way, nor can we work out the solutions we outline in elaborate detail. We concentrate on a substantial fraction of the overall text-grammatical phenomena which is particularly pertinent to translation. This fraction

is the **coherence of texts**. Put concisely, the question we keep in mind throughout this study is:

- What is the difference between a set of unrelated sentences and a text, and how can we avoid destroying a text by translating it into unrelated sentences?

In this chapter we account for the prerequisites on which the present study is built. Section 1.1. gives a preliminary sketch of what text coherence is and which role it plays in machine translation. Section 1.2. reviews some of the main grammatical theories germane to text coherence, which may be taken to be the theoretical sources of our applications. The study is written for a general audience in theoretical and computational linguistics, and we do not proceed very far beyond the point which would make it necessary to get involved with the technical details of the linguistic design of a machine translation system. This general guideline notwithstanding, our work derives from one particular machine translation project, the **Distributed Language Translation** system (**DLT**), and builds on earlier work done for that system. It may therefore be in order to provide a brief sketch of the main characteristics of DLT and, also of the grammatical model we refer to when speaking about translation not only at the text level, but also at the lower levels. This is done in sections 1.3. and 1.4., respectively.

Having outlined in chapter 1 the theoretical and implementational setting of our study, we immediately leave it in order to approach text coherence in chapter 2 without any bias, as far as possible. In view of the explorative and at times seemingly somewhat groping way in which we deal with this highly intricate problem, we have chosen to address text coherence initially with a good deal of deliberate, creative eclecticism. The chapter draws on various grammatical text theories and computational text models, trying to apply appropriate elements of them in combination with our own findings. This is done with an orientation toward practical application, specifically by closely examining a sample translation. In particular, chapter 2 deals with various grammatical and extragrammatical text coherence phenomena such as pronominal reference, reference identity of content words, thematic progression and rhetorical patterns.

In chapter 3 we return to the theoretical model set out in 1.4. and attempt to extend it to cover text coherence, incorporating the insights and the practical devices considered in chapter 2. We outline the position of text coherence in an overall translation-oriented text grammar (3.1.) and describe three different types of coherence and of relevant coherence-translating devices (3.2. to 3.4.). It is well known that the rules and the knowledge that pertain to translation are not only of a grammatical nature. Section 3.5. is accordingly dedicated to extragrammatical, pragmatic approaches and section 3.6., finally, outlines a number of steps towards implementation of coherence-preserving devices in machine translation.

1.1. Text and context

To translate means to express in another language the content of a given text. The elements from which the target language text is composed are words and their grammatical relations (or more precisely, as we argue in 3.1., morphemes and their relations). Words, the most self-evident text elements, are well suited for illustrating the decisive role the **context** has to play in translation.

Texts and many of their elements are linguistic signs of **form** and **content**. The objective of translation is to replace the form and to preserve the content of the text. Translation is thus form manipulation with reference to content. The most outstanding obstacle to accurate translation is **ambiguity**. As far as translation is concerned, ambiguity in linguistic signs such as words is of a twofold nature, at least: Firstly, a word does not have a single meaning, but refers to a number of concepts. This is especially obvious in words such as English *ball*, that refer to quite distinct concepts, but it is no less true for any arbitrary word, even technical terms, since a concept can always be divided up into several subconcepts. There is in principle no end to this process of dissecting meaning into ever more subtle elements. Saying that words are ambiguous in this way is accordingly almost a tautology. It should, however, be borne in mind that translation is not directly concerned with disambiguating words on this monolingual basis. Translation has to do with resolving the second, translation-relevant, type of ambiguity and may in the course of such a disambiguation indirectly revert to monolingual ambiguity of the first type.

The second type of ambiguity is bilingual or **contrastive ambiguity**, which also may be called *translation ambiguity*. Two words (or other text elements) from different languages that are often adequate translations of each other do not denote exactly the same concept or the same range of concepts. Virtually all bilingual dictionaries whatsoever bear witness to this fact, giving in many cases not only a single translation, but several **translation alternatives** for a single source language entry. In spite of this incongruence between source and target language, translation is nevertheless possible because a word in a given text does not denote the entire range of its possible meanings, but normally only a certain fraction of them, which often may be well covered by a single target language expression (Those who believe in countable distinct concepts might say that a word in a given text denotes only a single concept out of the set of concepts it can refer to.). The main goal of disambiguation for translation purposes is to find out which fraction of its possible range of meanings a word refers to in the given text. By contrast to monolingual ambiguity, the resolution of translation ambiguity is not an endless process. It has a well-defined

accomplishment condition: the choice of a single translation alternative. Translation-oriented disambiguation may thus be viewed as an analysis of the source language text, steered by the requirements of the target language. Translation ambiguity and translation-oriented disambiguation are **language pair-specific**.

It is in translation-oriented disambiguation that the role of **context** in translation is found: The context in which a word (and *mutatis mutandis* other text elements) is used restricts and determines its meaning in the particular occasion. As Henrik Nikula (1986: 41) puts it, context is co-text plus situation. The indications as to the precise meaning of a word in a text should thus be sought first in the surrounding text and secondly in extratextual information. It cannot be the purpose of a machine translation system to pursue this analysis into infinite depth. Translation, machine or human, is not concerned with resolving the infinite monolingual ambiguity as such, but only inasmuch as this is inevitable for resolving translation ambiguity. Since translation-relevant information may be found both in the text itself and in the extratextual situation, the devices for detecting and consulting it are of a **grammatical** and of a **pragmatic** character (see 1.4.).

The surrounding text contributes to a proper understanding of a given word or expression. This means that there is a **connection** between a given text element and other pieces of the text. It is this connectedness that makes up the **coherence** of a text. Obviously it is not enough for a good translation just to understand the source text properly and to detect its coherence. Because the target and source texts should be understood the same way by their respective readers, not only detection of the coherence of the source text, but also the creation of a similarly cohering target text are indispensable. Machine translation must therefore devise rules both for the **analysis** and for the **synthesis** of text coherence. We hope to take a few steps towards this aim in the present study.

1.2. Text linguistics at a glance

This book draws on the stock of theoretical insights about text coherence obtained in a number of linguistic schools. Especially in chapter 2 we attempt to apply some of these theories' ideas to specific translation problems. We do not refer to a specific source from which our idea is taken for every single step we take, nor do we always explicitly identify the ideas we have added ourselves when applying the models. It is therefore appropriate at this point to mention the main models we make use of.

It would be an interesting chapter in the history of science to investigate how the scientific interest in language in the course of the centuries shifted from texts to sentences, to words, to sounds, and back. The interest in texts is not a novelty of this century. On the contrary, the roots of modern linguistics (in the countries of European culture) are found, as far back as we can trace them, in ancient Greek language philosophy, where the text occupied a central position, for example also as the object of rhetoric.

In our century, this interest was revived. Elisabeth Gülich and Wolfgang Raible (1977: 60ff.) give a good account of this development. They list four main groups of general linguistic text models (plus a few models for literary texts, which we pass by here):

- The **Prague school**. The **theme-rheme** distinction originates from this school, in particular, from Vilém Mathesius (1929). From Mathesius's ideas an uninterrupted flow of theoretical and practical work has come, mainly under the heading of *aktuální členění věty*, in English usually called Functional Sentence Perspective.

- The **Copenhagen school**. The language theory developed by Louis Hjelmslev (1943) and his associates, **glossematics**, is the first model to choose the text as its central object, approaching other elements such as sentences, as it were, primarily as elements of texts.

- **Tagmemics**. This school, linked with the name of Kenneth Pike (1967: 287), started a text-grammatical revolution first of all against American Structuralism of Bloomfield's direction.

- A group of models which we should like to subsume under the heading of a **German school**, since they all are offsprings of an innovative approach to linguistics taken by Peter Hartmann in the 1960's (cf. Beaugrande/Dressler

1981: 24, note 8; Hartmann e.g. 1971). The main representatives of this school are in Gülich and Raible's account Roland Harweg (1968), Harald Weinrich (1972), Klaus Heger (1976) and János Petöfi (1969). Among the models of literary texts they also list Teun van Dijk (1978), whose interest in text linguistics starts off in the German school (Gülich/Raible 1977: 250). Although Gülich and Raible in 1977 do not yet give it such a prominent place as the other models, we may now, a decade later, assign also the work by Wolfgang Dressler (1972), later in association with Robert-Alain de Beaugrande (Beaugrande/Dressler 1981) to the German school.

Most of these authors have published enormous quantities of relevant works. We quote here only an initial or central work.

Several of these approaches have found applications in computational linguistics. The model of the Prague school has been developed by scholars such as Eva Hajičová, Petr Sgall (e.g. 1987 for a recent sample from their long list of publications) and others in Prague, but has also found a following at various other places in the world (a computational example: Vasconcellos 1986). The Prague approach as well as the approaches of Hjelmslev, Weinrich and Heger are of special interest to the present study, due to their dependency-minded view on grammar.

Both in general and in computational linguistics, many other ideas and solutions have been put forward. We refrain from an attempt to mention them all, but we refer to a few of them in the subsequent chapters where we deal with some of their ideas more thoroughly.

1.3. Distributed Language Translation

Although this study is written with a wider audience in mind, it is a part of the research endeavour undertaken within a particular machine translation project, the DLT system.

Distributed Language Translation is the name of a long-term research and development project of Buro voor Systeemontwikkeling (BSO/Research), a software house based in Utrecht (the Netherlands). DLT was initiated around 1980 by Toon Witkam. With funding from the European Community, Witkam's subsequent feasibility study (Witkam 1983) provides a detailed account of the linguistic, computational, and commercial aspects of DLT. Since the beginning of 1985, the DLT project has been in a seven-year research and development period with the aim of delivering a complete prototype for a single language pair (English to French) by the end of the period (1991). A restricted prototype was demonstrated in December 1987. This period is being jointly funded by the Netherlands Ministry of Economic Affairs and BSO, and has a total budget of 17 million guilders. Marketing of the DLT machine translation system is scheduled for about 1993.

DLT is a system for semi-automatic machine translation with a monolingual interactive dialogue with the user. It is designed for use in personal computers in data communication networks and is therefore set up to work without post-editing. The basic idea, from which the epithet "distributed" derives, is that the translation process is split up in two parts: The text is entered in, say, English and immediately translated into an **intermediate language**. Problems that cannot be resolved by the system are submitted to the user by means of an interactive dialogue. The dialogue is exclusively in the source language – English, in the example – so that the user need not have command of the target or the intermediate language and need not even know the identity of the target languages. The intermediate-language form is sent to receivers in the network, and only there the text is translated on into the final target language(s), without a dialogue with a user in the second half of the process. The system is **modular** in the sense that the intermediate form of a text lends itself to further translation into any target language and is thus in no way dependent on the source language. DLT is designed as a **multilingual** machine translation system which is easily **extensible** to include more languages of whatever kind. It is not restricted to typologically similar languages. These requirements presuppose a fully expressive intermediate language, and the need of translating **from** the intermediate language into the target language **fully automatically** presupposes an extremely clear and

translation-friendly intermediate language. DLT therefore uses a slightly modified version of **Esperanto** for this purpose.

Besides the feasibility study (Witkam 1983), a number of publications describe the DLT system as well as various of its facets. Overviews are given by Witkam (1985), Hutchins (1986: 287ff.) and Schubert (1986b). At the time of writing the most up-to-date one is by Witkam (forthc.). DLT's model of dependency syntax is suggested by Schubert (1986a) and later worked out in more detail, especially with respect to metataxis, a dependency-based system of translation syntax (Schubert 1987b). The (likewise dependency-minded) semantic-pragmatic model of DLT with its word expert system and its lexical knowledge bank is described in detail by Papegaaij (1986), in a concise sketch also by Papegaaij, Sadler and Witkam (1986). The ambitious quality test of DLT's semantic modules, that was announced by Alan Melby as an external assessor in 1986, has in the meantime been passed successfully. It was designed to measure the precision of lexical transfer, more precisely of content word choice at the syntagma level, in a first, preliminary version of DLT. (About the set-up of the test cf. Melby 1986, about the outcomes Melby forthc.).

The present study builds on these achievements in an attempt to extend the dependency-syntactic and -semantic devices to the text level. Since it is closely linked to the ongoing development of the DLT machine translation system, the present study focuses in its practical parts (chapter 2) mainly on the language pair that is currently worked with in DLT: English and Esperanto. The theoretical framework is then approached in chapter 3 in a more general way, with the aim of achieving cross-linguistic relevance.

This is not the place to repeat any of the discussions in the aforementioned publications, but since distinct usages of identical linguistic terms have become abundant during the past several decades, if not longer, it seems expedient to provide the reader of this study with a shortcut introduction to the way in which grammatical notions are conceived and grammatical terms are used throughout this book. This is done in the next section.

1.4. Some terms and concepts of grammar

Although a certain part of this study contains attempts to apply other scholars' findings, which makes it necessary at times to relate their analyses and to use even their terms, we try to describe grammatical and extragrammatical phenomena by means of a consistent set of appropriately termed concepts. How linguistic terms are used is partially a question of definition, and partially also of labelling. We thus do not maintain that the following terms are the best, nor do we claim that they are the only possible ones. But in view of the existing diversity of definitions, we find it necessary to make our own choices clear. Basically we stick to the definitions adopted earlier (Schubert 1987b: 14f.):

> **Language** is a system of signs for human communication. The signs consist of **form** and **content** (Saussure 1916: 32). **Grammar** is the theory about the internal system of language. Grammar contains the study of both the formal and the content side of the linguistic sign. The theory about the formal side is **syntax**, the one about the content is **semantics**. Language is not a closed system, but is subject to influence from outside factors. The theory about extra-linguistic influence on language is **pragmatics** (with **sociolinguistics** as an important sub-branch).

The terms *syntax*, *semantics* and in particular *pragmatics* are defined in many different and sometimes mutually contradictory ways by different authors. As for our definitions, it should be pointed out that we take *syntax* to comprise all levels of language, so that it deals not only with the arrangement of words in sentences, but with morphemes in words, words in syntagmata, syntagmata in clauses, clauses in sentences and sentences in texts. When we speak about syntax, syntactic features etc., these terms cover even morphology and the form-oriented side of word formation. *Semantics* is also taken to comprise all levels, including, for example, the content side of word formation. Our definition of *pragmatics* is among the broadest definitions found for the term. It is not confined to the persons participating in the speech act, nor to speakers' intentions, underlying communicative motivations or the like. Though emphasising throughout this and other studies a strict separation of form, content and extralinguistic implicature, we do not maintain that these separations are, as it were, observable facts beyond all discussion. On the contrary, we think that they have been drawn as purposeful, but ultimately arbitrary decisions within grammarians' model design. Accordingly we do not deny that there are semantic motivations for syntactic

regularities, borderline cases between pragmatics and semantics and the like (Schubert 1987b: 226).

For the design of a theoretical grammar model as well as for computational implementations of language-related rules clear-cut separations of syntax, semantics and pragmatics are useful. We keep them in mind throughout, although we find it purposeful in the beginning not to stick too strictly to a preformulated model in the more experimental parts of this study. Indeed this study is experimental in two respects. Firstly, it explores ways of translating texts as coherent units, and secondly, it investigates the applicability to the text level of the principles and insights previously established in the design of rule systems for the lower levels of language structure. It is therefore appropriate to give a very short account of some of these principles.

To translate means to transpose a message from an extremely intricate symbol system, whose internal regularity scholars have not been able to describe in any totally exhaustive way, into another such system. If machine translation is to attempt an automation of this intellectually demanding process, a maximal reduction of the complexity of the symbol systems and the transposition process is of prime importance. We therefore strive to detect the precise relevance of grammatical features to translation, which has much to do with the **cross-linguistic validity** of grammatical categories and their values. It is our aim to remove as many **language-specific** features and characteristics as possible from the translation process proper, using them in separate, monolingual processes before (source language) or after (target language) translation proper. In these monolingual processes language-specific features often serve as indicators for a more abstract, translation-relevant grammatical function.

It is on the basis of this reasoning that we have chosen to apply **dependency grammar** (Schubert 1986a: 11ff., 1987b: 17ff.). Dependency grammar – more so than the alternative, constituency grammar – focuses on the **function** various phenomena have in the grammatical system, and uses **form** only to detect or to express such a function. Translation proper, those steps of the overall translation process in which elements of the source language are actually replaced by elements of another language, can then more straightforwardly work on the basis of grammatical function rather than grammatical form. Word order, which seems to be a single phenomenon if viewed as form, but serves several quite diverse functions, is a good example of the usefulness of such an approach in a machine translation attempt. Word order is discussed in more detail in 2.4. and, with explicit reference to the form-function distinction, in 3.3.

The policy of maintaining a strict distinction between syntax and semantics has another consequence which is borne in mind throughout the discussions in this study. We imagine the role of syntax on the one hand and semantics and pragmatics on the other hand as follows: the syntactic parts of the overall translation process generate all **syntactically possible translation alternatives** of a given piece of text and that thereafter semantic and pragmatic processes make a **choice** out of the alternatives available. It is our working hypothesis in this study that this approach of a form-related preselection with a subsequent content-related choice is transferable to the text level as well.

To round off these introductory remarks, we should return to the concept of **text coherence**. Robert-Alain de Beaugrande and Wolfgang Ulrich Dressler (1981: 3ff.) suggest a distinction between *Kohäsion* and *Kohärenz*. Abbreviating their model to a minimum, we can say that they use *Kohäsion* to refer to the syntactic connectedness in a text, whereas *Kohärenz* describes semantic-pragmatic connectedness. Although such a distinction seemingly fits in well with our approach, we have chosen to investigate the overall connectedness in texts as a unit, but of course with reference to the interplay of syntactic, semantic and pragmatic devices for detecting and rendering it. We use the term *coherence* for all these aspects.

Clues and devices of text coherence

A practical analysis

Treating a text as nothing but a collection of individual sentences doesn't do justice to a text's structural complexity. Apart from the fact that sentences taken in isolation from their surrounding text will often be highly ambiguous, and sometimes practically meaningless, much of the message of the text itself can only be understood when seen as a single structure. No simple summing up of the separate sentences' meanings (if at all possible) can capture a significant part of the message which the text was intended to convey.

Text understanding involves complex linguistic processes on many levels. The sentences a text is composed of must not only be analysed internally (i.e., as self-contained units), but also as functional elements in the larger unit of the text. Many words in many sentences can only be properly understood when seen as markers of textual functions, or influenced by such functions. Syntactic variation (passive/active, nominalization of verbs, etc.) is not just free variation, but more often than not fulfils a definite textual function.

Translating one natural language into another can be seen as a form of analysis: explaining an utterance's meaning by rephrasing it in another language, or as Peter Newmark (1981: 23) puts it, "a translation is the only direct statement of linguistic meaning". Most people seem to agree that a considerable amount of surface processing (syntactic manipulation) is necessary to carry out such a task. There is much less agreement on the depth of understanding (semantic and pragmatic processing) required to translate from one natural language into another. Early machine translation proposals assumed understanding was unnecessary. Later proposals included "a mild form of understanding". And there are some (e.g. Tucker/Nirenburg 1984: 148; Nirenburg/Raskin/Tucker 1986: 627) who claim "full" or "deep understanding" is the only road to machine translation "in an interesting sense".

2.1. English to Esperanto – a sample translation

Before we take a position on the scale from "no understanding" to "full understanding" (see 3.5.2.), it is perhaps better to study in detail a sample translation, in order to see what kind of decisions are involved in translating a coherent piece of text. We will submit the sample text to a kind of "close reading", highlighting interesting points, with special emphasis on the kind of information that is available at each point, how that information is presented, and how it influences the translation decisions that are taken.

By going through a piece of text in this way, we hope to be able to extract a few of the more striking aspects of text translation, which will then be discussed in detail in subsequent chapters. Note that the view of translation presented here is of necessity **incomplete**. Many questions will be left unanswered, and much more work will have to be done to complete our understanding of the complex process of natural language translation. The analysis in this chapter is for a good deal biased towards English, which is compared with DLT's intermediate language Esperanto. Considerations of a wider cross-linguistic scope are included here and there, but they should mainly be sought in chapter 3.

2.1.1. The text

The sample text is taken from a well-known periodical, and is part of a regularly appearing column (*Computer Recreations* by Brian Hayes in *Scientific American*, Volume 250, 1984, No. 2). It was chosen as an instance of non-specialist informative text, aimed at educated readers not necessarily familiar with the subject matter. The translation is a high quality human translation in Esperanto, made for the purpose of comparison. The translation is certainly not meant as an example of what a machine translation system should be able to achieve, but serves to illustrate the kind of choices a translator must make to rephrase a source language text into a target language text, in this case DLT's intermediate language, Esperanto.

Figure 1 shows the original text. We have left out the first three paragraphs and added sentence numbers. The text was translated into Esperanto by one of DLT's linguists. The translation is shown in the right column of the table. In DLT's intermediate language the morpheme boundaries in words are normally indicated with a special token (e.g. *komput'il'a'j*) which we have omitted here to make the text more

Figure 1: A sample translation

Computer Recreations	**Komputilaj Ekzercoj**
Turning turtle gives one a view of geometry from the inside out	*Testudiĝo permesas rigardi la geometrion elinterne*
by Brian Hayes	*de Brian Hayes*
1. The new way of thinking about geometry has come to be known as "turtle geometry".	*1. La nova aliro al geometrio diskonatiĝis sub la nomo "testuda geometrio" (turtle geometry).*
2. It is closely connected with the programming language Logo, which in turn has its roots in the Massachusetts Institute of Technology.	*2. Ĝi estas intime ligita al la programlingvo Logo, kiu siavice devenas de la Teknologia Instituto de Massachusetts (Massachusetts Institute of Technology).*
3. Logo was conceived in the 1960's by Seymour Papert of M.I.T., primarily as a language for introducing children to computers.	*3. Logon elpensis, en la sesdekaj jaroj, Seymour Papert ĉe M.I.T., unuavice kiel lingvon por eklernigi infanojn pri komputiloj.*
4. Many others have since contributed to its development and to its applications both in education and in other fields.	*4. De tiam, multaj aliaj kontribuis al ĝiaj evoluo kaj aplikadoj, ne nur en edukado sed ankaŭ sur aliaj kampoj.*
5. Among them are Harold Abelson and Andrea A. diSessa of M.I.T., who have set forth the ideas underlying turtle geometry in a remarkable expository work: *Turtle Geometry: The Computer as a Medium for Exploring Mathematics.*	*5a. Inter ili troviĝas Harold Abelson kaj Andrea A. diSessa, de M.I.T.* *5b. Ili eksplikis la ideojn, sur kiuj baziĝas la testuda geometrio, en elstara enkonduka verko:* Turtle Geometry: The Computer as a Medium for Exploring Mathematics (Testuda geometrio: la komputilo kiel rimedo por esplori la matematikon).
6. The original turtle was a mechanical device: a small wheeled vehicle whose movements could be controlled by instructions typed on a computer keyboard.	*6. Origine, la testudo estis mekanikaĵo, nome radohava veturileto, kies moviĝado povis esti regata per tajpado den instrukcioj ĉe komputila klavaro.*
7. The first such creature was built by the British neurophysiologist W. Grey Walter in the late 1940's.	*7. La unuan tian estaĵon konstruis la brita neŭrofiziologo W. Grey Walter en la fino de la kvardekaj jaroj.*
8. It had a dome-shaped cover somewhat like a turtle's shell.	*8. Ĝi havis kupoloforman kovrilon iom similan al kiraso de testudo.*
9. A mechanical turtle can move forward or backward and can change direction by pivoting in place.	*9. Mekanika testudo povas moviĝi antaŭen aŭ malantaŭen kaj povas ŝanĝi sian direkton per simpla pivotado.*
10. A pen can be mounted on the undercarriage, so that when the turtle is made to wander over a sheet of paper, it leaves a record of its path.	*10. Sub la testudo eblas munti skribilon, tiel ke ĝi postlasas spuron de sia vojo, kiam oni devigas ĝin travagi folion de papero.*
11. Today such "floor turtles" are less common than "screen turtles", which move and draw on the surface of a cathode-ray tube.	*11. Nuntempe, tiaj "planktestudoj" estas malpli oftaj, ol "ekrantestudoj", kiuj moviĝas kaj desegnas sur la surfaco de katodradia tubo.*
12. The turtle itself is represented on the screen by a simple triangular form, which moves in response to commands or programs entered at the keyboard.	*12. La testudon mem reprezentas sur la ekrano simpla triangula formo, kiu moviĝas laŭ instrukcioj aŭ programoj tajpitaj per la klavaro.*