
Nili Mandelblit
University of California, San Diego
Dep. of Cognitive Science,
La Jolla, CA 92093-0515
e-mail: mandelbl@cogsci.ucsd.edu

GENERAL TOPIC: What sort of information is to be captured by an adequate interlingua?

In my work, I investigate instances of linguistic expressions which provide only **PARTIAL** information about a conceived event (i.e., only some aspects of a conceived event are expressed explicitly, while others are understood implicitly). I argue that languages differ in the type of information they most often express explicitly (i.e., each language explicitly communicates **DIFFERENT** aspects of the same generic

event).

In such cases, a direct transfer of the source text into the target language cannot provide a correct translation, and it is the role of the interlingua to COMPLETE the missing (implicit) information from the source text, before the target text can be generated.

A crucial aspect of my research is in identifying how languages actually differ in the information they explicitly communicate. I propose that a primary factor in defining which aspects of generic event-types are commonly expressed in a particular language is the inventory of CONSTRUCTIONS available in the language.

A Construction is a syntactic (or morphological) pattern which is independently associated with a semantic structure (cf., Fillmore & Kay, ms.; Goldberg, 1995). A specific construction is used by language speakers to express a novel conceived event only if correlation is found between the semantic structure of the conceived event and the semantics associated with the grammatical pattern (the construction). Since constructions vary not only in the semantic structures associated with them but also in the partial information they highlight, variations in the inventory of constructions across languages also imply variations in the type of information explicitly communicated in each language (Mandelblit 1995a).

Example:

Goldberg (1995) analyzes the Caused-Motion construction in English. Its syntactic form is [NP V NP PP], and the semantic structure associated with it (according to Goldberg) is a generic Caused-Motion event (i.e., X causes Y to move in direction Z). Sentence (1-3) below are instances of the Caused-Motion construction. Note that the caused-motion semantics does not exist in any of the lexical items independently, and is hence assumed to exist in the syntactic structure itself.

- (1) Martha trotted the horse into the stable.
- (2) The wind blew the ship off course.
- (3) The audience laughed the poor actor off the stage.

An important point to note is that only PARTIAL information about the conceived caused-motion event is actually expressed in examples (1-3).

In (1), the event being communicated is one in which Martha is causing the horse to trot (and move) into the stable. However, nothing is said explicitly about HOW Martha made the horse trotting (what was the

CAUSING event). Did Martha lead the horse into the stable, or did she hit the horse, thereby causing the horse to trot in the direction of the stable?

In (2), the event being communicated is one in which the wind blowing causes the ship to move away from its original course. In this example, the sentence provides explicit information about the causing force that made the ship change its location (the wind blowing). However, no explicit information is given about the resulting motion event (i.e., in what manner and where did the ship move: was the ship being SHIFTED into another course? or was it drowning down into the sea?). In both examples (1-2), a default scenario is commonly imposed by the listener to complete the missing information.

Example (3) provides again explicit information about the causing event (the audience laughing), but the resulting motion event is left implicit. In what manner did the actor move off the stage? Was the actor passively SHIFTED off the stage (like the ship in example 2), or was the actor voluntarily RUNNING AWAY from the stage? Again, background knowledge of default scenarios imposes a specific interpretation.

What happens when we try to translate English Caused-Motion sentences into other languages (Hebrew or French, for example)? Hebrew and French do not have an independent Caused-Motion construction. Rather to express a caused-motion event as in sentences (1)-(3), Hebrew and French speakers make use of a GENERIC CAUSATIVE construction that exists in the language (i.e., the *faire* construction in French, or the morphological *hifUil* construction in Hebrew). However, while the main verb in the Caused-Motion construction in English may express either the resulting motion event (as in example 1), or the causing event (as in example 2-3), the main verb in the causative *faire* construction in French and the *hifUil* construction in Hebrew ALWAYS denotes the RESULTING event of a causal sequence of events (and the CAUSING event is left implicit). Hence, clearly a translation of sentences (2)-(3) into Hebrew and French cannot be a direct function of the main verb in the source text.

Below are the Hebrew and French translations for sentences (1-3). The English examples (i) are followed by an Hebrew translation (ii), a word-to-word transfer of the Hebrew version into English (iii), and a French translation (iv).

- (1) (i) She *trotted* the horse into the stable.
(ii) Hi hidhira(d.h.r-hifUil) et hasus letoch haurva.

- (iii) She TROT-CAUSE(past) the-horse into the-stable.
- (iv) Elle a fait trotter le cheval dans 1 ecurie.

(2) (i) The wind *blew* the ship off course.

- (ii) Haruax hesita(n.s.t-hifUil) et hasfina mimaslula.
- (iii) The wind SHIFT-CAUSE(past) the-ship off-its-course,
- (iv) Le vent a ecarte le navire de sa trajectoire.

(3) (i) The audience *laughed* the actor off the stage.

- (ii) Hakahal hivrix(b.r.x-hifUil) et hasaxan min habama (besxoko).
- (iii) The audience RUN-AWAY-CAUSE(past) the-actor off the-stage.

The main verb in the Hebrew (and French) translations in example (2)-(3) is not a function of the main verb (or any other lexical item) in the source sentence. To perform the translation of sentences (2)-(3), a translator (human or machine) must first reconstruct the original causal sequence of events communicated in the source language. The representation of the whole causal sequence of events (the causing event and the implicit effected motion event) forms the INTERLINGUAL representation. From the interlingual representation, a target text can be generated (the translation) using available grammatical constructions in the target language (i.e., the *faire* or *hifUil* constructions in French and Hebrew respectively, both explicitly communicating only the effected motion event).

What kind of information is needed to construct an interlingual representation from the source text, and to generate a target sentence from the interlingual representation?

In addition to language-specific and interlingual lexicons, a translation system must have information about:

(1) The inventory of constructions available in the source and target languages. For each construction we need to specify the generic event type associated with the construction, and the partial information most commonly highlighted (or explicitly expressed) by the construction.

(2) Background knowledge - common scenarios of causally related events.

REFERENCES:

For further information about this work, please refer to:

- (1) Mandelblit, N. (1995a). RCognition, Translation, and NLPS.

Technical Report, Cognitive Science, University of California, San Diego.

(2) Mandelblit, N (1995b). "Beyond Lexical Semantics: Mapping and Blending of Conceptual and Linguistic Structures in Machine Translation", In Proceedings of the 4th Int. Conf. on the Cognitive Science of Natural Language Processing, Dublin, Ireland, July 1995.

(3) Blending: Creative Aspects in Grammar and Translation. Presentation at the Computing Research Laboratory, New Mexico State University, April 1996.