

モダリティ表現の翻訳における問題点

Problems in Translating Modal Expressions

Christoph NEUMANN

Tokyo Institute of Technology

neumann@cs.titech.ac.jp

Francis BOND

NTT Communication Science Laboratories

bond@cslab.kecl.ntt.co.jp

ABSTRACT

機械翻訳で使用するためのモダリティの適切な形式化を妨げる従来のモダリティの理論について議論し、モダリティの新たな形式的な概念の要点を述べる。

We identify the points in conventional modality theory that seem to hinder a proper formalisation of modality for the use within machine translation and outline the essentials of a formal concept of modality.

1 INTRODUCTION

Our goal is a formalisation of the theory of modality as a base for machine translation. The importance of modality is shown by the abundance and variety of modal expressions in most languages. However, a formalisation of the translation of modal expressions especially between Japanese and European languages is a difficult task for several reasons. This is why machine translation systems usually fail when translating modal expressions. We identify the problem points and extract solutions to overcome them. A formal model of modality will have to be an interlingua module consisting of equivalency classes of modal trigger elements or combinations of such trigger elements.

2 FREQUENCY OF MODAL EXPRESSIONS

Modality is a semanto-pragmatic category that is well examined in linguistics and that seems intrinsic to language production. Modality designates the way in which the speaker qualifies the validity of the proposition of an utterance. Usually, linguists distinguish four types of modality: epistemic (judgement), deontic (obligation), dynamic (ability) and conditional (hypothesis).

Modality is intrinsic to language production. A survey of the occurrence of 13 modal elements in the EDR-corpus of Japanese (EDR 1995) shows that modal elements are not marginal, despite first appearance. The element *ta* alone occurs in 31 % of the corpus sentences. That *ta* may mark (non-modal) past tense or (modal) hypothesis, doesn't undermine the importance of modality, but underlines the need to distinguish it clearly from other linguistic phenomena.

Modality type	Modal element	Occurrence
Epistemic	<i>yo.</i>	437
	<i>darou</i>	3010
	<i>souda</i>	1317
	<i>kitto</i>	50
	<i>osoraku</i>	112
	<i>youda</i>	1148
Deontic	<i>tai</i>	2777
	<i>noda, nodesu</i>	3119
	<i>kudasai</i>	245
Dynamic	<i>dekiru</i>	4305
Hypothesis	<i>ta.</i>	59729
	<i>naraba</i>	130
	<i>toshitara</i>	78

Table 1: Frequency of modal elements in the EDR-corpus (195,000 sentences)

3 CONVENTIONAL STRATEGIES

TRANSLATION

Modality research, especially in European linguistic tradition, has been rather monolingual. That is, research tried to extract the modal functions a given modal form in one language performs. In this way, the German verb mode "KonjunktivII" was recognised as establishing a hypothetical statement and it was noted that *must* (E¹) as

¹ Languages have the following abbreviations throughout the paper: D: German, E: English, F: French, J: Japanese.

well as *müssen* (D) can express either an obligation or a objective-based (epistemic) assumption (Lyons 1977).

However, there has been relatively little research in comparing the establishment of modality in different languages, and – as a consequence – on discovering how the universal cognitive need to express modality is mapped into concrete expressions, which parallels there are between modal forms in different languages, and how these parallels can be formalised.

In lack of a proper formalisation framework, current machine translation systems seem to use two strategies of translating modal expressions.

3.1 Lexical strategy

A modal expression is translated by the equivalent dictionary entry (belonging to the same word category) in the target language. This works sometimes within European languages, as modal auxiliaries are here a major modal element, and as the Germanic and Romanic languages all have rather similar sets of about half a dozen such auxiliaries with similar meanings (e.g. *must* = *müssen* (D) = *devoir* (F)).

The weak point of this strategy becomes evident, for instance, when T1² always maps *würde(n)* (D) to *would* (here, the conditional clause in English would require past tense)

I D: *Ich würde dir helfen, wenn ich nicht arbeiten würde.*

E: *I would help you if I wasn't working.*

System: **I would help you if I would not work.*

In fact, even in close languages neither lexical equivalents nor word categories are isomorphic for modal expressions in both languages. Aijmer (to appear) shows that in Swedish translations of English literature, *may* in epistemic auxiliary usage is translated only in 31 % of the cases into the Swedish auxiliary *kan*, but in 47 % of the occurrences into the adverb *kanske*.

3.2 Strategy of equivalent expressions

A modal form in one language is translated into a form or combination of forms in the target language. The target form(s) may be syntactically divergent. For instance, in ALT-JE, a Japanese-English machine translation system, the abstract category "Wish (1st person)" can be triggered by the adjective *tai*, but is mapped to the English verb *want to*.

However, this strategy still suffers from the restriction to certain target expressions.

4 PROBLEMS IN TRANSLATING MODALITY – AND SOLUTIONS

Five reasons can be stated why modality is difficult to grasp for human translators and why machine translation systems tend to mistranslate modal expressions. We present the reasons and propose ways to solve them:

4.1 Which Linguistic Categories Can Trigger Modality?

Most modality theories do not tempt to cover all linguistic categories involved in modalisation. Even recent works like Metzler (1993) define modality mainly through two verbal categories, modal auxiliary verbs and verbal mode. However, even straightforward examples like the epistemic adverb *kanske* in Swedish above show that modality is expressed by a wide variety of different linguistic categories, of which auxiliaries and mode are only two options.

In fact, while mode as a marked phenomena of Indo-European languages is completely absent in Japanese, modality in Japanese sentences is often established by discourse particles like *yo* or *ne* (cf. Masuoka 1991, Ueno 1989)

How can we identify the element that adds modality to a given sentence? All identifiable elements and categories in a modal sentence have to be left out one after another until it is clear without which one the sentence has only a plain propositional meaning. Focussing on examples in Japanese, English, German and French, with this method, we have identified so far 24 linguistic categories responsible for modalisation, including also "exotic" ones like word order, Aktionsart or person.

A side effect of this method was the finding that in many cases, two or more elements are jointly responsible for modalisation, so-called "trigger combinations".

4.2 Modal "Functions"

The linguistic search of modality emerged from the notion of modality in logic; initially, linguists classified modal expressions along the logic notions of alethic³, deontic and epistemic modality (cf. Lyons 1977).

The connection between linguistic and logic modality is not justified. This connection seems implicitly assumed by the inference that the human mind is logically organised, that language expresses the human mind, thus, language must express logic. This deduction itself may not be true; but even if it is true, this does not mean that it is modality, of all linguistic phenomena, that is expressing this logic.

² "T1 Professional 3.0" by Langenscheidt, commercial version and derivative of Siemens' "Metal" German-English system

³ alethic modality = "the necessary or contingent truth of propositions" (Lyons 1977: 791)

Alethic modality, e.g., has no autonomously identifiable existence in language. On the other hand, the distinction into four categories is too rough to grasp the difference between (both examples deontic)

2a *You must go.*

2b *You may go.*

This can be resolved by establishing two sub-classes, obligation and permission, but then, why is the imperative in

3 *Come in!*

used in a permissive sense?

In a similar way, Japanese modality research focuses on explaining the function "hypothesis" (cf. Akatsuka/Tsubomoto 1998).

The linguistic interest in modality is to grasp the different functions of modal forms. However, these findings are made possible by already assuming *a priori* the different (cf. Klinge [to appear]).

If we want to try to formalise translation of modal expressions, we have to ignore the wish to recognise actual cognitive function classes behind modal expressions in order to get out of this vicious circle. Rather than functions, we need to define equivalency classes by listing all overt elements that can be interchangeably used to produce a certain effect of modality (=function). Overt elements belonging to such a class will be called trigger elements of this class.

Such monolingual equivalency classes must be intersected with equivalency classes in other languages, resulting in (a probably increased number of) interlingual equivalency classes. For instance, while English surprisingly seems to have one single class for both "normative" obligation⁴ and objective assumptions (both expressed by *must*), forms in Japanese (e.g. *no-da* vs. *hazu-da*) or German (*müssen* vs. *einfach müssen*) are not interchangeable, thus establish two separate monolingual classes, so that there will be also be two separate interlingual equivalency classes.

This way of defining equivalency classes can be automatised; manually, we have identified so far 16 different classes for the four languages in focus.

4.3 Mono-functional Approach

Even if we can identify new form-based functions, many forms may trigger two or more function, while only few forms unequivocally identify a single function. In fact, it is the "classical" modal auxiliaries like *must*, *may* and *müssen*, *sollen* (D) that seem to cover a whole range of

modal functions. Unfortunately, clearly assignable forms seem to occur much less often than the ambiguous ones. This is shown by the frequent occurrence of *ta*, *deshou* or *no-da* compared to the rarity of *kitto* or *kudasai* (cf. table 1).

However, most functions can be clearly identified because of characteristic combinations of trigger elements, called trigger combinations. Thus, *ta* is identified as introducing a hypothesis, if the same sentence contains, for instance, *naraba* or *toshitara*.

In fact, the above stated tendency of languages to use ambiguous modal forms rather than clearly identifiable ones leaves no choice but to look for second elements that help identify the whole modal class. Thus, the assignment of trigger combinations is a central part of formalising the translation of modal expressions.

4.4 Syntactically Divergent Structures

Modality is not a property of the verb, but of the sentence. Modalisation can be done in many ways and on different levels of the sentence construction. This is self-evident in the light of the number of linguistic categories which are able to establish modality.

This means also that it is hardly more than (lucky) coincidence if one modal class has syntactically similar trigger elements on both sides of the translation. Equivalent modal forms are very often syntactically divergent. This divergence can only be dealt with, if the modal information of a sentence is completely separated from the syntactical analysis and generation, i.e. if it is transmitted to the target language in the abstract terms of an interlingua, even in mere two-language systems.⁵

If modal information was left to even very refined transfer mechanisms, we still would have to establish precise mapping rules for every single trigger into every single trigger in both languages. It is not feasible, though, to list all possible combinations..

4.5 Different Degree of Modalisation

The last difficulty in translating modal expressions lies in the phenomenon that different languages seem to require a different degree of modalisation. This means that a "weak" modal expression may be better not be translated at all in the target language, resulting in a "null" modal expression.

In this respect, Japanese seems to be higher modalised than European languages. E.g.

4a J: *Oishi-sa!* ("It looks tasty.")

⁴ "Normative" duty means that the duty is only related, but not imposed by the speaker (e.g. by society, religion).

⁵ Similarly, Bond et al. (1997) show that nominal and adverbial time expressions (in J→E translation) cannot be handled in direct (lexical) mapping, but require an abstract "intermediate" representation.

with *so* indicating that this is the speakers impression, not knowledge, would be come a mere

4b D: *Lecker!*

in German without any modalisation.

The Japanese modal particles *yo* and *ne* (indicating knowledge res. asking for hearer confirmation) are also often simply ignored when translated into English or German.

5 CONSTRUCTIVE SOLUTIONS

Translation needs an independent approach towards modality as modal expressions represent an important part of language production and as modal expressions seem not be translatable in an "isolated" way, be it lexically or even on a more complex level.

We have established the essential features above that are required for a proper formalisation of modality translation. Without any theoretical claim, modality must be formulated as (a module of) an abstract interlingua. The structure of this module will be determined by the intersections of the modal classes established by all participating languages. These modal classes are defined as equivalent classes consisting of modal "trigger" elements that are interchangeable in a given modal sentence without altering the meaning. The trigger elements may as well be combinations of elements.

We focus on the notion of "modalisation" rather than on "modality", as "modalisation" stresses the fact that there is a process of materialisation of an abstract concept (i.e. the interlingua, here: modality) through linguistic elements, in the same way the notion of "lexicalisation" indicates that an abstract concept has been materialised through a lexicon entry. As for the flexible range of lexicalisation among different languages, the degrees of modalisation can also vary from language to language.

These requirements have resulted in the proposal of a Module of Modality (MoM), first presented in Neumann (1998), which is constantly evolving and modified.

As a first step, ALT-J/E already separates "subjective" verbal information, i.e. tense, aspect and mode, from "objective" verbal information (the proposition) (cf. Ikehara et al. 1991). What the system needs now, is a more abstract representation of the subjective parts. Thus, parts of the MoM are being currently implemented as part of the system.

Finally, while we the immediate goal of our research is the formalisation of modalisation in order to be able to improve the quality of practical modality translation, it is at hand that a abstract interlingua successful in mediating modal expressions could serve back as a model for a theory of modality. As it is well known, such a coherent theory does not exist yet, either.

REFERENCES

- Aijmer K. (to appear), "Epistemic Possibility in An English-Swedish Contrastive Perspective" in Proceedings of the Colloque International, 10.-12.12.1998 Anvers, Les verbes modaux dans les langues germaniques et romanes
- Akatsuka N., Tsubomoto A. (1998) *Modariti to Hatsuwa-Kodo* [Modality and Utterance Behaviour], Tokyo: Kenkyusha
- Bond F., Ogura K., Uchino H., (1997), ""Temporal expressions in Japanese-to-English Machine Translation" in Seventh International Conference on Theoretical and Methodological Issues in Machine Translation: TMI-97, 55-62
- EDR (1995), *Electronic Dictionary Technical Guide* (In Japanese), Tokyo, Japan Electronic Dictionary Research Institute, Ltd.
- Ikehara S., Shirai S., Akio Y., Nakaiwa H. (1991), "Toward an MT system without pre-editing – effects of new methods in ALT-J/E" in Proceedings of MT Summit III, 101-106
- Klinge A. (to appear), "Modal underspecification and Problems of Metalanguage" in Proceedings of the Colloque International, 10.-12.12.1998 Anvers, Les verbes modaux dans les langues germaniques et romanes
- Lyons J. (1977), *Semantics*, Cambridge. Cambridge University Press
- Masuoka T. (1991): *Modariti-no bunpo* [Grammar of Modality], Tokyo, Kuroshio
- Metzler Lexikon Sprache* (1993) (editor: H. Glück), Stuttgart/Weimar: J. B. Metzler
- Neumann C. (1998), "Modality Expressions in Japanese" in IEICE Technical Report Vol. 98 No. 209, 9-16 (Natural Language Understanding and Models of Communication - NLC 98-8)
- Ueno T. (1989), "Doshi to Ho" [Verb and Mode] in Gengo:18/9, 66—69