Abduction and Mismatch in Machine Translation

Jean Mark Gawron, SRI International[1]
San Diego State University

December 9, 1999

1 Intro

There are a number of quality and correctness issues connected with state-of-the-art machine translation (MT) systems. The first requirement, which we will not be addressing in this paper, is to produce a correct analysis of the input text, where correctness is defined by the needs of the particular system. Assuming the correct analysis is found, the next step is search for a correct representation of a target. Our chief concern is this paper is to examine those cases in which the search fails to produce a satisfactory result. There are two possible reasons for this failure. First the search result may be suboptimal. Second, the search space may be too confined. In the first case, the search space is big enough, but search is being terminated too soon for practical reasons. In the second, all acceptable results lie outside the search space. In this paper we propose an architecture that allows the search space in translation to be greatly expanded and provides practical mechanisms for early termination. The goal is a more flexible translation system with better tools for discriminating among candidate solutions.

A chief reason why the definition of the search space is an issue in MT systems is that there is no exact definition of what translation is. There is a reasonably well-defined sense in which a translation may be said to be exact. But as a practical matter exact translation is rarely acheived. The reason for this is the problem of *translation mismatch*, which is discussed in next section.

2 Mismatch

Table 1, adapted from Nirenberg (1987), illustrates some simple lexical mismatches between English and Russian taken from the domain of kinship terms, well-known for its rich variety cross-linguistically.

Note, for example, that the English lexeme *father-in-law* collapses the Russian distinction between wife's father and husband's father. More seriously, *brother-in-law* collapses a three-way distinction. Russian is sensitive to whether the in-law is spouse of a sybling or sybling of a spouse, as well as to whether the spouse is male or female. Finally English simply has a lexical gap for daughter's husband's father, which is monomorphemically expressed in Russian as *svat*. Thus in translating English in-law terms to Russian, information must be added to what is lexically encoded. While, in translating *svat* from Russian, either information must be lost, as in choosing *in-law*, or a paraphrase must be chosen from among several options, either *son-in-law's father*, or *daughter's father-in-law*, or *daughter's husband's father*.

Exampe (1) illustrates a lexical gap posing a somewhat different problem. The English word *partnership* has no exact counterpart in Japanese, though two renderings seem likely. Thus, if the relationship has a more formal legal basis, *teikei kankei* is likely; however, if it is a more informal pairing, *kyoroku kankei* is better. Thus, in contrast to the kinship example, in translating a concept with lexical rendering in the target, it is natural to use a paraphrase that requires that some information be added:

[1]This is a report on joint work with Megumi Kameyama (SRI) and Masayo Iida (Inxight), whose insights and suggestions I am greatly indebted to.

Interlingua	Russian	English
wife's father	test'	father-in-law
husband's father	svyokor	father-in-law
wife's brother	shurin	brother-in-law
husband's brother	dever'	brother-in-law
sister's husband	zyat'	brother-in-law
daughter's husband's father	svat	[GAP]

Table 1: Some Kinship Terms in English and Russian; Nirenberg 1987

(1) a. partnership
 b. kyoroku kankei
 collaboration relation
 c. teikei kankei
 cooperation relation

Example (2) illustrates a case of non-lexical, or constructional, mismatch. French has perfectly good translations of the English words *knife* and English *wound*, but does not justapose them directly, as English does, to express the concept of a knife wound. The most literal rendering of *knife wound* is (2c), *blessure à couteau*, but this is collocationally odd, nor does changing the preposition seem to help very much. What seems to be required is an insertion.

(2) a. knife wound
 b. blessure à coup de couteau
 wound to/of blow of couteau
 c. ? blessure à couteau

Thus, (2b), *blessure à coup de couteau*, is much better.[2] The individual words in this case have direct translations but the means of assembling those translations in the target are not readily available. Collocational constraints intervene. The solution in this case is to insert some extra material, as in (2b). The question such examples naturally raise is this: Is there any non ad hoc organization of the linguistic descriptions in an MT system that will allow a search algorithm to discover such solutions to mismatch problems? If not, then certainly an approach like example-based transfer(cites ??), which at least promises to handle some example-supported subset of the ad hoc cases, looks appealing; in Section **??**, we examine an alternative, frame-based transfer.

This example also illustrates why the translation relation in general is not symmetric. In mapping from source to target some small adjustment is called for and a small change in meaning results. The target now maps back to something new in the source. Thus, (2b) would most naturally (and simply) be rendered back into English as *stab wound*.

(3) a. syooka ki kei
 digestion organ system
 b. digestive system

[2] Constructions which are semantically broad or vague, such as the English noun noun compound construction, seem to be particularly susceptible to constructional mismatch, because languages don't all accommodate vagueness in the same way. It remains significant that the French phrase *blessure à couteau* would probably be comprehensible. Our concern in this paper will be to explore a proncipled account that makes both translations possible, and allows the trade-offs in producing them to be explored.

(4) a.

chu goku	gawa	no	yousei
china	side	ADNOM	request

b. Chinese request

Example (5) illustrates another sort of non-lexical mismatch. In this case the Japanese constituents have direct English counterparts; however, the modification relations must be changed to produce a fluent English phrase.

(5) a.

gijutsu	men	de	teikei	kankei	o	musubu	nda
technology	aspect	INST	cooperation	relation	ACC	tie/bind	PAST

b. ? form a partnership with the technology aspect

c. form a partnership involving technology

d. form a technology partnership

There is no way to have the analogue of *gijutsu men de* (*with the technology aspect*) modify the main verb as it can in Japanese.

Table 1 (from Kay, et al. 1994) provides an example of what might be called *conceptual mismatch*, taken from the domain of bus and train travel. The table gives expressions describing the cancellation of a transit ticket in four languages. There is considerable indeterminacy as well as considerable variation, as the glosses show. We see that some languages focus on the fact that stamping the ticket makes passage valid, some on the fact that stamping the ticket guarantees it is not reusable (invalid), and that one allows both ways of looking at the matter (Italian). Consider the issue of translating from German to English. Although German *entwerten* generally glosses as English *invalidate*, English *invalidate* is not an option in this case. In most contexts an MT system will not have enough information about the type of machine used to decide between *cancel*, *stamp* and *punch*, so the best translation choice appears to be *validate*. Thus in this context, we translate the German word meaning *invalidate* as *validate*.

It would be odd to say that German *entwerten* sometimes means *validate*. A better description of this state of affairs is to say that *enterwerten* means *invalidate* as it usually does, but that the conventional ways of describing ticket cancellation in English and German highlight different aspects of the situation. A conceptual mismatch occurs. Fluent translation involves not just recognizing the meanings that are used, but also recognizing the kind of situation that is being described, as well as knowing the conventions in each language for describing that kind of situation.

The occasion for conceptual mismatch arises when the concept referred to is one for which neither language has any specialized lexical apparatus. The concept can then be referred to either by free composition of the expressions of the language or by some conventional expression. In the latter case, the case of an idiom, both languages may choose a transparent encoding (referred to as an *idiom of encoding*, Makkai 1972). When they choose transparent encodings that highlight different aspects of the situation, conceptual mismatch occurs.

Wherever there may be lexical mismatch there may also be conceptual mismatch, for the simple reason that languages B and C may choose to express the content lexicalized in language A in different ways. This is illustrated in (6).

(6) a. Kim stabbed Sandy.

b.

Kim	ga	Sandy	ni	sasikizu	owaseta
Kim	Nom	Sandy	Dat	stab wound	give PST.

c.

Kim	a	donné	un	coup	de	poignard	à	Sandy.
Kim	Aux	give PST PART	det	blow	of	knife	to	Sandy

Kay, Gawron, Norvig 1994		
Language	Word	Gloss
English	validate cancel stamp punch	
French	valider composter obliterer	validate punch cancel
German	entwerten	invalidate
Italian	validare invalidare cancellare	validate invalidate cancel

Figure 1: Bus Ticket Cancellation

d.	Kim	pchnela	Sande	z	nozem.
	Kim	thrust at/strike	Sandy ACC	with	knife INST.

Where English has a lexical item *stab*, French and Japanese offer the option of two similar but slightly differing constructions. The French option glosses as *give a knife blow* and the Japanese option as *give a stab wound.* French and Japanese here thus provide a clear case of conceptual mismatch. Whether there is mismatch in French and Japanese with respect to English depends on one's analysis of the lexical semantics of *stab.* One may decompose *stab* so that its meaning is represented as *give a stab wound*, as proposed in Dorr 1994.

A similar point may be made with the Polish construction, which glosses as *strike with a knife.* One may lexically decompose *strike* so that it means *give a blow*, in which case French and Polish may plausibly be analyzed as using constructions with the same meaning. We return to the issue of lexical decomposition and its interaction with mismatch in Section **??**.

3 Approach

The unifying theme in all our examples of translation mismatch is that an adequate or even perfect translation does not have the same meaning as the source. In some cases the reason for the mismatch is a lexical gap; in some cases it is the collocational requirements of the target language.

To accommodate both cases, we define translation mismatch as *the lack of a target exactly corresponding to the source and satisfying some given model of the target language.* The fact that mismatch exists is not surprising. The interesting point is that mismatch does not preclude adequate translation. Indeed, when all the nuances of lexical and grammatical meaning are taken into account, mismatch as we have defined it is so common that translation is only possible if mismatch can be coped with. The truth is that most translation involves some kind of compromise, some adjustment to the needs of the target involving the loss or gain of linguistic information. Our goal in this paper is to characterize precisely how this happens by laying out a framework for approximate translation.

We have two tasks:

- characterize when a target language model rules something out; and
- characterize what can be added in or left out and at what price.

To this end we will introduce the notion of the *cost of a target text with respect to a source.* Two factors will enter in to calculating the cost of a target:

- Fidelity: For this, we require some characterization of the information in the source text and the information in the candidate target and some way of measuring how close they are. Fidelity, then, has to do with satisfying *content constraints.*
- Fluency: closeness of fit of the candidate target to a model of the target language. What can appear in the target, whether it is added information or a faithful rendering of something in the source, will depend, for example, on the particular valence and collocational possibilities of the predicate it appears with. Fluency has to do with satisfying *language constraints.*

What we have called mismatch arises in precisely those cases where the two principles clash. Our goal is to investigate the workings of a system that tries to strike a balance between the two.

For fidelity we will assume that the content of source and target must be as close as possible, identical in the ideal case. In order to account for the cases where the contents are not identical, we will need a theory capturing the logical relations of different predicates and a forgiving variety of inferencing that will allow us to infer an acceptably fluent target from the source, while keeping track of the assumptions we needed to make to do so. We argue below that the right kind of forgiving inferencing is a variety of abduction and propose a framework called Translation by Abductive Proof (TAP).

For fluency, we will assume a statistical model of sortal restrictions of the sort (Resnik, proposal refs[Dagan, et al.]. Andry, et al.). and we show how the system of rules proposed here can be integrated with such models. Critically, the kind of language modeling assumed here is gathered from monolingual corpora. Rare bilingual corpora, with the heightened sparse data problems they carry along, are avoided.

Our strategy will be the following: in cases of conflict between content constraints and language constraints, search for contents *near* the content of the source that satisfy the language constraints. The idea will be that given a set of axioms, a theory describing semantic properties of source and target languages, a cost-based abductive inferencing scheme gives us a characterization of the distance between contents.

A TAP MT system can be classified as an interlingua system which uses limited inferencing to handle mismatch.[3] It is an interlingua system because it represents the meanings of both target and source in a single representation language (called an *interlingua*). In the rare case of exact translation target and source will have have the same interlingua representation. In cases of mismatch source and target will have distinct interlingua representations, reflecting a meaning difference in the translations.

Successful interlingua translation depends on the ability to compute a detailed analysis of the meaning of a source text. In this sense improved monolingual descriptions and improved computational techniques for analysis, especially statistical techniques, promise improved performance for interlingua systems.

There are two chief motivations for pursueing an interlingua approach. The first is specifically related to the problem of mismatch; in order to pursue an approach which measures fidelity and

[3] In Section **??**, where interlingua approaches are discussed in more detail, we will see that most interlingua systems make some provision for mismatch.

allows lapses in fidelity, some form of reasoning is required. But unless one wants to write distinct reasoning axioms for each language, that means source and target have to be translatable into some single representation that can be reasoned upon.[4] As we shall see this provides an important constraint on our interlingua but still allows a fair degree of flexibility in its design.

The second motivation for pursueing the interlingua approach is that it breaks the problem of translation down into relatively independent and manageable components. Every natural language processing task requires that some amount of linguistic structure be correctly described and correctly analyzed at runtime. The interlingua approach divides the translation process into two parts, deep analysis and deep generation, based on independent descriptions of the source and target languages. This amounts to a very desirable property we will call *target-source independence*. Consider an interlingua MT system with an analysis description S1' of a source language S1 and and a generation description T1' of a target T1. The usefulness of S1' and T1' is tested when a new target language T2 places new descriptive demands on S1. If S1' has to be entirely redone in the light of T2, then it was incorrect to to have called it a description of S1 in the first place. What S1' captured, at best, was a set of facts particular to the particular language pair S1 and T1. The converse argument can be made for the addition of a new source language S2. If T1' needs extensive overhaul in the light of S2, then it is not an independent target language description.

The interest of the claim of target-source independence thus rests on the extent of the overlap in descriptions of source language S1 from target to target and of descriptions of target language T1 from source to source. If the overlap is significant then the claim is interesting. If it is predictable—if we know exactly which part of a language's description is likely to need amendment when a new target is added—so much the better.

Proponents of interlingua sometimes assert the strongest possible version of target-source independence. Source description S1' is constant for all targets; target description T1' is constant for all sources. We argue below that this ideal is unattainable in systems that makes deal realistically with mismatch. Our point here is that even weaker versions of target-source independence are of interest. Partial autonomy of source and target descriptions—especially partial predictable autonomy—is sufficient. This is what the TAP approach offers.

4 Translation by Abductive Proof (TAP)

We begin with a brief example characterizing translation as a case of deduction. Given a set of axioms relating predicates in our interlingua, in particular axioms that license direct translation, we can view translation as a deduction that maps a source-oriented representation to a target-oriented representation.

We then present a variant of the same system that allows abductive inferencing, using it to analyze variants of the same example.

4.1 Translation by Deduction

Figure 2 shows the architecture of our deductive translation system in very abstract form. Parsing is assumed to produce an interlingua semantic representation. Our chief requirement for an interlingua is that it provide a representation on which inferencing is possible. All we need to need to satisfy this requirement is a disambiguated source semantics, that is, one in which distinct

[4] This is a point which is emphasize by [MCI] in their discussion of their DRT-based interlingua system.

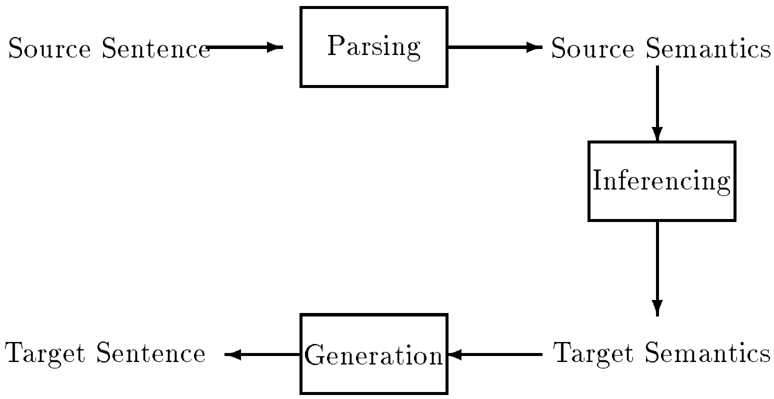

Figure 2: TAP Architecture

word senses have distinct interlingua predicates.[5]

The source semantics is fed into an inferencing component which produces a target semantics which can be fed to the generation module. The target semantics must be an interlingua representation which can be used to generate a target sentence. In the case of isomorphic translation where each word in the source has a direct translation in the target, the source and target semantics are identical.

The example we will begin with is (7), exhibiting a Japanese-English translation pair.

(7) a. John wa tokei o siteita.
b. John was wearing a watch.

Example (7) is very simple, yet on close examination it involves at least two cases of mismatch. First the Japanese word *tokei* corresponding in the output to *watch* actually has a more general word sense corresponding to either *watch* or *clock*. So the English translation requires specializing the Japanese word sense. The Japanese form *siteita* is actually the past progressive form of *suru*, a very general verb which can mean *do* or *make*, which takes the specific sense of *wearing-an-accessory* when its object is a clothing accessory.[6] Thus, translation requires generalizing this specific kind of wearing to the generic English clothing-and-accessory verb *wear*.

In (8), we have shown the simplified source and target semantics we will assume for this example.

[5] Other issues of disambiguation arise in this connection too, in particular the issue of scope disambiguation. It has been argued in a number of places that translation or meaning transfer should be defined on a scopally underspecified representation(CLE, MRS, refs). The principal motivation for this view is that scope ambiguities are often preserved in translation. Why, then, the argument goes, resolve them during analysis only to reintroduce them during generation? This argument carries a great deal of weight, but it needs to be balanced against the fact that fluent translation in general requires inferencing. Both these observations taken together mean that the right kind of scope neutral representation for MT is one on which inferencing can be soundly defined. At present, this is still an unsolved problem.

[6] There is another factor which enters in. To get the English translation *wear* the Japanese verb must be in the progressive. If the past perfective form is used, the correct translation would be *put on*. Thus the axioms shown below need to have an extra condition of progressivity placed on the Japanese predicates. For simplicity we have abstracted away from this interesting feature.

(8) John wa tokei o siteita

$$\downarrow$$

Parsing

$$\downarrow$$

$$\text{wear-accessory}(e) \wedge \text{wearer}(e, \text{j}) \wedge \text{worn}(e, x) \wedge \text{timepiece}(x)$$

$$\downarrow$$

Inference

$$\downarrow$$

$$\text{wear}(e) \wedge \text{wearer}(e, \text{j}) \wedge \text{worn}(e, x) \wedge \text{watch}(x)$$

$$\downarrow$$

Generation

$$\downarrow$$

John was wearing a watch

We now proceed to sketch the inferencing steps required to relate the source and target semantics. We will divide the axioms used into two bins *linguistic axioms* and *world knowledge* axioms. The distinction is a common one in linguistics. It corresponds roughly to the lexicon encyclopedia boundary. Not much will hinge computationally on this distinction, but a great deal hinges on it practically. An MT system that relies extensively on real world knowledge is one that will probably never get built. We argue below that the core knowledge in a TAP system is linguistic knowledge (of a very specific kind); and that the world knowledge, while helpful, can always be dispensed with.

The linguistic axioms needed to handle our simple example are shown in (9):

(9)
- a. $\text{timepiece}(x) \wedge \text{wearable}(x) \longleftrightarrow \text{watch}(x)$
- a'. $\text{wearable}(x) \rightarrow [\text{timepiece}(x) \longleftrightarrow \text{watch}(x)]$
- a". $\text{wearable}(x) \rightarrow [\text{timepiece}(x) \leadsto \text{watch}(x)]$
- b. $\text{wear-accessory}(e) \longleftrightarrow \text{wear}(e) \wedge \text{worn}(e, x) \wedge \text{apparel}_{\text{accessory}}(x)$
- b'. $\text{worn}(e, x) \wedge \text{apparel}_{\text{accessory}}(x) \rightarrow [\text{wear-accessory}(e) \leadsto \text{wear}(e)]$

The axioms that the lexicon writer would be expected to write are shown in (9a) and (9b). Consequences that will be discussed below are shown in the primed variants. We assume that Japanese *tokei* will be translated as the interlingua predicate *timepiece*, which is more general than *watch* and related to *watch* roughly as in axiom (9a). The relevant sense of *siteita* is translated as *wear-accessory*, which is more specific than *wear* and related to it as in axiom (9b).

The world knowledge axioms are shown in (10): **World-Knowledge Axioms**

(10)
- a. $\text{watch}(x) \rightarrow \text{apparel}_{\text{accessory}}(x)$
- b. $\text{worn}(e, x) \rightarrow \text{wearable}(x)$

Axiom (10a) simply categorizes a watch as an accessory, and axiom (10b) tells us that if something fills the *worn* role of a wearing event, as the watch does in (8), then it is wearable, which is the distinguishing characteristic of watches among timepieces in (9a).

The structure of the proof required for translation is shown in Figure 10. The Japanese oriented representation (the Source semantics) is shown at the top, and the new conjuncts in target semantics at the bottom. The arrows represent not logical consequence but simply proof dependence. So what is pictured is the network of logical dependencies invoked to derive the new conjuncts of the target semantics from the source semantics, given the axioms in (9) and (10).

The intuition motivating this treatment of the mismatches in (7) through logical inference is that source and target, *as a whole*, convey the same information. In isolation, the Japanese word

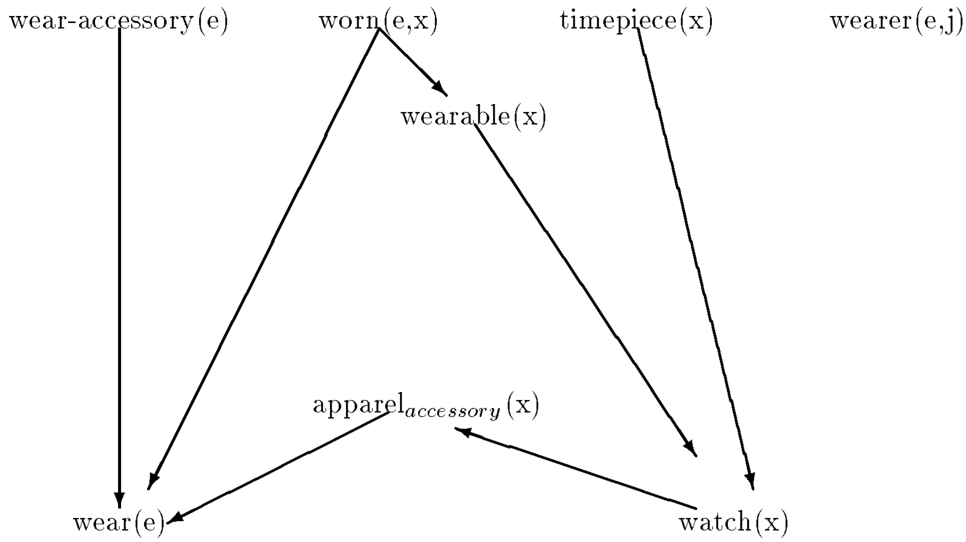

Figure 3: Proof Tree for example (7)

tokei can be translated either as *watch* or *clock*, and the word *siteita* encodes only a very specific kind of wearing; therefore translating *tokei* in isolation as *watch* risks adding misinformation, and translating *siteita* as *wear* risks losing information. Yet in (7) the linguistic context contains enough information to ensure that a watch is being talked about, and once it is clear that a watch is what is being worn, it follows that the kind of wearing at issue is accessory-wearing. Therefore, in this context, translating *watch* as *tokei* adds no misinformation and translating *siteita* as *wear* loses no information. Figure **??** shows the proof steps involved in establishing the informational equivalence of the source and target.

The key idea in characterizing translation as deduction follows Rayner 1993. We assume that what is required for exact translation is a proof that some expression composed entirely of target-oriented predicates is equivalent to the source. In Rayner's work such proofs always ground out in the kind of linguistic knowledge axioms illuistrated in (9), equivalence axioms relating source and target oriented predicates.

There are some serious practical objections that can be raised to the idea of translation based on provable equivalence which will be addressed in the next section. But one objection can be dispensed with quickly. One might worry that *logical* equivalence is too weak a notion to be the basis for translation. For example, in many modal systems with intuitive theories of the notion of *three sided square* and *four-sided triangle*, these predicates will have the same extensions in all possible worlds, the empty extension, and thus *three-sided-square*(x) and *four-sided-triangle*(x) will be provably equivalent in all possible worlds. But we don't want to translate an expression meaning *three-sided square* with one meaning *four-sided triangle*.

This is correct. But in the TAP system sketched below *translation equivalence rules* like (9a) are distinct from other kinds of axiomatic knowledge, and all candidate translations are proposed because some translation equivalence rule licenses that proposal. Thus, *three-sided-square*(x) will not be proved equivalent to *four-sided-triangle*(x) unless these two goals can be connected by translation equivalence rules.

In a TAP architecture, there is no reason to identify the kind of equivalence encoded in translation equivalence axioms and (9b) with equivalence in all possible worlds or even with some (possibly) stronger notion like logical equivalence. For our purposes, translation equivalence needs to be some equivalence relation *stronger than* logical equivalence. That is, we need to be able to assume that if A and B are translationally equivalent, then A entails B and B entails A.

4.2 Translation by Abduction

Example (7) shows how a very simple variety of inferencing can help resolve translation mismatches. But even this very simple example has some worrisome properties. First, (7) is obviously a case in which local linguistic context contains *exactly* the information to guarantee informational equivalence. The local linguistic context had to be exactly right, and exactly the right axioms had to be available to provide the connections in Figure **??**. In particular, the predicate *siteita* happens to decide the issue between clocks and watches. But often the local linguistic context will lack such information, and there will be no translation that is guaranteed to be information-preserving. Example (11), for instance, provides a case in which neither *clock* or *watch* is completely justified as a translation of *tokei*, or putting it another way, in which either *clock* or *watch* may be the correct translation, depending on context.

(11) a. teeburo no ue ni tokei ga arimasu.
b. There was a watch on the table.
c. There was a clock on the table.

More troubling perhaps, even when an inference can be made, the information necessary to make the inference is often extremely subtle.

(12) a. There was a clock on the wall.
b. I asked the conductor what time it was. He took out his watch and told me it was three.

Something attached to a wall is probably a clock rather than a watch. Something that is "taken out" and looked at on a train or in a train station is probably a watch rather than a clock. To know either of these things is to know something about the way things usually are in the world. It is not linguistic knowledge in any conventional sense of the term. Nor is it something easily captured by classical axioms. Common sense default reasoning is involved. It is a reasonable assumption for a translator to make to choose *clock* in the context of (a) and em watch in the context of (b).

However it was implemented, it is difficult to imagine a practical MT system representing and using the kind of information needed to handle examples like those in (12) by reasoning.

Thus there are major problems for an inferencing-based approach:

- Sometimes the resolving information simply isn't available.
- Sometimes the resolution involves common sense or default reasoning rather than classical deduction.
- Knowing what needs to be known to do common-sense reasoning involves a vast amount of knowledge.

We propose to deal with all of these issues by augmenting the deductive system described in the last section with *abduction*. The primary benefit is that abduction is a kind of reasoning that can be performed with partial or incomplete information.

We first present some background on abduction. Then we return to the issue of reasoning with partial information and show how a general abductive mechanism helps. Then we show how abduction helps in the specific case of the kind of missing information involved in translation mismatch.

Abudction is often loosely characterized as follows:

Abduction

Abduction is inference to the best explanation. From Q and $P \longrightarrow Q$, infer P.

Of course this step is not logically valid. There are in general many explanations for any Q, in general many P one could assume as a provisional explanation. But what is valid is a conditional proof. That is:

$$P \longrightarrow [\,[P \longrightarrow Q] \longrightarrow Q]$$

That is, if we have P, then $P \longrightarrow Q$ entails Q. The computational consequence of this is that an ordinary theorem prover may be turned into an abductive theorem prover simply by allowing it to make and keep track of assumptions. Where the deductive theorem prover returns a proof, the abductive theorem prover returns a proof together with the assumptions necessary to make it go through.

Of course once the power to make assumptions is granted, anything can be proved, since at the limit, any proof goal is proven by being assumed; so what abduction requires to be practical is some criteria for selecting from all the candidate Ps, some standards according to which "best guesses" are made.

Some classic criteria for selecting P (from Kay, et al. 1994) are:

a. **Consistency:** P should be consistent with what is already known (and assumed).

b. **Simplicity:** P should be as specific as possible. (Thagard 1978)

c. **Consilience:** Explain as many Qs as possible. (Thagard 1978)

The standards can be illustrated through the paradigm example of medical diagnosis. The doctor seeks by the observation of symptoms to come up with a best guess as to the disease. The Ps are candidate diseases, the Qs are symptoms. *Consistency* requires that the candidate disease have effects consistent with observed symptoms. If blood pressure is normal and some candidate disease P entails elevated blood pressure, P is, all things being equal, out of the running. *Simplicity* expresses the requirement that a specific disease is sought, even though a general characterization such as kidney malfunction, may account for all observed symptoms. *Consilience* says that a diagnosis that accounts for all the observed symptoms is better than one that accounts for only half.

Assumption consistency is guaranteed by attempting to prove $\neg$P before assuming P. Working abduction systems typically implement one or both of the other assumption heuristics through the device of assumption costs. Some finite set of goals is declared assumable and for each assumable goal a cost is assigned. Costs are sometimes assigned by an arbitrary assumption scheme (Hobbs, et al. 1988). They are sometimes interpreted as a probabilities (Charniak and Goldman 1988).

Some conceptual difficulties in interpreting abduction costs as probabilities, specifically within the context of using abduction for text interpretation, are discussed in Norvig and Wilensky (1990). The use of abduction in disambiguation is discussed in Kay et al. (1990).

We will assume the following:

(13) a. Only literals [atomic formulae and their negations] declared to be assumable are assumable.
b. Everything declared to be assumable is assigned an *assumption cost* (abductive proofs have costs).
c. Proofs are charged for their assumptions and their length.(longer proofs are more costly than shorter proofs).

All things being equal, practical consilience will be guaranteed by (13c). If the costs of assumptions generally all fall in the same neighborhood, the cheapest proof will be the one that makes the fewest assumptions, so an assumption that helps prove more than one goal will, all things being equal, come out ahead. Specificity is guaranteed by adopting, as a principle of cost assignment:

$$[\mathrm{P} \longrightarrow \mathrm{Q}] \longrightarrow \mathrm{Cost}(\mathrm{Q}) \geq \mathrm{Cost}(\mathrm{P})$$

Note that this principle is inconsistent with a probability interpretation.

The cost of a translation is a measure of its *fidelity*. The more costly the assumptions, the less faithful the translation.

We adopt the following heuristics for translation cost.

(14) a. All things being equal, combine information into a single word whenever possible.
b. Grammatically or paradigmatically enforced distinctions are assumable cheaply. There is a good chance they are being encoded because the language requires it and not because the speaker's communicative goals do. [examples: number and pronoun gender in English, classifier information and distinction between wearing verbs in Japanese]

If we assume that every axiom used adds something to a proof's cost, then heuristic (14a) will generally be satisfied. A proof that fires one axiom that consumes two source language predicates will be favored over one that consumes those predicates with two axioms.

A key point about the abduction framework is that losing and gaining information in translation are treated symmetrically. In Section 4.1, we presented a deductive translation of (7). Consider now a hypothetical abductive translation system that attempts the same translation without the benefit of axiom (10a), the world knowledge axiom that captured the fact that a watch is a kind of accessory. The proof given in section 4.1 no longer goes through, but an alternative abductive proof in which accessoryhood is assumed is possible. This is shown in (15), where the assumption is given a cost of \$1.

(15)

$$\mathrm{wear\text{-}accessory}(e) \wedge \mathrm{wearer}(e, \mathrm{j}) \wedge \mathrm{worn}(e, x) \wedge \mathrm{timepiece}(x)$$

$$\downarrow$$

$$\text{Inference}$$

$$\downarrow$$

$$\mathrm{apparel}_{\mathrm{accessory}}(x)[\$1] \rightarrow [\mathrm{wear}(e) \wedge \mathrm{wearer}(e, \mathrm{j}) \wedge \mathrm{worn}(e, x) \wedge \mathrm{watch}(x)]$$

$$\downarrow$$

$$\text{Generation}$$

$$\downarrow$$

$$\mathrm{apparel}_{\mathrm{accessory}}(x)[\$1] \rightarrow [\text{John was wearing a watch}_x]$$

This yields the proof tree shown in Figure 15, where the assumed step is enclosed in a box. The proof involved losing information because the assumption allows a specific Japanese predicate *wear-accessory* to be transalted by a more general English predicate.

Similarly, an abductive MT system could derive the translation of (7) without the benefit of axiom (10b), the world knowledge axiom that captured the relation between being worn and being wearable with an abductive proof in which wearability is assumed. This is shown in (16), where the assumption is given a cost of \$1.

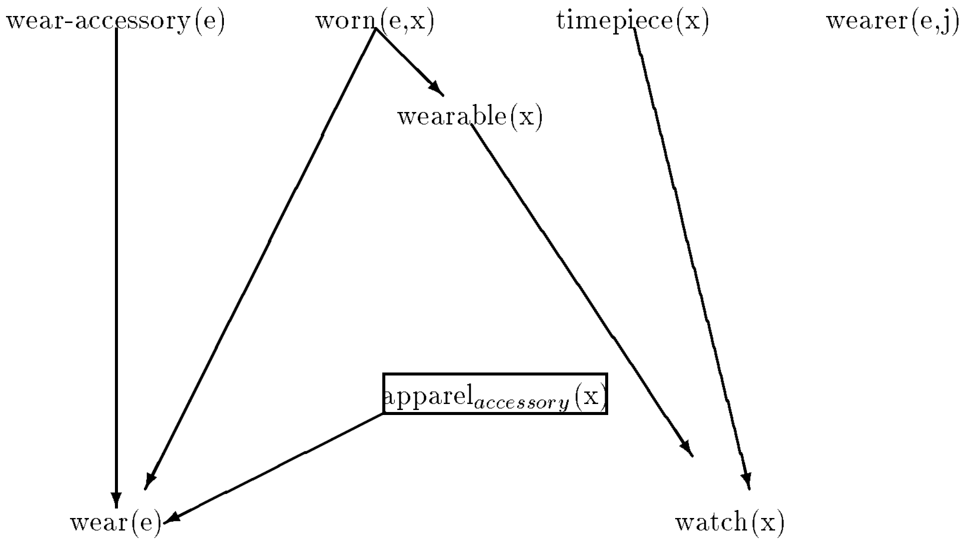

Figure 4: Proof Tree from Figure 10 without axiom (10a)

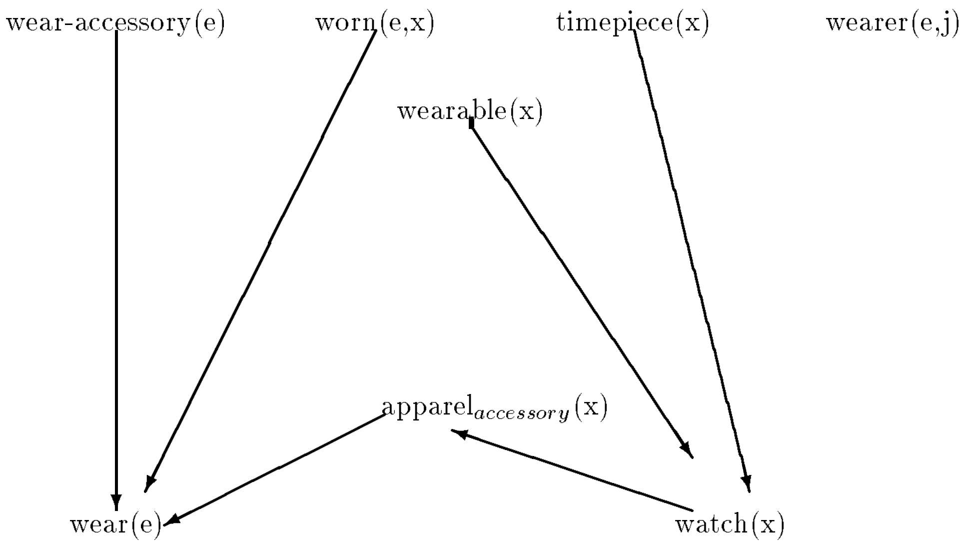

Figure 5: Proof Tree from Figure 10 without axiom (10b)

(16)

$$\text{wear-accessory}(e) \wedge \text{wearer}(e, \text{j}) \wedge \text{worn}(e, x) \wedge \text{timepiece}(x)$$
$$\downarrow$$
$$\text{Inference}$$
$$\downarrow$$
$$\text{wearable}(x)[\$1] \rightarrow [\text{wear}(e) \wedge \text{wearer}(e, \text{j}) \wedge \text{worn}(e, x) \wedge \text{watch}(x)]$$
$$\downarrow$$
$$\text{Generation}$$
$$\downarrow$$
$$\text{wearable}(x)[\$1] \rightarrow [\text{John was wearing a watch}_x]$$

This assumption allows the general predicate *timepiece* to be translated by the more specific predicate *watch*, so information is gained. The proof tree is shown in Figure 16.

5 Semantic Decomposition

The fact that word-for-word translation is so rare means that translation systems must map between distinct word senses in source and target. For interlingua systems the device that has played the central role in this mapping has been *lexical decomposition*.

The standard picture is that IL makes a set of core semantic distinctions to which the word senses of all languages can be reduced. Once the source is decomposed into its IL *components*, the parts can be recombined, possibly in different ways, to give the word senses of the target language.

Typically concrete proposals for IL have not made the strong claim that that there is a single IL representation that can capture all the fine lexical distinctions of all languages. Rather they have claimed that there are important dimensions along which meanings can be decomposed, and that, relative to any target language, each IL representation captures an equivalence class of meanings. This means that, given an IL representation, realization in the target requires a non-trivial step of lexical choice.

An example of this view of IL is the view of Lexical Conceptual structure taken by Jackendoff (Jackendoff 1983, 1990) and pursued for MT by Dorr (Dorr, 1983, 1994). In principle, LCS's are only intended to capture those aspects of meaning that have syntactic consequence (although Jackendoff's notion of syntactic consequence is fairly fine-grained); so there are going to be distinctions, for example, those between natural kind terms like *dog* and *horse*, which are not represented in LCS, or which are represented trivially, by some unanalyzed distinguisher.[7]

There are numerous cases where some kind of decomposition of semantic components seems indispensable.

1. Aspect and tense
2. Singular/plural and mass/count
3. Genericity
4. Quantifiers, partitives, classifiers

All of these cases share the property that they take us outside basic predicate/argument structure for open class predicates. Often these components of meaning come fused with lexical predicates in complex ways; and standard semantic analyses do factor these parts of meaning out.

There are a number of other cases where special syntactic circumstances take us beyond an intuitive notion of word senses. Special constructions that add argument structure, discussed in Goldberg (1995), among others, are an obvious case:[8]

(17) a. My father frowned away the compliment and the insult.
b. Sharon was exactly the sort of person who'd intimidate him into a panic.
c. I cannot inhabit his mind nor even imagine my way through the dark labyrinth of its distortion.
d. Pauline smiled her thanks.
e. The truck rumbled down the street.

These examples all introduce extra arguments as well as semantics beyond the basic word senses of the verbs involved. Examples (a) and (b) are basically resultative: The father causes the insult

[7] Compare the Katz/Fodor notion of distinguisher. Nirenberg (1988) provides a fairly careful discussion of the lexical choice problem for an interlingua system.

[8] Goldberg (1995), p. 55.

to go away by frowning, Sharon causes the narrator to go into a panic by intimidation. Example (c) is a complex blend of metaphorical motion and the basic verb sense. To imagine one's way into something is to proceed toward conceiving/understanding it by imagination. Examples (d) and (e) (attributed by Goldberg to Levin and Rappaport) involve meaning blends. The manner of thanking is by smiling. The truck moved down the street while rumbling. Many of these generate word salad when translated directly, and can often be best translated using more than one clause. Thus breaking the meaning apart into the pieces that are blended in the constructions seems essential.

The patterns exhibited in (17) are generalizable and merit some non-idiosyncratic treatment. For example, verbs of sound emission like *rumble* can regularly be used to convey motion:

(18) a. The man with the wooden leg clumped into the room.
b. The train screeched into the station.
c. The fly buzzed out of the window.

And the *way* construction of (17c) can impose metaphorical or real motion, changing argument structure, in a variety of contexts:

(19) a. I knitted my way across the Atlantic.
b. Anyone watching would have thought he was scowling his way along the fiction shelves in pursuit of a book.

In general any construction which adds something to what might be considered the basic sense of a verb offers a good motivation for decomposing the resulting semantics into a basic sense plus added meaning. Goldberg's most productive example of this sort is what she calls the Caused-motion construction.

(20) a. Sam pushed him within arm's length of the grenade.
b. Mary urged Bill into the house.
c. Mary hit the ball over the fence.

In each case the action described by the main verb can be inferred to cause a change of motion, resulting in the referent of the direct object being in the location described by the prepositional phrase, which may constrain either the goal, as in (a) and (b), or the path as in (c). Of particular relevance for translation is the fact that, for many languages, these constructions do not translate into a single clause. The meaning must be taken apart into cause and resultant motion and rendered in two clauses. For example, for Japanese:

Thus far we have argued two kinds of "decomposition" of meaning beyond what comes from word senses: grammatical meaning and constructional meaning.

But these cases can be handled Kwithout decomposotion of word senses. Are there cases where decomposition of word senses is justified? Dorr (1983, 1994) argues that there are and presents a number of compelling examples. One kind of case is illustrated by the following French/English pairs:

(21) a. Pierre hit Jean.
b. Pierre a donné un coup à Jean.
c. Pierre punched Jean.
d. Pierre a donné un coup de poingà Jean.

In bopth cases the French may be seen as making explicit the correct decompotion of the English. Hitting is the giving of a blow, with the instrument unspecified; punching is the giving of a blow with the fist. This rendering makes translation easy (the French makes explicit the interlingua semantics) and it seems to capture the relation

Mismatch TWO:

But talk about specific situations may be mixed up with “general purpose” talk in complicated ways. M. Kay’s punching ticket example. entwerten.

Our point here ito emphasize that the case of *entwerten* used to describe ticket-cancellation is not that different from the case of *tokei*. Used of tickets, *entwerten* describes a particular kind of social act, linked to rather complicated conventions. Yet *entwerten* is a perfectly real word with a perefectly good meaning. Presumably one that could be decomposed. [[CITE VM quote]]

Thus the kind of word meaning that may be decomposed may also play a role in describing a particular kind of social or real world category. And in general neither the decomposition, nor the word sense itself, is sufficient to predict the categories of things it may refer to. The two kinds of semantic phenomena are intertwined. Either may affect how something is translated.

Blood sugar.

My homework =¿ mes devoirs.

Partnership ↔ kyoroku kankei

[other phrase closer to social/legal fact meaning: teikei kankei]

One might claim that *partnership* has been insufficently decomposed, that it really should be rendered as *collaboration relationship* in interlingua. But Japanese offers other ways of expressing partnership, for example, ??, *cooperation relationship*.[9] What privileges one of these ways over the other? [Move: The question is essentially the same one we were asking in section ??. What privileges one way of conceptualizing scenes over another?]

Again the idea of simultaneously valid descriptions arises. The words used to pick out social kind have an independent semantics, but different descriptions may pick out the same social fact, described with one word in English. The Japanese description is compositional in the trivial sense that more than word is involved, and perhaps in the deeper sense that the words involved may be decomposed into even deeper units, but that decomposition is irrelevant to the translation here, as it was in the entwerten and blood sugar cases. In all three cases what was relevant instead was what we will call the *referential* function of the language used, what kind of thing —whether event or object—was being picked out in the world. We will say that in these cases what translation preserves is *reference*.

We have defined translation mismatches as cases of valid translation where the contents of source and target are not the same. In the system described below source and target have the same content if the proof of their equivalence has no undischarged assumptions. It will turn out that cases of what we call reference preservation are sometimes cases of mismatch (assumptions will be required) and sometimes not.

The case of *kyoroku kankei* is a real case of mismatch. The phrase *kyoroku kanakei* may be used to describe relationships that are not partnerships, and there are partnerships (for example, silent partnerships that are basically just capital investments with certain entailed ownership rights) that are not the sort of collaboration relationships described by *kyoroku kankei*. To use this *partner* as the English rendering of the Japanese phrase thus entails making assumptions. [other possibility]

On the other hand, whenever translation is exact, the reference function is of course preserved. This holds even when the pieces are not assembled isomorphically. Thus, for example,

[9] Discuss why Japanese words for cooperation and collaboration are not synonyms. Rather the difference is neutralized in this context, because both provide ways of describing the same social kind.

the translation of the compound *bed bug* as French *punaise* preserves the reference to a particular kind of bug and appears to be exact.

The important point here is that when when fidelity can't be preserved along with fluency, sometimes preserving reference preserves the most essential thing. Although preserving content preserves reference, the converse is not true, and thus translations that preserve reference may involve mismatch.

We can roughly capture the notion of preserving reference within our logical framework by defining a class of interlingua predicates which capture situation or object classes. Call this class SIT. We assume universal SIT includes one predicate for every noun and verb sense in every language. Of course numerous senses will be shared by multiple languages. But SIT will also include other predicates as well. For example, we propose to analyze the case of *entwerten* as follows:

(22) [a]$Germanentwertenaxiom$
[b]$frenchentwertenaxiom$

What is striking here is that we are proposing [Concept Name] as a predicate of SIT when [Concept Name] is not the sense of any word in any language.

What is the role of [Concept Name], if not to capture a word sense? The answer is quite simply that it captures a concept that plays a role across several languages. We could replace the two axioms in (22) by a single axiom that directly relates the French and German predicates. But this solution would not preserve target source independence. We would need to add two new axioms when we added [the third language]. Using [Concept Name] we only need to add one new axiom. Without [Concept Name] we need [N choose 2] axioms to handle N languages. With it, we need N. The same target-source independence argument that motivates interlingua in general motivates [Concept Name].

What this shows is that interlingua benefits not just from abstract sub-lexical concepts like those in CS. It also benefits from having very specific concepts for very specific kinds of situations and things in the world. Essentially, whatever kind of reference may be conventionalized will motivate an interlingua concept.

We place no particular requirements on the relationship between SIT and CS, the set of conceptual structure predicates discussed in Section 6. In particular we make no claim that they need to be disjoint. Thus, SIT, which includes all word senses, may include all or some elements of CS. We discuss the special place of CS axioms in such a system in Section **??**.

A larger and more difficult class of cases is where translation merely preserves communicative goal. These cases may be thought to pose a challenge to a logic-based theory of translation, since there is no apparent logical relation holding between the source and target.

A very clear case of this is the translation of formulas. When one translates *once upon a time* with *il etait une fois* one is aligning two expressions that serve the same function, that of starting a story. It seems to be unnecessary to go beyond that to claim that meaning is preserved as well. Similarly the translation of a formula from a wedding ceremony or an oath may be best rendered by choosing an equivalent social formula in an equivalent social context in the target culture. Finally there is the well-known case of proverbs and idioms. One may sometimes best translate a proverb or an idiom by a target expression that expresses the same moral or concept.

To capture the equivalence of formulae like *once upon a time* and *il etait une fois* we would need a conditional equivalence. Roughly, the conditions would be:

$$\text{discourse-genre}(\text{tale}) \wedge phase(\text{opening}) \Longrightarrow \text{`Il etait une fois'}(e) \leftrightarrow \text{`Once upon a time'}(e)$$

We return to the important subject of conditional equivalence in Section ??.

5.1 Reference and Apparent Cases of Mismatch

We turn in this section to some more elaborate cases that may usefully be analyzed as reference preservation.

contribute to the capital of participate (capitalwise) in =¿ participate in a joint venture with

We propose to treat these cases just like the cases of natural and conventional kinds we discussed in Section **??**. That is, we map through SIT. This again requires that SIT include predicates that are more specific than word-senses usually are taken to be:

(i) contribute to the capital of participate (capitalwise) in ¡= joint-venture-partnership$_{SIT}$

(ii) joint-venture-partnership$_{SIT}$ ¡= participate in a joint venture with

Thus we treat this as completely parallel to the *entwerten* case. Note that we do not write something more like a conventional semantic transfer rule (iii):

(iii) contribute to the capital of participate (capitalwise) in ¡= participate in a joint venture with

The idea is to avoid language-pair particular statements like (iii), prefering a treatment consistent with the methodology that motivated interlingua in the first place. Of course direct mappings like (iii) have certain efficiency advantages. We show in Section 9 how rules like (iii) are derived from rules (i) and (ii)

Where does this leave us with respect to source-target independence? The claim is that rules like (i) are part of the description of the source language. They interpret source language predicates wrt to interlingua. Rules like (ii) are part of the description of the target.

Often the fragility of the translation process can be observed by considering a set of close parapahrases in the target and observing the contrasting results of piece-by-piece in the target.

(23) a. Stab John in the heart
b. stab John through the heart

Only one has a "direct" correspondent in French.

donner qn un coup de couteau dans le coeur

Very few, if any, bilingual dictionaries will propose a translation relation between *dans* and *through*.

Again rather than a description that directly relates *dans* and *through* here, in effect positing a special meaning for *dans*, we will propose that the two English sentences be recognized as two different ways of describing the same kind of situation. The idea is to capture the translation facts while stopping short of the claim that the two English sentences mean the same thing.

$$\text{in}_{\text{E}}((E;[STABBING]),BP) \leftarrow \text{body-part}_{\text{SIT}}((E;[STABBING]),BP)$$
$$\text{through}_{\text{E}}((E;[STABBING]),BP) \leftarrow \text{body-part}_{\text{SIT}}((E;[STABBING]),BP)$$

Reminding reader of Dorr's analysis of stab. But Japanese:

(yy) Donner un coup de couteau ¡=¿ give a knife wound

Oops but it isn't by giving a knife-wound. It's by giving a knife BLOW. A perfectly reasonable alternative conceptualization of the same kind of stabbing situation. The supreme difficulty of this problem. Decomposition of the action. Seeing that scenes in which a knife wound is given can also be grasped as scenes in which a blow is given with the knife.

How this could be done:

```
donner un coup de couteau
   |
   [a]
```

```
    |
GIVE KNIFE-BLOW
  |
  [b]
  |
STAB
  |
  [c]
  |
GIVE STAB-WOUND
    |
    [d]
    |
[give a stab wound]
```

Imagine the links are bidirectional. Links (a) and (b) would be motivated in a French-English system, relating the giving of a knife blow to a stabbing situation. Links (c) and (d) would be motivated in an English-Japanese MT system, relating the giving of a knife wound to a stabbing situation. But now, given the links are bidirectional, we have the information needed for the French Japanese system, linking the giving of a stabbing wound to the giving of a knife blow. We give more explicit rules in Section 8, but clearly what is going on when all the axioms are invoked is that this translation is represented as a case of reference-preservation. The source and target languages have chosen slightly different ways of looking at the same event type.

The striking thing about this picture is that our interlingua system as a whole relates the SIT predicate STAB to two distinct decompositions. This example shows how the kind of decomposition allowed on our logic-based approach differs from that of a classical interlingua system, where decomposition is unique per word sense and may therefore profitably be a lexical property. In this system a single sense of the English verb *stab* has been associated with two distinct decompositions. The idea was already implicit in our treatment of *entwerten*. STAB is just another SIT predicate, and the same situation may be looked at in different ways in different languages (or indeed, in a single language).

A third example with "stab".

John was stabbed in the heart.

Arguably

John a recu a coup de couteau dans le coeur

is better than:

John a ete donne un coup de couteau dans le coeur

Here to get from English *stab* to French *donner un coup de couteau* some amount of decomposition is required. But that decomposition won't directly get us to *recevoir*. LCS includes (with ample motivation) a CAUSE predication in the analysis of stab, which (again with excellent motivation) is lost in the representation of recevoir (in fact, give is analyzed as the causative of recive). So we can't directly match a stabbing event to receiving a stab blow.

[Note for later: Abduction: causality easily assumed. Leave it at low cost. A leaving-out axiom needed. An issue about how such "mixed" predicates are represented. Coindexing of eventvars between CS and FL reps.]

Summarize: We have argued that a significant number of translation mismatches preserve reference. In these cases semantic analysis needs to provide some account of the kinds of objects or situations the predicate can be used to described. In Section 6, we saw, on the other hand, a number of cases where decomposition is useful. The challenge is to find a descriptive framework that provides both kinds of information.

6 Interlingua and Semantic Decomposition

The basic commitment in an interlingua system is to a single representation in which the semantics of all languages can be rendered. But this leaves some important questions unanswered. In particular, once it is acknowledged that translation is almost never word-to-word, the following question becomes urgent:

> How will the relation between different word senses be captured?

Every interlingua system (and even some transfer systems) has attempted to answer this question, in part, through a strategy of semantic decomposition. Word meanings are decomposed into components which can be combined in some highly restricted ways. The relationships between word meanings are then accounted for by the components they share and differ in.

Certain kinds of "decomposition" of the semantics of words are standard. For example, English word meanings include information such as tense and number which are standardly separated out. English *went* and *goes* differ in a tense component and share a root lexical meaning. But once we move beyond meaningful inflectional morphology, the ground is less steady.

When translation is one-to-one decomposition is unnecessary. As noted in works like (**?**) and (**?**), the virtues of decomposition emerge when the meaning of a single lexical item in the source is distributed over several words in the target or when several words in the source make contributions to a single word in the target. Dorr acknowledges that the occurrence of such configurations is language-pair particular, but she argues that doing decomposition in every case acheives maximal source-target independence. On this basis, Dorr proposes an interlingua system in which meanings are decomposed into a specific universal semantic representation called Lexical Conceptual Structure (LCS) in the lexicon. We will call this kind of lexically-based interlingua system with uniform decomposition a classical interlingua system.

In essence, without necessarily endorsing all the details of LCS, we accept the premise of Dorr and others that decomposition needs to be a stable part of a language's semantic description. But our picture of interlingua will depart in certain respects from that of a classic interlingua system. First we will assume that a complete interlingua contains distinct predicates for every word sense in every language. What captures the relations between the predicates is a set of axioms we call the interlingua theory. Thus, an interlingua is useless without the accompanying theory.

There are two chief motivations for not assigning decompositions in the lexicon. First, we will propose an algorithm for abduction on which the cheapest proofs are found first. Proof cost will roughly correlate with proof length. In cases where senses can be mapped one-to-one, the shortest proof will bypass decomposition and simply use the same word sense for source and target. In cases where only a single conceptual predicate needs to be separated off —for example, causatives when the target has a causative morpheme —the shortest proof will do partial decomposition. Thus if decomposition is deferred, it is possible to arrange things so that decomposition is done only as needed.

Second, this way of setting things up does not commit us to a unique decomposition for a word sense. We argue below that different languages, or even different constructions in a single language, may factor a word sense into different incompatible parts.

The first idea, lazy decomposition, poses no particular theoretical challenge to the kind of LCS-based account of lexical semantic meaning outlined in Jackendoff (**?**), the theory on which

Dorr's system is based. As far as the theory goes, what we give here is a simple notational variant in which we have factored lexical semantic description into two pieces, word sense assignment and sense decomposition.

The second idea, that a single sense may have more than one decomposition, is potentially a major weakening of the theory. The question is whether "decomposition" has the same meaning for us here as it does in LCS, where it plays a specific theoretical role. Putting this another way: We represent LCS decompositions by logical axioms licensing replacing a single predicate by a collection of predicates. But there are many such axioms possible, not all them proposing the same "decompositions" as those licensed by LCSs. Even if Jackendoff's theory is completely right, it may be that some non-LCS "decomposition" is useful for translation. We argue that there are such cases, requiring distinct "decompositions" of the same sense.

There has within theoretical linguistics been some tendency to oppose a logical-based axiomatic approach to lexical semantics to all decompositional approaches, including approaches employing deep case or thematic roles, which attempt to capture only a small set of lexical semantic relations. Logic-based approaches may be typified by the work of Dowty, for example Dowty (**?**), in which the lexical facts attributable to thematic roles are instead accounted for by entailment sets. The idea is that a notion like *actor* or *agent* is really a cluster of entailment properties, only a portion of which will hold for any given predicate. Note that whether a thematic role is notated by a logical predicate is not important. Dowty's objections will apply to any Davidsonian acount that uses predicates like *agent* and *patient* to distinguish roles. This is a specific challenge to a specific decompositional idea.

But there is no fundamental incompatibility between decomposition and an axiom-based approach. In principle, as Jackendoff himself notes (**?**), any decompositional account can be reformulated in logical axioms. And presumably a decompositionalist like Jackendoff will admit that his LCS representations are not intended to capture all entailment relations,[10] and that therefore some auxiliary "logical theory" of lexical relations may be required.[11].

Do "stab" example here.

Throughout this paper we will look at interlingua analysis rules as part of a unified 'theory' (in the logician's sense) relating source and target language predicates to interlingua and interlingua predicates to each other. A strict separation will be maintained between source and target lnaguage predicates and interlingua predicates, even in the case where an interlingua predicate is exactly equivalent to a source language predicate.[12] This not only avoids confusion. It also allows to layer our axioms in a way that promotes efficiency, as described in Section 8.

CAUSE GO TOWARD

Note that GO parameterizes in two ways in LCD, possession movement and spatial movement. Presumably in stab it is specialized toward spatial movement.

Write axiom. English to French direction.

Define a special set of Conceptual structure predicates we call CS. No commitment to any "pure" CS representation of entire meaning. None in Jackendoff either. We return to this issue in the next section.

[10]That decompositions can not capture all entailment relations is easy to show, because of the existence of mutually definable pairs like *require* and *forbid*. One may choose one as primitive, say, *forbid* and define *require* as FORBID NOT, but then to capture the fact that *require not to go* entails *forbid to go* one needs at least enough logic to derive FORBID from FORBID NOT NOT.

[11]There does appear to be one minor theoretical advantage to a predicate-based approach

[12]The subscript IL will be used to distinguish interlingua predicates.

7 Comparison to Transfer Approach

The need for language-pair specific information. Along with this the difficulty that writing a covering set of transfer rules appears to require re-stating the structure of source and target, independently defined in source and target grammars.

One could argue that the MRM architecture described below is really a semantic transfer system. On this view the MRM is really a semantic transfer module to which the analysis modules provides the input and from which the generation module accepts the output. At some level this view is formally correct. There are n properties which distinguish the MRM architecture from a semantic transfer system

- "Transfer" is resorted to only in cases where there is some kind of mismatch.
- "Transfer" is on a representation common to all languages.
- "Transfer" is semantically motivated. It is licensed by the closeness of the contents of source and target structures.

It is this last feature which only some form of inferencing can guarantee, which in turn motivates the use of a common representation.

It is also the last feature which reformulates the problem. The problem of valid transfer is seen to depend on deep semantic descriptions of the source and target, resulting in reduction to som e common representation language. This is the way that interlingua systems factor the problem. The central descriptive goal is to minimize the amount of description of facts particular to a pair of languages. The rest of this paper is devoted to showing how that can be done.

Nevertheless the fact that the MRM can be viewed as a transfer module is important. In Section 9, we show how to compile MRM ineterpretation and generation axioms linked to source and target language predicates into a transfer system bypassing IL.

8 Choosing an Abduction Model

We assume a interlingua theory which has axioms which capture the relations among word senses. In certain cases source and target use exactly the same senses and a single interlingua representation serves both as source analysis and target semantics. The picture in Vauqouis's translation triangle is born out. In all other cases an equivalence proof must be found between two interlingua translations.

The problem as we have seen is that in many important cases the source has no equivalent fluent target representation. In such cases no unconditional equivalence proof can be found; instead what we desire is a proof of equivalence with undischarged assumptions. Moreover we desire a "good" proof, a proof that gives a fluent result with maximally likely assumptions. These criteria do not fix a particular notion of best translation, because they involve maximizing two criteria, but they give us a framework within which to explore trade-offs.

The general class of theorem-proving approaches which fit these requirements are called *abductive*. An abductive theroem prover is allowed to return a proof together with a list of the assumptions needed to make the proof valid. Moreover, because allowing abductive proofs allows anything to be proved (if necessary, the goal itself can be assumed), abductive theorem provers typically employ a search strategy amenable to bounding (such as iterative deepening) and some cost scheme to control the search.

Abduction is sometimes called inference to the best explanation. The idea is that given some set of principles and a phenomenon to be explained, an explanation consists of an application of the principles that derives the phenomenon, together with some set of 'reasonable" assumptions to complete the derivation. In the simplest case, given 'p implies q', we can explain 'q' by assuming 'p'. In the next simplest, given 'p1 implies q', 'p2 implies q'..., 'pn implies q', we explain 'q ' best by assuming the least costly of the ps.

For the translation problem, we will assume that source, target, and interlingua predicates are all disjoint. We will also assume an interlingua theory expressed in a set of equivalence axioms that relate interlingua expressions to either source or target expressions. Finally, we will need a pre-defined set of assumable interlingua predicates, with every combination of assumable predicates and arguments assigned an assumability cost. We will call the total; cost of the interlingua predicates assumed the *fidelity cost*.

In trying to perform translation by abduction. the explanation sought must come in the form of a target language expression. This actually simplifies the task because it defines a set of ground facts, realization facts, which require no further explanation. For our purposes a realization fact will simply be an atomic formula of the target language semantics.

Ground steps in a proof may either be facts in a knowledge base or they may be assumed. We will require that all realization facts be assumed, with their assumability costs determined by a target language model, as explained below. We will call the total cost of the target language expressions assumed the *fluency cost*.

This leaves two large problems for the characterization of cost of a translation.

- What is the relationship between the fluency cost and the fidelity cost? That is, how are they to be weighted?
- What determines the relative assumability costs of different interlingua expressions?

We regard both of these questions as a matter for exploration, but certain reqirements are quite natural.

Not really important here whether the system is committed to the interpretation as abduction view.

(?) Advantage of view this as logical process:

- A natural account of redundancy [particularly important in some of the Japanese examples below]
- A natural account of generalization (tokei)
- A natural account of specialization (tokei)
- A natural account of Adjustment

(?)

Explain KH.

Explain the problem with KH.

Explain AET. [Use Handout appendix]

A theorem prover for proving equivalences, in the version adopted here, equivalences within first order formulas.

- conditionalize equivalences.
- inference rules allowing us to infer equivalences on subparts of complex expressions.

- Natural termination condition, as long as we have some reasonable characterization of what a target language expression is.

- Abduction falls out naturally as equivalences with undischarged conditions.

Reasonable to have assumability costs. Reasonable for source language expression to be presented with assumability costs. Thus a portion of the source may end up untranslated, one the assumed conditions under which a translation is proposed, at a cost. [examples of where this is very feasible]

9 Compiling a Polylingua System Into a Transfer System

We discuss efficiency-motivated tasks that an interlingua axiom compiler can perform, axiom composition and axiom filtering.

For composition consider the entwerten case.

Consider for the joint venture case described in example ??. Only a subset of the entire interlingua "theory" needs to be active for any given language-pair. It would thus be desirable to have some way of filtering In this section we discuss how this is to be acheived.

Only decomposition rules mapping to IL. Natural bound on these. Length 1.

Put an arbitary bound of 1 on composition derivations. as follows:

PSS =¿ SIT

SIT semantics can't feed another composition rule.

This means we have a bound on all derivations not involving arbitrary paraphrase. Can compile out a new transfer module.

The problem:

using language models.

References

Alshawi, H. and Carter, D. 1994. Training and Scaling Preference Functions for Disambiguation. Computational Linguistics 20.4. 635-648.

Anderson, J. M. *On Case Grammar: Prolegomena to a Theory of Grammatical Relations.* Croom Helm. London.

Barnett, J., Mani, I., Martin, P., and Rich, E. 1991. Reversible Machine Translation: What to do when the Languages Don't Line Up. Microeletronics and Computer Technology Corporation. MCC Technical Report Number ACT-NL-068-91.

Brown, P. F., Della Pietra S. A, Della Pietra, V.J., and Mercer, R.L. 1993. The Mathematics of Statistical Machine Translation. Computational Linguistics 19. 2. 263-312.

Charniak, E, and Goldman, R. 1988. A Logic for Semantic Interpretation. *Proceedings of the 26th Annual Meeting of the ACL.*

Choi, S.and Bowerman, M. 1991. Learning to express Motion Events in English and Korean: the influence of loexical specific lexicalization patterns. Cognition 41. 83-121.

Dagan, I. and Itai, A. 1994. Word Sense Disambiguation Using a Second Language Monolingual Corpus. Computational Linguistics 20.4. 563-596.

Dorr, B. 1993. *Machine Translation: A View from the Lexicon.* MIT Press. Cambridge.

Dorr, B. 1994. Machine Translation Divergences: A Formal Description and Proposed Solution. Computational Linguistics 20.4. 597-634.

Downing, P. 1977. 'On the Creation and Use of English compound nouns. *Language* 53:4.810-842.

Fillmore, C.J. 1968. The Case for Case. In Bach E. and Harms, R. *Universals of Linguistic Theory* Holt, Rinehart, and Winston, New York.

Fillmore, C.J. 1977. The Case for Case Reopened. In P. Cole (ed). *Syntax and Semantics 8: Grammatical Relations*, 59-81. Academic Press, New York.

Fillmore, C.J. 1997. *Lectures on Deixis.* CSLI Publications. Stanford.

Gawron, J.M. 1986. Situations and Prepositions. *Linguistics and Philosophy* 9:327-382.

Goldberg, A. 1995. *Constructions: A Construction Grammar Approach to Argument Structure.* University of Chicago Press, Chicago.

Gopnik, A. and Choi, S. 1995. Names, Relational Words, and Cognitive Development. In Tomasello, M. (ed). *Beyond Names for Things.* 63-80. Lawrence Erlbaum.

Gruber, J. 1965. Lexical Structures in Syntax and Semantics. Ph.D. Diss. Massachusets Institute of Technology.

Herskovits,A. 1986. *Language and Spatial Cognition: An Interdisciplinary Study of the Prepositions in English.* Cambridge University Press, Cambridge.

Hobbs,J.R., Croft, W., Davies, T., Edwards D., and Laws, K. 1987, Commonsense Metaphysics and Lexical Semantics. Computational Linguistics 13: 241–250.

Hobbs,J.R., Stickel, M., Appelt D., Martin P. 1988. Interpretation as Abduction. *Proceedings of the 26th Annual Meeting of the ACL.*

Hobbs,J.R., Stickel,M., Appelt D., Martin P. 1993. Interpretation as Abduction. *Artificial Intelligence* 63: 69-142.

Hobbs,J.R. and Kameyama, M. 1990. Translation by Abduction. *Proceedings of the Seventh ICCL, COLING-90.* Helsinki, Finland.

Hutchins, W. J. and Somers, H. L. 1992. *An Introduction to Machine Translation.* Academic Press, Londin.

Jackendoff, R. 1983. *Semantics and Cognition.* MIT Press. Cambridge.

Jackendoff, R. 1987. On Beyond Zebra: The relation of Linguistic and Visual Information. *Cognition* 26: 89-114.

Jackendoff, R. 1990. *Semantic Structures.* MIT Press. Cambridge.

Jackendoff, R. 1996. The Architecture of the Linguistic Spatial Interface. In *Language and Space.* Bloom, P., Peterson, M.A., Nadel, L., and Garrett, M.F. (eds) MIT Press. Cambridge.

Jackendoff, R. and Landau, B. 1995. Spatial Language and Spatial Cognition. In *Languages of the Mind.* Jackendoff, R. MIT Press. Cambridge.

Johnson-Laird. P.N. 1983. *Mental Models.* Harvard University Press. Cambridge.

Kameyama, M., Ochitani, Ryo, Peters, S., and Hidetoshi, Shirai. 1991. Resolving Translation Mismatches with Information Flow. In *Proceeds of the 29th Annual Meeting of the ACL.* 193-200. Berkeley.

Kaplan, R.M., Netter, K., Wedekind, J., and Zaenen, A. 1989. Translation by Structural Correspondences. *Proceedings of the Fouthg Conference of tyhe European Chapyer of the ACL.* 272-281. Manchester.

Kay, M., Gawron, J. M., and Norvig, P. 1994. *Verbmobil: A Translation System for Face-to-Face Dialog.* CSLI Lecture Notes No. 33. CSLI. Stanford

Lakoff, G. and Johnson M. 1980. *Metaphors we Live By.* University of Chicago Press. Chicago.

Lakoff, G. *Women, Fire, and Dangerous Things.* University of Chicago Press. Chicago.

Landau, B. and Jackendoff, R. 1993. "What" and "Where" in Spatial Language and Spatial Cognition. *Behavioral and Brain Sciences* 16: 217-238.

Levinson, S. C. 1996. Frames of Reference and Molyneux's Question: Cross-Lingustic Evidence. In *Language and Space.* Bloom, P., Peterson, M.A., Nadel, L., and Garrett, M.F. (eds) MIT Press. Cambridge.

Makkai, A, *Idiom Structure in English.* Mouton. The Hague.

McCarthy, J. 1986. Applications of Circumscription to Formalizing Common-Sense Knowledge. *Artificial Intelligence* 26.

Nagao, M. 1989. *A Japanese View of Machine Translation in Light of the Considerations and Recommendations reported by ALPAC, U.S.A.* Japan Electronic Dictionary Development Association.

Nirenberg, S. 1987. A Distributed Generation System for Machine Translation: Background, Design, Architecture, and Knowledge Strtucres. Carnegie Mellon University. Center for Machine Translation Technical Notes. CMU-CMT-87-102.

Nirenberg, S., McCardell, R., Nyberg, E., Werner, P., Huffman, S., Kerschaft, E., and Nirenberg, I. 1988. Diogenes-88. Carnegie Mellon University. Center for Machine Translation Technical Notes. CMU-CMT-86-107.

Nirenberg, S. Carbonell, J., Goodman, K. 1992. *Machine Translation: A Knowledge Based Approach.* Morgan Kaufman. San Mateo, California.

Norvig, P. and Wilensky, R. 1990. A Critical Evaluation of Commensurable Abduction Models for Semantic Interpretation. *Proceedings of COLING-90.* Helsinki.

Nunberg and Zaenen 1992. 'Systematic Polysemy in lexicology and lexicography.' *Proceedings of Eurolex 1992.* Tampere, Finland.

Pierce, C.S. 1955. *Abduction and Induction.* Dover. New York.

Pusteyovsky, J. 1991. The Generative Lexicon. Computational Linguistics 17.4. 409-441.

Rayner, M. 1993. Abductive Equivalential Translation and its Application to Natural Language Database Interfacing. Ph. D dissertation Royal Institute of Technology, Sweden.

Sato, S. and Nagao, M. 1990. Memory-Based Translation. *Information-Processing Society of Japan.* Working Group - NL 9(70). In Japanese.

Stickel, M. 1989. "Rationale and Methods for Abductive Reasoning in Natural Language Interpretation" in R. Studer (ed) *Natural Language and Logic.* Proceedings, international scientific symposium. Hamburg, FRG

Talmy, L. 1978. Figure and Ground in Complex Sentences. In *Universals of Human Language* Vol. 4, Greenberg, J.H. (Ed.) Stanford University Press. Stanford.

Talmy, L. 1983. How Language Structures Space. In *Spatial Orientation: Theory, Research and Application,* Pick, H.L. and Acredolo, L.P. (Eds.) Plenum Press. New York.

Talmy, L. 1985. Lexicalization Patterns: Semantic Structure in Lexical Forms. In *Language, Typology, and Syntactic Description* , Vol. 3, Shopen, T. (Ed.) *Grammatical Categories and the Lexicon.* Cambridge University Press. Cambridge.

Thagard, P. R. 1978. The Best Explanation: Criteria for Theory Choice. The Journal of Philosophy 75: 76-92.

Zimmer, K. 1971. 'Some General Observations about Nominal Compounds.' *Working Papers on Language Universals* 8:C1-21.