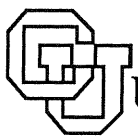
Proceeding of the IJCAI Workshop on Computational Approaches to Non-Literal Language: Metaphor, Metonymy, Idiom, Speech Acts and Implicature

Edited by Dan Fass, Elizabeth Hinkelman and James Martin

CU-CS-550-91 August 1991



University of Colorado at Boulder
DEPARTMENT OF COMPUTER SCIENCE

Proceedings of the IJCAI Workshop on Computational Approaches to Non-Literal Language: Metaphor, Metonymy, Idiom, Speech Acts and Implicature

CU-CS-550-91

held at the
Twelfth International Joint Conference on Artificial Intelligence

Edited by
Dan Fass,
Elizabeth Hinkelman
and
James Martin

24 August 1991
Sydney, Australia

Contents

Some Interplay between Metaphor and Propositional Attitudes
John Barnden 1

Sense Extension as Lexical Rules
Ted Briscoe and Ann Copestake 12

Metaphor and Literal Meaning
Kenny Coventry 21

Metaphor Comprehension Model: Theory and Implementation
Kouichi Doi, Hirohiko Sagawa, Hidehiko Tanaka 32

Metonymy, Case Role Substitution and Sense Ambiguity
Dan Fass 42

Metaphor and Abduction
Jerry Hobbs 52

Can AI Systems Generate Creative Metaphors?
Bipin Indurkhya 62

MetaBank: A Knowledge-Base of Metaphoric Language Conventions
James H. Martin 74

Bi-directional Parsing for Idiom Handling
Yuji Matsumoto, Katsuyoshi Yamagami and Makoto Nagao 83

A Formalization of Metaphor Understanding in Situation Semantics
Tatsunori Mori and Hiroshi Nakagawa 92

Metaphor vs. Anomaly: Conceptual Constraints on Verbal Metaphoric Extension
Sylvia Weber Russell 102

Register and Speech Act Theory in Text Generation
Cameron Shelley, Neil Randall and Chrysanne DiMarco 114

Learning Metaphorical Relationships Between Concepts Based on Semantic Representation Using Abstract Primitives
Masaki Suwa and Hiroshi Motoda 123

Beyond Literal Meaning: Conversation Acts
David R. Traum and Elizabeth A. Hinkelman 132

Idiomatic Expressions, Non-Literal Language and Knowledge Representation
Eric van der Linden 141

A Connectionist Model of Literal and Figurative Adjective Noun Combinations
Susan H. Weber 151

Extending the Lexicon by Exploiting Subregularities
Robert Wilensky 161

Some Interplay between Metaphor and Propositional Attitudes

John A. Barnden

Computing Research Laboratory & Computer Science Dept
New Mexico State University
Box 30001/3CRL
Las Cruces, NM 88003-0001
(505) 646-6235 jbarnden@nmsu.edu

ABSTRACT

Evidence from real discourse suggests that beliefs and other propositional attitudes are often viewed by speakers and other agents in a metaphorical way. Typical metaphors that are involved are MIND-AS-CONTAINER and IDEAS-AS-INTERNAL-SPEECH. It is therefore necessary for AI systems for propositional-attitude representation/reasoning to represent and reason within such views. This approach contrasts with the highly abstract logical approach adopted in most propositional attitude research. The paper concentrates on showing how the metaphorical views are revealed and on the effect they can have on discourse understanding. A formal representation scheme based on the various metaphors has been partially developed, but has been described elsewhere and is not presented here. A further, more general issue discussed is that of the embedding of metaphors (of any sort) within propositional attitude contexts. This embedding is not normally considered in treatments of metaphor or of propositional attitudes.

1. Introduction

There is a lot of discussion in the philosophical literature about (i) what beliefs, desires, intentions and other propositional attitudes *are*, and about (ii) what a formal, scientifically objective semantics of attitude reports should be like. (An attitude report is a natural language sentence, such as "John believes that Sally is clever," that ascribes a particular attitude to some agent.) Some recent and comprehensive discussions are offered by Schiffer (1987b) and Richard (1990). The two questions are of course intimately linked. On question (i), one stance is that the notions of belief and so on are merely parts of our folk psychology — in other words our commonsense views on the nature of mind — and have no substantive role in any scientific psychology or fully worked out philosophy of mind [see, e.g., Stich 1983, Churchland 1981]. This eliminativist stance then has strong consequences for how one approaches (ii). On the other hand, on the part of non-eliminativist investigators, there are numerous different positions on the nature of beliefs. For instance, someone's believing something is often cast as the person being in a certain relationship to some content or content-bearing object, but there is strong disagreements about what these objects are — propositions, formulae, concepts, natural language utterances, or whatever.

However, whether or not the eliminitivist stance is adopted, what almost all discussions lack is an appreciation of the following point. It is precisely our commonsensical views of mind, *not* any scientifically/philosophically accurate views, that are of paramount importance when it comes to question (iii): how should AI systems interpret incoming attitude reports?[1] The reason for this paramount importance is basically simple: there are always speakers behind attitude reports, and *those speakers'* commonsense views of attitudes may have an important role in how the speakers report the attitudes and in what they

[1] To this we could add question (iv): how do, in fact, people interpret incoming attitude reports? For brevity I will view this here as subsumed under question (iii).

expect the listeners to infer. This is so no matter how far from any scientific/philosophical truth those views are.

The main connection I wish to draw to non-literal language is that commonsense views of attitudes are almost entirely *metaphorical*. I therefore claim that metaphors lurk behind attitude reports, and in many cases these metaphors have to be addressed by a listener in order to make sense of what a speaker is saying. Further, the metaphors are not always implicit, but often make an explicit appearance in attitude reports, as illustrated below.[2]

This paper will not attempt a definition of metaphor, other than to say that metaphor is a matter of looking at one thing in terms of another and often involves an analogy between the two things. My overall view of metaphor is like that of Lakoff, Johnson and Sweetser [Lakoff & Johnson 1980, Johnson 1987, Lakoff 1987, Sweetser 1990], except that I place no reliance on the notion of image-schematic structure in Johnson 1987 and the related Invariance Hypothesis of Lakoff (1990) (in collaboration with Turner 1990). In particular, I concur with the claims that

(a) metaphor is an important phenomenon in *thought*, not just in language,

(b) metaphor in language derives from metaphor in thought,

(c) metaphor is often creative in the sense of *imposing* objects, properties and relationships on a domain, as opposed to merely highlighting ones that are already objectively known of the domain (cf. the notion of creativity in interactional theories of metaphor, which are reviewed in Waggonner 1990),

(d) indeed, much, perhaps most, perhaps all, of our understanding and reasoning about abstract domains, including the mind, is based on such metaphorical imposition from commonsensically-understood, sociophysical domains.

By virtue of this last point, the metaphors studied in this paper are probably best classed as "suggestive" ones in the terminology of Indurkhya (1991). I also hold (in company with, for example, Gibbs & O'Brien 1990) that conceptual metaphor need not be *consciously* entertained in thought.

The examples in this paper are mostly (inconsequentially edited) versions of examples observed in novels and magazines, and many of these original examples are collected, with detailed citations, in Appendix A of Barnden (1990). I am currently engaged in a data collection effort to provide statistics on the frequency of different modes of expression in a large amount of text. For the moment, I claim that the examples should strike readers as reasonable, mundane, readily understandable pieces of English. Since the examples are taken out of context, and I have included in them information that might normally be conveyed in a more implicit, contextual manner, they may give the appearance of being contrived. However, the ways in which propositional attitudes are conveyed in the examples are all quite frequent in real text and speech.

2. Fairies and Beliefs

Before considering any example attitude reports, let's consider a fairy report, such as:

"John danced with the fairies yesterday."

Assume that the speaker, Susan say, is speaking literally and sincerely. Then, even though there are no fairies, and even though the listener, Luke say, doesn't believe there are any, Luke needs nevertheless to construct an internal representation of what the sentence is saying. The situation so far is not much

[2] Another type of connection between propositional attitudes and metaphor is discussed in Ballim, Wilks & Barnden (1991) and Wilks, Barnden and & Wang (1991: this conference). I am also much concerned with the role of *metonymy* in belief reports, and the way it interacts with the role of metaphor; these matters are briefly discussed in Barnden (1990).

different from what happens with fictional literary entities, as in the sentence "Sherlock Holmes was arrogant." The need to be able to represent fictional entities — granting them existence in some sense — has been discussed recently from the AI standpoint by Hirst (1989). However, Luke may have to do much more than simply represent fairies in some way. He may have to reason about fairies, *within the context of what he conjectures Susan's view of fairies to be*, in order to make the sentence cohere with others. Suppose, for instance, Susan continues by saying

> "The Queen fairy was cross with them, and banned John from going to the wood ever again."

The listener must presumably reason as follows, to make sense of this.

> Fairies often hang about in woods, and typically do their dancing there. John danced with them in such a wood.

Thus, Luke is ascribing a view of fairies to Susan and reasoning within that view. Furthermore, Susan surely *expects* Luke to reason thus (even though there are other possibilities known to both Susan and Luke, such as that fairies *might* dance in discos). By default, the view Luke ascribes to Susan will be his own view of fairies (or a view he takes to be the normal one), but in particular contexts he might have to ascribe some more special view to Susan (perhaps because of previous statements of hers about fairies).

I have belaboured these rather obvious points because their close analogues when propositional attitude talk is substituted for fairy talk have hardly been noted, and have consequences for propositional attitude research which cannot be obvious, since philosophers of note have not observed them. Suppose Susan says the following, reporting a rather familiar sort of scenario:

> "John knew that he had to meet Peter at 10am. But he arranged an hour long meeting with Mary starting at 9.30am."

The most natural interpretation of this is probably that John was forgetting, at the moment of arranging the Mary meeting, that he had a 10am appointment.[3] In other words, he (consciously?) believed, at the moment of the arranging, that he was free for at least an hour from 9.30am. Nevertheless, I conjecture that most people would agree that he still, in some sense, knew at that moment that he had a 10am appointment. It's just that the knowledge was unconscious, suppressed, or inoperative in some sense. Thus we have to ascribe to Susan (our speaker) a commonsense view of attitudes that allows it to be the case that someone can have unconscious/suppressed/inoperative knowledge that (easily) contradicts other beliefs.[4] Now, there are many views of attitudes that would allow this to be the case, and some of them will be touched on as we proceed. However, it is important to observe that a particular view can be *explicitly* signalled, as follows.

3. Explicit Signalling of Mind Views

Consider the following variant of our example:

> "With one part of his mind, John knew that he had to meet Peter at 10am. But he arranged an hour long meeting with Mary starting at 9.30am."

The "one part of his mind" type of locution is very common in real discourse about propositional attitudes. It reveals that Susan holds, or at least is currently adopting, a view of the mind whereby it has parts that are to some extent independent: for, surely, she expects Luke to infer that the part of

[3] There are of course other reasonable interpretations, such as that John is deliberately snubbing Peter.

[4] I make nothing of the distinction between knowledge and belief, as such, in this paper. For my purposes, knowledge is simply true belief.

John's mind that had the 10am knowledge was not the part operative in the Mary arrangement; and that *that* part believed that there was a free hour from 9.30am. The inferences that Susan expects Mike to make are a functional analogue of the ones she expected Luke to make in the fairy example, although of of course the details of the inferences are widely different.

The idea of the mind as having parts or regions fits in with at least two distinct metaphors of mind: the MIND-AS-CONTAINER metaphor and what I call MIND-PARTS-AS-PERSONS. The former is well-known from (Lakoff & Johnson 1980, Johnson 1987, Lakoff 1987). A version of it is also one of the two metaphors that Wellman (1990: pp.268–271) identifies as central in the development of the child's theory of mind, and it plays an important role in the psychological study of idioms in Gibbs & O'Brien (1990).

Under MIND-AS-CONTAINER, the CONTAINER can be large and the beliefs distant from each other in the CONTAINER. The beliefs can consequently fail to interact. Beliefs must be near to each other in order to interact, as is is evident from the common idea of "bringing ideas/beliefs together," as in

"John didn't notice the clash because he didn't bring the two arrangements together in his mind."

Actually, the idea of mind regions and of belief movement are consistent not just with MIND-AS-CONTAINER but also with more general metaphor, MIND-AS-PHYSICAL-SPACE.

By default, the CONTAINER is thought of as being fairly small, and beliefs are already close together. (In their psychological study, Gibbs & O'Brien (1990) observe that subjects' mental images related to MIND-AS-CONTAINER-based idioms strongly tend to take the container to be a mundane physical container about the size of a human head.) However, these defaults are easy to defeat, as they implicitly are in the example sentence just displayed.

As for MIND-PARTS-AS-PERSONS, this casts the mind as being made up of sub-persons that have their own beliefs etc. and who may or may not communicate with each other, much as is the case with real persons.[5] Although the metaphor is not particularly strongly signalled in the above example, it is more strongly signalled in the following variant:

"One part of John knew that he had to meet Peter at 10am. But ... "

And it is yet more strongly signalled in:

"One part of John was insisting that he had to meet Peter at 10am. But ... "

"One voice inside John was insisting that he had to meet Peter at 10am. But ... "

I claim that the use of "insisting" is metaphorically casting the relevant part of John as being a person uttering natural language statements directed at other sub-persons. These other parts can be deaf to the insistence, just as real people can. And this deafness is what Susan expects Luke to infer in order to understand her sentences.

A view of someone's mind as populated by sub-persons does not illuminate the mental life of those sub-persons themselves. In principle, it would be possible for *them* to be viewed through MIND-PARTS-AS-PERSONS, yielding a recursive application of the metaphor. However, I have not encountered this. My current approach is to the sub-persons to be viewed, by default, through MIND-AS-CONTAINER. This metaphor has a broad role as a default in my approach, as will become clear below.

The second of the two metaphors for mind that are central in child development, according to Wellman (1990: pp.268–271), is MIND-AS-HOMUNCULUS. Under this metaphor, the mind is viewed

[5] I am *not* alluding here to cases of extreme schizophrenia, although such cases are not inconsistent with the discussion. A complication is that in these cases one *might* be inclined to say that the person *literally* contains several sub-persons.

as a person whose jobs is to actively interpret incoming information, manipulate ideas, and so on. Our MIND-PARTS-AS-PERSONS model may well be a natural developmental extension from the MIND-AS-HOMUNCULUS metaphor.

4. Ideas-As-Internal-Speech

Because of the frequent appearance of internal natural language in uses of MIND-PARTS-AS-PERSONS, it is then actually a special case, or elaboration, of a further common and commonsensical metaphor of mind: IDEAS-AS-INTERNAL-SPEECH. This metaphor can also appear alone, without any notion of sub-person. The metaphor has it that when a person is entertaining an idea — notably as a belief, intention or desire — the idea takes the form of a *natural language* sentence internally uttered by that person. The metaphor is very clearly signalled in examples such as the following:

> "John said to himself, 'I have an appointment at 10am'. "

This sort of example is very common in mundane English text, such as in popular novels.[6] Now, I suggest the sentence displayed conveys *more* than simply that John believed that he had an appointment at 10am. In particular, it conveys that John *consciously* believed this, and that it is *not* the case that John is suppressing or forgetting the fact that he had the appointment. For evidence for this claim, consider:

> "Saying to himself, 'I have an appointment with Peter at 10am,' John arranged to be unavailable at that time.

The natural interpretation here seems to be that John is deliberately ignoring his 10am appointment, and it would be very unnatural to take John to be suppressing or inactivating the knowledge of that appointment. This interpretation is in complete contrast to the the fact that it was *natural* in our original example to take John to be doing this.[7]

However, we should stop short of taking the IDEAS-AS-INTERNAL-SPEECH metaphor (or other views that imply that the beliefs in question are conscious or otherwise operative), from implying that the agent necessarily draws all the straightforward consequencies the beliefs have. Consider:

> "Saying to himself, 'I have an appointment with Peter at 10am,' John arranged a ninety-minute meeting with Mary starting at 8.45am."

A possible interpretation which should be considered here is that John was consciously believing that he had a free ninety minutes starting at 8.45am as well as consciously believing that he had an appointment at 10am, but simply failed to to the right arithmetic (even though perhaps perfectly capable of doing it in other situations). And, this needn't involve any difficulty in bringing *arithmetic* to bear, but rather in a failure simply to bring together the two beliefs and arithmetic process in the right way. I think we all have experienced strange failures to draw simple consequences, in ourselves as well as in others.

One theoretical difficulty, for my purposes here, that surrounds IDEAS-AS-INTERNAL-SPEECH is that someone might claim that it wasn't *metaphorical* at all. Perhaps we *literally* say things to ourselves internally. Certainly, the conscious experience one has when one is reported as saying something to oneself *feels* pretty much the same as hearing someone else say something and/or saying something to someone else. Also, saying-to-oneself may commonly be accompanied by sub-vocal muscular movements. My own intuition is nevertheless that one only metaphorically says things to oneself. In any case, the

---

[6] Here and below, I ignore the possibility that John actually utters out loud to himself the sentence "I have an appointment at 10am," so that he is saying it to himself in exactly the same sense as he might say it to someone else.

[7] We cannot take saying-to-oneself to be equivalent to *realizing*, although the latter also may well serve to convey conscious belief, because realizing is factive whereas the saying-to-oneself is not. The factive quality of realizing is that if X says Y realizes P, then X must be taking P to be true.

fact that *some* commonsense views of propositional attitudes might not be metaphorical does not reduce the importance of the role of metaphor in such views in general. Most are clearly metaphorical.

Interestingly, the following ways of reporting thoughts are extremely common:[8]

"John thought, 'I have an appointment at 10am'."

"John thought to himself, 'I have an appointment at 10am'."

(The latter is particularly interesting, since thinking is, surely, always to oneself!) I conjecture that these sentences are almost or completely synonymous with the "John said to himself" variant, and have the same implications as regards consciousness.

Strangely, the (commonsensically assumed) link between thought and speech can work the other way round, as in the following common sort of example:

"John thought aloud, 'I have an appointment at 10am'."

Here, a thinking and a literal saying are simultaneously conveyed; and seems reasonable to suggest that the thinking is conscious. The fact that the inclusion of "aloud" conveys literal saying seems to add weight to the claim that we commonly view thinking as *saying* that happens *not* to be aloud.

Altogether: *saying* can be used metaphorically to describe *thinking*, and *thinking* can be used to describe *saying*. But I am not sure whether or not the description in the latter case is metaphorical.

The IDEAS-AS-INTERNAL-SPEECH metaphor is reminiscent of *meta-linguistic* approaches to the representation of propositional attitudes. These approaches, well-known in the philosophical literature but hardly ever seriously entertained in AI, go back at least to Carnap (1947). A good recent example is the proposal of Elgin (1985). See also Schiffer (1987a,b) and Elugardo (1989) for discussion of some meta-linguistic techniques, notably that of Davidson (1968); and Richard (1990) for a proposal that is partially meta-linguistic. The essence of pure, simple forms of the approach is to render the sentence

"John believes that Sally is clever"

by means of a formula like

```
bel-nl(John, 'Sally is clever')
```

which contains a logic term that is a quotation of a natural-language sentence. There are problems with the approach, such as the ambiguity of the quoted sentence [Schiffer 1987b, p.120]. However, some of these are catered for by Elgin's proposal, in which formulae analogous to the one just displayed are accompanied by logic expressions that clarify the meaning of the quoted English sentence. Also, the problems discussed in the literature are largely to do with the question of whether it is philosophically respectable to claim that to believe something is to be in a certain relationship (denoted by the predicate symbol `bel-nl`) to a natural language sentence. However, this issue is only weakly related to the question of whether, in practice, it is *heuristically good enough* for a cognitive system to express belief states of other agents by means of such a relationship.

The meta-linguistic strategy comes with no necessary commitment to the idea that John's believing that Sally is clever is a matter of John's having the sentence *Sally is clever* in his head (perhaps by saying it to himself). Rather, the intuitive interpretation of `bel-nl` could be that that sentence merely paraphrases some mental substate or representation within John. This, incidentally, takes care of the common objection that John might not be an English speaker, or might be a non-human animal (or robot, for that matter). However, the sentences-in-the-head special case of the meta-linguistic

[8] For instance, I have seen many such examples in the Noddy children's stories of Enid Blyton, as well as in more weighty literary works.

approach is especially interesting to me in that it conforms to IDEAS-AS-INTERNAL-SPEECH. This special case of the meta-linguistic approach has obvious problems as a *general* approach to representing beliefs. For instance, it obviously fails to apply to agents that do not communicate in the natural language in question, or in any natural language. However, this objection evaporates if the IDEAS-AS-INTERNAL-SPEECH approach is merely one weapon in the representational armoury. Moreover, if the agent in question is, say, a Spanish speaker, one might still use IDEAS-AS-INTERNAL-UTTERANCES-in-*English*, since speaking in English can be taken as a mere *metaphor* for speaking in Spanish. Equally, IDEAS-AS-INTERNAL-UTTERANCES-in-English could be used to describe the beliefs of a dog, since speaking in English can be taken as a mere metaphor for canine cognition.

5. Other Metaphors

I have pointed out that various metaphors of mind carry different implications for what inferences the listener should or should not make, and that such inferences can be important in achieving a coherent understanding of natural language discourse. I discuss the implied inference patterns much more extensively in Barnden (1990), where I also look at various other metaphors for mind. These include MIND-AS-BATTLEGROUND, MIND-AS-CHEMISTRY, MIND-AS-ANALOGUE, MIND-AS-TERRAIN, and AGENT'S-SEEN-WORLD. Of these, MIND-AS-BATTLEGROUND is related to the ARGUMENT-AS-WAR metaphor discussed by Lakoff & Johnson (1980), and AGENT'S-SEEN-WORLD and is related to the UNDERSTANDING-AS-SEEING metaphor addressed by Johnson (1987), Lakoff (1987), Lakoff & Johnson (1980) and Sweetser (1990)). AGENT'S-SEEN-WORLD underlies examples like

"In John's view, Sally is clever."

The idea is that John's beliefs are taken to portray situations that are visible in the world as seen by John.

It is also worth noting here that it is very common to express a belief in the following sort of way:

"To me, the war was a shambles."

Similar options are available in other languages: in Spanish one can say "Para mí, que la guerra era un fracaso" [M. Gonzalez, personal communication]; and according to Langacker (1990: p.233), in the language Nerawi one can use a construction directly corresponding to "To me." It is not clear yet exactly what sort of model of mind underlies these modes of expression, but, according to a suggestion in Langacker (1990), the speaker in our example is to be taken as a *setting* for the war being a shambles. We may take the notion of setting here to be a metaphorical extension of the notion of a physical setting.

In Barnden (1989b, 1990) I also present part of a detailed formal representational framework that allows an AI system to represent attitudes in a way the exploits the ontologies assumed by the various metaphors. Thus, if John's mind is cast as a container, then the system treats beliefs as physical objects and *uses the ordinary physical "in" predicate* to express their being in John's mind. If, on the other hand, John's beliefs are cast in terms of the IDEAS-AS-INTERNAL-SPEECH metaphor, then a formula similar to the `bel-nl` formula displayed above can be used, but using, in place of `bel-nl`, *the same predicate symbol for saying as would be used for real cases of saying.* Thus, the representational approach is highly eclectic and commonsense-imbued, unlike the highly uniform but commonsensically impoverished styles of representation used in other approaches.

There are many alternative commonsense views of mind. This is apparent from Tomlinson (1986), Larsen (1987), Talmy (1988), Lakoff, Espenson & Goldberg (1989), Richards (1989), Martin (1990), Sweetser (1990), and other works on metaphor, as well as from the above list. A natural language understanding system must be able to choose amongst them in accordance with context, in the most general sense of this term. Of course, in many cases an attitude report is nude of any explicit signalling

of particular commonsense views. This is true of the type of attitude report on which the propositional attitude field is almost entirely fixated, namely reports of the form

"X believes that ..."

Nevertheless, textual or environmental context may indicate that the speaker is using a particular commonsense view of mind. For instance, a recent attitude report by the speaker may have explicitly signalled a view, and unless there is evidence to the contrary it is reasonable to assume that the speaker is still using it.

However, what happens if there are no contextual clues as to what commonsense view of mind the speaker is adopting? One answer would be to have a view-neutral mode of representation. Although this may look like the obvious approach, I propose instead that the system assume that the speaker is using a *default* view. In fact, the default view I currently propose is MIND-AS-CONTAINER (although this choice may well have to be revised). The argument for this approach, which has several strands, is given in (Barnden 1990); and one strand depends on the highly technical considerations explained in detail in Barnden (1986, 1987a,b, 1989a). In any case, for the purposes of the present paper, it is enough to consider cases where the attitude report itself, and/or its context, does indicate that the speaker is adopting a particular commonsense view of mind.

One major focus of attention in most work on propositional attitudes is the question of the extent to which agents can be held to believe the consequences of their beliefs. Clearly, people don't always believe such consequences.[9] Propositional attitude researchers try to impose general, uniform constraints on what consequences are believed, if any. However, as argued in detail in Barnden (1990) and as hinted at above, the question of what sorts of consequence our speaker Susan takes John to be making from his other beliefs is answered in part by the nature of the commonsense view of mind that Susan is adopting. Thus, under IDEAS-AS-INTERNAL-SPEECH it seems that John is taken by default to draw all consequences that are sufficiently simple; whereas under MIND-PARTS-AS-PERSONS he will by default fail to do that in the case of beliefs held by different sub-persons.

Further than this, I *conjecture* that on all occasions, in real discourse, when it is actually important to reason as to what consequences an agent John has drawn from his beliefs, the attitude reports in question or features of their context indicate that a specific commonsense view of mind is being entertained by the speaker. If this is true — and I only have preliminary evidence for it — then the many occasions in real discourse when no specific view is indicated are simply occasions of reporting free-standing beliefs that need not be connected to other beliefs.

6. Nested Attitudes

The above considerations are amplified in complexity, and interest, when we turn to *nested* (alias *iterated*) attitudes. Consider

"George thinks that John knew that he [John] had to meet Peter at 10am. But ..."

We now have to consider not only Susan's possible views of mind, but also views she may be ascribing to George. As to the latter, it may be that context indicates that Susan is adopting a certain view, but there is no indication of what view she is ascribing to George. In this case, it may be reasonable to assume by default that she is ascribing her own current view to George. Or, it may be reasonable to assume by default that she is ascribing the default MIND-AS-CONTAINER view. I have not resolved these issues, but the formal representation scheme in (Barnden 1989b, 1990) allows different views to hold sway at different levels of the nesting. For instance, George's own beliefs can be viewed through MIND-AS-CONTAINER, but he may be viewed as viewing Mike through AGENT'S-SEEN-WORLD.

[9] Unless, that is, one interprets belief in the highly artificial sense of *implicitly* believing that is current in AI – a notion concocted precisely to make it come out that one believes all the consequences of one's beliefs, but which has no bearing on any piece of real discourse I have ever encountered.

Consider now a case in which a view is explicitly signalled:

"George thinks that one part of John knew that he [John] had to meet Peter at 10am. But ..."

Is the speaker claiming that George is adopting (say) a MIND-PARTS-AS-PERSONS view of John, or is this view merely the speaker's way of portraying *some*, possibly unknown, view of George's that allows contradictory beliefs to coexist in John's mind? This question is actually a special case of a much more general *metaphor scoping* issue, as follows.

7. Metaphor Scoping and Conclusion

Metaphor scoping is discussed in (Barnden 1990) and (Wilks, Barnden & Wang 1991: this conference), and I take the liberty of paraphrasing from the latter. Consider the sentence

"John believes that a cure for terrorism is needed."

The complement of this belief report — the clause following the word "that" — can be construed as involving a terrorism-as-disease metaphor. The "inner scope" reading involves the idea that John himself thinks of terrorism as a disease (and he might report his belief by means of the sentence "A cure for terrorism is needed"). The "outer scope" reading is that John believes something about terrorism that is being portrayed *by the speaker* in terms of disease-curing. John does not necessarily have a belief couched in these terms, nor would he necessarily report his belief in these terms. Perhaps John would say: "Something needs to be done to eliminate terrorism and repair the damage it has done to society." It is also pointed out in Wilks *et al.* (1991) that the metaphor scope distinction is precisely analogous to one drawn in the case of definite descriptions within belief contexts. The latter distinction is one of the most important topics in the propositional attitude field, yet its extension to metaphor appears not have been studied either by attitude theorists or by metaphor theorists. An interesting sub-issue, concerning inner scope readings of metaphors, is the effect the agent's (John's) beliefs about snakes have on the understanding of what is being conveyed by the metaphor. John's beliefs may differ significantly from the speaker's and listener's.

The issue is more complex than simply [sic] a matter of scoping of a metaphor, since the very question of whether a metaphor is involved in the first place interacts with it. Thus, in

"John believes that Mike is a snake"

it may be that John believes that Mike is *literally* a snake, although, let us assume, the speaker and listener know Mike to be a person. Thus, John's overall belief state must be considered not only in understanding what John believes about Mike *given* that we take John to believe Mike to be a person and we take the inner-scope reading of the snake metaphor, but also in deciding whether a metaphor is involved at all.

The question of what metaphor-scoping to adopt in the second George nested-attitude sentence displayed above is merely a special case of the general metaphor-scoping issue. However, it serves the purpose of illustrating especially clearly my main thesis, namely that, in propositional attitude research, what is important are the various agents' views of mind, not any scientifically correct view.

ACKNOWLEDGMENTS

I have benefited from discussion with Afzal Ballim, Saskia Barnden, Tom Eskridge, David Farwell, Margarita Gonzalez, Louise Guthrie, Steve Helmreich, Eric Iverson, George Lakoff, Anthony Maida, James Martin, Paul McKevitt, Jin Wang, and Yorick Wilks. I am also indebted to Wendy Dare and Margarita Gonzalez for data collection activities.

REFERENCES

Ballim, A., Wilks, Y. & Barnden, J.A. (1991). Belief ascription, metaphor, and intensional identification. *Cognitive Science, 15* (1), pp.133–171.

Barnden, J.A. (1986). Imputations and explications: representational problems in treatments of propositional attitudes. *Cognitive Science, 10* (3), pp.319–364.

Barnden, J.A. (1987a). Interpreting propositional attitude reports: towards greater freedom and control. In B. du Boulay, D. Hogg & L. Steels (Eds), *Advances in artificial intelligence – II,* Amsterdam: Elsevier (North–Holland). pp.159–173. (Procs. of 7th. European Conf. on Art. Int., July 1986.)

Barnden, J.A. (1987b). Avoiding some unwarranted entailments among nested attitude reports. *Memoranda in Computer and Cognitive Science*, No. MCCS–87–113, Computing Research Laboratory, New Mexico State University, NM 88003, USA.

Barnden, J.A. (1989a). Towards a paradigm shift in belief representation methodology. *J. Experimental and Theoretical Artificial Intelligence 2*, pp.133–161.

Barnden, J.A. (1989b). Belief, metaphorically speaking. In *Procs. 1st Intl. Conf. on Principles of Knowledge Representation and Reasoning.* San Mateo, CA: Morgan Kaufmann. pp.21–32.

Barnden, J.A. (1990). Naive Metaphysics: a metaphor-based approach to propositional attitude representation. (Unabridged version.) *Memoranda in Computer and Cognitive Science*, No. MCCS–90–174, Computing Research Laboratory, New Mexico State University. Under revision for submission to *Cognitive Science.*

Carnap, R. (1947). *Meaning and Necessity.* Chicago: University of Chicago Press.

Churchland, P.M. (1981). Eliminative materialism and propositional attitudes. *J. Philosophy, 78*,pp.67–90.

Davidson, D. (1968). On saying that. *Synthese, 19*, pp.130–146.

Elgin, C.Z. (1985). Translucent belief. *J. Phil., 82* (2), pp.74–91.

Elugardo, R. (1989). Representationalism and Church's translation argument. *Phil. Studies, 56*, pp.107–125.

Gibbs, R.W., Jr. & O'Brien, J.E. (1990). Idioms and mental imagery: the metaphorical motivation for idiomatic meaning. *Cognition, 36* (1), pp.35–68.

Hirst, G. (1989). Ontological assumptions in knowledge representation. In *Procs. 1st Intl. Conf. on Principles of Knowledge Representation and Reasoning.* San Mateo, CA: Morgan Kaufmann.

Indurkhya, B. (1991). Modes of metaphor. *Metaphor and Symbolic Activity, 6* (1), pp.1–27.

Johnson, M. (1987). *The body in the mind.* Chicago: Chicago University Press.

Lakoff, G. (1987). *Women, fire and dangerous things.* Chicago: University of Chicago Press.

Lakoff, G. (1990). The Invariance Hypothesis: is abstract reason based on image-schemas? *Cognitive Linguistics, 1* (1), pp.39–74.

Lakoff, G., Espenson, J. & Goldberg, A. (1989). Master metaphor list. Draft manuscript, Cognitive Linguistics Group, University of California at Berkeley.

Lakoff, G. & Johnson, M. (1980). *Metaphors we live by.* Chicago: University of Chicago Press.

Langacker, R.W. (1990). Settings, participants, and grammatical relations. In S.L. Tsohadzidis (Ed.), *Meanings and Prototypes: Studies in Linguistic Categorization.* London and New York: Routledge. pp.213–238.

Larsen, S.F. (1987). Remembering and the archaeology metaphor. *Metaphor and Symbolic Activity, 2* (3), 187–199.

Martin, J.H. (1990). A unified approach to conventional non-literal language. Talk at 5th Rocky Mountain Conference on Artificial Intelligence, New Mexico State University, Las Cruces, New Mexico. (Content is distinct from that of his paper with same title in the Proceedings.)

Richard, M. (1990). *Propositional attitudes: an essay on thoughts and how we ascribe them.* Cambridge, U.K.: Cambrdige University Press.

Richards, G. (1989). *On psychological language and the physiomorphic basis of human nature.* London: Routledge.

Schiffer, S. (1987a). Existensionalist semantics and sententialist theories of belief. In E. LePore (Ed.), *New Directions in Semantics.* London: Academic Press.

Schiffer, S. (1987b). *Remnants of meaning.* Cambridge, Mass.: M.I.T. Press.

Stich, S.P. (1983). *From folk psychology to cognitive science: the case against belief.* Cambridge, MA: MIT Press.

Sweetser, E.E. (1990). *From etymology to pragmatics: metaphorical and cultural aspects of semantic structure.* Cambridge, U.K.: Cambridge University Press.

Talmy, L. (1988). Force dynamics in language and cognition. *Cognitive Science, 12,* 49–100.

Tomlinson, B. (1986). Cooking, mining, gardening, hunting: metaphorical stories writers tell about their composing processes. *Metaphor and Symbolic Activity, 1* (1), 57–79.

Turner, M. (1990). Aspects of the Invariance Hypothesis. *Cognitive Linguistics, 1* (2), pp.247–255.

Waggoner, J.E. (1990). Interaction theories of metaphor: psychological perspectives. *Metaphor and Symbolic Activity, 5* (2), pp.91–108.

Wellman, H.M. (1990). *The child's theory of mind.* Cambridge, MA: MIT Press.

Wilks, Y., Barnden, J. & Wang, J. (1991). Your metaphor or mine: belief ascription and metaphor interpretation. Forthcoming in *Procs. 12th Int. Joint Conf. on Artificial Intelligence* (Sydney, Australia, Aug. 1991). San Mateo: Morgan Kaufmann.

Sense extensions as Lexical Rules

Ted Briscoe and Ann Copestake

Computer Laboratory, University of Cambridge, Pembroke Street, Cambridge, CB2 3QG, UK
ejb@cl.cam.ac.uk aac@cl.cam.ac.uk

Abstract

We define a declarative and computationally tractable concept of lexical rule as a component of a unification-based lexicon employing (default) inheritance and typed feature structures. We argue that this concept of a lexical rule is expressive enough to cover both derivational morphological processes and metonymic and metaphoric sense extensions, and that the application of such rules is productive within finely specified subsets of the lexicon identifiable via the type system. In addition, we argue that formulation of these processes in a uniform framework uncovers similarities between them; such as the blocking of metonymic sense extension processes in a fashion similar to the blocking of derivational morphological ones. We illustrate these points principally with reference to a family of increasingly specific rules of "grinding" and indicate how machine-readable versions of conventional dictionaries can be utilised as a source of information to guide identification and specification of such rules.

1 Introduction

In this paper, we take the position, persuasively argued in Pustejovsky (1989a,b), that the lexicon is not a passive list of unrelated entries, but rather a highly structured and generative device. The rule-governed processes which can be expressed as lexical rules extend far beyond those discussed under the rubric of lexical redundancy rule in classical generative grammar. We present a very general, though declarative and computationally tractable, definition of a lexical rule, formalised within the framework of unification-based approaches to lexical organisation employing typed feature structures and (default) inheritance. We are currently exploring the possibility that this account of lexical rules is sufficiently expressive to describe processes of derivational morphology, conversion or zero-derivation, and metonymic as well as metaphoric sense extension. Linguistic research has emphasised the phonological / graphological consequences of derivational processes to the extent that conversion or zero-derivation has often been treated in terms of "zero-morphemes" and "zero-affixation". In our work, we are concentrating on the syntactic and semantic consequences of such processes and treat all of them as mappings between lexical entries expressed as typed feature structures. In this framework, it is affixation, rather than conversion, which is the marked operation since it requires changes to the phonological or orthographic part of the lexical entry (see, e.g., Cahill (1990) for a theory of such operations couched in a unification-based account of the lexicon). We argue that there are sufficient similarities between these processes to treat them as lexical rules and that the distinction between metonymic processes of sense extension and more clearly derivational rules is not as clear-cut as it is sometimes assumed.

A standard example of metonymic sense extension is the use of a word denoting a place to refer to (some of) the people inhabiting that place (e.g. "village", "palace"). This process seems to involve the foregrounding of one component of the meaning of the place denoting word – we follow Pustejovsky (1989a,b,c) in assuming that the lexical semantic representation of nouns is richer than is standardly assumed in generative linguistic theory and that, in particular, the "qualia structure" for such nouns will contain (telic) information which allows direct access to the information that they are inhabited. A well-known example of metaphoric sense extension is that involving use of a word denoting an animal to refer to humans ("John is a pig", "John is a wombat" etc). Although the sense extension from animals into metaphorical senses denoting humans with some particular characteristic is apparently productive, the actual characteristics involved, and even whether the word can be applied to men or to women or both, cannot be predicted from knowledge of the animal sense. Thus, the properties ascribed to a person by "pig" are arguably no more than stereotypical associations with the animal, rather than central aspects of its meaning / qualia structure. In the case of "wombat" we would argue that the association of foolishness derives from the phonological form of the word, rather than beliefs about the animal. Despite the more associative or analogical nature of metaphorical sense extension, we would argue that there is a core component to such processes which should be expressed in terms of a lexical rule, rather than in terms of general purpose reasoning. As with the metonymic cases (Pustejovsky, 1989c; Briscoe et al., 1990), we believe that the notion of coercion during syntactic and semantic interpretation provides an account of when a metaphorical interpretation will be adopted, and we would like to characterise the limits of coercion in terms of possible mappings defined by lexical rules.

An example of a derivational morphological process is the addition of the "er" suffix to verbs, typically creating a noun denoting the agent of the action denoted by the verb (e.g. "teach", "teacher"). There are several apparent differences between this type of process and the metonymic and metaphoric sense extensions considered above. The derivational rule involves a change of syntactic class, it affects the argument structure of the derived predicate, it involves affixation, and although there is a foregrounding of one aspect of the verb meaning, the result would not traditionally be described as a metonymic, or indeed metaphorical, usage. Nevertheless, there are clearly derivational processes which do not affect syntactic class (e.g. "re-program", "un-reprogrammable") and sense extensions which do; for example, countability of nouns changes depending on whether they are interpreted as types, substances or portions (e.g. "There was beer all over the table", "John drank a beer"). Not all derivational processes affect argument structure (e.g. "un-kind"), whilst metonymic sense extensions (e.g. "John enjoyed the film", John finished the beer") can; at least when given the analysis of Pustejovsky (1989c) and Briscoe et al. (1990). Finally, processes of conversion and derivation can be identical; for example, both "purchase" and "replace" have deverbal nominal forms "purchase" and "replacement", both nouns can denote the action involved and take appropriate complements ("Bill's purchase of his new car", "Bill's replacement of Sue with Mary"), and both can denote the result of the action ("Bill's purchases were many and varied" "Bill's replacement was Sue"). Traditionally, this latter resultative meaning would be described as metonymic and probably specialised and non-productive. We think that our definition of lexical rule will allow an account of both as productive syntactic and semantic operations mapping between lexical entries. The difference between metaphorical and metonymic operations is a matter of the degree to which the lexical rule determines or circumscribes the eventual interpretation.

There are other similarities between sense extension and derivational morphology; clearly, productivity is an issue in both, and in particular, sense extension processes may apparently be blocked (preempted by synonymy), in a way comparable to the situation in derivational morphology (see e.g. Bauer 1983:87f). For example, the regular form "stealer" does not generally occur, apparently because of the availability of "thief". Another productive metonymic sense extension is that of animal denoting (count) nouns to (mass) nouns denoting their meat (e.g. "lamb"), but this process too is blocked by the presence of a synonymous lexeme with different form ("pig", "pork"). By representing such processes in terms of lexical rules mapping between entries, we hope to account for blocking in terms of syntactic and semantic identity with an entry defined without recourse to the relevant lexical rule. In addition, we hope to express sense extension processes, and indeed derivational ones, as fully productive processes which apply to finely specified subsets of the lexicon, defined in terms of both syntactic and semantic properties expressed in the type system.

As part of the ACQUILEX project[1] we are investigating the occurrence of such extended and derived senses in machine readable dictionaries (MRDs). Although many regular relationships are recognised by lexicographers (Ostler and Atkins 1991), the representation of these extended and derived senses in existing MRDs is unsystematic. Thus our lexical knowledge base has to be able to deal with extensions which are never given a separate sense ("John is a wombat"), even if it is not possible to represent them completely. In this paper, we first examine in more detail some of the classes of sense extensions and their realisation in *The Longman Dictionary of Contemporary English* (LDOCE). We then describe the representation language of the lexical knowledge base (LKB) which is being used to represent information extracted from MRDs on the ACQUILEX project, and illustrate how it is possible to represent two of the examples mentioned. The LKB is unification-based; it supports a restricted range of operations; (default) unification, (default) inheritance and lexical rule application and is intended as a lexical rather than a general purpose knowledge representation. However, we are using it to represent a relatively rich range of semantic information which allows us to indicate some of the shifts in meaning exemplified both in sense extension and in derivational morphology and which we hope will allow us to investigate the examples where productive processes are blocked.

2 The representation of sense extensions in dictionaries

We will consider two types of sense extension, both of which apply to words which denote animals. The first is the metaphorical use, mentioned earlier. The second involves the extension to a non-count sense with a meaning which can be paraphrased as "some substance derived from that animal", where this will normally be either fur/skin or meat/flesh. We in fact think of this as a special case of a more general

[1] 'The Acquisition of lexical knowledge for Natural Language Processing systems' (Esprit BRA-3030)

sense extension rule which we will refer to as "grinding". It is well known that any count noun denoting a physical object can be used in a mass sense to denote a substance derived from that object, when it occurs in a sufficiently marked context. We refer to this as 'grinding' because the context normally suggested is the "Universal Grinder" (see Pelletier and Schubert 1986). So if "a table" is ground up the result can be referred to as "table" ("there was table all over the floor"). Several regular sense extensions can be regarded as special cases of 'grinding', where the extension may have become lexicalised. Thus besides the animal/meat examples, trees used for wood ("beech") have a sense denoting the wood, and so forth (see Copestake and Briscoe (1991) for further discussion).

We are investigating the representation of these sense extensions in LDOCE. Both extensions appear productive, at least in a sufficiently marked context; for example in the LOB corpus we find the use of "mole" as a mass term:

> Badger hams are a delicacy in China while mole is eaten in many parts of Africa.

but even when the use seems conventional it may or may not be represented as a separate sense in LDOCE ("lamb" is given a meat/flesh sense but "haddock" is not). Since space for definitions is very restricted in printed dictionaries, senses may be omitted even when the animal is conventionally used for food (the position is also confused by lexicographers' use of various techniques such as bracketing to "collapse" two senses into one). Furthermore, words which primarily denote the meat (eg "pork") normally block the sense extension process. Nevertheless this can often still apply in a more "marked" fashion. For example using "pig" instead of "pork" is possible but not lexicalised, and its use is marked, perhaps suggesting that the meat is of very inferior quality.

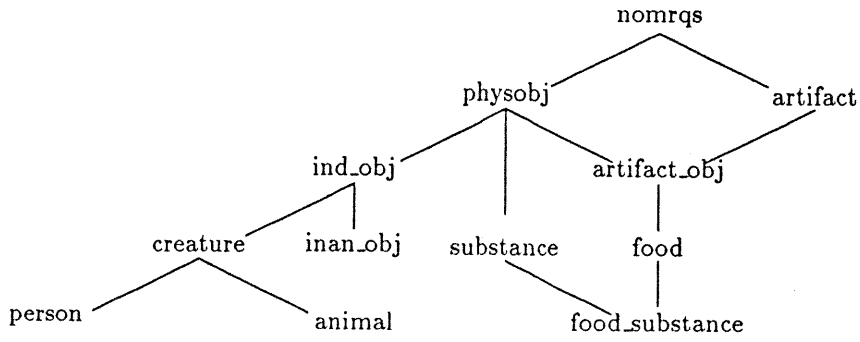
One technique which we are developing to investigate the extent to which regular sense extensions can account for polysemy in dictionary definitions, is to examine exhaustively all cases where a word has a primary sense in one semantic class (such as animal) and a secondary sense in another (such as person or substance). We can achieve this by deriving relatively complete sense-disambiguated taxonomies from the dictionary (using techniques described in Copestake (1990)), starting from the primary senses of words such as "person" and "animal" and intersecting the resulting taxonomies. Thus in the animal/person case there are over 100 words with senses in both taxonomies, the majority of which appear to be metaphorical uses. The other major class involves nominalisations such as "scavenger" being given separate senses (in this case one denoting a creature and the other a person). Furthermore the majority of the polysemy, at least for the class of words denoting animals, appears to be regular.

3 Representation in the LKB

The LKB is designed to represent any information which may be extracted from MRDs as part of the ACQUILEX project, including both syntactic and relatively complex semantic information. The LKB's representation language is based on the use of typed feature structures similar to those described in Carpenter (1990). Feature structures must be well-formed with respect to types and particular features will only be appropriate to specified types and their subtypes. Types are hierarchically ordered; the association of *constraints* with types allows non-default inheritance. We augment this with a restricted concept of default inheritance (allowing only 'orthogonal' multiple inheritance (Touretzky 1986)); default inheritance is formalised in terms of default unification of feature structures ordered by an inheritance hierarchy. The type system constrains both default inheritance and lexical rule application. The LKB's representation language is described in detail in Copestake et al (1991); the following sections are an informal description illustrated with relevant examples.

3.1 The type system

The type hierarchy defines a partial ordering (notated $\sqsubseteq$) on the types and specifies which types are *consistent*. Only feature structures with mutually consistent types can be unified — two types which are unordered in the hierarchy are assumed to be inconsistent unless the user explicitly specifies a common subtype. Every *consistent* set of types $S \subseteq \mathsf{TYPE}$ has a unique greatest lower bound or meet (notation $\sqcap S$). This condition allows feature structures to be typed deterministically — if two feature structures of types **a** and **b** are unified the type of the result will be $\mathbf{a} \sqcap \mathbf{b}$, which must be unique if it exists. If $\mathbf{a} \sqcap \mathbf{b}$ does not exist unification fails. Thus in the fragment of a type hierarchy shown in Figure 1 **artifact** and **physobj** are consistent with $\mathbf{artifact} \sqcap \mathbf{physobj} = \mathbf{artifact\_obj}$.



Our system differs somewhat from that described by Carpenter (1990) in that we adopt a different notion of well-formedness of typed feature structures. In our system every type must have exactly one associated feature structure which acts as a constraint on all feature structures of that type; it subsumes all well-formed feature structures of that type. The constraint also defines which features are *appropriate* for a particular type; a well formed feature structure may only contain appropriate features. Constraints are inherited by all subtypes of a type, but a subtype may introduce new features (which will be inherited as appropriate features by all its subtypes). A constraint on a type is a well-formed feature structure of that type; all constraints must therefore be mutually consistent. Constraints can be seen as extending the PATR-II notion of templates in that the inheritance of constraints allows concise definitions of all feature structures, not just lexical entries; however the type system also prevents inappropriate features occurring.

For example the constraints associated with the types **artifact** and **physobj** might be:

$$\left[\begin{array}{l}\textbf{artifact}\\ \text{PURPOSE} = \textbf{formula}\end{array}\right]$$

$$\left[\begin{array}{l}\textbf{physobj}\\ \text{FORM} = \textbf{physform}\end{array}\right]$$

Bold case indicates types; thus, for instance the "value" **formula** can be any feature structure of this type. The constraint on **artifact_obj** will contain information inherited from both parents, thus:

$$\left[\begin{array}{l}\textbf{artifact\_obj}\\ \text{FORM} = \textbf{physform}\\ \text{PURPOSE} = \textbf{formula}\end{array}\right]$$

Further examples of constraints and features which we will use in this paper are:

$$\left[\begin{array}{l}\textbf{ind\_obj}\\ \text{FORM} = \left[\begin{array}{l}\textbf{physform}\\ \text{SHAPE} = \textbf{individuated}\end{array}\right]\end{array}\right]$$

$$\left[\begin{array}{l}\textbf{creature}\\ \text{AGE} = \textbf{scalar}\\ \text{SEX} = \textbf{gender}\end{array}\right]$$

$$\left[\begin{array}{l}\textbf{animal}\\ \text{EDIBLE} = \textbf{boolean}\end{array}\right]$$

$$\left[\begin{array}{l}\textbf{substance}\\ \text{FORM} = \left[\begin{array}{l}\textbf{physform}\\ \text{SHAPE} = \textbf{unindividuated}\end{array}\right]\end{array}\right]$$

$$\left[\begin{array}{l}\textbf{food}\\ \text{PURPOSE} = \left[\begin{array}{l}\textbf{formula}\\ \text{PRED} = \textbf{eat}\end{array}\right]\end{array}\right]$$

All these types are intended to represent that part of the lexical entry which contains relatively complex semantic information, the "relativised qualia structure" (Calzolari 1991).

We shall also make use of the following types to define syntactic properties etc:

$$\textbf{lex-sign} \sqsubseteq \textbf{top} \quad \begin{bmatrix} \textbf{lex-sign} \\ \text{ORTH} = \textbf{string} \end{bmatrix}$$

$$\textbf{noun} \sqsubseteq \textbf{lex-sign} \quad \begin{bmatrix} \textbf{noun} \\ \text{SYNTAX} = \begin{bmatrix} \text{COUNT} = \textbf{boolean} \end{bmatrix} \\ \text{RQS} = \textbf{nomrqs} \end{bmatrix}$$

$$\textbf{count-noun} \sqsubseteq \textbf{lex-sign} \quad \begin{bmatrix} \textbf{noun} \\ \text{SYNTAX} = \begin{bmatrix} \text{COUNT} = + \end{bmatrix} \end{bmatrix}$$

$$\textbf{mass-noun} \sqsubseteq \textbf{lex-sign} \quad \begin{bmatrix} \textbf{noun} \\ \text{SYNTAX} = \begin{bmatrix} \text{COUNT} = - \end{bmatrix} \end{bmatrix}$$

The feature structure below is well-formed since it contains all the appropriate features and no inappropriate ones, it is subsumed by the constraints on its type and all its substructures are well-formed.

$$\begin{bmatrix} \textbf{count-noun} \\ \text{ORTH} = \text{“haddock”} \\ \text{SYNTAX} = \begin{bmatrix} \text{COUNT} = + \end{bmatrix} \\ \text{RQS} = \begin{bmatrix} \textbf{animal} \\ \text{SEX} = \textbf{gender} \\ \text{AGE} = \textbf{scalar} \\ \text{EDIBLE} = \textbf{boolean} \\ \text{FORM} = \begin{bmatrix} \textbf{physform} \\ \text{SHAPE} = \textbf{individuated} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Given the type system introduced above, a lexical entry in the LKB:

```
haddock 1 count-noun
      <rqs> = animal.
```

would be expanded out into such a feature structure[2].

3.2 Lexical rules

A lexical rule in the LKB is a feature structure of type **lexical-rule**. The expanded constraint for the type is:

$$\begin{bmatrix} \text{lexical\_rule} \\ 0 = \text{lex\_sign} \\ 1 = \text{lex\_sign} \end{bmatrix}$$

thus all lexical rules have to have the features 0 and 1 which must both have values which are of type **lex_sign**.

New lexical signs may be generated by unifying a copy of the lexical entry with the feature structure at the end of the path <1> in a copy of the lexical rule — the feature structure at the end of the path <0> is then the new lexical sign. Lexical rules are indexed by the type of their "input" and "output" feature structures, so they will only be applied to entries of the appropriate type.

A general type for grinding lexical rules could be specified in the LKB as follows:

[2] The actual type system being employed is considerably more complex, since only the relevant features are being shown in these examples.

$$\textbf{grinding} \sqsubseteq \textbf{lexical\_rule} \quad \begin{bmatrix} \text{grinding} \\ 1 = \begin{bmatrix} \text{count-noun} \\ \text{ORTH} = \boxed{1} \\ \text{RQS} = \text{ind\_obj} \end{bmatrix} \\ 0 = \begin{bmatrix} \text{mass-noun} \\ \text{ORTH} = \boxed{1} \\ \text{RQS} = \text{substance} \end{bmatrix} \end{bmatrix}$$

The effect of the lexical rule is to transform a count noun with the "relativised qualia structure" (RQS) properties appropriate to an individuated physical object **ind_obj** into a mass noun with properties appropriate for a substance **substance**.

We specialise the grinding rule to allow for cases such as the animal/meat extension explicitly. The typed framework provides us with a natural method of characterising the subparts of the lexicon to which such rules should apply. The lexical rules can, in effect, be parameterised by inheritance in the type system. For example given the type hierarchy shown in Figure 1 we can then give rules which inherit information from **grinding** such as animal_grinding:

$$\text{animal\_grinding} \quad \begin{bmatrix} \text{grinding} \\ 1 = \begin{bmatrix} \text{RQS} = \begin{bmatrix} \text{animal} \\ \text{EDIBLE} = + \end{bmatrix} \end{bmatrix} \\ 0 = \begin{bmatrix} \text{RQS} = \text{food\_substance} \end{bmatrix} \end{bmatrix}$$

Thus given the lexical entry for "haddock" shown above we can apply the lexical rule to generate a sense meaning "haddock-flesh" (partially represented as):

$$\begin{bmatrix} \text{mass-noun} \\ \text{ORTH} = \text{haddock} \\ \text{SYNTAX} = \begin{bmatrix} \text{COUNT} = - \end{bmatrix} \\ \text{RQS} = \begin{bmatrix} \text{food\_substance} \\ \text{PURPOSE} = \begin{bmatrix} \text{formula} \\ \text{PRED} = \text{eat} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

(where the specification of the value **eat** for the purpose role arises from the constraint on the type **food_substance**, inherited from **food**, and the type mass-noun arises from the grinding type.)

The lexical rule for the metaphorical use of animal words might be something like:

$$\text{animal\_metaphor} \quad \begin{bmatrix} \text{lexical\_rule} \\ 1 = \begin{bmatrix} \text{SYNTAX} = \boxed{2} \\ \text{ORTH} = \boxed{3} \\ \text{RQS} = \begin{bmatrix} \text{animal} \\ \text{SEX} = \boxed{1} \end{bmatrix} \end{bmatrix} \\ 0 = \begin{bmatrix} \text{SYNTAX} = \boxed{2} \\ \text{ORTH} = \boxed{3} \\ \text{RQS} = \begin{bmatrix} \text{person} \\ \text{SEX} = \boxed{1} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

In principle, multiple lexical rules may be applied in sequence. For example another lexical rule which we might specify is "portioning" which would convert food (or drink) denoting mass nouns into count nouns denoting a portion of that substance (eg "three beers", and even "three haddocks" in the context of a restaurant). All outputs of lexical rules must be potentially valid lexical entries. In the case of zero-derivational processes, like those we have been defining, we wish to restrict the set of lexical rules so that application may not be circular — that is if there is a lexical rule which could generate the set of feature structures F2 from the set F1, no other lexical rule or sequence of lexical rules may be specified which

could generate any member of F1, or a feature structure subsuming any member of F1, starting from any member of the set F2, since lexical rule application would not then terminate. One straightforward way of ensuring this condition is met is to constrain the set of lexical rules so that if a rule which transforms feature structures of type t_1 to t_2 exists no sequence of rules exists which transforms t_2 to t_1 or any supertype of t_1. However, this condition is overrestrictive because some types of derivational rule can apply to their own output iteratively ("meta-meta-theory", "anti-anti-missile", "great-great-grandmother", "re-re-program"). Such differing properties of lexical rules can be formalised as they emerge by creating distinct types of lexical rule and placing well-formedness constraints on each type. Expressing constraints of this type will require a more expressive language than we currently deploy in the LKB, in the spirit of that described in Carpenter et al. (1991). The richer these constraints the more a substantive theory of lexical processes will emerge within the formal computational framework outlined above.

3.3 Default inheritance

To allow default inheritance we introduce the concept of *psort*; a feature structure from which another feature structure inherits information, by default. The hierarchical ordering on psorts (which must be consistent with the type hierarchy) provides an order on defaults. Default inheritance is implemented by a version of default unification. Only orthogonal multiple inheritance (Touretzky 1986) is allowed; information inherited from multiple parents must not be contradictory. A default inheritance hierarchy which connects semantic parts of lexical entries can be derived semi-automatically from taxonomies extracted from MRDs (Copestake 1990). Defaults may also be useful in the representation of syntactic information in the lexicon (e.g. Flickinger, 1987)

The use of default inheritance in which we are primarily interested here is to allow more specific information about an extended word sense extracted from an MRD to augment and possibly override information inherited from the result of application of a lexical rule. In cases where the lexical rule predicts the extended sense correctly, the specific information will duplicate information already present. If the rule is correct, but incomplete, the specific information will augment the inherited information. If it is partially incorrect, the more specific information will override that inherited from the result of lexical rule application.

For example, if the result of applying the lexical rule to a sense is notated as sense+rule-name (eg lamb_1+animal_grinding) then the representation of the LDOCE sense lamb (2) (the meat) in the LKB might be:

```
lamb 2 < lamb_1+animal_grinding.
```

where < is used to indicate default inheritance from a psort. In this case no extra information need be added. In contrast the entry for lamb (3) ("a young gentle person") might augment the information inherited from the lexical rule:

```
lamb 3 < lamb_1+animal_metaphor
< rqs : age > = low.
```

In the case of haddock, where no LDOCE entry is found, the structure derived from the lexical rule alone would be used.

Both the types of sense extension that we have been considering seem to be productive. Examples such as:

> Don't be such a wombat!

(said to someone behaving stupidly, in the speaker's opinion) seem to be perfectly acceptable, but such a use of "wombat" is probably not found in any dictionary. If we are to attempt to process such utterances, the structure generated by the lexical rule itself has to be sufficient. This can give the information that "wombat" may denote a person, but not what the characteristics imputed are, since these are not predictable. However, most such uses occur in marked contexts where it is relatively obvious (to a human reader) what characteristics are meant (and in any case information from the lexical rule alone would be sufficient for many NLP applications). Such lexical rules should only be invoked in a suitably marked context and we have argued elsewhere, on the basis of corpus data, that coercion is triggered by predicational environment and subsequent overriding of the default metonymic interpretation (or perhaps elaboration of the metaphorical one) is triggered by marked, informationally-rich contexts (see Briscoe et

al., 1990; Copestake and Briscoe, 1991). Below we give an example of metaphorical use of "shark" taken from the SEC corpus which illustrates these (and some other) processes. (Note that the sense of "shark" involved here is not related to the arguably distinct homograph relating to (financial) malpractice or criminality.)

> Nowhere more so than on the world's golf courses, where a man they call "the great white shark" was on the prowl... That one little tap in at Turnberry put paid... to the cruel jibe that Greg Norman was just a great white fishfinger. His transformation to a genuine killer shark began on that windswept course...

4 Conclusion

We have illustrated how some cases of metonymic and metaphorical use of words might be at least partially represented in a lexical knowledge base system, with much the same apparatus as is needed for the representation of derivational morphology. Our work on using MRDs to investigate this behaviour is at an early stage; we have illustrated how relatively crude techniques like taxonomy intersection can be used to get at some of the classes of words which undergo this behaviour, but in order to do this more completely we need to use more sophisticated extraction techniques. The (partial) representation of information derived from MRDs in the LKB should give us a way to investigate such sense extensions as well as being practically useful in NLP systems.

It is at least arguable that much of the information that we are representing here is real world knowledge rather than lexical information, and it is clear that our representation is only partial; both because we are simplifying information from the definitions and forcing it into limited templates provided by the type system and because the LKB's limited operations fall far short of an inference system that is capable of reasoning about real world knowledge. We have argued elsewhere (Briscoe et al (1990), following Pustejovsky (1989)) that lexical processes do involve far more semantic information than is generally represented in the lexicons of NLP systems, but a further practical argument for attempting this sort of representation is that it is needed to detect when lexical rule application may be "blocked" by the existence of other word forms.

As mentioned above the phenomenon of blocking appears to occur with some cases of regular sense extension in a way that seems similar to derivational morphology. In the cases where a word sense exists such as "pork" we need to recognise its equivalence with the sense obtained from "pig" using the lexical rule. To do this we need a rich representation which indicates information such as "origin" which has not been mentioned above. Because words like "pig" can be used in the extended sense, in marked contexts, the correct way to represent blocking in the LKB is not to prevent application of the lexical rule but the generated occurrence should be noted as being in some way peculiar. Sense extensions may also be blocked if a phonologically similar word exists which could be confused with the extended sense; for example it appears unusual for "deer" to be used in a metaphorical sense presumably because it could be confused with "dear"; again this phenomenon apparently also exists in derivational morphology (Bauer 1983:97). However these effects are not well-understood — for example blocking does not seem to apply to the metaphorical sense extensions in the same way (for example **colt (2)**, **fledgling (2)** and **greenhorn (1)** are given very similar definitions in LDOCE). Rather than attempting to decide on a representation at this point we intend to investigate the effect on some comprehensive set of examples, using the MRDs, to try and establish a suitable treatment.

References

Bauer L(1983) *English Word-formation*, CUP

Briscoe E J, Copestake A A and Boguraev B K(1990) 'Enjoy the paper: Lexical semantics via lexicology', *Proceedings of the 13th Coling*, Helsinki, pp.42–47

Cahill L(1990) 'Syllable based morphology', *Proceedings of the 13th Coling*, Helsinki, pp.48–54

Calzolari N(1991) 'Acquiring and representing semantic information in a lexical knowledge base', *Proceedings of the ACL SIGLEX Workshop on Lexical Semantics and Knowledge Representation*, Berkeley, California, pp.188–197

Carpenter R(1990) 'Typed feature structures: Inheritance, (In)equality and Extensionality', *Proceedings of the Workshop on Inheritance in Natural Language Processing*, Tilburg, pp.9–18

Carpenter R, Pollard C and Franz A(1991) 'The specification and implementation of constraint-based unification grammars', *Proceedings of the Second International Workshop on Parsing Technologies,* Cancun, Mexico, pp.143–153

Copestake A A(1990) 'An approach to building the hierarchical element of a lexical knowledge base from a machine readable dictionary', *Proceedings of the Workshop on Inheritance in Natural Language Processing,* Tilburg, pp.19–29

Copestake A A, de Paiva V C V, Sanfilippo A and Briscoe E J(1991) *Functionality of the LKB,* Ms. Computer Laboratory, University of Cambridge

Copestake A A and Briscoe E J(1991) 'Lexical Operations in a Unification Based Framework', *Proceedings of the ACL SIGLEX Workshop on Lexical Semantics and Knowledge Representation,* Berkeley, California, pp.88-101

Flickinger D(1987) *Lexical rules in the hierarchical lexicon,* PhD dissertation, Stanford University

Ostler N and Atkins B T S(1991) 'Predictable Meaning Shift: Some Linguistic Properties of Lexical Implication Rules', *Proceedings of the ACL SIGLEX Workshop on Lexical Semantics and Knowledge Representation,* Berkeley, California, pp.76–87

Pelletier F J, and Schubert L K(1986, forthcoming) 'Mass Expressions.' in Gabbay and Guenthner (eds.), *Handbook of Philosophical Logic, Vol 4,* Reidel, Dordrecht

Pustejovsky J(1989a) *The Generative Lexicon,* Ms. Brandeis University

Pustejovsky J(1989b) 'Current issues in computational lexical semantics', *Proceedings of the 4th European ACL,* Manchester, pp.xvii–xxv

Pustejovsky J(1989c) 'Type coercion and selection', *Proceedings of the West Coast Conference on Formal Linguistics,* Vancouver

Touretzky D F(1986) *The mathematics of inheritance systems,* Morgan Kaufmann, Los Altos

Metaphor and Literal Meaning*

Kenny Coventry
Centre for Cognitive Science,
University of Edinburgh,
2 Buccleuch Place,
Edinburgh
EH8 9LW
Telephone 031 650 4409
031 650 1000 (switchboard)
Fax 031 662 4912
E-mail : kenny@uk.ac.ed.cogsci

Abstract

The relationship between literal meaning and metaphor are examined in light of work on the semantics of spatial prepositions. The theories of Searle, Davidson and Hesse on the relationship between metaphor and literal meaning are considered, and it is argued that their positions are not in fact as distinct as is generally believed. In particular, their conception of literal meaning is criticised as being inadequate (with the possible exception of Davidson). Some work from the semantics of spatial prepositions is used to illustrate the creativity of literal language. Finally, it is argued that metaphor can be viewed as exhibiting the creative properties of literal meaning, but to a greater extent. This leads to a new perspective on metaphor.

*This work has been supported by an ESRC studentship, award No. C00428822003. My sincere thanks go to Ehud Rahat for his many helpful discussions and other members of the Centre for Cognitive Science Metaphor Workshop and members of the Human Communication Research Centre Spatial Expressions Group. Thanks also to Steve Finch for reading and commenting on an earlier draft. All errors, of course, are entirely my own.

Metaphor and Literal Meaning

Our aims in this paper are to focus on metaphor as a linguistic entity, although the delineation of what a linguistic entity is will be an issue which will emerge in our discussion. In particular we will demonstrate that the three main theories of metaphor, those of Searle, Davidson and Hesse, are not in fact as distinct as is generally believed*. We will argue for centrality of metaphor in the construction of reality, and in the understanding and use of language as a means of communication. Thus, Aristotle's view that metaphor is not essential, but nice, somewhat reduces the significance of metaphor. Furthermore, theories which emphasise the importance of literal language over more figurative language fundamentally misconceive what the complexities of literal language are. One of our conclusions will be that literal language and metaphor are not as distinct as one may think.

Searle : Literal Meaning and then Metaphor

Searle 's (1979) position on metaphor is very much derived from the assumptions he makes about literal language, which he views as primary. His account rests largely on the distinction between sentence/word meaning and utterance/speaker's meaning. The literal meaning of a sentence for Searle is the sentence meaning which is realised in the semantic structure of a sentence and is truth conditional. The literal use is characterised by the case where what the speaker intends to convey is the same as what the sentence means. Thus a hearer understands, in the literal case, what the speaker means to convey by computing the sentence meaning. By contrast, the metaphorical meaning of a sentence is different from the literal meaning, but is still the speaker's meaning. For example, the sentence, 'Tom is a chicken' uttered by a speaker literally means, according to Searle, that Tom is a farmyard animal/bird with the label chicken. The metaphorical meaning, which is quite different from the literal meaning, somehow has to communicated by the speaker and comprehended or worked out by the hearer, For Searle, the problem of metaphor is to specify how this occurs.

Presumably there are some common principles shared by speakers and hearers that enable a speaker to communicate to the hearer the metaphorical utterance. The problem for Searle is one of, "how is it possible for the speaker to say metaphorically that 'S is P', and mean 'S is R', when P does not mean R" (Searle, p. 103). A series of steps are presented which the hearer can go through is order to get to the correct paraphrase for the metaphorical expression, or intended meaning of the speaker. This begins with the literal meaning of the sentence (truth conditional meaning). With our case of 'Tom is a chicken', the hearer has to first realise that the truth conditions of this utterance do not hold, so the literal meaning is false. Having rejected the literal meaning, the hearer can apply principles of metaphorical interpretation. One such step is to look for substitutions for 'chicken' which can be salient properties to do with chickens. Having examined properties of chickens, and possible substitutions, the hearer then reexamines the S term ('Tom' in our example) and works out which of the properties narrowed down of R fit with S.This three-step strategy involving principles of metaphorical interpretation are "individually necessary and collectively sufficient to enable speaker and hearer to form and comprehend utterances of the form 'S is P', where the speaker means metaphorically that S is R, where P≠R" (Searle, p. 112) .

Searle's account is riddled with problems. He relies very much on a distinction between sentence meaning and speaker's meaning. However, can metaphorical expressions, like literal expressions, not have meaning connected with the words uttered, rather than being mysteriously branded speaker's meaning ? Cooper (1986) thinks little of this distinction. He argues that the problems lie as a clash between the idea that a metaphor means what a speaker wants it to mean and the specification of principles that should reveal what the speaker means. The idea that speakers use such principles to form metaphorical utterances

** Limitations of space render it impossible to review the abundance of approaches to metaphor in the literature. This leads to the omission of consideration of the work of Black (1978), Gibbs (1984), Lakoff and Johnson (1980), Miller (1979), Kittay (1987), MacCormac (1985) among others (see reference section below).*

is wrong as speakers may intend to convey anything by using a linguistic expression. Examples of these are malapropisms (see p.3), in which the literal meaning (in Searle's terms) bear absolutely no resemblance to the speaker's intended meaning. It follows from this that the hearer can't be using the same principles. We must therefore look elsewhere to explain how successful communication takes place. The understanding of metaphors isn't obviously guided by rules.

Searle, as should be clear from above, presents the principled relations between the metaphorical meaning of an expression and its literal meaning as having to do with the comprehension process. It is not at all clear why Searle does this. It is far simpler just to state that the relations map metaphorical meaning on to literal meaning. Furthermore, this illustrates the fact the Searle's account only deals with dead metaphors. Novel metaphors are not dealt with. There may indeed be something systematic about metaphors that have stuck, or may stick in our language, but there are also metaphors that are used only once, and may exhibit considerable irregularity. The focus on irregularity leads us into a discussion of Davidson's account. We will return to Searle later.

Davidson : Passing Theories and First Meanings

Davidson observes that there are many utterances on-line which are irregular without communication being effected in any way. Hence, Davidson reacts against the definition of meaning as relating to generalisations in the speakers' community. Literal meaning has to be explained in a way that doesn't depend solely on regularities in communication. Standard meanings for Davidson as those one would find in a dictionary. Literal meaning is what he calls 'first meaning'. This is the first intention of a speaker when he utters words and sentences on a particular occasion.This first intention is that the hearer will interpret the utterance if, and only if, the utterance is true. For example, if the speaker utters the sentence, 'clowns are funny', the literal meaning is that 'clowns are funny'. Thus the first intention is to have the speaker getting the hearer to interpret an utterance in a particular way. After there are other intentions, which are at a higher level, such as causing the hearer to buy tickets to go to the circus. Semantics, in Davidson's view, concerns itself only with first intentions.

Communication for Davidson is very much like a game (e.g., poker, involving bluffs, varying playing strategies, etc.). Before the start of a conversation, the hearer has a semantic theory which he believes will be adequate for the interpretation of a specific speaker's utterances. This is what is termed the hearer's prior theory. If the speaker is unfamiliar to the hearer, the hearer may use standardisations across a linguistic community as a prior theory (i.e., the literal meaning in Searle's terms). However, the prior theories vary greatly depending on who the hearer is about to communicate with. Once the communication is underway the prior theory originally adopted by the hearer may not be working sufficiently well, so adjustments have to be made. A new theory is then adopted by the hearer (a new semantic theory) which Davidson calls a passing theory. If the passing theory is successful, it may become a new prior theory (for further communicating with that person). If not, the hearer can resort back to the previous prior theory. Thus successful communication occurs when speaker and hearer have the same (or equivalent) passing theories. It is not a big step to see that linguistic competence for Davidson is flexibility in the adjustment of passing theories.

Metaphor involves explanation at higher levels of intention than that of first intention. When a speaker uses a metaphor, he knows what the words standardly mean, and intends for them to interpreted in this way. The interpretation in the metaphorical sense occurs at a higher level of intention, and therefore is outwith the realm of semantics. Metaphors do not always result in changes in semantic theories. These can be contrasted with malapropisms, (e.g., when Mrs Malaprop utters ' a nice derangement of epitaphs' when she means, 'a nice arrangement of epithets') in which a change in semantic theory is necessary in order to understand the speaker's first intention. The change necessary here is in response to one utterance, and therefore takes the form of a passing theory. With metaphor, however, a change is likely to occur in the prior theory as the hearer notices the repetition of the metaphor. Another contrast between metaphor and malapropisms is that metaphors are intended to be interpreted as a false sentence by a speaker. Malapropisms, slips of the tongue, etc, are not intended to be interpreted in the wrong way.

Davidson is keen to point out the creative processes involved with the understanding of metaphor, malapropisms and similar figurative language. In the case of malapropism, the hearer has to be creative to adopt a passing theory in order to get at the speaker's first intention. In the case of metaphor the creativity comes in understanding the speaker's meaning (the higher levels of intention). This position is very different form the systematic rules for metaphorical interpretation proposed by Searle. Davidson, in stark contrast, places metaphor firmly outside the role of semantics, and argues that metaphor has to do with language use (rather than meaning). Metaphors use false sentences to achieve something different from the expression of the belief in what is uttered. The creation in the understanding of metaphor is separate form the actual understanding, the semantics, of the first meaning. This creation evades systematicity, and thus firmly places itself in the domain, as already stated, of language use.

Hesse : The Explanatory Function of Metaphor

Hesse views metaphor as fundamental to scientific explanation and development. Her thesis centres on the belief that 'the deductive model of scientific explanation should be modified and supplemented by a view of theoretical explanation as metaphoric redescription of the domain of the explanandum'. She proposes (following Max Black) an interaction theory of metaphor. Such a view consists of 'primary' and 'secondary' systems in that properties from one system are attributed to another. Both primary and secondary systems carry with them a set of associated ideas or beliefs that come to mind when the system is referred to. These beliefs and ideas are common to speakers of a given language.

Like Searle and Davidson, Hesse points out the necessity of literal falsehood in order for the metaphor to be successful. From this there is initially some principle of assimilation between primary and secondary systems. The metaphor works by transferring the associated ideas and implications of the secondary system to the primary system. Even the original literal description is shifted in meaning. The primary and secondary systems are thus seen to interact with each other and the result is that each individual system is somehow changed. This viewpoint is central to the interaction view and is inconsistent with the comparison view of metaphor which assumes that the literal descriptions for both systems are, and remain, independent of the use of the metaphor and the metaphor is irreducible to them.

Hesse's work centres round the desire to understand how explanation works and changes in science. To give an example, one can look at the model in physics where, "Sound (primary system) is propagated by wave motion (taken from a secondary system)." Hesse argues that the secondary system allows the primary system to be viewed in a new way. Similarly, the secondary system is itself seen in a new way as a result of the application to the primary system. Thus wave motion is seen is a new light, as well as sound (excuse the pun).

A metaphorical expression used for the first time, according to Hesse is intended to be understood. A metaphor is nonsense if it communicates nothing. Indeed, this implies the rejection of all views that make metaphor a wholly noncognitive, subjective, emotive, or stylistic use of language. Intelligible metaphor also implies the existence of rules of metaphorical use, and since literal meanings are shifted by their association with metaphors, it follows that the rules of literal usage and metaphoric interpretation, though they are not identical, are nonetheless related.

General Discussion : Metaphor and Literal Meaning

The above accounts appear to treat metaphor quite differently. On the one hand, Searle and Hesse converge on the view that there are principles of metaphoric interpretation (although they differ as to how these should work) and on the other hand Davidson views metaphor as outside the domain of semantics completely and as a phenomenon of language use. However they all share what we view as a simplistic notion of literal meaning. In this discussion we will not start with the assumption that literal meaning is different from metaphor. Instead we will focus on the similarities between the two. This will lead us to a new theory of why metaphors exist, which furthermore resurrects inference and experientialism in lexical semantics.

Most semantic theories assume that sentence or utterance meaning is computed by combining the meaning of its parts some set procedures. This is the principle of compositionality devised by Frege. In order to make the principle precise, one must give a specification of at least the following :

(a) The nature of the meaning of the smallest parts - i.e., a theory of lexical semantics :

(b) The relevant whole-part structure of each complex expression - i.e., a theory of the semantically relevant level or levels of syntax ;

(c) The "functions" in question - i.e., a theory of what combinatorial semantic operations there are, and how the rules for combining meanings operate on lexical meanings and syntactic structure to produce the meaning of the whole - in short, a theory of compositional semantics.

When spelt out like this the complexities of literal meaning begin to emerge. Let us begin with attempts to give a theory of lexical semantics. Firstly, we can examine cases where the principles do not seem to hold. Expressions such as *to have a bee in one's bonnet, to kick the bucket, up the creek*, etc. are semantically peculiar, and are usually described as idioms. Idioms are traditionally viewed as expressions whose meaning cannot be inferred from the meanings of its parts. The part meanings in question are assumed to be lexical entries. As the individual lexical entries (i.e., the semantic part of the entries) cannot be combined to come up with the right meaning, the whole idiom must be stored in the lexicon separately. This, of course, assumes the notion of standard meanings of some sort in a speakers' community. Indeed, linguistics takes as its subject matter standardisation in speakers' communities, which is reflected in the search for universal grammar, a set of principles common to all natural languages. Idioms in languages are quite common, but at this point we may not see literal language as threatened in any way. Idioms can be stored separately in the lexicon, and that is that.

Let us turn our attention to two examples which are not clearly idiomatic :

(1) 'the ham sandwich walked across the room'
(2) 'that cat sure can play'

The first example (cited by Nunberg, 1978) is a case of metonymy in which a waiter uses the description 'ham sandwich' to refer to a customer who has ordered a ham sandwich. The lexical entry for 'ham sandwich' obviously can't contain the entry 'man'. Therefore, the literal meaning of the sentence (in the conventional sense) is false. However, this expression is not an idiom, and will not be stored separately in the lexicon. This is true especially by virtue of the fact that we understand a multitude of novel utterances which are instances of metonymy (the customer could have ordered anything!).

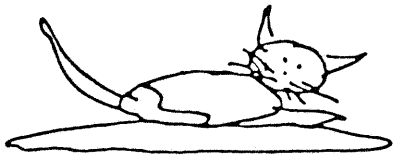
The second example is interpreted differently depending on whether one is in a jazz club or a cattery. However, it is unlikely that the lexical entry for 'cat' contains the entry 'man'. This case is interesting as it is not clear whether or not it an instance of metonymy, idiomatic or whatever. In either context the meaning is obvious, and it appears unlikely that one has to go through the literal meaning (if the lexical entry for cat is consulted) to get to the meaning in the context of a jazz club. Therefore, it would appear more sensible to view inference as playing a role in understanding the utterance, in other words, one has to work out in both cases what the word 'cat' refers to. Thus one can reject the standard pragmatic model of comprehension as embodied by Searle. Additionally there a number of reaction time studies which suggest that metaphorical comprehension is no slower than comprehension of literal meaning (e.g., Gerrig, 1989: Gerrig and Healy, 1983: Gildea & Glucksberg, 1983: Ortony, Schallert, Reynolds & Antos, 1978 ; see Hoffman & Kemper, 1987 for a review.).

From these examples it is clear that inference plays an important part in the understanding of sentences in contexts. It is our viewpoint that this is more important than the specification of lexical entries for words. It is simply impossible to list in a lexicon all the possible senses of a word as there are an infinite number (see Clark, 1983 for a discussion) and the lexicon is surely finite. We must view creativity as a necessary feature of language use and understanding. It is our view that the application of the computer

metaphor of mind has led to the simplification of natural language processing through the need to represent meaning in terms of simple rules and relation. Before examining an alternative to the standard pragmatic model of comprehension, we will turn to some more examples.

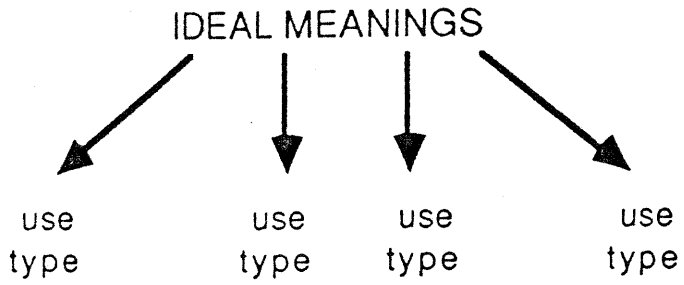
The now cliched case of 'the cat is on the mat' surely represents one of the most obvious and secure of literal expressions. Despite this, locative expressions like this provide many difficulties which are not merely products of our lack of knowledge about how lexical entries are combined. Searle acknowledges some of these problems in that he recognises 'background conditions' with literal language. The truth conditions of sentences (i.e., for a large class of unambiguous sentences like "the cat is on the mat") vary with variations in these background assumptions. Thus, depending upon the satisfaction of the background conditions a sentence may or may not have a determinate truth value. For example, in this case of "the cat is on the mat" one assumes that there is gravity present. Thus is one was on a spaceship near the moon and a cat and mat passed the window in the spatial relations as depicted in figure 1, one would be uncertain as to whether one can say that 'the cat in on the mat'. However, what does *on* mean in the sentence. To take an equally banal example, one can take the example of 'pear is in the bowl'. What does *in* mean in this sentence ?

Figure 1



Various theorists have posited different mental representations for the meaning of *in*. Let us examine one or two of these theories. Firstly. Annette Herskovits (1986) proposes that word meaning is defined in an ideal world - in the spatial domain a world of lines, points, surfaces and of definite relations of inclusion, contact, intersection and so forth. These 'ideal meanings' are stretched to describe and communicate facts about the complex and imperfect world in which we live. Thus, the basic idea is that the preposition *in* means 'inclusion of a geometric construct in a one-, two-, or three-dimensional construct'. From this ideal meaning 'use types' can be derived (see figure 2). Herskovits proposes many different use types for *in*. Two examples of these are "spatial entity in container" and "gap/object 'embedded' in physical object".

Figure 2



Hence, the preposition has two different uses (meanings) in the following sentences ;

The flowers are in the vase
The crack is in the vase

This analysis is flawed in a number of ways. The notion of ideal meaning lacks justification. There would appear to be little point is using such an abstract construct as central in the representation of meaning, particularly as use types vary often considerably from the ideal meaning. However, there are problems which are more readily demonstrable. *On* and *in* for example have a use type that is identical, namely "accident/object part of physical object", but the prepositions don't have the same distribution :

The freckles on his face
*The freckles in his face

The handle on the basket
*The handle in the basket

The mark on the scale
*The mark in the scale

Furthermore, many of the use types are not in fact separate. With the examples in above we argue that the meaning of *in* in each case is the same. It is the meaning of the whole expression which leads to a different spatial conceptualisation, and not the meaning of the preposition in isolation. This has been demonstrated empirically by Coventry (forthcoming), cited in Coventry and Ludwig (1991). Subjects' pattern of responses to the pictures and sentences below (figure 3) were as follows.

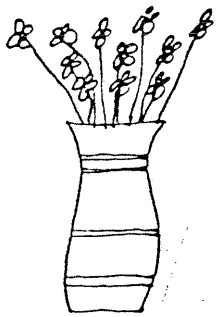
People were most likely to complete sentences (1) and (2) with the preposition *in*. When presented (on a different occasion) with the complex picture in (3), they were quite happy to use *in* in both sentences, although Herskovits cites them as distinct use types. The explanation for this is that they are the same use type. It is not the meaning of the preposition which is varying in the sentences, but the meaning of the whole locative expression. It is the nature of cracks and flowers at the conceptual that leads to what appears to be different meanings for in. Cracks can only be in surfaces in one particular spatial relationship, and flowers can be viewed in the same way (a different spatial relationship with the vase to the crack). Thus it is ridiculous to imagine a crack sticking out of the top of a vase in mid air in the way that flowers do when they are in vases.This finding is indeed stronger when one compares this to other prepositions in which subjects did switch when presented with the complex picture (*at* used in the sense of functional control (i.e., man interacting with piano) versus simple geometric relations (positioning) : see figure 4.

Figure 3

(1)

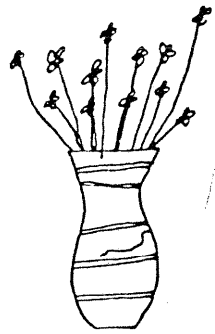
The crack is___the vase.

(2)



The flowers are___the vase.

(3)



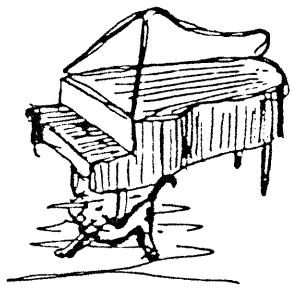
The flowers are____the vase.
The crack is____the vase.

Figure 4

(1)

The man is___the piano.

(2)



The cat is___the piano.

(3)

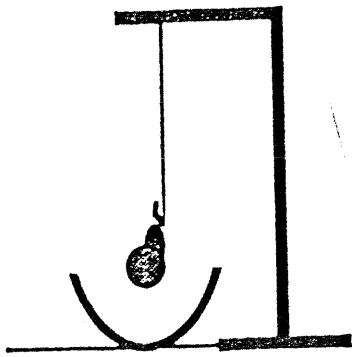
The man is___the piano.
The cat is___the piano.

This evidence concurs with the theoretical position proposed by Lang (1990). He argues for a distinction between the levels of semantic form and conceptual structure. With our example, this means that how we conceptualise flowers, cracks and vases will determine at one level how the whole sentence or utterance is interpreted. This enables a considerable amount of flexibility in the understanding and use of language. If we return to our example with the meaning of the preposition *in* we can think of contexts in which one can say that 'the crack is in the vase' and it is sticking out the top of the vase. This would be at home in a Monty Python sketch for instance.

We are now arriving at the point where the boundary between what is generally referred to as literal meaning and metaphor is fudged. Literal language, following the Lang distinction, involves processes of inference which are necessary in order to combine lexical units in a sentence. This needs to be the case as the lexicon is only finite. Even if it was infinite, and all the different senses were stored in the lexicon (i.e., lexemes; see Cruse, chapter 3), it is not clear how one would select the right sense to combine with other elements in a sentence. Having a level of conceptual structure allows different conceptualisations of objects to be combined in a sentence without some senses being prior in the traditional way.

To move to another example (taken from Garrod and Sanford, 1990), let us look at the sentence, 'the pear is in the bowl'. We can assume that *in* means something like functional containment. However, we can imagine a context in which things look a little more complex. We can imagine a game in which one has to move a frame (attached to a piece of string and a pear) in such a way as to get the pear and bowl in the spatial relations as depicted in figure 5. In these circumstances one can say at the end of the game that the pear is *in* the bowl. This is problematic as the sentence is false in the null context. The meaning of *in* is simply not met.

Figure 5



More explicitly, without the context of the game subjects will not say say that the 'pear is in the bowl' when simply presented with figure 5. The pear simply is not *in* the bowl in this case. We have to ask how it can possibly be *in* the bowl in the context of the game. This is more problematic than Searle's background conditions as the sentence in isolation is false, not simply indeterminate. Two possibilities exist to deal with this. Firstly, the sentence may be interpreted as false in the null context and then somehow put into context through incorporation into some sort of discourse model. Alternatively, the model may be used to interpret the sentence in the first place. A discourse model is some sort of representation we build up of a text or discourse (we will have no more to say here ; see Garrod and Sanford for a discussion).

What is clear is that both options involve conceptualisation and inference procedures. Thus we can view our discussion of conceptual information as a necessary factor involved with meaning across sentence boundaries, and not just within sentence boundaries. Consequently we can return to the characterise of literal meaning and metaphor. As already stated, Searle, Davidson and Hesse all view the literal falsehood of a metaphor as an essential feature. However, our examples demonstrate that literal sentences are not obviously true or false. This is especially true of cases where the meaning of a sentence is governed by the context, as is the case of our pear and bowl example.

The work presented here differs from the approaches to spatial prepositions of Bennett (1975), Brugman (1981), Lakoff (1987), Miller and Johnson-Laird (1976) and Herskovits (1986). The main difference is that, with the exception of Bennett (1975), our approach favours minimal specification of the lexical entries for spatial prepositions. The other difference is that the current approach endeavours to come up with psychologically motivated distinctions between senses of words, rather than merely sketching occurrences purely descriptively. One such factor which we argue provides a motivation for a distinction between senses is that of whether the preposition described dynamic relations or static relations. Thus we found an experimental difference with at(functional/dynamic) versus at(positioning/static) above, but no difference with two difference occurrences of in(functional/dynamic) and on(functional/dynamic).

The conception of literal meaning we subscribe to is nearer Davidson's characterisation than that of Searle and Hesse. Standard meanings can be used to understand an utterance if unfamiliar (and indeed go towards an understanding of familiar utterances too), but with familiarity with the use of language of a particular speaker (i.e., a context) the context is liable to be used in the first instance. This is a creative process, which I believe can be formalised, but doesn't obviously lie within the framework of truth-conditional semantics. One can view the understanding of an utterance as a combination of contextualised effects on lexical meanings (such as 'cat' in the context of a jazz club) plus standard meanings ('cat' in the null context).

Moving to metaphor, it should be obvious by now that the principles of metaphoric interpretation mentioned above can easy be seen to apply in the cases of literal language at the conceptual level of representation. Thus the difference between literal and metaphorical language in this respect is one of degree. In the case of metaphor the sentence is obviously false, and it is especially obvious to the hearer that inference procedures are going to have to be used in order to get to the speaker's meaning. Furthermore, the viewpoint of Hesse that primary and secondary systems which mutually effect each other holds with literal language. With pears and bowls, the locative relation *in* is determined by the conceptualisation of pears, bowls, and the transfer and integration of this conceptual material from one nominal to another.

An example will help to illustrate our argument. We can consider the sentences below:

(i) The cat is on the mat.
(ii) The pear is in the bowl.
(iii) Stan is on the bottle.
(iv) Ollie is in a drunken stupor.

What we wish to claim is that the prepositions in the above sentences all have minimally specified entries. Furthermore, the entries in (i) and (iii), and (ii) and (iv) are identical. *On* means functional support, and *in* means functional containment (cf Garrod and Sanford, 1990). The lexical entries will embody this information. With (i) and (ii) what one must do is to get to the interpretation of the sentence by examining the information about the subject and object with respects to the control relation specified in the lexical entry for the preposition. This requires, as has been already said, knowledge about the function of objects, etc. Moreover, it is necessary to cross check information about both subject and object before the meaning of the whole sentence can be reached.

In the cases of (iii) and (iv) exactly the same procedure will ensue, only it is harder to arrive at the full expression. In the case of (iii), Stan's behavioural demeanour is being supported by an alcoholic binge (functional support). In the second state, the rather stronger notion of control (that of functional containment) suggests that one is less likely to be able to stop the habit. Now, in both cases the lexical entries are the same as in (i) and (ii), but the interpretation of the whole expression is harder.

This viewpoint gives us a new perspective on metaphor, and why they exist. The thesis is that metaphors exist in order to develop and hone the inference procedures and conceptualisations necessary for the understanding of literal language (which isn't obviously truth-conditional). One can draw an analogy with

playing a musical instrument. In order to develop technical expertise on a piano, for example, it is necessary to practice specific exercises such as scales and arpeggios. These devices have little musical content, and simply serve as a mechanism to improve expression and communication of musical ideas which are formed differently. One can take this metaphor further and argue that the scales and arpeggios themselves form part of the musical content which is present in a musical communication.

Metaphor has one role of developing inference procedures which are necessary for the understanding of everyday language which serves a non--figuratuve function. One can imagine taking a magnifying glass to these processes and arriving at metaphor. However, we do not claim that this role for metaphor can explain wholly the existence of metaphors. Hesse's emphasis on metaphor as a fundamental tool for development in science is, we believe, true. Nevertheless the use of metaphors in science as a basic tool goes hand-in-hand with the view that within an individual theory there are themselves metaphors and assumptions which cannot be falsified. With respects to language, we hope to have demonstrated that literal meaning and metaphor are similarly not far apart. Thus Davidson's quote that metaphor is 'the dreamwork of language' is misleading. It is maybe best to view literal meaning instead as analogous to daydreaming while metaphor is dreamtime during sleep. As humans we sleep to dream, and we simply cannot survive without sleep.

References

Bennett, D. C. (1975). *Spatial and temporal uses of English prepositions: an essay in stratificational semantics*. London: Longman.

Black, M. (1979). Metaphor. In A. Ortony (ed.), *Metaphor and Thought*. Cambridge University Press.

Brugman, C. (1981). *The story of 'over'*. M.A. Thesis, University of California at Berkeley. Reprinted by the Indiana University Linguistics Club.

Carbonnell, J. G. (1982). Metaphor: An inescapable phenomenon in natural language comprehension. In W. G. Lehnert & M. H. Ringle (Eds.), *Strategies for natural language processing*. Hillsdale, NJ: Erlbaum.

Clark, H. H. (1983). Making Sense of Nonce Sense. In d'Arcais, G. B. F. and Jarvella, R. J. (eds.). *The Process of Language Understanding*. Chichester: John Wiley and Sons. 297-331.

Cooper, David E. (1986). *Metaphor*. Aristotelian Society Series, Volume 5. Basil Blackwell:Oxford.

Coventry, K. R. (Forthcoming). *The semantics of spatial prepositions*. PhD thesis, Department of Cognitive Science, University of Edinburgh.

Coventry, K. R. and Ludwig, A. (1991). Semantics of prepositions: a literature review and proposed framework for future treatment. *Cognitive Science Research Paper No 45*, Edinburgh University. In Press.

Cruse, D. A. (1986) *Lexical Semantics*. Cambridge University Press.

Davidson, D. (1984). What metaphors mean. In *Inquiries into Truth and Interpretation*. Clarendon Press, Oxford.

Davidson, D. (1986). A nice derangement of epitaphs. In E. LePore (ed.) *Truth and Interpretation : Perspectives on the Philosophy of Donald Davidson*. Basil Blackwell, Oxford.

Garrod, S. C. and Sanford, A. J. (1990). Discourse models as interfaces between language and the spatial world. *Journal of Semantics 6*:147-160.

Gerrig, R. J. & Healy, A. F. (1983). Dual processes in metaphor understanding: Comprehension and appreciation. *Journal of Experimental Psychology: Learning, Memory and Cognition, 9*, 667-675.

Gerrig, R. J. (1989). Empirical constraints on computational theories of metaphor:comments on Indurkhya. *Cognitive Science 13*, 235-241.

Gibbs, R. W. (1984). Literal meaning and psychological theory. *Cognitive Science, 8*, 275-304.

Gildea, P. & Glucksberg, S. (1983). On understanding metaphor: The role of context. *Journal of Verbal Learning and Verbal Behaviour, 22*, 577-590.

Herskovits, Annette (1986). *Language and Spatial Cognition. An interdisciplinary study of the prepositions in English.* Cambridge University Press.

Hesse, M. (1963). *Models and Analogies in Science.* London: Sheed and Ward.

Hesse, M. (1963). The explanatory function of metaphor. In *Models and Analogies in Science.* London: Sheed and Ward.

Hoffman, R. R. & Kemper, S. (1987). What could reaction-time studies be telling us about metaphor comprehension ? *Metaphor and Symbolic Activity, 2*, 149-186.

Kittay, E. A. (1987). *Metaphor: Its Cognitive Force and Linguistic Structure.* Clarendon Press, Oxford.

Lakoff, G. (1987). *Women, fire and dangerous things.* Chicago: Chicago University Press.

Lang, Ewald (1990). Primary perceptual space and inherent proportion schema: two interacting categorization grids underlying the conceptualization of spatial objects. *Journal of Semantics 7*: 121-141.

MacCormac, E. R. (1985). *A Cognitive Theory of Metaphor.* A Bradford Book: MIT Press.

Miller, G. A. (1979). Images and Models, Similes and Metaphors. In A. Ortony (ed.), *Metaphor and Thought.* Cambridge University Press.

Nunberg, G. D. (1978). *The Pragmatics of Reference.* Doctoral dissertation, City University of New York. Reproduced by the Indiana University Linguistics Club.

Lakoff, M. and Johnson, M. (1980). *Metaphors We Live By.* University of Chicago Press.

Miller, G. A. & Johnson-Laird, P. N. (1976). *Language and perception.* Cambridge University Press.

Ortony, A., Schallert, D. L., Reynolds, R. E. & Antos, S. J. (1978). Interpreting metaphors and idioms: Some effects of context on comprehension. *Journal of Verbal Learning and Verbal Behaviour, 17*, 465-477.

Ortony, A. (1979). *Metaphor and Thought.* Cambridge University Press.

Searle, J. R. (1979). *Expression and Meaning: Studies in the Theory of Speech Acts.* Cambridge: Cambridge University Press.

Searle, J. R. (1979). Literal meaning. In *Expression and Meaning: Studies in the Theory of Speech Acts.* Cambridge: Cambridge University Press.

Searle, J. R. (1979). Metaphor. In *Expression and Meaning: Studies in the Theory of Speech Acts.* Cambridge: Cambridge University Press.

Metaphor Comprehension Model: Theory and Implementation

Kouichi DOI *, Hirohiko SAGAWA †, Hidehiko TANAKA ‡

Hidehiko TANAKA Lab. Dept. Electric eng., University of Tokyo
7-3-1, Hongo, Bunkyo-ku, Tokyo, 113, JAPAN

Phone: +81-3-(3812)-2111 ex.(7413)

{doi, sagawa, tanaka}@mtl.t.u-tokyo.ac.jp

Abstract

Metaphor comprehension is important for natural language understanding by computers. Metaphor has delicate nuances which depend on a given natural language or a given person. Metaphor comprehension is essential for better man machine interface or more refined machine translation. Our objective is to find a method of metaphor comprehension for computers similar to the way a person comprehends.

We review well-known theories. We then formulate a principle that metaphor comprehension is divided into detection and comprehension based on Sperber's symbol analysis model. We adopt the interaction view as our metaphor comprehension theory. Further, in order to find appropriate meaning of metaphor in a given context, mixture theory is adopted to solve the context problem.

Our metaphor comprehension model is based on these theories. In this paper, we classify metaphor and context. Associations among contexts are obtained by our psychological experiments. It is assumed that a metaphor's meaning is expressed as lists of words, obtained by these experiments.

In the implementation of our metaphor detection and comprehension system, a metaphorical sentence in a given context is translated into a set of plain sentences.

1 Introduction

Metaphor comprehension is very important for natural language comprehension on computers. For example, a metaphor is found in a sentence like "Programs run". That is to say, the word "run" has the original meaning of an object's moving physically by itself. The meaning of "execute" is added to the original meaning. In state-of-the-art machine translation a simple multivocal word[1] can be treated, and the system can use a simple knowledgebase. But more complicated multivocal words, as in metaphors, are not yet discussed. Journals and newspapers which are published daily contain many metaphorical expressions which are special to their own language. If journals and newspaper are to be translated by machines, it will be necessary to comprehend metaphor automatically. In particular, a metaphor must first be detected. A metaphor alters its meaning according to its context. It must then be comprehended. Furthermore it is not completely understood now to a complicated knowledgebase. In this paper, we propose a method of solving the multiple-meanings, and a method of building a

*International Institute for Advanced study of Social Information Science, Fujitsu Laboratories Ltd.
†Central Research Laboratory, Hitachi, Ltd.
‡Dept. Electric eng., University of Tokyo

[1]e.g. the word "right" has several meanings. One is "correct"; another is "what one is properly entitled to"

complicated knowledgebase. In the field of man-machine interface, when the system encounters an unknown expression, the expression should be comprehended automatically, as far as possible.

Our approach is to make the computer metaphors similarly to the way human does.

We propose a method of metaphor comprehension using, *associative lists*, obtained by psychological experiments. We also analyze the relation between metaphor and context which determines the meaning of the metaphor. We then describe our implementation.

2 Background and Related Works

2.1 Background Theories

Our system is based on Sperber's symbol analysis model[1]. Grice's conversation principle could also be considered as the basis for this system, but there are some problems in adopting it[1]. According to Sperber, when we think of something strange, *evocation* and *focusing* occur. The former corresponds to our metaphor detection routine, and the latter, to our metaphor comprehension routine.

In this paper, we deal with *live metaphor*[2]. A *live metaphor* an expression which is not stereotyped as a *dead metaphor* is. A *live metaphor* changes meaning according to the context. Therefore the context should be treated as well as the meaning of the metaphor. We believe that the essential problem of metaphor comprehension is in the case of *live metaphor*.

Our metaphor comprehension model reflects also the interaction view of certain philosophers[3]. According to this view, words in a metaphor interact with one another, and then alter their meaning. For example, in the sentence "Man is wolf", both "man" and "wolf" have new meanings, "cruel" and "solitude" respectively. This view lead us to adopted the associative list: a list of words which builds up an image which constitutes a *concept*, but which cannot be expressed by only one word. This list is made by psychological experiment.

Finally our metaphor comprehension model involved the context is based on the mixture theory[4]. According to this theory, a hearer comprehends a metaphor as follows:

Phase1 the hearer computes several interpretations when he hears a word or a phrase with multiple meanings.

Phase2 He tries to select the best meaning according to the context.

Phase3 If the multiplicity of meaning is not solved when the sentence finishes, he selects one of the meaning as the best interpretation, and fixes on that meaning.

Phase4 If the selected interpretation does not fit the context, then he recalls the former phrases, and tries to find another new meaning.

In our system, a metaphor is detected and the meaning is computed, and recomputed when a new context appears.

2.2 Related Works

There are many related works on dead metaphor or analogy. In Martin[5], metaphor which is related to the computer is treated using a kind of semantic network, and transformation of metaphor is also treated. In Iwayama[6], metaphor is explained using the concept of entropy.

But in their research, metaphors dependent on the context and brand new metaphors[2] are not handled. Neither of them discusses obtaining the data of metaphor comprehension. A complete semantic network cannot be constructed because links between concept and attribute is ambiguous. A semantic network cannot represent a "meaning" which cannot be represented by only one word. Associative relations among concepts and attributes cannot be represented in their systems.

3 Classification of Metaphor and Context

3.1 Why a Person Uses Metaphor

We consider three reasons why a person uses a metaphor: emphasis, confirmation and endowment with new meaning.

Thes may be cases combined in some cases.

3.2 Classification of Metaphor

Metaphors can be classified into five patterns according to the point of **detection** as follows[7]:

1. The metaphor violates the inclusive relation in the semantic network.
2. The metaphor violates the attribute relation on semantic network.
3. The metaphor violates the actual situation in the external world.
4. The metaphor the consensus level of the word.
5. the metaphor is part of proverb or idiom.

In this paper, first four of these are treated.

A metaphor has three components ,*tenor*, *vehicle* and *ground*[8]. For example, in "man is wolf", "man" is *tenor*, "wolf" is *vehicle* and "cruel" is *ground* which is unexpressed. In many actual sentences, only *tenor* and *vehicle* appear. Therefore considering where the *vehicle* appears, metaphors are also classified as follows:

Only in subject: "Stone[3] cries."

Only in predicate: "Man is wolf."

[2] i.e. *live metaphor*
[3] which is compared to man

Both in subject and predicate: "Blood is thicker than water."

These relate to the sentence patterns, that is to say, for example, "(noun) is (noun)", or "(noun)(declinable word)".

In this paper, only the first two of these patterns are treated.

3.3 Classification of Context

The metaphor sentence, "he is a stone"[4], is treated in our system. This sentence is *live metaphor* in Japanese . The context which surrounds this sentence, will affect the metaphorical sentence in the following ways:

1. the case which the attribute of the tenor[5] is given literally. For example, "He is cold"[6].
2. the case which the sentence includes the word associated with tenor. For example, "He does not help me".
3. the case which the sentence is proverb or idiom. For example, "Fine clothes make the man".
4. the case which the sentence tells common sense. For example, "He does not know even chess".
5. the case which the sentence associated with the situation. For example, "He runs down the slope".
6. the case which the sentence is metaphorical sentence. For example, "He is ice".

In this paper, only 1, 2 and 4 are treated. The 3, 5 and 6 are cases of *dead metaphor* or allegory.

4 Assumption and Approach

As our approach is based on symbolism, a semantic network approach is adopted in our implementation. The input of our system is natural language and output is also natural language. The language is Japanese, but the essence of system does not depend on any particular language.

To simplify matters, we assume that

1. a speaker does not tell lies,
2. a speaker does not speak ambiguously,
3. a speaker does not say inappropriate things,
4. a speaker does not say unnecessary things.

The result of our metaphor comprehension system is not a replacement of a word, but a replacement of a *meaning*. The *meaning* is constructed as a set of words which constitute a concept or an image. These words are obtained by the psychological experiments.

[4]This sentence is strange in English, but it is normal in Japanese. This is the one of the reason why metaphor comprehension is necessary. In English "he is made of stone" is more natural.

[5]"he" in this case

[6]It is "dead metaphor", but it is *literal* meaning

5 Psychological Experiment

We made a few psychological experiments to discover associative relation among words. Two experiments were done. The first experiment looked for the associative relation of a word. Subjects of the experiment were presented a word[7], then they were asked to write words in the order in which they associate with the word. The second experiment is to categorize the words which came from the first experiment. This category is a *meaning* which cannot be expressed as only one word.

When "stone" is an entry point, the results are shown in section 7.

6 System Composition

Our principle is that when the *vehicle* is found, the associative lists are searched by using the *vehicle* as the entry, and context is interpreted as the sorting of the associative lists. This is our interpretation of the interaction view. Our system is implemented using SICStus Prolog on UNIX.

The semantic network consists of templates of meaning before sentences come into the network. A node represents a word or a concept. The necessary links are an *inclusive relation* link, an *attribute relation* link, a *possibility* link, an *opposite* link and an *association* link. The necessary operations are to check the inclusive relation, to check the attribute relation, to check the inheritance of attribute relation, to check the possibility of attribution, to check the opposite relation between attribution, to check the associative words between concepts and to declaration of a new attribution. New information from the input sentence is stored in the semantic network as an instance, the latest information being the first in the list[8].

The elements of an associative lists are lists of words. These are based on our psychological experiments discussed above.

6.1 Metaphor Comprehension Routine

Metaphor comprehension is realized by searching the associative list of the *vehicle*. Words which correspond to attribution of tenor are selected. The components of the list are sorted in order as follows:

1. whether the component has the attribute of the instance is the semantic network or not.
2. whether the component is related to the common sense.
3. whether the component has the inherited attribution.

Lists are then sorted in the condition of each the first component also in order above.

[7] i.e. "stone"
[8] i.e. "asserta" is used

7 Implementation

7.1 Overview

Total system is shown in figure 1.

The numbers in the figure represent the flow of data. At number 1, there are sentences in natural language, for example, a Japanese sentence, 「彼は石だ」 (He is a stone.). At number 2, there are parsed trees which are analyzed for morphology and syntax. At number 3, independent words are selected, and an internal representation is made. At number 6, the kind of metaphor is returned. At number 8, the *vehicle* word is used. At number 9, the list of lists which is associated with the vehicle is used. At number 15, 16, and 17, an internal representation about proverb is used. At number 21, there are sentences in natural language, for example, 「彼は冷たい」 (He is cold.). At the all of other numbers, internal representations are used.

Input part and extracting independent words make a internal representation from the sentence in natural language. The sentence is analyzed in morphology and in syntax using DCG grammar.

7.2 Examples

In this paper, our system is explained taking for example "stone". Other examples are shown in [9]. Six patterns of sentences in Japanese are executed in our system. They are

1. 彼は冷たい。(He is cold.) 彼は石だ。(He is a stone.) 彼は冷たい。(He is cold.)

2. 彼は静かだ。(He is silent.) 石は叫ぶ。(The stone cries.) 彼は静かだ。(He is silent.)

3. 彼は役に立たない。(He is useless.) 彼は石だ。(He is a stone.) 彼は役に立つ。(He is useful.)

4. 彼は助けてくれない。(He does not help me.) 彼は石だ。(He is a stone.)

5. 彼はチェスを知らない。(He does not know chess.) 彼は石だ。(He is a stone.)

6. 彼は石だ。(He is a stone.) 彼は役に立たない。(He is useless.)

Let us give an illustration from example 6. In example 6, the metaphorical sentence is computed twice. When the first sentence comes into our system, system computed reasonable meaning at first. Metaphor comprehension routine searches the associative list using "stone" as the entry point. In this case,

```
[[[無機質 (inorganic matter)],
        [動かない (remain still), 殺風景 (inelegance), 冷たい (cold), 静かだ (quiet), つまらない (be not interested in)]],
[[武器 (weapon)],
        [攻撃に使える (can be used as attack), 当たると痛い (be hurt when it hits), 投げる (throw), 道具 (tool), 鋭利な (sharp), ガラスを割る (break window), 投げられる (can be thrown)]],
```

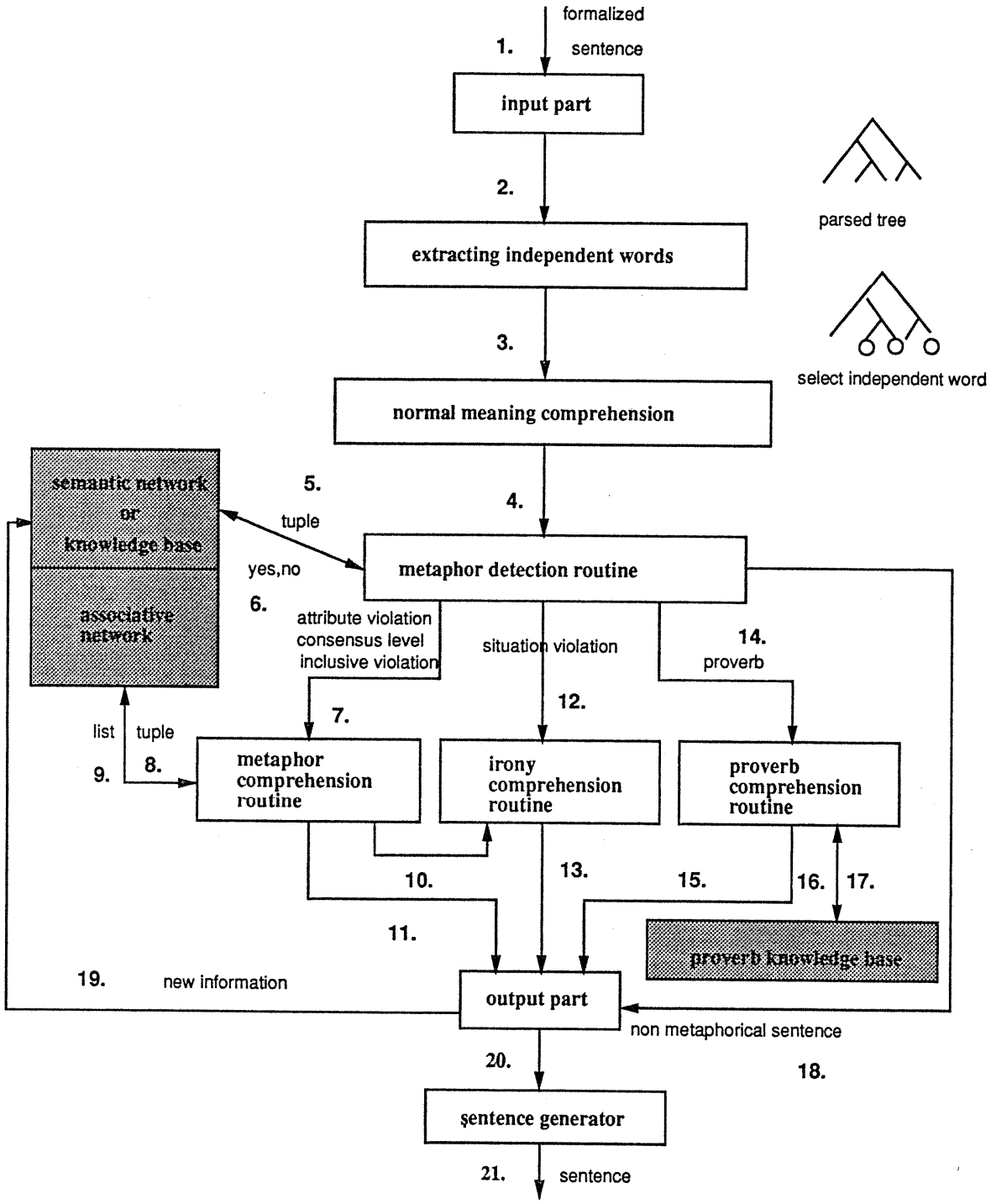


Figure 1: System Composition(in detail)

[[宗教 (religion)],

[美しい (beautiful), 不動 (remain still), まもってくれる力 (might of protection)]],

[[ダメなやつ (useless one)],

[紙になぜか負ける (be lost paper), 融通がきかない (be not flexible), 自分から動かない (does not move in itself), 役に立たない (useless), 価値がない (worthless), 無知だ (ignorance)]], [[食べる (eating)], [かめない (cannot be bitten), 食えない (cannot be eaten)]],

[[色 (color)],

[灰色 (gray), 白 (white)]],

[[物質的特徴 (feature as a substance)],

[かたい (hard), 頑丈だ (strong), 重い (heavy), 割れる (crack), 砂より大きく岩より小さい (bigger than sand but smaller than rock), ころがる (roll), 化石がある (there is fossil), 10cm くらいの楕円形 (an oval about 10cm, さめにくい (be not easy to be cold down)]],

[[形 (shape)],

[大きさがいろいろ (big or small), 小さい (small), 丸い (round), 美 (beauty), ゴツゴツしている (rugged)]],

[[場所 (whereabout)],

[川 (river), 海 (sea), 川辺にたくさんある (there are many in river side), どこにでもある (everywhere)]]]

Then attributions which correspond with attribution about "man" are selected, and sorted using the associative attribution between 助けてくれない (do not help me) and 冷たい (cold). At first, the components of sublist are sorted by the word(in this example, 冷たい (cold)) as a key. Then the categories are sorted in the same way above. The result is,

[[[無機質 (inorganic matter)],

[冷たい (cold), 動かない (remain still), 静かだ (quiet), つまらない (not be interested in)]],

[[宗教 (religion)],

[美しい (beautiful)]],

[[ダメなやつ (useless one)],

[融通がきかない (be not flexible), 役に立たない (useless)]],

[[物質的特徴 (feature as a substance)],

[重い (heavy)]],

The remains are empty(abbreviated). The temporary output sentences are

彼は冷たい、動かない、静かだ、つまらない。(He is cold, remains still, is quiet and is not interested in.)

彼は美しい。(He is beautiful.)

彼は融通がきかない、役に立たない。(He is not flexible and useless.)

彼は重い。(He is heavy.)

Each sentences are candidates for the meaning of this metaphor. The first sentence is the first candidate, and so on.

The second sentence gives context. Then our system computes the meaning again, using associative link described as above. In this case, 役に立たない (useless) is the key for the sorting. Therefore the final output sentences are

彼は役に立たない、融通がきかない。(He is not useful and flexible.)

彼は冷たい、動かない、静かだ、つまらない。(He is cold, remains still, is quiet and is not interested in.)

彼は美しい。(He is beautiful.)

彼は重い。(He is heavy.)

Thus the mixture theory[4] is implemented.

8 Conclusion

In the field of machine translation or man machine interface, metaphor comprehension will be very important.

We reviewed the well-known theories. Metaphor detection and comprehension model was built based on Sperber's symbol analysis model, the interaction view of philosophy and the mixture theory.

We constructed a new model of metaphor detection and comprehension. Metaphor and context were classified into several kinds for computers processing to treat on computers.

In this paper, metaphor is classified, in the point of detection and in the point of where *vehicle* appears, and its relation with sentence pattern.

Data of associative information were derived from psychological experiments. It is considered that the method of correcting data is the same in all language essentially.

Then we implemented these data by Prolog with a technique that metaphor was treated as multivocal word, and also treated related context which restricts the meaning of the metaphor sentence. Input sentences are written in natural Japanese language. Metaphor is detected by using semantic network. Associative lists are searched using *vehicle*("stone") as an entry. Output sentences are also written in natural language. Then new attribution is asserted in this system. This is one of the method of knowledge acquisition.

But there are many sentence patterns in natural language. In this paper, we treated only the three patterns. In each case, metaphor can be detected and comprehended appropriately. More complicated sentence patterns should be treated after this.

In this paper, context is also classified. New kind of context may be found after this, almost all contexts, however, are given in this paper.

The usages of metaphor are classified into three patterns in this paper. The future work is to distinguish these patterns automatically.

The methods of metaphor comprehension is only in scope of words. Metaphor comprehension through image which cannot be told only by words is the other future research theme.

We do not treat an image in the present model. So new knowledge representation which is appropriate for images is the next target of this study.

9 Acknowledgments

I would like to thank professor J. A. Robinson for his comment and for reading this paper.

References

[1] Dan Sperber. *Le Symbolisme en Général.* Hermann, éditeurs des sciences et des arts, Paris, 1974.

[2] Paul Ricœur. *la Métaphore Vive.* Éditions du Seuil, 1975.

[3] M. Black. *Metaphor*, volume 55 of *Proceedings of the Aristotelian Society*, pp. 273–294. Harrison & Sons Ltd. London, 1954.

[4] H.H. Clark and E.V. Clark. *Psychology and Language.* Harcourt Brace Jovanovich, Inc., 1977.

[5] James H Martin. A computational theory of metaphor. Technical Report UCB/CSD 88/465, Computer Science Division (EECS), University of California, Berkeley California 94720, 1988.

[6] M. Iwayama, T. Tokunaga, and H. Tanaka. A method of calculating the measure of salience in understanding metaphors. In *AAAI'90*, 1990.

[7] Kouichi Doi and Hidehiko Tanaka. Metaphor detection based on sperber's symbol analysis model. *Trans. IPS Japan*, Vol. 30, No. 10, pp. 1265–1273, October 1989. in Japanese.

[8] I. A. Richards. *Philosophy of Rhetoric.* Oxford University Press, 1936.

[9] DOI kouichi. *Theory and Implementation for Computational Model on Metaphor Comprehension - Detection and Comprehension of Connotation* -. PhD thesis, University of Tokyo, 1991. in Japanese.

Metonymy, Case Role Substitution and Sense Ambiguity

Dan Fass

Centre for Systems Science
Simon Fraser University
Burnaby, British Columbia, Canada V5A 1S6.
fass@cs.sfu.ca

Abstract

Metonymies are viewed as substitutions of related case roles, for example, the metonymy PRODUCER FOR PRODUCT substitutes an agent role (PRODUCER) for a related patient (PRODUCT). This view of metonymy is applied to some problems of sense ambiguity, notably, sense coverage (the domain of a word sense) and sense extension (how new senses are created). A potential implementation is outlined within the metonymic inferencing component of Fass's Collative Semantics. Finally, the view is compared against mechanisms for sense coverage and sense extension by Stallard, Martin, and Pustejovsky.

1. Introduction

> "Well, I'd like to make one thing quite clear at the outset – when you speak of a train robbery, this in fact involved no loss of a train. It's merely what I like to call the *contents* of the train which were pilfered – we haven't lost a train since 1946, I think it was, the year of the great snows – we mislaid a small one" (Peter Cook, 1964).

Peter Cook is incorrect here – he is confusing the sense coverage of 'rob' and 'steal'. The difference in sense coverage between 'steal' and 'rob' is captured by a CONTAINER FOR CONTENTS metonymy: you steal something and rob the *contents* of something (Mish, 1986), hence a robbed train refers only to things "lost" from the train whereas a stolen train refers to a "lost" train.

Sense coverage and metonymy are two of the topics addressed in this paper. The outline of the paper is as follows. Section 2 describes metonymy, section 3 outlines the relationship between metonymy and case role substitution, and section 4 points out the application of the "metonymy as case role substitution" view to the treatment of sense ambiguity. Section 5 briefly outlines how Fass's Collative Semantics (hereafter CS) theory might be improved by the addition of the metonymy as case role substitution view. In section 6, the view of metonymy is compared with some proposed mechanisms for explaining sense coverage and sense extension.

2. Metonymy

Metonymy is a form of non-literal language in which one entity is used "to refer to another that is related to it" (Lakoff and Johnson, 1980, p. 35). Metonymy is often confused with metaphor (see Fass, 1988a, to appear).

> (1) "The ham sandwich is waiting for his check" (Lakoff and Johnson, 1980, p. 35).

Sentence (1) contains a metonymy. The metonymy is that the concept for ham sandwich is related to an aspect of another concept, for "the male person who ordered a ham sandwich."

Instances of metonymy have been organized by a number of researchers (e.g., Stern, 1931; Lakoff and Johnson, 1980; Yamanashi, 1987) into categories or "metonymic concepts," as Lakoff and Johnson call them. A common metonymic concept is PART FOR WHOLE, otherwise known as synecdoche.

(2) "Dave drank the *glasses*" (= the liquid in the glasses).

(3) "The *kettle* is boiling" (= the liquid in the kettle)
(Waldron, 1967, p. 186; Yamanashi, 1987, p. 78).

CONTAINER FOR CONTENTS, another metonymic concept, occurs in (2) between 'drink' and the sense of 'glasses' meaning "containers," and also in (3). The verb 'drink' prefers a potable liquid as its patient but in (2) there is a preference violation because glasses are not potable liquids. It is not glasses that are drunk, but the potable *liquids* in them. There is a relationship here between a CONTAINER (a glass) and its typical CONTENTS (a liquid): this relationship is the metonymic concept CONTAINER FOR CONTENTS.

While Lakoff and Johnson (1980) deserve credit for popularizing the idea of the metonymic concept, Stern (1931) should be given recognition for a much more comprehensive classification of metonymic concepts produced almost 50 years before. Appendix A contains some of the metonymic concepts distinguished by Stern (1931), with examples of each. Appendix B lists all of Lakoff and Johnson's (1980) metonymic concepts, with examples of each. Stern organized his metonymic concepts into hierarchies. Lakoff and Johnson observed the same, for example, that THE FACE FOR THE PERSON is a special case of THE PART FOR THE WHOLE.

3. Metonymy and Case Role Substitution

Stern's classification in Appendix A includes some very general metonymic concepts which contain terms that look like case roles, for example, Instrument for Action, and Instrument for Product. Indeed, metonymy and case seem very closely related because the terms in every metonymic concept (e.g., PRODUCER and PRODUCT) are readily identifiable as instances of case roles (e.g., Agent and Patient), hence every metonymic concept appears to specify a relationship between two case roles. Below, the metonymic concepts of Appendix B and those used in CS are listed in terms of the case roles to which they belong.

Agent	for	Patient
PRODUCER	FOR	PRODUCT
ARTIST	FOR	ARTFORM
CONTROLLER	FOR	CONTROLLED
Patient	**for**	**Agent**
INSTITUTION	FOR	PEOPLE RESPONSIBLE
Instrument	**for**	**Agent**
OBJECT USED	FOR	USER
OBJECT OPERATED	FOR	OPERATOR
Instrument	**for**	**Patient**
CONTAINER	FOR	CONTENTS
Patient	**for**	**Patient**
THE PART	FOR	THE WHOLE
THE FACE	FOR	THE PERSON
PROPERTY	FOR	WHOLE
Location	**for**	**Agent**
THE PLACE	FOR	THE INSTITUTION
Location	**for**	**Patient**
THE PLACE	FOR	THE EVENT

A number of observations can be made about the above list. First, some metonymic concepts link two instances of the same case role, the patient role. These are all forms of synecdoche (PART FOR WHOLE) and their function is fairly transparent: to "zoom" outwards from a part of a whole to either the

whole or a larger part of the whole.

Second, some metonymic concepts link a pair of different case roles, but substitute in opposite directions, for example, CONTROLLER FOR CONTROLLED is an instance of Agent for Instrument, while INSTITUTION FOR PEOPLE RESPONSIBLE is an instance of the reverse, Instrument for Agent. (Note also in Stern's classification that [3.4] Instrument for Action is the reverse of [4.2] Action for Instrument or Means of Action.) Both directions serve a useful function. The Agent for Instrument substitution serves to highlight the persons responsible for some action, e.g., "Nixon bombed Vietnam" (see CONTROLLER FOR CONTROLLED in Appendix B). Conversely, the Instrument for Agent substitution serves to hide those responsible, e.g., "Exxon has raised its prices again" (see INSTITUTION FOR PEOPLE RESPONSIBLE in Appendix B).

Third, the Location and Instrument case roles play a useful role in "objectifying" agents and patients. For example, in Location for Agent, locations are used to substitute for agents, again downplaying the people responsible for some action, as in "The White House isn't saying anything" (THE PLACE FOR THE INSTITUTION).

Fourth, a serious problem with the notion of "metonymy" is that it is so open-ended:

> "Nunberg (1978) ... has shown that there is no finite set of possible coercion functions. The relation between the explicit and implicit referents can be virtually anything" (Hobbs and Martin, 1987, p. 521).

Maybe by viewing metonymy as case role substitutions, we can constrain what can and cannot be a metonymy. Conversely, perhaps metonymy can be used to explain observations in the case grammar literature that certain case substitutions are common, e.g., that instruments can be agents (see Fillmore, 1968).

4. Metonymy and Sense Ambiguity

Another application of metonymy viewed as case role substitution is to sense ambiguity. In dictionary definitions, it is common to divide the meaning of a word into numbered senses, and those senses into lettered subsenses. Each sense covers a different domain. For example, *Webster's Ninth New Collegiate Dictionary* (Mish, 1986) distinguishes five numbered senses of the transitive verb 'to play' and then divides each sense into lettered subsenses, then divides the lettered subsenses into sub-subsenses numbered in brackets.

> **4 a:** to perform (music) on an instrument <~ a waltz> **b** to perform music on <~ the violin> **c** to perform music of (a certain composer)

Similarly, in the definition of 'to play' in the *Longman Dictionary of Contemporary English* (Proctor et al, 1978), referred to hereafter as LDOCE, play16 has two subsenses:

> **16 a:** to perform a musical piece (e.g., a march) **b:** to perform music by a particular composer

The relationship between these subsenses is readily understandable in metonymic terms: the metonymy CONTAINER FOR CONTENTS relates musical pieces (CONTAINER) to music (CONTENTS).

Note though that this metonymic relationship is not always consistently used by dictionary makers. For example, in the definition of the verb 'compose' in LDOCE, compose3 is:

> **3:** to write (music, poetry etc)

This word sense compose3 is like play16b in that both share the object preference, music. However, there is no listed sense of 'compose' equivalent to play16a, i.e., there is no related subsense given that means: to write a musical piece or poem.

Notwithstanding this observation, there are plenty of other examples of regular metonymic relationships between subsenses of a word. Consider, for example, some "operator" verbs like 'to fly', 'to plough', and 'to sail'.

(4) "The plane flew to Cuba."
(5) "The pilot flew to Cuba."
(6) "Susan flew to Cuba."

In all of these sentences, it is apparent that a plane went to Cuba. In sentences (5) and (6), however, no plane is mentioned. Instead, in (5), we are led to understand that a pilot operated the plane to Cuba. The most common understanding of (6) is probably that Susan was a passenger on a plane to Cuba though another interpretation is possible, that Susan was the plane's pilot.

Now, do we want to say that different senses of 'fly' are being used in these sentences? Mish (1986) lists separate senses of 'fly' that correspond to the uses of fly in (4), (5) and (6):

> **1 b:** to operate (as a balloon, aircraft, rocket, or spacecraft) **c** to journey over or through by flying **3 :** to transport by aircraft or spacecraft

A case-frame equipped with suitable optional cases – (instrument plane), (agent pilot) and (patient passenger) – could probably capture the uses of 'fly' in the three sentences. A complementary approach is to note the metonymic relationships between the subject nouns in the sentences. Between (4) and (6), the relationship between plane and Susan is VEHICLE FOR PASSENGER, a type of CONTAINER FOR CONTENTS. Between (4) and (5), the relationship between plane and pilot is OBJECT OPERATED FOR OPERATOR, a species of Lakoff and Johnson's OBJECT USED FOR USER. It seems that Agent-Instrument substitutions are common in "operator" verbs, e.g., who flies when flying: the pilot (agent) or the aeroplane (instrument)?

Again, a caveat is necessary here because consider other "operator" verbs like 'to drive' and 'to bicycle'.

*(7) "The car drove to Seattle."
(8) "The driver drove to Seattle."
(9) "Susan drove to Cuba."

Note that (7) doesn't make sense in the way that (4) does.

5. Better Processing of Metonymy

This section discusses how the metonymy as case role substitution view might be incorporated into the metonymic inferencing component of CS (for an extended description of the CS approach to metonymy processing, see Fass, 1988a, to appear). Briefly, in CS metonymy is treated as a kind of domain-dependent inference. Metonymic concepts are represented by *metonymic inference rules* and the process of finding metonymies is called *metonymic inferencing*. These rules have been implemented in meta5, a computer program containing CS.

• CONTAINER FOR CONTENTS

metonymic_inference_rule(Source, Target):-

find_cell(Target, [it1, contain1, Contents]), [1]
find_sense_network_path(Source, Contents). [2]

This metonymic concept is target-driven. The target is the "container" in a container-contents relation ([1]). The "contents" is the substitute metonym that replaces the target. The next path through the "sense-network" (a hierarchically organized semantic network) is sought between the source and the contents ([2]).

If a preference violation occurs (indicating non-literal language) when meta5 analyzes a sentence, a metonymy is recognized if a metonymic inference is found; conversely, if no inference is found then no metonymy is recognized.

The case role constraints outlined in section 3 can be straightforwardly added to metonymic inference rules. The rules presently contain type restrictions on the pairs of terms represented in each

metonymic concept. These rules would be modified to have the form *metonymic_inference_rule*(Source_term, Source_case, Target_term, Target_case). In the case of CONTAINER FOR CONTENTS, the left hand side of the rule would have the form *metonymic_inference_rule*(Source_term, patient, Target_term, instrument), thereby specifying the case roles of this particular metonymic concept.

These case role restrictions might be used to limit what can and cannot form a chain of metonymies (Reddy, 1979). For example, below is a chain of two metonymies constructed by meta5 when processing (10).

(10) "Ted played *Bach*" (= the music of Bach).

The chain consists of ARTIST FOR ARTFORM and CONTAINER FOR CONTENTS metonymies between 'Bach' and the verb sense play12 (which means "to play music") in meta5's lexicon. Below are the successful metonymic inferences made by meta5, together with the case roles of those inferences.

ARTIST	FOR	ARTFORM
johann_sebastian_bach	composes	musical_pieces
Agent	for	Patient
CONTAINER	FOR	CONTENTS
musical_pieces	contain	music
Instrument	for	Patient

As can be seen, the two metonymic concepts impose two sets of views on musical_pieces: ARTIST FOR ARTFORM views them as ARTFORMs and Patients; CONTAINER FOR CONTENTS views them as CONTAINERs and Instruments. Perhaps certain sequences of case roles like (Agent-Patient, Instrument-Patient) are acceptable in metonymic chains while others are not.

Finally, sentence representations might be modified to contain metonymic inferences, including not only the substituted terms but also the substituted case roles. Different subsenses, or sub-subsenses, of a word then "fall out" as a word sense in combination with different metonymies. For example, suppose that meta5 and some dictionary both record the verb sense play12 as meaning "to play music," and the dictionary records a verb subsense play12a as meaning "to play a musical piece." Suppose further that meta5 analyzes a sentence containing play12 and a CONTAINER FOR CONTENTS metonymy. This analysis would appear in the representation for the sentence as play12 + CONTAINER FOR CONTENTS, equivalent to "to play a musical piece," or the verb subsense play12a from the dictionary.

6. Related Work and Discussion

This section discusses related computational approaches to sense coverage and sense extension. Like Martin (1990), the approach to metonymy outlined in this paper is in opposition to "massively lexical approaches" in which all the senses of a word are recorded. Stallard (1987), Martin, Pustejovsky (1989) and others have all begun to develop mechanisms that remove the need for massively lexical approaches. Their three approaches are outlined below, together with brief comments on each. All of these mechanisms, it seems to me, fall within what Pustejovsky has called the "generative lexicon." The idea behind the generative lexicon is that a small number of rules, combined with the main senses of words, can be used to generate the semantic coverage needed for a lexicon. The approach I outlined in the previous section – word senses extended by metonymic inferences – also falls within the generative lexicon idea.

Stallard (1987) tackles the problems of sense extension, nominal compounds, metonymy, metaphor and anaphora resolution. Each word sense is assumed to have a core sense from which extended senses are derived by recursive application of polysemy operators. These polysemy operations include metonymic and metaphorical sense extension, "broadening" (which allows a word to refer to a wider class of items than before), "exclusion" (removes a subset of the members of the denotation of a word), and "narrowing" (narrows the denotation down to a particular subset). Metonymic extension "re-interprets a

predicate by interposing an arbitrary, sortally compatible relation between an argument place of the predicate and the actual argument" (Ibid., p. 182), thereby shifting the argument place of a predicate (Ibid., p. 183). Metaphorical extension operates by shifting the whole predicate.

Stallard (1987) does not describe metonymy processing in much detail; however, his approach would seem to be compatible with the views of metonymy, metonymy processing and sense ambiguity outlined in this paper.

Martin (1990) has been concerned with metaphor interpretation and extension using conventional metaphors (Lakoff and Johnson, 1980), also known as metaphorical concepts. Examples of metaphorical concepts include ARGUMENT IS WAR and SAD IS DOWN. Metaphorical concepts are analogous to metonymic concepts. Martin (1990) has written a computer program called MIDAS that contains a "knowledge base" of metaphorical concepts organized into hierarchies. MIDAS includes a Metaphor Extension System (MES) which is used in situations where neither the word senses nor the metaphorical concepts in MIDAS's lexicon have sufficient coverage for treating a sentence. In such a situation, the MES acquires "novel" metaphors by systematically extending, elaborating, and combining already known metaphors.

Note that Martin opts to resolve such situations by extending metaphorical concepts rather than word senses. This is presumably because he is opposed to proliferations of word senses, but it could be argued that he is substituting one kind of proliferation (word senses) for another (metaphorical concepts).

There are some interesting parallels between the MES in MIDAS and metonymic inferencing in meta5 that look worth investigating. The MES operates on metaphorical concepts, combining already known concepts to form extended metaphors. Metonymic inferencing in meta5 operates on metonymic concepts, combining already known concepts to form metonymic chains.

The metaphorical concepts used by MES are in a hierarchy. It also seems that the metaphorical concepts used by MES contain labels which are akin to case roles, like Kill-Victim (a patient role) and Terminated-Process (also a patient role). It could be that the mechanisms in MES that exploit the hierarchy and case-like roles might be useful for extending the metonymic inferencing component in meta5.

Pustejovsky (1989) has applied his generative lexicon idea to the category shifting of verbs (from process to state) and has sought principles of cocomposition that might govern this. He argues that a theory of lexical decomposition is needed that is richer than case roles and more than verbs. For verbs, his idea is *cocomposition*, where the aim is to specify "general principles of event composition," e.g., that a process can result in something, such as the process of baking can result in a cake, as in (11b) below.

(11a) "John baked the potato" (i.e., cooked something).
(11b) "John baked the cake" (i.e., cooked and thereby created something).

Hence, rather than have two senses of 'bake', simply have a single "change of state" sense of the verb viewed as a process, and allow this to shift to the process-result.

It might be that Pustejovsky's notion of "cocomposition" should fall within an account of metonymy: the kind of shift occurring between (11a) and (11b), for example, seems closely related to the Nomina Actionis metonymies of Stern (1931) listed in Appendix A.

7. References

Cook, Peter (1964). From a dialogue between Peter Cook and Alan Bennett about 'The Great Train Robbery' as part of the 1964 version of the show "Beyond the Fringe." In Roger Wilmut (1980). *From Fringe to Flying Circus: Celebrating a Unique Generation of Comedy 1960-1980.* London, England: Eyre Methuen, p. 26.

Fass, Dan C. (1988a). Metonymy and Metaphor: What's the Difference? *Proceedings of the 12th International Conference on Computational Linguistics (COLING-88)*, Budapest, Hungary, pp. 177-181.

Fass, Dan C. (1988b). Collative Semantics: A Semantics for Natural Language Processing. (PhD thesis) Memorandum MCCS-88-118, Computing Research Laboratory, New Mexico State University, NM.

Fass, Dan C. (to appear). Met*: A Method for Discriminating Metonymy and Metaphor by Computer. *Computational Linguistics.*

Fillmore, Charles J. (1968). The Case for Case. In E. Bach and R.T. Harms (Eds.) *Universals in Linguistic Theory.* New York, NY: Holt, Rinehart, and Winston, pp. 1-88.

Fillmore, Charles J. (1977). Scenes and Frames Semantics. In Antonio Zampolli (Ed.) *Linguistic Structures Processing.* Amsterdam, Holland: North-Holland, pp. 55-81.

Hobbs, Jerry R., and Paul Martin (1987). Local Pragmatics. In *Proceedings of the 10th International Joint Conference on Artificial Intelligence (IJCAI-87)*, Milan, Italy, pp. 520-523.

Lakoff, George, and Mark Johnson (1980). *Metaphors We Live By.* London, England: Chicago University Press.

Martin, James H. (1990). *A Computational Model of Metaphor Interpretation.* New York, NY: Academic Press.

Mish, Frederick C. (1986). *Webster's Ninth New Collegiate Dictionary.* Springfield, MA: Merriam-Webster Inc.

Nunberg, Geoffrey (1978). The Pragmatics of Reference. Ph.D. Thesis, City University of New York, NY.

Procter, Paul, Robert F. Ilson, John Ayto, et al (1978). *Longman Dictionary of Contemporary English.* Harlow, Essex, England: Longman Group Limited.

Pustejovsky, James (1989). Current Issues in Computational Lexical Semantics. In *Proceedings of the 4th Conference of the European Chapter of the ACL*, April 10-12 1989, Manchester, England, pp. xvii-xxv.

Reddy, Michael J. (1979). The Conduit Metaphor – A Case of Frame Conflict in Our Language about Language. In Andrew Ortony (Ed.) *Metaphor and Thought.* London, England: Cambridge University Press, pp. 284-324.

Stallard, David (1987). The Logical Analysis of Lexical Ambiguity. *Proceedings of the 25th Annual Meeting of the ACL*, Stanford University, Stanford, CA, pp. 179-185.

Stern, Gustaf (1968; first published in Sweden 1931). *Meaning and Changes of Meaning*, Bloomington, IN: Indiana University Press.

Waldron, Ronald A. (1967). *Sense and Sense Development.* London, England: Andre Deutsch.

Yamanashi, Masa-aki (1987). Metonymic Interpretation and Associative Processes in Natural Language. In Makoto Nagao (Ed.) *Language and Artificial Intelligence (Proceedings of an International Symposium on Language and Artificial Intelligence held in Kyoto, Japan, 16-21 March 1986).* Elsevier Science Publishers, B.V., (North-Holland), pp. 77-86.

Appendix A: Some of Stern's Metonymic Concepts

Stern's (1931) book is about sense-change. Seven classes of sense-change are distinguished. Two of those are "transfer" and "permutation." Transfers are based on a number of relations:

[1] Proper Names for Objects

Proper names are employed to denote scientific units of measurement (e.g., 'ohm', 'ampere', 'volt', 'coulomb') and inventions ('burberry', 'gatling', 'macadam', 'mackintosh', 'remington', 'wellingtons') (Ibid., p. 295).

[2] Places-Names for Products or Events

Such names include 'calico', 'camembert', 'champagne', 'china', 'java', 'mokka'. An alternative explanation is permutation (p. 295).

Permutations are listed for nouns, adjectives and adverbs, verbs, and particles. Permutations for nouns include the following:

[3] Objects' Names (Concrete and Abstract)

[3.1] Material for Object Made from It

The type is old. The Old English 'iren' is used for iron articles. 'Brass' means "a musical instrument of brass" and "brass money" (p. 362).

[3.2] Receptacle for Content

'Barrel' is used for "the contents of a barrel," a 'tub' is often used for the contents of the tub (p. 363).

[3.3] Part or Constituent Detail for the Whole, and Vice Versa

A 'hand' is a person employed in any manual work, a workman or workwoman. 'Blade' is often put for sword. The reverse: "A *dungeon* is originally a strong tower, the central tower of a fortress. Since such towers were generally furnished with cellars used as prisons, *to put a man in the dungeon* would mean putting him in prison, and this meaning of dungeon gradually became the most common one" (p. 364, italics in original).

[3.4] Instrument for Action

Stern has only one good instance. A 'ballot' is a "voting paper" and later "the act of voting" (p. 365).

[3.5] Instrument for Product

A 'whistle' is an instrument and also the sound of a whistle (p. 365).

[3.6] Article of Dress or Equipment for Person

Words like 'drum', 'bag-pipe', 'banner' and 'ensign' can all denote the people using them (p. 367).

[4] Nomina Actionis

[4.1] Action for Product, Result, or Object

"A very common type" (p. 369). 'Batch' is "the process of baking" and "the quantity of bread produced at one baking." 'Cast' is "the act of casting or throwing" and "the distance which anything can be thrown." 'Lending' is "the act of lending" and "something lent." 'Lift' is "the act of lifting" and "the thing lifted" (all on p. 370).

[4.2] Action for Instrument or Means of Action

'Aid' is "an action of aiding" and "something by which assistance is given" (p. 371).

[4.3] Action for Agent

'Help' is "the action of helping" and "any thing or person that affords help." 'Aid' is "the action of aiding" and "the person aiding." 'Failure' is "the fact of failing to effect one's purpose" and "a thing or person that proves unsuccessful" The same with 'success'. "The last two [failure and success] may be metaphorical" (all on p. 371).

[4.4] Action for Place of Action

"The type is very common, and also very old" (p. 372). 'Crossing' is "the action of crossing" and "the place of crossing."

[5] Names of Persons for Products etc.

A Milton or Shakespeare is "a book or play by ...," and likewise with a picture, e.g., a Rembrandt (p. 373). Likewise with wines, e.g., Cliquot and Johannisberger.

[6] Place-Names

[6.1] Place-Name for Action or Event

'Church' is "a church building" and also "a divine service" as in phrases like to attend church, go to church, be at (in) church, after church, etc (p. 373). Similarly for chapel, college, school and others.

[6.2] Place-Name for Inhabitants or Frequenters

We use "the City" for its inhabitants, "the gallery" for those sitting there during a performance, as in "to play to the gallery" (p. 374). 'House' means "legislative assembly" but, originally, the building or room where they assemble. "Downing Street" is used to mean the government of the United Kingdom (p. 375).

Appendix B: Lakoff and Johnson's Metonymic Concepts

The following metonymic concepts are from Lakoff and Johnson (1980, pp. 37-39, italics in original). The square brackets contain equivalent noun permutations from Stern (1931).

THE PART FOR THE WHOLE [Part or Constituent Detail for the Whole]

Get *your butt* over here!
We don't hire *longhairs.*
I've got a new *four-on-the-floor V-8.*
I've got a new *set of wheels.*

THE FACE FOR THE PERSON [Part or Constituent Detail for the Whole]

She's just a *pretty face.*
There are an *awful lot of faces* out there in the audience.
We need some *new faces* around here.

PRODUCER FOR PRODUCT [Names of Persons for Products]

I'll have a *Löwenbräu.*
He bought a *Ford.*
He's got a *Picasso* in his den.
I hate to read *Heidegger.*

OBJECT USED FOR USER [Article of Dress or Equipment for Person]

The *sax* has the flu today.
The *BLT* is a lousy tipper.
The *buses* are on strike.

CONTROLLER FOR CONTROLLED

Nixon bombed Vietnam.
Ozawa gave a terrible concert last night.
Napoleon lost at Waterloo.
A Mercedes rear-ended *me.*

INSTITUTION FOR PEOPLE RESPONSIBLE

Exxon has raised its prices again.
You'll never get the *university* to agree to that.
The *Army* wants to reinstitute the draft.
The *Senate* thinks abortion is immoral.

THE PLACE FOR THE INSTITUTION [Place-Name for Inhabitants or Frequenters]

The *White House* isn't saying anything.

Paris is introducing long skirts this season.
Wall Street is in a panic.

THE PLACE FOR THE EVENT [Place-Name for Action or Event]
Let's not let Thailand become another *Vietnam*.
Remember *the Alamo*.
Watergate changed our politics.

Metaphor and Abduction

Jerry R. Hobbs
Artificial Intelligence Center
Menlo Park, California

Abstract

In this paper a recent approach to inference in text understanding based on abduction is applied to the problem of metaphor interpretation. The fundamental ideas in the "interpretation as abduction" approach are outlined. A succinct characterization of interpretation is given, along with a brief example, and the principal features of a weighted abduction scheme that is used are described. Two examples are analyzed in the abductive framework to determine what problems arise. The examples are a conventionalized metaphor schema and a standard category metaphor contextually interpreted. The primary problem that arises for metaphor interpretation in the abductive framework is dealing with the fact that metaphors are not literally true. Two perspectives on this difficulty are offered.

1 Introduction

The use of inference in text understanding has been an important theme in computational linguistics for at least two decades, and there have been a number of inference-based approaches to metaphor interpretation (e.g., Carbonell, 1982; Fass, 1988; Hobbs, 1983; Indurkhya, 1987; Martin, 1986). In recent years, a particularly elegant formulation of the use of inference in text understanding has emerged, based on abduction, or inference to the best explanation (e.g., Charniak and Goldman, 1988; Hobbs et al., 1988, 1990; Norvig, 1987). A great many problems in interpretation, such as reference resolution, the expansion of metonymies, and the resolution of some lexical and syntactic ambiguities, have been shown to be subsumed under this new approach. This paper is an attempt to extend the abductive approach to cover metaphor interpretation as well, by redoing an inference-based approach to metaphor in light of one of the abductive frameworks.

In Section 2 the fundamental ideas in the "interpretation as abduction" approach are outlined, though necessarily briefly. A succinct characterization of interpretation is given, along with a brief example. The principal features of a weighted abduction scheme are described.

In Section 3 two examples, analyzed in an older framework in Hobbs (1983), are reanalyzed in the abductive framework to determine what problems arise. The examples are a conventionalized metaphor schema and a standard category metaphor contextually interpreted.

The primary problem that arises for metaphor interpretation in the abductive framework is dealing with the fact that metaphors are not literally true. In Section 4 two perspectives on this difficulty are offered.

2 Interpretation as Abduction

2.1 Characterizing Interpretation

Abductive inference is inference to the best explanation. The process of interpreting sentences in discourse can be viewed as the process of providing the best explanation of why the sentences would be true. More precisely,

To interpret a sentence:

(1) Prove the logical form of the sentence,
together with the constraints that predicates impose on their arguments,
allowing for coercions,
Merging redundancies where possible,
Making assumptions where necessary.

By the first line we mean "prove from the predicate calculus axioms in the knowledge base, the logical form that has been produced by syntactic analysis and semantic translation of the sentence."

In a discourse situation, the speaker and hearer both have their sets of private beliefs, and there is a large overlapping set of mutual beliefs. An utterance straddles the boundary between mutual belief and the speaker's private beliefs. It is a bid to extend the area of mutual belief to include some private beliefs of the speaker's. It is anchored referentially in mutual belief, and when we prove the logical form and the constraints, we are recognizing this referential anchor. This is the given information, the definite, the presupposed. Where it is necessary to make assumptions, the information comes from the speaker's private beliefs, and hence is the new information, the indefinite, the asserted. Merging redundancies is a way of getting a minimal, and hence a best, interpretation.

This approach to discourse interpretation has been implemented in the TACITUS system (Hobbs et al., 1990) at SRI International, using Stickel's Prolog-Technology Theorem Prover (Stickel, 1989). The system has been employed in several moderate-scale applications.

2.2 An Example

This characterization, elegant though it may be, would be of no interest if it did not lead to the solution of the discourse problems we need to have solved. A brief example will illustrate that it indeed does.

But first a notational convention that is used throughout this paper: We will take $p(x)$ to mean that p is true of x, and $p'(e,x)$ to mean that e is the eventuality or possible situation of p being true of x. This eventuality may or may not exist in the real world. The unprimed and primed predicates are related by the axiom schema

$$(\forall x)p(x) \equiv (\exists e)p'(e,x) \wedge \mathit{Rexists}(e)$$

where $\mathit{Rexists}(e)$ says that the eventuality e does in fact really exist. This notation, by reifying events and conditions, provides a way of specifying higher-order properties in first-order logic. See Hobbs (1985) for further explanation of this variant of Davidsonian (1967) notation.

The example is

(2) The Boston office called.

This example illustrates three problems in "local pragmatics", the reference problem (What does "the Boston office" refer to?), the compound nominal interpretation problem (What is the implicit relation between Boston and the office?), and the metonymy problem (How can we coerce from the office to the person at the office who did the calling?).

Let us put these problems aside, and interpret the sentence according to characterization (1). The logical form is something like

(3) $(\exists e, x, o, b)call'(e, x) \wedge person(x) \wedge rel(x, o) \wedge office(o) \wedge nn(b, o) \wedge Boston(b)$

That is, there is a calling event e by a person x related somehow (possibly by identity) to the explicit subject of the sentence o, which is an office and bears some unspecified relation nn to b which is Boston.

Suppose our knowledge base consists of the following facts: We know that there is a person John who works for O which is an office in Boston B.

(4) $person(J)$, $work\text{-}for(J, O)$, $office(O)$, $in(O, B)$, $Boston(B)$

We also know that *work-for* is a possible coercion relation,

(5) $(\forall x, y)work\text{-}for(x, y) \supset rel(x, y)$

and that *in* is a possible implicit relation in compound nominals,

(6) $(\forall y, z)in(y, z) \supset nn(z, y)$

Then the proof of all but the first conjunct of (3) is straightforward, by backchaining on axioms (5) and (6) into the ground instances of (4). We thus assume $(\exists e)call'(e, J)$, which constitutes the new information in the sentence.

Notice that all the local pragmatics problems have been solved. "The Boston office" has been resolved to O. The implicit relation between Boston and the office is the *in* relation. "The Boston office" has been coerced into "John, who works for the Boston office."

This is of course a simple and purely illustrative example. More complex examples and arguments are given in Hobbs et al. (1990).

2.3 Weighted Abduction and Biconditionalizing Axioms

The scheme for weighted abduction that we use is described in Stickel (1989) and Hobbs et al. (1990). Briefly, in proving an expression, one may make assumptions at a cost. One then searches for the cheapest proof. Moreover, cheaper proofs may often be obtained by unifying literals in goal expressions, and this is an important mechanism in resolving coreferences.

Exactly how the costs should be assigned is discussed further in Hobbs et al. (1990). In the remainder of this paper, this issue will be ignored. Our concern will rather be to show that the correct interpretations of metaphors are *possible* in the abductive approach.

In the current TACITUS implementation, whenever an assumption is made, it is checked for consistency. The extension of the abductive approach to metaphor interpretation suggests that this check should be soft rather than hard. Inconsistent assumptions should be allowed if that will result in an otherwise very good interpretation. This is the topic of Section 4.

The abduction scheme allows us to make an important modification in the way axioms are written—all axioms can be biconditionalized. In general, an axiom of the form

species $\supset$ genus

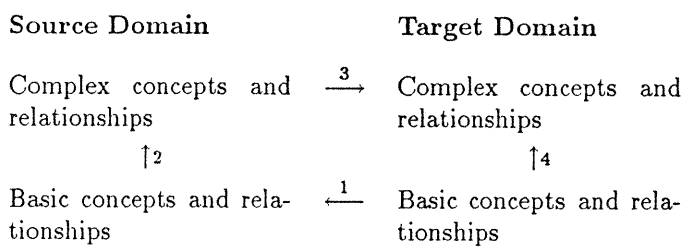


Figure 1: Analogical Processes Underlying Metaphor.

can be converted into a biconditional axiom of the form

genus $\wedge$ differentiae $\equiv$ species

Often we will not be able to prove the differentiae, and in many cases they cannot even be spelled out. But in our abductive scheme, this does not matter; they can simply be assumed at a cost. In fact, we need not state them explicitly at all. We can simply introduce a predicate that stands for all the remaining properties. It will never be provable, but it will be assumable. Thus, the axiom

$$(\forall x)elephant(x) \supset clumsy(x)$$

may be turned around into

(7) $(\forall x)clumsy(x) \wedge etc_1(x) \supset elephant(x)$

The "et cetera" proposition $etc_1(x)$ can never be proved—there will be no other axioms involving that predicate. But it can be assumed, thus allowing us to use this axiom as a way of using the fact that something is clumsy as (perhaps weak) evidence for its being an elephant.

3 Interpreting Metaphors by Abduction

3.1 The Schema for Metaphor

The basic schema for metaphor (and analogy) is that shown in Figure 1.

There are two domains of knowledge, a source domain that is generally very well understood, expressed as a highly elaborated set of axioms, and a target domain, that is generally less well understood. We wish to reason or describe something in the target domain. Rather than doing so directly, we map a basic concept in the target domain into a corresponding basic concept in the source domain. We reason or describe in the source domain, with its richer vocabulary and set of axioms, yielding a complex concept in the source domain. Then we map the result back into the target domain, thereby expressing a complex concept there.

Interpreting a metaphor is a matter of reversing this process. We are given a complex concept in the target domain, expressed in the vocabulary of the source domain. The problem is to discover what it means by determining how it is composed out of basic concepts in the target domain. To do this, we decompose the complex concept into basic concepts in the source domain, and then undo the analogical mapping to determine the meaning in the target domain.

A computational account of metaphor must specify precisely how each of the arrows in this commuting diagram is realized in a formal system. Our answer is essentially as follows: The relation between domains is taken to be simply identity. Predicates from the source domain will simply be predicated of entities from the target domain. This of course brings with it problems of logical consistency, and how to deal with that is the subject of Section 4. The relations between basic and complex concepts will be simply those implicational relations encoded in the axioms. Interpreting a metaphor by abduction will then be a matter of back-chaining along arrows 3, 2, and 1 to an account in terms of the basic concepts in the target domain.

We show how this works for two examples, a conventionalized metaphor schema and a standard "category" metaphor whose interpretation depends on context.

3.2 A Conventionalized Metaphor Schema

The first metaphor to be examined is

(8) The variable N goes from 1 to 100.

Here, the target domain, computer science, is being modelled in terms of the domain of spatial or perhaps more abstract, topological relations. This metaphor rests on the core metaphor that identifies a variable having a value with an entity being located *at* some place. This conventionalized identification can be expressed by the following axiom:

(9) $(\forall e, x, y) variable(x) \wedge value'(e, y, x) \supset at'(e, x, y)$

That is, if x is a variable and e is the condition of y's being its value, the e is also the condition of x being *at* y. The relation between the complex predicate *go* and more basic predicates is expressed by the following axioms, capturing the fact that a change in location is a going event:

(10) $(\forall e, e_1, e_2, x, y, z) change'(e, e_1, e_2) \wedge at'(e_1, x, y) \wedge at'(e_2, x, z)$
$\supset go'(e, x, y, z)$

That is, if e is a change from state e_1 to state e_2 where e_1 is the state of x being at y and e_2 is the state of x being at z, then e is a going by x from y to z.

Now consider the example. Its logical form is

$$(\exists e_0) go'(e_0, N, 1, 100) \wedge variable(N)$$

This is a statement in the target domain, computer science. But we treat it as though it were a statement in the source domain and use the source domain's axiom (10) to decompose the complex concept *go* into the more basic concepts of *change* and *at*. We then use axiom (9) to interpret the *at* relation. The two atoms $variable(N)$ generated in this way are unified with the identical atom from the logical form and that, together with the change and the two value relations are assumed for the minimal interpretation. We thereby have interpreted sentence (8) as asserting a change in value for the variable N.

3.3 A Category Metaphor

The next metaphor we will examine is

(11) John is an elephant.

A number of suggestions have been made about the appropriate inferences to draw in cases such as this. Ortony et al. (1978) said that it is high salience properties that should be transferred, such as size in the case of elephants. Glucksberg and Keysar (1990) say it is diagnostic properties; that is, in (11), we look for some property of elephants for which an elephant is the prototypical examplar—such as large size. Carbonell (1982) has argued that abstract properties, rather than physical properties, should be transferred; thus, "has a trunk" should not be. Gentner (1983) has argued that relations (predicates with two or more arguments) are more frequently transferred than monadic properties.

One difficulty with all these suggestions is that they do not depend on context, whereas we know that interpretation always depends on context. Consider the following sentence:

(12) Mary is graceful, but John is an elephant.

The most reasonable interpretation is that John is clumsy. This is not an especially high salience property of elephants. It is not clear that elephants are prototypical exemplars of clumsiness. It seems to be intermediate between an abstract and a physical property. And it is not a relation.

The non-abductive analysis of this example was relatively clean in Hobbs (1983). There was an axiom that said elephants are clumsy:

$$(\forall x)elephant(x) \supset clumsy(x)$$

That inference was selected because it led to the recognition of a contrast relationship between the two clauses, as signalled by the conjunction "but".

In the abductive approach, the analysis is complicated somewhat by the fact that we can only backchain. We must use the converse form of the axiom, as given in (7):

(7) $(\forall x)clumsy(x) \wedge etc_1(x) \supset elephant(x)$

That is, if something is clumsy and some other unspecified properties hold, then it is an elephant.

We will need to introduce a further complication as well, since we will have to refer explicitly to the properties of clumsiness, elephanthood, and grace. Axiom (7) must be rewritten as follows:

(13) $(\forall e_3, x)clumsy'(e_3, x) \wedge etc_1(e_3, x) \supset elephant'(e_2, x) \wedge gen(e_3, e_2)$

That is, if e_3 is the condition of x's being clumsy and some other unspecified things are true of e_3 and x, then there is a condition e_2 of x's being an elephant. Furthermore, there is a very tight relation between e_3 and e_2—x is an elephant *by virtue of* its being clumsy and the other things being true. We encode this relation with the predicate *gen*, since it is similar to the "generates" relation common in the philosophical literature. This sort of complication of axioms is necessary for most axioms used for interpreting complex discourse in any case.

In Hobbs (1983) the interpretation of (12) was driven by the recognition of a coherence relation between the clauses. In many cases in the abductive approach, especially where a conjunction occurs explicitly, this can be subsumed under the general characterization of interpretation. In (12), the "but" relation is part of the information conveyed by the text, and consequently part of what needs to be explained. We can say that a "but" relation holds between two eventualities e_1 and e_2 if they are contradictory properties p and $\neg p$ of two entities x and y that are similar by virtue of sharing some other property q:

$$(\forall p, q, x, y, e_1, e_2, e_4)p'(e_1, x) \wedge not'(e_2, e_4) \wedge p'(e_4, y) \wedge q(x) \wedge q(y) \supset but(e_1, e_2)$$

This however is too strong. It may be that the contrast is between not e_1 and e_2 but between eventualities related to e_1 and e_2. We therefore (for this example) rewrite the above axiom as follows:

(14) $(\forall p, q, x, y, e_1, e_2, e_4)p'(e_1, x) \wedge not'(e_3, e_4) \wedge p'(e_4, y)$
$\wedge gen(e_3, e_2) \wedge q(x) \wedge q(y) \supset but(e_1, e_2)$

That is, a "but" relation holds between e_1 and e_2 if there is a p such that e_1 is p's being true of some x, and there is an e_3 that generates e_2 that is the negation of an e_4 which is p's being true of some y, and there is some q true of x and y. (This axiom is second-order, but not seriously so, if we restrict the instantiations of the predicate variables to predicate constants.)

Next we need an axiom relating clumsiness and grace.

(15) $(\forall e_3, e_4, y)not'(e_3, e_4) \wedge graceful'(e_4, y) \supset clumsy'(e_4, y)$

That is, if e_3 is the condition of e_4 not being true, where e_4 is the condition of y's being graceful, then e_3 is the condition of y's being clumsy.

Suppose we also know that Mary and John are people:

$person(M)$, $person(J)$

Now we are ready to interpret sentence (12). Its logical form is

$(\exists e_1, e_2)graceful'(e_1, M) \wedge elephant'(e_2, J) \wedge but(e_1, e_2)$

We can then backchain on axiom (13) from "elephant" to "clumsy", assume $etc_1(e_3, J)$, backchain on axiom (15) from "clumsy" to "not graceful", and assume $not'(e_3, e_4)$ and $graceful'(e_4, J)$. We also assume $graceful'(e_1, M)$. Then we have a proof of $but(e_1, e_2)$, using axiom (14), with p instantiated as $graceful$ and q instantiated as $person$.

4 Metaphor and Truth

There is a problem in our account of how example (12) is interpreted. It requires the assumption of propositions that are not true. John is not an elephant. Here I will sketch two possible solutions to this problem.

1. Predicate Coercion: When a predicate is applied to an argument for which it is not appropriate, there are two possible interpretive moves. We can decide that the argument actually refers to something other than it denotes explicitly—we can coerce the argument into something related to it. This is metonymy. Or we can decide that the predicate actually denotes a property other than the one it denotes explicitly—we can coerce the predicate to a related predicate. Metaphor is one example of this. Thus, we interpret

I can't read James Joyce.

metonymically. We coerce from James Joyce to the books written by James Joyce. By contrast, we are likely to interpret

I can't read Saddam Hussein.

metaphorically. We coerce the predicate, taking "read" to mean "understand".

In metonymy, one applies a coercion function k to the argument, transforming $p(x)$ into $p(k(x))$, in order to achieve congruence between the predicate and the argument. In our functionless notation, this becomes

$$p(y) \wedge Req(p, y) \wedge rel(y, x)$$

where $Req(p, y)$ is the requirement that p imposes on its argument, and $rel(y, x)$ expresses the coercion relation between y and x.

By analogy, this would suggest that metaphor be handled by applying a coercion function to the predicate, transforming $p(x)$ into $k(p)(x)$. In our functionless notation, this becomes

$$q(x) \wedge Req(q, x) \wedge Rel(q, p)$$

The predicate q would be new. The meaning of a predicate is determined by the axioms it occurs in. To construct a new predicate q we would have to specify the axioms it occurs in. But this is essentially what we are doing when we sort through the properties related to p that hold in this case. We are picking out the subset of p's axioms that hold for q. The uncertainties in this process correspond to the uncertainties we sometimes experience in interpreting metaphors; they are one of the sources of the metaphor's power.

In this approach *elephant*(J), in the context of example (12), is coerced into another predication, say, *elephant*$*(J)$. If something is an elephant$*$, it is clumsy and perhaps large, but it does not have a trunk.

2. Interpretation and Judgment: In the abductive approach to interpretation, we make assumptions when we are unable to prove something (for less cost). But these assumptions can play a number of different roles. In many cases assumptions are used to accept new information. More generally, however, they can be used to accommodate speakers, whatever they say.

We have said that to interpret a text is to find the best explanation for why it *would be* true, not why it *is* true. Deciding whether something *is* true is a logically (though not necessarily chronologically) separate process, one that we can call *judgment*.

In the case of metaphor, we make certain assumptions in order to interpret the metaphor, such as that John is an elephant, and then in a logically separate judgement step, we decide which of our assumptions we are in fact prepared to believe.

Let us carry this approach one more step both toward formalization and toward embedding it in a larger framework. In Hobbs et al. (1990), it is suggested that a rational agent can be seen as going through the world, continuously trying to prove abductively the proposition "The situation I am in is coherent and I will thrive in it". The first clause generates interpretation, the second action. One kind of coherent situation, involving both interpretation and action is a turn in a conversation. Here there is a speaker S, a hearer H, and an utterance u. The utterance is an action on the part of S that serves in the achievement of S's goals. The utterance has an interpretation ϕ, which we may think of as a set of propositions. The hearer makes some kind of judgment about the information contained in ϕ. This can all be expressed by the rule

$$(\forall u, \phi, H, S)Serve\text{-}Goal(u, S) \wedge Interp(u, \phi) \wedge Judge(H, \phi) \supset Turn\text{-}in\text{-}Conversation(u, S, H)$$

That is, if an utterance u serves a goal of the speaker S, the interpretation of u is ϕ, and the hearer H judges ϕ, then there is a turn in a conversation in which S utters u to H.

A small set of axioms enable backchaining from $Interp(u, \phi)$ into the whole abductive framework of interpretation described in this paper. One may think of this as the entry into the *informational* aspect of discourse.

Other axioms having $Serve\text{-}Goal(u, S)$ as their consequent would tap into the whole *intentional* aspect of discourse, as elucidated in the work of Cohen and Perrault (1979) and many others. Thus, there might be an axiom that says that if H's believing ϕ serves a goal of S, then u serves a goal of S.

The conjuncts $Interp(u, \phi)$ and $Serve\text{-}Goal(u, S)$ share variables, so the informational and intentional aspects can influence each other. What might otherwise be the best interpretation of an utterance could be rejected if there is no way to relate it to the speaker's goals.

Finally, a first cut at an expansion of $Judge(H, \phi)$ might go as follows: To judge ϕ one must judge each proposition P in ϕ. There are three possibilities for P. P may be mutually known already, the given, in which case there is nothing to do. P may be inconsistent with what is already known, in which case it is judged false and rejected. Otherwise, P will be entered into the knowledge base, as mutual knowledge. This of course is oversimplified. In fact, the conjunct $Judge(H, \phi)$ taps into the whole question of belief revision.

In this account, it would be perfectly normal in the course of interpretation to assume a proposition that is known to be false. The judgment as to whether it should become a permanent belief is part of a logically separate step.

The predicate coercion solution to the problem of metaphor and truth has the advantage of giving an analogous treatment to metaphor and metonymy. Its disadvantage is that it involves a significant increase in notational complexity. The judgment solution has the advantage of requiring nothing that is not already required in a larger framework for discourse interpretation and generation anyway, but of course means that the details of that framework must be worked out.

Acknowledgements

I have profited from discussions on this work with Douglas Appelt, Elizabeth Hinkleman, Joanna Moore, and the participants of the NATO workshop on Computational Models of Communication in Trento, Italy, November 1990. The research was funded by the Defense Advanced Research Projects Agency under Office of Naval Research contracts N00014-85-C-0013 and N00014-90-C-0220, and by a gift from the Systems Development Foundation.

References

[1] Carbonell, Jaime, 1982. "Metaphor: An Inescapable Phenomenon in Natural-Language Comprehension" In W. Lehnert and M. Ringle (Eds.), *Strategies for Natural Language Processing*, pp. 415-434. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

[2] Charniak, Eugene, and Robert Goldman, 1988. "A Logic for Semantic Interpretation", *Proceedings, 26th Annual Meeting of the Association for Computational Linguistics*, pp. 87-94, Buffalo, New York, June 1988.

[3] Cohen, Philip, and C. Raymond Perrault, 1979. "Elements of a Plan-based Theory of Speech Acts", *Cognitive Science*, Vol. 3, No. 3, pp. 177-212.

[4] Davidson, Donald, 1967. "The Logical Form of Action Sentences", in N. Rescher, ed., *The Logic of Decision and Action*, pp. 81-95, University of Pittsburgh Press, Pittsburgh, Pennsylvania.

[5] Fass, Dan, 1988. "Collative Semantics: A Semantics for Natural Language Processing", MCCS-88-118, Computing Research Laboratory, New Mexico State University, Las Cruces, New Mexico.

[6] Gentner, Dedre, 1983. "Structure-Mapping: A Theoretical Framework for Analogy", *Cognitive Science*, vol. 7, pp. 150-170.

[7] Glucksberg, Sam, and Boaz Keysar, 1990. "Understanding Metaphorical Comparisons: Beyond Similarity", *Psychological Review*, Vol. 97, No. 1, pp. 3-18.

[8] Hobbs, Jerry R., 1983. "Metaphor Interpretation as Selective Inferencing: Cognitive Processes in Understanding Metaphor", *Empirical Studies in the Arts*, Vol. 1, No. 1, pp. 17-34, and Vol. 1, No. 2, pp. 125-142.

[9] Hobbs, Jerry R. 1985. "Ontological Promiscuity." *Proceedings, 23rd Annual Meeting of the Association for Computational Linguistics*, pp. 61-69. Chicago, Illinois, July 1985.

[10] Hobbs, Jerry R., Mark Stickel, Paul Martin, and Douglas Edwards, 1988. "Interpretation as Abduction", *Proceedings, 26th Annual Meeting of the Association for Computational Linguistics*, pp. 95-103, Buffalo, New York, June 1988.

[11] Hobbs, Jerry R., Mark Stickel, Douglas Appelt, and Paul Martin, 1988. "Interpretation as Abduction", SRI Technical Note 499, SRI International, Menlo Park, California. December 1990.

[12] Indurkhya, Bipin, 1987. "Approximate Semantic Transference: A Computational Theory of Metaphors and Analogies", *Cognitive Science*, vol. 11, no. 4, pp. 445-480, October-December 1987.

[13] Martin, James H., 1986. "Learning by Understanding Metaphors", *Proceedings*, Eighth Annual Conference of the Cognitive Science Society, Amherst, Massachusetts.

[14] Norvig, Peter, 1987. "Inference in Text Understanding", *Proceedings, AAAI-87, Sixth National Conference on Artificial Intelligence*, Seattle, Washington, July 1987.

[15] Ortony, Andrew, R. Reynolds, and J. Arter, 1978. "Metaphor: Theoretical and Empirical Research", *Psychological Bulletin*, Vol. 85, No. 5, pp. 919-943.

[16] Stickel, Mark E., 1988. "A Prolog-like Inference System for Computing Minimum-Cost Abductive Explanations in Natural-Language Interpretation", *Proceedings of the International Computer Science Conference-88*, pp. 343–350, Hong Kong, December 1988.

[17] Stickel, Mark E., 1989. "Rationale and Methods for Abductive Reasoning in Natural-Language Interpretation", in R. Studer (ed.). *Proceedings*, Natural Language and Logic, International Scientific Symposium, Hamburg, Germany, May 1989, *Lecture Notes in Artificial Intelligence* #259, Springer-Verlag, Berlin, pp. 233–252.

Can AI Systems Generate Creative Metaphors?

Bipin Indurkhya
Computer Science Department
Boston University
111 Cummington Street
Boston, MA 02215
email: bipin@cs.bu.edu

Abstract

Poetry is filled with creative metaphors—metaphors that appear anomalous at first, but produce meaningful deep insights after they are understood. Yet, these creative metaphors have remained outside the mainstream of cognitive science and AI research. In this article I outline an account of creative metaphors that sees the underlying process as that of change of representation, and explore its computational implications. I will argue that there are existing AI systems that are sufficiently rich to be able to generate instances of interesting and creative metaphors in the perceptual domains. I will address some of the philosophical issues related to creative metaphors in the context of AI systems, and then conclude by emphasizing the need to study the creativity of AI systems systematically so as to provide a better handle on human creativity, and to lead us closer towards developing a model of creative metaphors in poetry.

1 Introduction

Imagine a poet comparing hawthorn (a wild flower shrub) to water. What are the similarities between the two? Suppose someone tells you that the ocean is like a harp. What can she possibly mean? How can a harp be similar to the ocean? In each of these cases, there seem to be no similarities between the source and the target. Yet, in the poem *The Hawthorn in the West of Ireland,* Eavan Boland boldly likens hawthorns with water: "...So I left it [hawthorn]// stirring on those hills// with a fluency// only water has. And, like water, able// to redefine land. ..."; and in his classic poem *Seascape,* Stephen Spender compares the ocean to a harp: "There are some days the happy ocean lies// Like an unfingered harp, below the land.// Afternoon gilds all the silent wires// into a burning music for the eyes." Each of these metaphors is quite unconventional and strikes a jarring chord when we encounter it at first. But after the metaphor is assimilated, we are left with a beautiful, fresh and vivid imagery. In fact much of the beauty of these two poems comes from the novelty of their metaphors.

Anyone who reads some poetry will easily affirm that poetry is filled with such novel metaphors—metaphors which almost always appear discordant at first.[1] In fact, novel metaphors are so pervasive and their cognitive force is so different from conventional metaphors[2] and comparison-theoretic

[1] Of course, a fair number of conventional metaphors also permeate the poetry [Lakoff & Turner 1989].

[2] Conventional metaphors are those metaphors that are so much a part of the accepted convention that they are almost polysemic. For instance, 'theory as building' metaphor as in "she buttressed her theory with further evidence".

or similarity-based metaphors[3] that they have been treated separately in Wheelwright's theory of metaphor [1962] under the name 'diaphor'; and the so called anomaly theory of metaphor[4] has been proposed to account just for them. Yet, novel metaphors have received little attention in cognitive science and artificial intelligence research. Much of the cognitive science research has remained focused on similarity-based metaphors, where different formalisms have been proposed to articulate exactly how, and in which respects, the similarities between the source and the target underlie a metaphor. Artificial intelligence research has followed in the same vein by exploring the mechanisms to compute this underlying similarity; and there has been some work on conventional metaphors where the meaning of a metaphor is explicitly represented as polysemy.

One of the reasons for at least the AI researchers to have shirked from addressing the creativity of metaphors may have been that metaphor is thought to be primarily a linguistic phenomenon, and the state of the art of the Natural Language Processing Systems, with all its developments of the past few years, is still in its infancy. There are many facets of the literal and conventional use of the language that are not understood well enough to be incorporated in an AI system. Consequently, it seems that it is much too unrealistic to try to address creative metaphors with the existing technology.

However, I have argued elsewhere [Indurkhya 1991a] that that the cognitive mechanism underlying creative metaphors is quite different from the one underlying similarity-based or conventional metaphors; and therefore, the research on creative metaphors need not await a full understanding of the similarity-based or conventional metaphors. While a formal and comprehensive account of creative metaphors (along with conventional and similarity-based metaphors) is presented elsewhere [Indurkhya, forthcoming], in this article I will examine the AI implications of my approach.

In doing so, I will focus on what I will refer to as perceptual metaphors instead of linguistic ones, for the following reason. I believe that metaphor is not primarily a linguistic phenomenon; rather, it plays a fundamental role in all cognition, language being merely one such cognitive activity. This, together with the fact that, unlike the NLP technology, the AI technology on machine perception, especially visual and auditory perception, has become quite mature, suggests that the research on creative metaphors may be more fruitfully pursued if we start with perceptual domains. Indeed, I shall argue in this article that there are existing machine perception systems that are capable of generating instances of creative metaphors, though they have not been studied properly because perceptual metaphors have not been in the mainstream of research.

This article is organized as follows. In the next section I will spell out exactly what I mean by the term 'perceptual metaphor'. In Section 3, I will present a brief account of my approach to creative metaphors, which identifies the underlying cognitive mechanism as that of change of representation. Then, in Section 4, I will identify two different modes of creative metaphor, each with its cognitive force. Following that, in Sections 5 and 6, I will discuss the implication of my approach to creative metaphors for the existing AI systems. In Section 7, I will make some remarks on extending the account of creative metaphors presented here to linguistic metaphors. Finally, Section 8 will summarize the main points of this article and discuss the current state of implementation.

See Lakoff & Johnson [1980] for more examples.

[3]These are metaphors that are based on some existing similarity between the source and the target. For instance, in "the river is a snake."

[4]See Tourangeau & Sternberg [1982], pp. 209–212.

2 What are Perceptual Metaphors?

Let us say you are standing on the top of a hill on a beautiful Spring day, and you are asked to describe your ambience (a perceptual datum). If the Spring air has made you feel light-headed, carefree and joyous, you might reply something like: "The trees are dancing joyfully; the birds are singing merrily; and the Earth has put on its most colorful dress." This description would be clearly considered non-literal, for a less imaginative, though perhaps more accurate, description might have been: "I am standing on the top of a hill. The temperature is 70°F. The wind is 15 knots from NNE. Looking down the west side of the hill, there is a rectangular area of approximately 15 yards by 7 yards filled with red tulips, followed by..."; I am sure you get the idea.

The point here is that any perceptual datum may be described in various ways. As in the above example, while some of these descriptions would be considered literal, there would be many other metaphorical ones as well. Then the metaphorical descriptions may be further grouped into three classes—conventional, similarity-based and creative—depending on the nature of the metaphors involved. All such instances of metaphors I will consider as perceptual metaphors.

Thus, a distinguishing feature of the perceptual metaphors is the presence of the perceptual datum of which the metaphor is a description. Obviously, we may not make this assumption in a purely linguistic information processing situation. For instance, an NLP system can be expected to 'understand' the metaphor when presented with the text of a poem alone. There is no perceptual data involved here.

Though assuming the presence of perceptual datum does help me in articulating an account of creative metaphors, and it is crucial to my argument that there are existing AI systems capable of generating instances of creative metaphors, it is not as limiting an assumption as it may appear. In fact, all NLP systems end up creating some sort of world model that represents their understanding of the text, and this world model can very well be used in place of the perceptual datum. I will elaborate further on this remark later in Section 7.

3 Creative Metaphor as Change of Representation

How do creative metaphors create similarities? Where do the created similarities come from? While I have treated these issues at length elsewhere [Indurkhya 1990; 1991c; forthcoming], the crux of my ideas will be presented here briefly. The key points are to make a distinction between an object and its representation (or description); and to recognize that every object has its autonomous structure, which, while it can be represented in many ways, resists being represented in *any* way. These points can be appreciated in the context of the example I introduced at the beginning of the last section. The view from the top of the hill exists independently of any observer and independently of any way to describe it. Moreover, the view has its own autonomous structure in that though it can be described in several different ways—poetically or scientifically, in English or in Hopi, etc.—it cannot be described in any arbitrary way. For instance, it would be simply incorrect to say: "It is overcast and freezing cold."

In the rest of this paper, I will refer to the perceptual datum as *Sensory Motor Data Set (SMD),* since it corresponds to the structured set of stimuli that one receives from the perceptual datum; and I will refer to the set of concepts used in the description as *Concept Network (CN).* The term 'network' is included in CN to reflect the fact that the terms of descriptions are not independent set of labels, but are interrelated to each other. For instance, the words 'hot' and 'cold' are opposite

to each other in that they cannot be applied to the same datum at the same time in the same way.

Now a major part of our cognitive activity consists of converting the constant flow of SMDs that we receive from the external world into representations in terms of CNs. In this process, we group chunks of SMDs together, and identify them with the concepts from CNs. It is this process which makes our concepts referential, by connecting them with the objects in the outside world through our sensory-motor organs; and it is this process that brings the outside world within our cognitive access by reducing it to a handful of concepts and categories with which we are familiar and which we can manipulate internally.

Some of this conceptualization process is hard-wired in the structure of the brain so that we have no choice but to 'see' the world in certain ways. For example, our visual system is so arranged that our eyes prefer to see straight edges in the visual field, even when they are not there. Our language and culture further conditions us so that our world view favors certain ways of looking at the world, while blocking other modes of reality. In other words, given an SMD, we often choose to represent it in one way, and not in another, due to biological or cultural reasons. All such representation, we will refer to as 'conventional'.

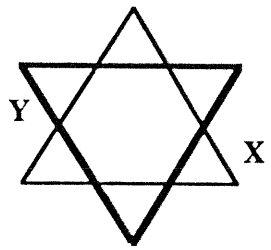
If all such constraints on the conceptualization process were rigid for biological reason, then metaphors would be impossible. There would be only one conventional description of every SMD, and that would be it. However, the effects of cultural and linguistic conditioning can sometimes be suspended, resulting in 'non-conventional' representations—or metaphors. Thus, at the heart of every creative metaphor can be seen a change of representation—a change from the conventional representation to an unconventional one.

This account of creative metaphors, and how they create similarities, can be better elucidated with an example that was first mentioned by Black [1979]. Suppose that the conventional representation of the figure of Star of David [Figure 1(a)] in a culture was based upon seeing it as two equilateral triangles [Figure 1(b)]. Now when someone compares it with Figure 1(c), which is conventionally represented as a hexagon with an osculating circle on each side [Figure 1(d)], initially there are no similarities between the two figures. Then, after a moment of introspection a flash of insight occurs. Aha! The conventional representation of the figure of Star of David is replaced with a new one that sees in it a hexagon with an equilateral triangle on each side [Figure 1(e)]. *Now* there are similarities between the two figures.

A couple of features of my account of metaphor must be emphasized here. Firstly, similarities are seen as characteristics of pairs of representations, and not of pairs of objects. Thus, we may not ask if these two SMDs are similar are not, since SMDs are cognitively inaccessible, except by means of conceptualization but that turns them into representations. This is where similarity-based approaches to metaphor reach their limitation because they rely on similarities from existing representations, but the most creative part of metaphor is in changing representation, and not in figuring out similarities from existing representations.

Secondly, this account explains why creative metaphors are so important cognitively. In changing representation of the object, *new* information is created. Features of the object that were not available in the conventional representation may become available in the new representation. Thus, creative metaphors allow us to partially recover the loss of information that is an inevitable by-product of any conceptualization.[5]

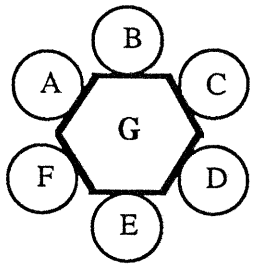
[5]See also Indurkhya [1991b], pp. 393–394.



(a) The figure of Star of David.

equilateral-triangle (X),
equilateral-triangle (Y),
large (X), large (Y),
flip-horizontal (X,Y),
vertex-up (X),
base-horizontal (X),
base-horizontal (Y),
concentric (X,Y).

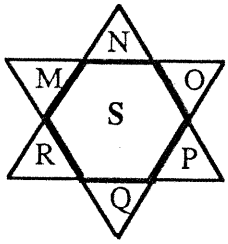
(b) A representation of Figure (a).



(c) Another geometric figure.

circle (A), small (A), circle (B), small (B),
circle (C), small (C), circle (D), small (D),
circle (E), small (E), circle (F), small (F),
regular-hexagon (G), large (G),
osculate (A,G), osculate (B,G),
osculate (C,G), osculate (D,G),
osculate (E,G), osculate (F,G).

(d) A representation of Figure (c).



triangle (M), small (M), triangle (N), small (N),
triangle (O), small (O), triangle (P), small (P),
triangle (Q), small (Q), triangle (R), small (R),
regular-hexagon (S), large (S), osculate (M,S),
osculate (N,S), osculate (O,S), osculate (P,S),
osculate (Q,S), osculate (R,S),
rotate-60 (M, N), rotate-60 (N,O),
rotate-60 (O,P), rotate-60 (P,Q),
rotate-60 (Q,R), rotate-60 (R,M).

(e) Changed representation of Figure (a).

Figure 1: An example of how similarities are created through change of representation. There are no similarities between the figures (a) and (c) as long as the similarities are sought with respect to their existing representations in (b) and (d) respectively. The similarities are apparent at once, however, when the representation of figure (a) is changed to the one shown in (e).

4 Modes of Creative Metaphor

It would be useful here to distinguish between two modes of creative metaphors to facilitate the later discussion. Following the terminology of Gordon [1961], who introduced them as creative problem-solving heuristics, we will refer to them as *making the familiar strange,* and *making the strange familiar.* The key to distinguishing these two modes is how one determines whether a representation (or description) is metaphorical or not.

All the examples of creative metaphors we have presented so far fall in 'making the familiar strange' category. Here the subject (I mean the person who is experiencing the perceptual datum and coming up with a representation of it) is familiar with the perceptual datum in the sense that she can come up with a conventional representation of it (which, of course, would be different from the metaphorical representation). In other words, the perceptual datum is familiar, yet its representation is strange (unconventional), hence the name 'making the familiar strange'. Notice that in this case while the conventional representation of the object is not needed in order to come up with the metaphorical representation, it (the conventional representation) is needed if the subject were asked, "Is your representation metaphorical?" This is because metaphorical representation is defined in contrast with the conventional one. So, in order to determine if a representation is metaphorical, the subject has to rule out that it is not conventional.

The second mode of creative metaphors contains those instances when the perceptual datum is so unfamiliar to the subject that she has no existing representation for it. We all have had experiences that we felt at a loss of words to describe. One's first ride in a roller coaster might serve as a good example. In such situation, one is either speechless, or communicates by not very informative exclamations like 'awesome!!', 'thrilling!!', etc. However, if one were forced to come up with a more informative description (for instance, if one were writing for a newspaper column to describe the thrills of bobsledding), the descriptions are invariably metaphorical—in fact, they often involves most creative metaphors. In all such cases, the object of description is strange (unfamiliar), yet the description is in familiar terms, hence the name 'making the strange familiar'. Since no conventional representation of the perceptual datum exists, any representation is metaphorical. Technically, the change of representation here is from non-existent to something.

In order to figure out if a representation is metaphorical or not in this mode, the subject can no longer rely on comparing the new representation with the conventional one, since the latter does not exist. Instead, the subject needs to have a meta-awareness of her conceptualization process. That is, she has to be able to recognize a 'strange' perceptual datum when she saw one. This may be implicit in the amount of difficulty involved in coming up with a representation in the first place.

5 Creative Metaphor in Artificial Intelligence Systems

Consider a machine vision system for recognizing objects [Figure 2]. It accepts the bit-map image from a digital camera (the SMD) and produces semantic-net type of representation of the input image. It has two concept networks: one for recognizing animals and the other for recognizing houses. The former has concepts—which are essentially semantic-net primitives—'horse', 'camel', 'trunk', 'hump', 'legs', 'head', etc. The latter has concepts 'roof', 'chimney', 'wall', 'door', 'yard', etc.

On being presented with the image of a camel, the machine vision system recognizes it as such, and represents it as a network of concepts from the 'animals' concept networks. The system would

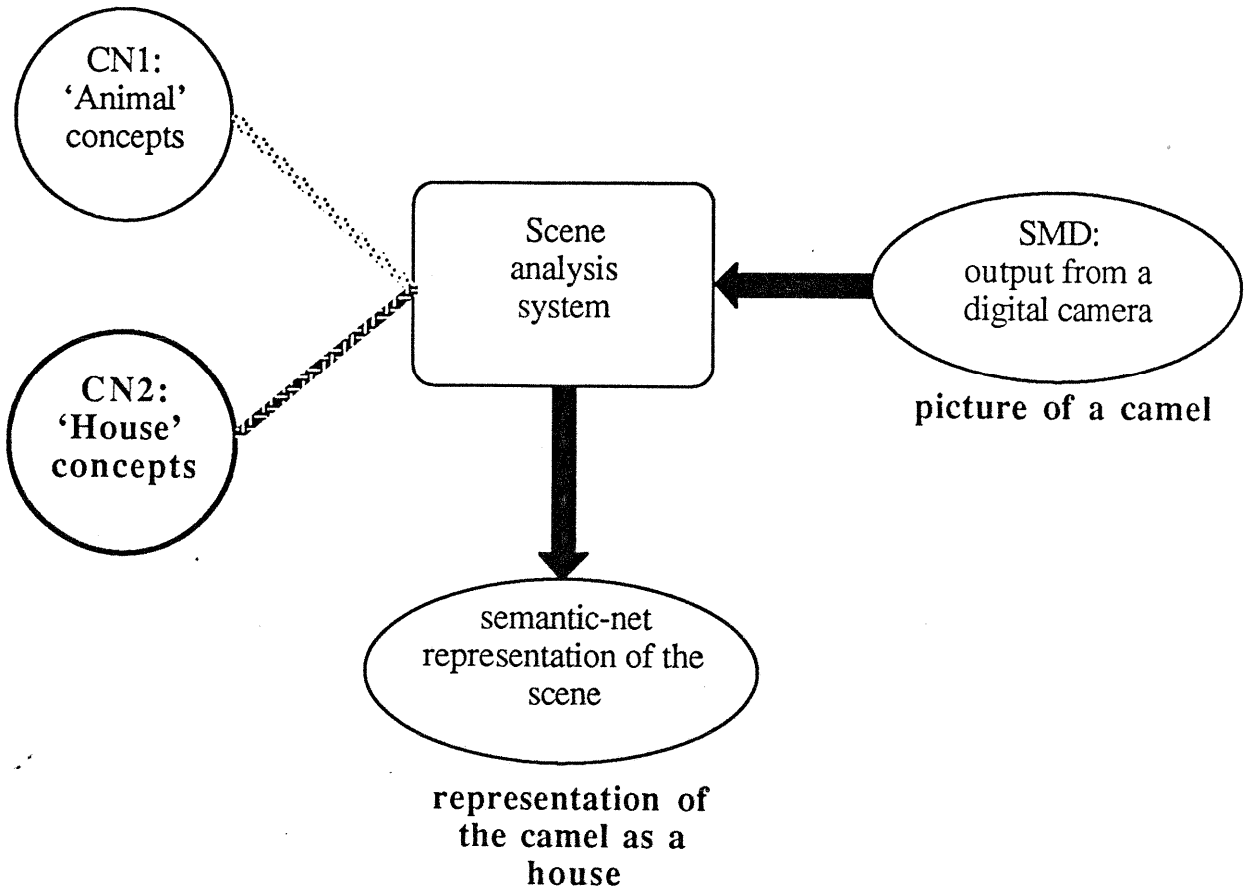


Figure 2: An example of creative metaphor in a computational setting. The scene analysis system would have normally represented the camel as an animal concept. But forced to use the house concepts, it could meaningfully represent it as a house. The process creates the similarities between a camel and a house.

record the fact that it has a 'hump', a 'neck' which is 'long', 'legs' which are 'four' and are 'long' and 'skinny'; its color is 'ochre', etc. All these concepts are identified with the appropriate regions, or the attributes of the regions, of the image in the process of recognition. This would be the conventional representation of the input image.

Now suppose that the machine vision system takes the same image as input, and tries to recognize it as a house. Now very different regioning and labeling routines will take over and try to reorganize the same SMD in a way so that the 'house' concepts may be instantiated. If the machine vision system is sufficiently rich, and is designed to handle noisy images, it may actually come up with such a representation. Moreover, the representation would be far from arbitrary, but would reflect a coherent match between the features of the image and the meanings of the house-related concepts. For instance, it would be the hump of the camel that would be labeled as 'roof', and the neck as 'chimney'.

This situation is no different from the one I described earlier, when a person standing on the top of a hill could choose to describe essentially the same view in two different ways—one literal and one metaphorical. Here we have the same image being represented in two different ways, one conventionally and the other creatively.

My example of recognizing the picture of a camel as a house may not be the best one, and my discussion of what a machine vision system might do is clearly hypothetical; but it is certainly not unrealistic given the existing state of the art of machine vision systems. Also, an analogous scenario can be created for a speech recognition system as well, where the same output of a microphone could be interpreted in two different ways depending on which concept network (vocabulary corresponding to what is thought to be the domain of discourse) is being used. The upshot of all this is that the machine perception systems can come up with creative representations. What is needed is to study this creativity systematically, and design machine perception systems that can not only work well in the domains they were designed for, but produce interestingly meaningful representations when operated in domains for which they were not intended. This latter part would essentially involve modeling human creativity. An architecture for modeling creative metaphors based on this idea is proposed in Indurkhya [1990].

The computational mechanism underlying the creation of similarity by a metaphor can be further elucidated here. If the machine vision system is given the picture of a camel and the picture of a house, both of which it represents conventionally, and then asked to find similarities between the two, then there might be none as long as the similarities are sought from the conventional semantic-net representations of the two figures. However, in order to create the similarity, the system would have to re-represent one of the images by using concepts from the conventional representation of the other image. Interestingly, the asymmetry of metaphors follows naturally in this account, since to represent the house as an animal will, in general, have a very different result (it may not even be possible!) than representing the camel as a house.

6 Determining When is Metaphor

We can now address the issue of self-awareness. So far, determining whether a representation generated by a machine perception system is conventional or metaphorical required us to have the God's eye view, for it is we who know that the picture is really that of a camel. How can the system be given this capability? To answer this we need to consider separately the two modes of creative metaphors that we introduced earlier.

Consider the case of 'making the familiar strange' first. Notice first that coming up with a metaphorical representation does not require that the conventional representation be arrived at first or be available. However, to determine whether a representation is metaphorical or not is another matter. For this, the system would need to know that the given perceptual datum is *not really* what it has made it to be; just as the person on the hill needs to know that the earth is not really wearing a colorful dress in order to realize that her description is metaphorical. Notice also that a metaphorical representation would almost always require some kind of bias (whether by setting some internal parameter or by giving it explicitly as an input) to force the representation to be a different from the conventional one. So, all that is needed to determine whether a representation is metaphorical or not is to remove this bias, generate another representation of the perceptual datum, and then compare it with the original representation. If they are the same, then the original representation is conventional; if they are different, then the original representation is metaphorical.

The other mode of 'making the strange familiar' is somewhat more complex. To understand this, consider the situation when the machine vision system is designed to recognize house scenes only. Suppose that on being presented with the picture of a camel, the system ends up representing it anyway. Now we, having the God's eye view, know that the picture is not really a house, and since the system has no concepts to represent animals, the datum is 'strange' to it. However, from the system's point of view, it is really a house, for it knows no other way of conceptualizing it.[6]

To incorporate this mode of creative metaphor, the only suggestion I have at this point is to have the system generate a confidence factor along with the representations. A high confidence factor would mean that the system is quite confident that the representation is what the perceptual datum really is. A low confidence factor would mean that though the perceptual datum has been represented using familiar concepts, it may be really something else altogether.

7 Some Remarks on Linguistic Metaphors

So far we have addressed the creativity issue for those metaphors where the perceptual datum was given. How about linguistic metaphors, where we do not have the access to perceptual datum? Surprisingly, the process is not as different as it may seem at first. Consider the fragments of the poems introduced at beginning of this article. What is going on when we read and understand the verse. Are you really comparing the meanings of the words 'sea' and 'harp' when you read Spender's *seascape?* Or is there some other crucial piece of information involved? Where does that piece of information come from? Remember that we only have the text of the verse available to us.

While I am currently extending my account of creative metaphors to wrestle with precisely these questions, I will offer a tentative explanation here. When we are reading any text, as we understand each word, phrase, or sentence, we imagine or partially recreate what might be called a 'world model'.[7] This world model is more detailed—we fill in the details with imagery based on our past perceptual experiences—than what might be gleaned from the dictionary meanings of the words occurring in the text and its syntactic analysis. Now when a metaphorical expression is

[6]This precise point, made in the context of human cognition, is the theme of Colin Turbayne's excellent book *Myth of Metaphor.*

[7]This model of text understanding not only is compatible with some previous psychological research on language understanding [Neisser 1976] and perceptual theories of metaphor [Johnson and Malgady 1980; Verbrugge 1980], but is also quite consistent with what most existing NLP systems actually do, though there are wide variations in how exactly the world model is represented and what it is called.

encountered, it is the existing world model at that point that invalidates a literal application of the expression and constrains the possible metaphorical interpretations. In other words, it is the world model that plays the role of the perceptual datum.

For instance, consider the verse from Spender's poem again. When you read it, you have to imagine more than what is explicitly mentioned in the poem. You will have to imagine the waves in the ocean, the wind making patterns on the waves, the sunlight reflecting on it in certain ways, etc. All this information—which comes from your prior perceptual experiences with watching waves, and cannot be extracted from a dictionary and/or encyclopedia—becomes part of your world model. When you read about harp, you now have to re-represent this world model as a harp. The pattern on the waves now becomes harp-strings. It is in this change of representation that the creativity of metaphor lies, and it is this change of representation that is responsible for creating similarities where none existed before.

Thus, the model of creative metaphors will be built upon a model of linguistic information processing where with every word and phrase is associated a World Model Builder that encodes the perceptual knowledge about the referent of the word or phrase. A major part of linguistic information processing would then involve using the world model builders of individual words to build a world model that represents an imagined perceptual datum of which the text is a representation. Creative metaphors would involve using the world model builders of metaphorical words and phrases in new ways so that parts of these world model builders are instantantiated in novel ways and the existing world model is represented differently in the process.

Notice that in this account, perceptual acquaintance with the referents of the words is absolutely necessary. I believe that this is a crucial piece of information in understanding creative metaphors. If someone has never seen sunlight playing on the waves (or watched it on TV, or seen photographs of it) I am sure that they would not be able to appreciate Spender's poem at all.

8 Conclusions and the State of Implementation

In this article I have argued that there exist machine perception systems that are capable of generating instances of creative metaphors. However, since perceptual metaphors have not been given enough attention, this creativity has not been studied properly. What is needed is a systematic study of how machine vision systems and speech recognition systems represent their perceptual datum in novel and interesting ways. This study will not only further our understanding of human creativity, but also pave the way for modeling creative metaphors of the language, such as the ones in Boland's and Spender's verses.

A study of the change of representation mechanism underlying creative metaphors in somewhat artificial albeit semantically rich microworlds has already begun in at least two independent projects. The Copycat program of Hofstadter and Mitchell [1990] provides a probabilistic, emergent model for generating appropriate representations for proportional analogy relations involving strings of characters, as in "What is to 'xyz' as 'abc' is to 'abd'". Indurkhya and O'Hara [forthcoming] are implementing a model for solving proportional analogy relations involving geometric figures—such as the ones shown in Figure 1 (a) and (c)—that require the representation of at least one of the figures to be changed. Unlike previous approaches to proportional analogy, such as the well-known ANALOGY program of Evans [1963], the emphasis in both these models is on finding the appropriate representations (or descriptions) of the terms of the analogy relations in different

contexts.[8] Perhaps what is needed now is to apply the insights from these models to some real-world domains such as machine vision and speech recognition, and study the change of representation taking place there.

9 References

Black M., 1979, "More about Metaphor", in A. Ortony (ed.) *Metaphor and Thought,* Cambridge Univ. Press, Cambridge, UK, pp. 19–45.

Boland E., 1989, *The Hawthorn in the West of Ireland,* appeared in *New Yorker,* March 27, 1989, p. 111.

Evans T.G., 1963, *A Heuristic Program to Solve Geometric-Analogy Problems,* Ph.D. Dissertation, Dept. of Mathematics, M.I.T., Cambridge, Mass.

Gordon W.J.J., 1961, *Synectics: The Development of Creative Capacity,* Harper & Row, New York, NY.

Hofstadter D.R. and Mitchell M., 1991, "An Overview of the Copycat Project," Technical Report CRCC-52-1991, Center for Research on Concepts and Cognition, Indiana University, Bloomington, Ind.

Indurkhya B., 1990, "Metaphor as Change of Representation," submitted for publication.

Indurkhya B., 1991a, "Modes of Metaphor," *Metaphor and Symbolic Activity 6 (1),* pp. 1-27.

Indurkhya B., 1991b, "On the Role of Interpretive Analogy in Learning," *New Generation Computing 8,* pp. 385–402.

Indurkhya B., 1991c, "A Computational Perspective on Similarity-Creating Metaphors," to appear in S. Hockey and N. Ide (eds.) *Research in Humanities Computing,* Oxford University Press, Oxford, U.K.

Indurkhya B., forthcoming, *Metaphor and Cognition,* Kluwer Academic Publishers, Dordrecht, The Netherlands (Fall 1991).

Indurkhya B. and O'Hara S., forthcoming, "Modeling the 'Redescription' Process in the Context of Proportional Analogies," Technical Report, Computer Science Department, Boston University, Boston, Mass.

Johnson M.G. and Malgady R.G., 1980, "Toward a Perceptual Theory of Metaphoric Comprehension," in R.P. Honeck and R.R. Hoffman (eds.) *Cognition and Figurative Language,* Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 259–282.

Lakoff G. and Johnson M., 1980, *Metaphors We Live By,* Univ. of Chicago Press, Chicago, Ill.

Lakoff G. and Turner M., 1989, *More than Cool Reason: A Field Guide to Poetic Metaphor,* University of Chicago Press, Chicago, Ill.

Neisser U., 1976, *Cognition and Reality,* W.H. Freeman, San Francisco. Calif.

Spender S., 1986, *Seascape,* in S. Spender *Collected Poems 1928–1985,* Oxford University Press, New York, NY.

Tourangeau R. and Sternberg R.J., 1982, "Understanding and Appreciating Metaphors", *Cognition 11,* no. 3, May 1982, pp. 203–244.

[8] Evans, in fact, was quite aware of the need for redescription in solving proportional analogies, but the architecture of his program was ill equipped to handle this. See Indurkhya [forthcoming], Chap. 10, for a detailed discussion.

Turbayne C.M., 1962, *The Myth of Metaphor,* Yale Univ. Press, New-Haven; revised edition with an appendix by R. Eberle "Models, Metaphors, and Formal Interpretations", Univ. of South Carolina Press, Columbia, 1970.

Verbrugge R.R., 1980, "Transformations in Knowing: A Realist View of Metaphor," in R.P. Honeck and R.R. Hoffman (eds.) *Cognition and Figurative Language,* Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 87–125.

Wheelwright P.E., 1962, *Metaphor and Reality,* Indiana Univ. Press, Bloomington.

MetaBank: A Knowledge-Base of Metaphoric Language Conventions

James H. Martin
Computer Science Department and
Institute of Cognitive Science
University of Colorado,
Boulder, CO
80309-0430
martin@cs.colorado.edu
303-492-3552

July 29, 1991

Abstract

The frequent and conventional use of non–literal language has been a major stumbling block for natural language processing systems since the early machine translation efforts. Metaphor, metonymy and indirect speech acts are among the most troublesome phenomena. Recent computational efforts addressing these problems have taken an approach that emphasizes the use of systematic knowledge about non–literal language conventions. This paper describes our current efforts to supply this knowledge in the case of conventional metaphor. We are constructing *MetaBank: an empirically derived and theoretically motivated knowledge–base of English metaphorical conventions.*

1 Introduction

The frequent and conventional use of non–literal language has been a major stumbling block for natural language processing systems since the early machine translation efforts. Metaphor, metonymy and indirect speech acts are among the most troublesome phenomena. This is both because they occur quite frequently and because they have resisted computational efforts to deal with them in a uniform and tractable manner.

Consider the following examples of conventional metaphor.

(1) It *came* to me that I had to prepare a talk for the conference.
(2) It *hit* me that I didn't have anything to say.
(3) It *struck* me that this wasn't a good situation.

The italicized words in these examples are being used in ways that clearly deviate from the physical or spatial meaning of the words. Nevertheless, there is little that is novel in these examples. In order to understand and generate uses like these, we make use of a rich set of underlying metaphorical conventions. In these examples, these conventions structure beliefs and believers as objects with locations. Correspondingly, change of location indicates a change in belief.

The main thrust of this *Metaphoric Knowledge* [8] approach to metaphor is that the interpretation of metaphoric language should proceed through the direct application of specific knowledge about the metaphors in the language. This approach has been embodied in MIDAS (Metaphor Interpretation, Denotation, and Acquisition System)[11]. MIDAS is a set of computer programs that can be used to perform the following tasks: explicitly represent knowledge about conventional metaphors, apply this knowledge to interpret metaphoric language, and learn new metaphors as they are encountered.

While the MIDAS project demonstrated some significant results, it nevertheless had a major shortcoming. The effectiveness of the approach for language interpretation, generation and acquisition was obviously

dependent on the size and the correctness of the knowledge–base of non-literal conventions. Unfortunately, the knowledge–base used by MIDAS does not have any kind of real coverage, nor does it have an empirically verifiable basis.

This paper describes our current efforts to overcome these problems in the case of conventional metaphor. We are constructing of *MetaBank: an empirically derived and theoretically motivated knowledge–base of English metaphorical conventions.* More generally, this effort can be seen as an attempt to develop methodologies for empirically capturing a wider variety of language conventions in computationally usable knowledge–base forms.

2 Knowledge-Based Linguistic Approaches

MIDAS can be seen as a part of a recent trend in the area of semantic interpretation that uses systematic knowledge about non–literal language conventions. Among the research projects following this general approach are Zernik's RINA system (phrasal idioms) [14], Fass's META5 system (metaphor and metonymy) [3], and Hinkelman's system (indirect speech acts) [4, 6]. These systems all attempt to leverage knowledge of systematic language conventions in an attempt to avoid resorting to more computationally expensive methods. Finally, like MIDAS they all use small knowledge–bases of conventions with no real coverage and little empirical validation. (Hinkleman's corpora-based research [5] is a notable exception to this).

3 Limitations to the Knowledge–Based Paradigm

As mentioned in section 0.1 the primary current limitation to systems following a knowledge–based approach to non–literal phenomena is the size and correctness of the knowledge–base itself. Consider the ways that this general problem manifests itself in MIDAS.

- There is no way of telling if the known metaphors that are included are represented correctly.
- There is no way of telling if a given metaphor is of high enough frequency and range to warrant significant effort.
- MIDAS lacks any empirical basis for both the abstract domain independent metaphors, and the domain specific UNIX metaphors, that are included in its knowledge-base.
- There is no way of determining the coverage of the current knowledge–base. How many conventional metaphors are really out there?
- There is no empirical basis for the particular hierarchical relationships embodied in the current knowledge–base.

4 MetaBank

The MetaBank project is an attempt to solve some of the above problems for the case of conventional metaphor and metonymy. To be specific, we are identifying resources and developing methodologies that will allow us to construct a robust knowledge–base of English metaphoric and metonymic conventions. This knowledge–base will be represented in a form that will make it generally useful to natural language processing applications. We will also show that it has direct relevance to various projects involved in the construction of what have been called large common-sense knowledge–bases.

The following sections describe the three–part approach we are following for the construction of MetaBank. The first part involves the collection of on–line textual resources and databases of linguistic generalizations. The second part is the development of a methodology for analyzing these resources and deriving generalizations from them. The final part is the actual construction and use of a knowledge–base based on the preceding analyses.

4.1 Resources

We are planning to make use of the following on–line resources as a part of this effort.

1. Berkeley Metaphor List: An on-line list of known metaphors with analyses.
2. Domain Specific Corpora: A variety of UNIX specific texts.
3. General Text Corpora: Selected texts from the Association for Computational Linguistics Data Collection Initiative.

The following sections describe our efforts to make use of these resources.

4.1.1 Berkeley Metaphor List

The Metaphor Project, at the University of California at Berkeley, under the direction of George Lakoff, has been collecting an on-line database of conventional English metaphors. The database entries for each metaphor include descriptions of the source and target domains, a set of example sentences using the metaphor, a short analysis, and where appropriate, pointers to a longer analysis from which the entry was derived. The entries are being compiled by hand from the published metaphor literature, and from an on-going series of graduate seminars at Berkeley. Currently, a total of approximately 50 metaphors of varying levels of specificity have been entered into the database. This collection will be used as the starting point for MetaBank.

4.1.2 Corpora

We are making use of two kinds of text corpora in this project: text having to do with the UNIX domain, and more general text selected from the ACL/DCI. The focus in our initial efforts will be on the UNIX text. More generic text will mainly be used to verify and extend results gained from the domain specific text.

The primary practical reason for the use of specialized text from UNIX domain is that the main testbed for MetaBank will be natural language application programs that operate in this domain. It is generally agreed that NLP systems will for some time to come be most useful and robust in limited, specialized domains. This has in the past entailed the construction by hand of specialized lexicons. What we are seeking is a means by which generic lexicons can be productively applied to new domains via knowledge of generic English conventional metaphors. Therefore the particular UNIX and computer system metaphors discovered, are not the primary accomplishment. We are primarily interested in the mechanisms by which generic English metaphors are specialized into particular domains.

Within the UNIX domain, we are using two kinds of text that can be characterized as expert prose and user prose. The first source is the standard UNIX documentation known as the MAN pages. The second source of expert prose that we are using is the full text of a textbook called *UNIX System Administrators Handbook.* This text is intended for readers who will be primarily responsible for administering and maintaining UNIX system.

The largest source of text for user's language is an archive of user mail to consultants for help in using UNIX. Mail to this alias is monitored by systems personnel and dealt with. A total of approximately 150,000 words of text (both user queries and consultants answers) have been collected.

To illustrate the kind of text being collected, consider the following example from the Boulder corpus.

```
>From rieman@tigger
To:  trouble@eclipse
Subject:  locked terminal
Cc:  rieman@tigger

My terminal hanging off of eclipse tty06 is locked again.    I can see
the processes (login to tigger) on eclipse, but I can't kill 'em --
eclipse says they aren't mine.

For your info here's how this happened (it has happened before):

I typed ''remote'' at the eclipse login prompt and logged into tigger.    My
default terminal setting is vt100 -- that's what I use most of
the time.    I only use the zenith when I'm not doing anything
important or time-consuming.
```

I forgot to tell tigger I was using a zenith and I more'd a file.

The zenith locked.

I hit shift-reset to knock some sense into the zenith, then pressed ctrl-z to get out of more. I killed the more process.

I typed ''setenv TERM z89'' to tell tigger I was using a zenith.

Then I tried again to use more. It locked again, and that's where I am now. Resetting the terminal has no effect, and killing my tigger process from another machine has no effect.

-john

Consider some of the metaphorical conventions illustrated by this message. (Eclipse and tigger are the names of computers in the following).

I can't *kill* 'em -- eclipse *says* they aren't *mine*

The use of *kill* here reflects a metaphor that views processes as being alive, where termination corresponds to killing The use of *say* reflects the metaphor that views computers as communicative agents capable of carrying on dialogs with users. In this case, program output is viewed as dialog by the computer. Finally, the use of the possessive *mine* illustrates the process as possession of the user metaphor, that underlies many of the security mechanisms in computer systems.

...ctrl-z to *get out of* more. I *killed* the more process.

This sentence illustrates a use of the process as enclosure metaphor. It manifests itself here in the notion of exiting or *get out of* as suspension of an ongoing process. Note that in the immediately following sentence the user switches metaphors to refer to a different aspect of the same process.

...to *tell* tigger I was using a zenith...

This sentence is another instance of the communicative agent metaphor. Here, however, we see simple input to a program as a speaking to the computer.

...and thats *where I am* now...

This final example is an illustration of a general metaphor structuring states as locations. Note that this final metaphor illustrates an important point. These domain specific corpora will also contain uses of non-domain specific metaphors. Therefore, domain independent abstract metaphors will be derived from all the textual sources we collect, including the technical ones.

Preliminary study of this data indicates that unconstrained user language is quite metaphorical in nature. Moreover, users are much freer in their use of metaphor than are the authors of corresponding manuals and text books.

4.2 Metaphor Analysis

The text analysis methodology we are developing uses the Berkeley metaphor list as a generator of probes into the various text corpora. The following steps illustrate our current methodology.

Step 1: Choose a metaphor of interest from the current list of metaphors. This is the target metaphor.

Step 2: Generate a set of linguistic probes for this metaphor. This involves selecting a set of words associated with the source domain of the chosen metaphor. The selection of these words will be guided by the analysis of the metaphor provided by the Berkeley list to ensure that the words are likely to occur with the metaphor. These words will typically be chosen from a well–defined and well–studied semantic field from a spatial or physical domain.

Step 3: Choose a corpus of interest appropriate to the metaphor being studied. This will typically involve an alternation between the domain specific corpora and the more general corpora. The purpose of this alternation is to determine that the behavior of the metaphor in the domain specific corpora is consistent with its behavior in other domains and with its description in the Berkeley list.

Step 4: Probe the selected corpus for uses of these verbs. This step simply involves searching the corpus for instances of the probes.

Step 5: Classify the resulting sentences according to their meaning. This typically involves classifying the use according the the semantic properties of the arguments to the probe verb. This classification step will initially be dependent on the subjective ability of the human classifier to identify various uses. This step will result in one the following possibilities.

1. A literal use of the probe word.
2. An instance of the metaphor in question.
3. Another metaphor in the known metaphor list.
4. Another conventional metaphor not yet in the list.
5. An isolated homonymous word sense with no metaphoric basis.
6. A novel use of some kind.

The result of each probe is, therefore, a combination of information about the metaphor in question, and information about other uses of the probe words. Some of the information that can be gleaned about the target metaphor is:

- How frequently does this metaphor occur?
- How frequently are the probe words used with the target metaphor?
- How many of the probe words from the source semantic field actually occur with this metaphor.
- What is the correct level of abstraction for the source and target domains of the metaphor.

Some of the information that can be gotten from negative probe sentences (sentences not relevant to the target metaphor) is:

- Identification of conventional metaphors not in the current database.
- Frequency information about other metaphors.
- Frequency information about the probe word with other meanings.
- Identification of novel metaphors. (Metaphors that would not be judged as conventional).
- Frequency of occurrence of novel uses.

Note that this method of classification of probe words based on the text accompanying probe words, is analogous to the methods used by Choueka [2] and Lesk [10] for word-sense tagging. In these efforts, local context (the words immediately surrounding the target word) was successfully used to classify a given use of a word in a text as an instance of a pre-determined dictionary sense. In this work, the role of the dictionary sense is replaced by a conventional metaphor.

This methodology can be seen as a combination of top-down and bottom-up techniques. On the top–down side it makes full use of the analytical linguistic work that has been done (as it is represented in the Berkeley database) to guide the search for metaphors in a text. The bottom-up side is the actual examination of large amounts of text. Note that methodology we are using allows us to probe the corpus in a focussed manner by using metaphors from the Berkeley list, while at the same time it allows us to discover metaphors not contained in the original list.

4.3 An Example: The Container Metaphor

In order to make this methodology more concrete, consider the following example analysis. In this example, the Boulder `trouble` mail archive is probed for instances of container metaphors. This is one of the more productive and well-studied kinds of metaphor in English. The Berkeley Metaphor list contains thirteen container metaphors (thirteen distinct target domains structured with the source concept container). In addition, significant effort was spent on computer system container metaphors as a part of the MIDAS project.

For the purposes of the this example, the probes *enter, get into, exit* and *get out of* from the source domain of containers or enclosures were used to probe the selected corpus. The following table shows the results.

choose "search" from the menu	enter "X", hit "apropos", and when the
BTW, when I try to login, it let's me	enter my password, then prints daveheib
Randomly in these two windows when I	entered the vi editor
I	enter my password when 'rlogin'g in from
prompt I have to be right there to	enter my passwd
Kermit tells me to	enter a Receive command and I do so.
I have a file called calendar that I	enter appointments
in. It seems that whenever I	enter an item in this file,
the appointment that I just	entered. It's not a huge
As soon as I	enter windows it fails (most of the time).
they	get into the queue, but they don't print
Oh, good grief! I can't	get into my home directory
log me into my home directory and I cannot	get into it by any other means.
a program that attempts to	get into computers around the Internet.
an intruder making persistent attempts to	get into Internet
I can then do a successfull :vi to	get into vi
hung up when I tried	exiting out of X windows. Keyboard, mouse, etc.
I tried to	exit, at which point I received the prompt
I was	exiting suntools when cashew choked in a
now hangs when I want to	exit and save using the VI editor
window, log into the machine. Now,	exit the login
i was on the console	exiting suntools with crtl-d crtl-q
pressed ctrl-z to	get out of more.
the only way I can	get out of it is to completely shut
would like to get my degree and	get out of here sometime

This probe yielded the following results from this probe:

1. All the probes occurred with the interactive–system as enclosure metaphor that was extensively studied with MIDAS. This was also the most frequent use.

2. A use of the metaphor with the MORE command which neither UC nor MIDAS had classified as an interactive–system.

3. A conventional use, structuring the home directory as an enclosure for the user, that was not known to MIDAS. It was used only with *get into.*

4. A conventional use of *enter* as in *to enter your passwd.* Again not known to MIDAS, and not discussed in the Berkeley list.

5. A relatively new use structuring programs as entering computers from the outside. As in *a program that attempts to get into computers around the Internet.* Mentioned in the Berkeley list, not known to MIDAS.

6. One generic use contained in the Berkeley list, *get my degree and get out of here*, that could be considered literally as leave, or metaphorically as graduate.

The net result of this was a strong confirmation of the Interactive-Process-As-Enclosure metaphor as currently known to MIDAS. It occurred quite frequently, with the characteristics predicted by the current representation. It also resulted in two conventional uses that have to be added to the metaphor list and added to MIDAS, and one new computer use that also needs to be added.

4.4 Knowledge–Base Construction

The first phase of construction involves integrating all of the metaphors from the Berkeley list into the current MIDAS knowledge–base. This involves representing the following kinds of knowledge.

- Knowledge about the source concepts of the metaphor. Particular care will be taken with the representation of the source concepts since it is assumed that they will occur across domains.
- Knowledge of the target concepts.
- Knowledge about the metaphors themselves. This information will largely come from the analyses of the metaphors from the Berkeley list.

The representation of this knowledge will entail significant effort since the Berkeley list does not give detailed information about the source and target domains. In this phase we intend to make extensive use of existing formalizations of physical and spatial domains. [7, 9].

This knowledge will be represented using the KODIAK [13] representation language. KODIAK is an extended semantic network language in the tradition of KL-ONE [1] and its variants. Briefly, MIDAS currently uses a hierarchical knowledge–base of structured associations linking source and target concepts that are conventionally related metaphorically. The details of KODIAK and the representation of metaphoric knowledge can be found in [11]. It should be noted that the KODIAK representation is sufficiently similar to other extended semantic network systems to ensure the transportability of MetaBank to other systems.

As noted above the Berkeley list currently contains approximately 50 high level metaphors. The current MIDAS knowledge-base contains 22 domain independent metaphors (11 of which are contained in the Berkeley list) and an additional 18 UNIX specific specializations. We believe that a stable MetaBank for English will contain no more than 200 high level generic metaphors.

5 Related Areas

There are two on–going areas of research that are of significant relevance to the work we are engaged in here. These are the construction of large machine usable lexicons, and the construction of large common–sense knowledge-bases.

5.1 Large Lexicons

In recent years, the lexicon has taken on a more central role, both in linguistic theories, and in computational efforts. Paralleling this trend, there has been increasing interest in the construction of large, robust, on–line lexicons. The advent of machine readable dictionaries and large text corpora has made much of this work possible. The notion of a word-sense, however, remains problematic [12] in much of this work. Many of the fine grained sense-distinctions made by lexicographers often seem unmotivated and arbitrary. On the other hand, many of the senses needed in the specialized domains in which these lexicons will be used are missing altogether.

In order to make this problem more concrete, consider again the belief metaphors discussed at the beginning of this paper.

(4) It *came* to me that I had to prepare a talk for the conference.
(5) It *hit* me that I didn't have anything to say.
(6) It *struck* me that this wasn't a good situation.

In each of these examples, a physical motion or locative word is being used to refer to either a belief state or a change in belief. Current lexical approaches are forced to simply attempt to list a meaning having to do with belief as a separate sense for each of the italicized words in these examples. The problem with this approach becomes obvious when one considers the productivity of the conceptualization underlying these senses. Consider the following examples.

(7) I was *led* to the conclusion that...
(8) I was *dragged kicking and screaming* to the conclusion...
(9) John was being *pulled* to the center of the debate.
(10) John would not *budge* from his *position.*
(11) Mike's *position* on this issue has not *shifted.*

In these examples, a relatively straightforward conventional metaphor representing four core mappings could account for all the surface lexical items: beliefs are objects with locations, believers are objects with locations, shared location entails active belief by the believer, and finally changes in location indicate changes in belief.

It is clear that a static lexical approach to this kind of phenomena is not appropriate. It places the impossible burden on the lexicon builder of capturing in a finite list a phenomena that is inherently generative.

5.2 Common–Sense Knowledge–Bases

CYC [9] and Tacitus [7] are two major recent efforts to construct large common–sense knowledge–bases. The MetaBank project relates to these projects in a number of ways. It obviously requires a common–sense conceptual representation for the various source and target domains that play roles in conventional metaphors. The first point of contact, therefore, is that we would like to avoid having to produce detailed representations of these domains ourselves. We plan to take advantage of the detailed analyses of various common–sense domains that have been done already.

The second point of contact with these efforts is a more subtle one. Builders of common–sense knowledge–bases often find themselves in one of two problematic situations when representing various domains. The first situation occurs when the designer has produced a logically consistent ontology for a domain based on non-language criteria, that bears no obvious relationship to the way that the domain is actually expressed in natural language.

Consider, as a concrete example, the various temporal logics that have been developed to reason about time. It is quite clear that none of these logics can predict the following conventional language uses.

(12) I had a lot of time to *kill.*
(13) I can't *waste* my time on things like this.
(14) Y ou shouldn't *spend* any time on that.

The ontology of time in any system based on these logics clearly must be augmented if they are to be able to deal with how we talk about time. In the case of English, the metaphor structuring time as a resource is necessary.

The second situation occurs when the knowledge–base designer is forced to rely heavily on linguistic evidence from a single language for the representation of a particular common–sense domain. In this case, the resulting representation may simply be a representation of a conventional metaphor for the domain for a given language. (Indeed for many applications this may be sufficient). This problem is most apparent when one considers the problem of translation.

The MetaBank project provides a partial solution to both of these problems. In the first situation, MetaBank provides the connection between the representations of domains and how the domains are expressed in language by augmenting domain representations with representations of conventional metaphors. In the second situation, MetaBank provides a motivated role for linguistic data that does not force an entirely linguistic representation.

6 Results

Following are the major results we expect of the MetaBank project:

- A knowledge–base of English metaphorical conventions that will be plausibly useful to many natural language processing applications.
- Corpora-based empirical information about the nature and distribution of metaphor usage in English.
- A methodology that will extend to the study of other forms of non-literal language conventions.

References

[1] Ronald J. Brachman and James Schmolze. An overview of the kl-one knowledge representation system. *Cognitive Science*, 9:346–370, 1985.

[2] Y. Choueka. Looking for needles in a haystack. In *Proceedings of the RIAO*, March 1988.

[3] Dan Fass. *Collative Semantics: A Semantics for Natural Language.* PhD thesis, New Mexico State University, Las Cruces, New Mexico, 1988. CRL Report No. MCCS-88-118.

[4] Elizabeth Hinkelman. *Linguistic and Pragmatic Constraints on Utterance Interpretation.* PhD thesis, University of Rochester, Computer Science Department, Rochester, NY, 1989. Report No. 288.

[5] Elizabeth Hinkelman. Surface signals of illocutionary intentions. In *Proceedings of the 5th Rocky Mountain Conference On Artificial Intelligence*, 1990.

[6] Elizabeth Hinkelman and James Allen. Two constraints on speech act ambiguity. In *Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics*, 1989.

[7] Jerry Hobbs, William Croft, Todd Davies, Douglas Edwards, and Kenneth Laws. Commonsense metaphysics and lexical semantics. *Computational Linguistics*, 13(3), 1987.

[8] George Lakoff and Mark Johnson. *Metaphors We Live By.* University of Chicago Press, Chicago, Illinois, 1980.

[9] Douglas B. Lenat and R.V. Guha. *Building Large Knowledge Bases.* Addison-Wesley, 1990.

[10] Michael Lesk. Automatic sense disambiguation using machine readable dictionaries: How to tell a pine cone from an ice cream cone. In *Proceedings of SIGDOC*, 1986.

[11] James H. Martin. *A Compututational Model of Metaphor Interpretation.* Academic Press, Cambridge, MA, 1990.

[12] James Pustejovsky. Towards a generative lexicon. *Computational Linguistics*, 17(1), 1991.

[13] Robert Wilensky. Some problems and proposals for knowledge representation. Technical Report UCB/CSD 86/294, University of California, Berkeley, Computer Science Division, May 1986.

[14] Uri Zernik. *Strategies in Language Acquisition: Learning Phrases from Examples in Context.* PhD thesis, University of California, Los Angeles, Computer Science Department, Los Angeles, CA, 1987.

Bi-directional Parsing for Idiom Handling

Yuji Matsumoto
Katsuyoshi Yamagami
Makoto Nagao

Department of Electrical Engineering
Kyoto University
Sakyo-ku, Kyoto 606 Japan
phone:+81-75-753-5345
email:matsu@kuee.kyoto-u.ac.jp

Abstract

The paper describes an idiom handling method based on a bi-directional extension of a concurrent Chart parsing system. Grammar rules are expressed either as global rules or as local rules. While the global rules are those for normal grammatical constructions and obey the normal left-to-right parsing strategy, the local rules are for idiosyncratic constructions like idioms and can operate bi-directionally, where both of the rules operate in a common concurrent parsing mechanism. A local grammar describes an idiom or an idiosyncratic construction and placed in the lexical entry of the word that characterizes the expression. Interaction of local grammar rules with global grammar rules as well as other local rules or meta-level parsing facilities like gap finding are realized very naturally.

1 Introduction

Idioms or other idiosyncratic expressions in natural languages pose a hard problem that cannot be bypassed for practical natural language processing systems. Numbers of ways for handling idiomatics expressions have been introduced. Preprocessing or postprocessing of idioms as an independent processing of the ordinary parsing has at least the following problems. Preprocessing cannot handle idioms that include grammatical constructions, which are available after the ordinary parsing process. Postprocessing requires some salvaging mechanism for the constructions that are ungrammatical according to the ordinary grammar rules, since some idioms have ungrammatical constructions in the ordinary sense. Idioms should not be expressed and treated as grammar rules from both computational and representational viewpoints. Representing idiomatic expressions by grammar rules and analyzing them using a general parsing strategy are unpractical since the number of idioms is numerous. On the other hand, in dictionaries, idioms are usually described in lexical entries of their characteristics words.

For handling idioms, the followings are natural requirements:

1. Ordinary grammar rules and rules for idioms should be independently described. This does not mean that they must be defined in different formalism.
2. At the analysis phase, ordinary grammar rules and idioms should be treated simultaneously.
3. The analysis of idioms should be invoked by a specific word or phrase that characterizes them.
4. The idiom handling should be amenable to an efficient parsing algorithm since the number of idioms is abundant.

As in usual dictionaries, we define an idiom in the lexical entry that plays the most characteristic role in the expression. We refer to such a lexicalized description as a local grammar rule. A local

grammar rule is activated only when the word appears in a given input string, and it tries to look for specified words or phrases in any directions according to the description.

We are currently under development of an natural language analysis system based on a parallel parsing framework [1] [2]. The parser is a parallel and bottom-up implementation of Chart Parsing [3]. In the system, all of nonterminal symbols in a grammar[1] are compiled into predicates in a logic programming language.[2] Our aim is to make the system flexible and useful to accept wide range of language phenomena as well as grammar formalisms. Although the program runs in parallel by processing multiple grammar rules simultaneously, a rule body is analyzed from left to right linearly. We first extend our parallel parsing mechanism so that it can handle rule bodies in both direction, to the left or to the right.

Usefulness of such an extended algorithm shows up for handling idioms, which are peculiar to specific words or phrases. We show how local grammars are described to represent and to handle idioms. Such local grammars operate in accordance with global grammar rules in the common parsing mechanism.

Most of the recent works on idiom handling represent idioms as patterns and put them at lexical entries. Our work share the same property. Some of the works represent them as grammar rules, e.g., [6], some represent them as lexicalized rules or patterns at lexical entries, e.g., [7], [8], [9], and others represent them as subcategorization, e.g., [10], [11]. Most of the work get the advantage by employing a unified processing of ordinary and lexical rules. Some of them devise a peculiar algorithm form their rules, and the other render it to a general treatment of the notation they use, e.g., the general treatment of subcategorization. We first show our concurrent parsing method for ordinary grammar rules, which is a concurrent implementation of Chart Parsing. Then, we show that a small augmentation makes is possible to handle lexicalized (local) grammars at the same level as the ordinary (global) rules, so that it achieves all of the requirements listed above.

2 Concurrent Parsing Mechanism

2.1 The basic mechanism

Our parser is a concurrent implementation of Chart Parsing. Definite Clause Grammar rules are compiled into a Prolog program that implements an efficient parsing program. In this section, we show the basic parsing mechanism only with context-free skeleton of grammar rules.

The parsing system itself is called the SAX system and consists mainly of a translator from grammar rules into a Prolog program of concurrent Chart parser. The main idea is to define the bottom-up Chart parsing operations concerning each nonterminal symbol directly by Prolog clauses.[3] The definition of a nonterminal symbol exactly implements what an inactive edge of the corresponding nonterminal symbol should do in Chart Parsing.[4] Active edges of Chart are represented as messages that are used in communication between nonterminal symbols (We will call them nonterminal processes or just processes). Shared variables are used as communication channels between nonterminal processes. We refer to such shared variables as *streams*. In the SAX system, the following grammar rule is translated into the subsequent set of Prolog clauses:

```
s --> a, b, c.
```

[1] We use Definite Clause Grammars [4] in our kernel grammar description.

[2] The current system is implemented in SICStus Prolog [5] utilizing its freeze mechanism to operate concurrently.

[3] In this sense, it is similar to the direct top-down backtracking implementation of a DCG in Prolog's default evaluation strategy. In our method, the compiled Prolog program achieves a bottom-up nonbacktracking parser in Prolog's default evaluation strategy.

[4] An inactive edge is a partial parse tree whose daughters have been completely determined, while an active edge is an incomplete partial parse tree that has some daughter left unanalyzed.

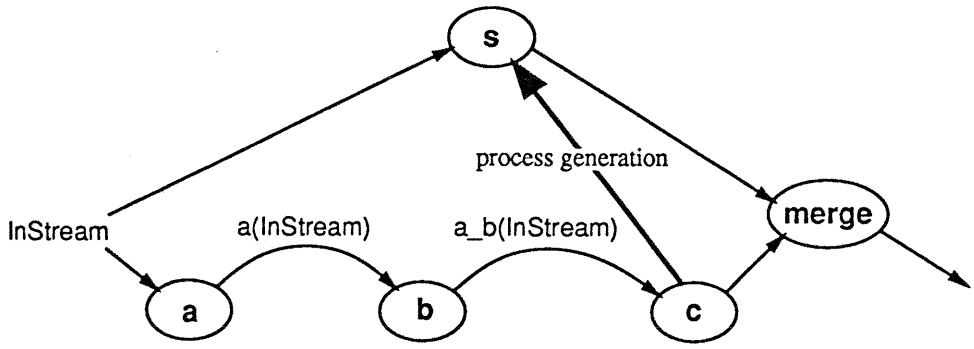


Figure 1: Process communication and generation

```
a(InStream, [a(InStream)]).

b(InStream, [b(InStream)|OutStream]) :-
     b2(InStream, OutStream).
b2([a(InStream)|R], [a_b(InStream)|OutStream]) :-
     b2(R, OutStream).

c(InStream, [c(InStream)|OutStream]) :-
     c2(InStream, OutStream).
c2([a_b(InStream)|R], OutStream) :-
     s(InStream, OutS1),
     c2(R, OutS2),
     merge(OutS1, OutS2, OutStream).
```

The first argument of a process is an input stream from processes existing to the left and the second argument is an output stream to processes appearing to the right. These are just the part in the whole program that corresponds to the above grammar rule. The whole SAX program consists of collection of all of such clauses together with some subtle programs that handle cases like receiving an empty stream.

Figure 1 shows a configuration where a consecutive occurrence of nonterminal processes, 'a', 'b' and 'c', generates another nonterminal process 's' through message communication. In the program above, 'a(InStream)' and 'a_b(InStream)' correspond to messages passed between these processes. Messages carry a stream within themselves so as to make sure that a newly generated process receives the proper input stream. In the figure, the new process 's' covers the three nonterminals below itself. So, it should receive the same messages the process 'a' receives, and passes its output to the right of the process 'c' via a merge program.[5]

The important feature of our parsing method is that once a partial parse tree (i.e. a nonterminal process) is constructed it deals with all of the adjacent incomplete partial parse tree, which appears in its input stream, and never backtracks to construct identical structures repeatedly. This advantage continues to hold in the subsequent extension. The detail of the parsing mechanism can be seen elsewhere.[1][2]

2.2 Bi-directional extension

We describe the bi-directional extension of the parser by assuming that grammar rules have annotations for specifying the order of analysis.

[5]The real implementation utilizes difference lists to get rid of costly merge operations.

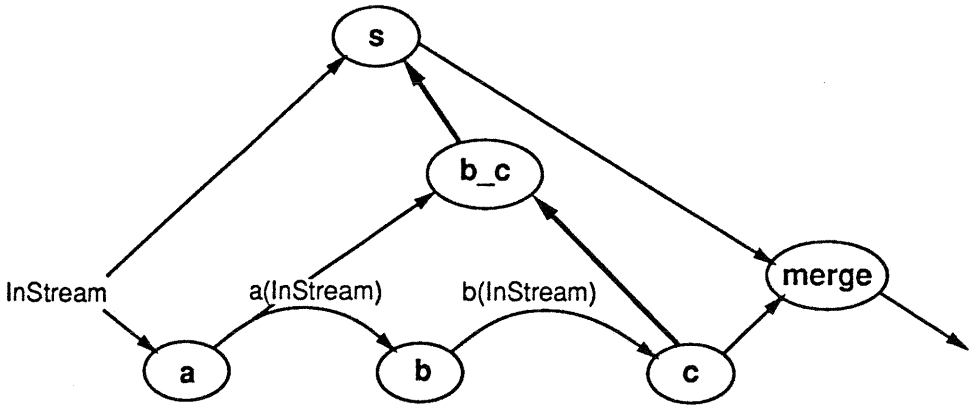


Figure 2: Analysis starting from the middle

```
s --> a^3, b^1, c^2.
```

This rule specifies that to parse an input string, 'b' should be recognized first, then 'c', and finally 'a'. It is done by translating this rule into the following SAX clauses:

```
a(InStream, [a(InStream)]).

b(InStream, [b(InStream)]).

c(InStream, [c(InStream)|OutStream]) :-
     c2(InStream, OutStream).
c2([b(InStream)|R], OutStream) :-
     b_c(InStream, OutS1),
     c2(R, OutS2),
     merge(OutS1, OutS2, OutStream).

b_c([a(InStream)|R], OutStream) :-
     s(InStream, OutS1),
     b_c(R, OutS2),
     merge(OutS1, OutS2, OutStream).
```

The translated program is self-explanatory. The configuration of parsing with these clauses is depicted in Figure 2. The difference from the basic method is that the clause 'b_c' appears instead of the definition of 'b2' in the program of the basic method. The clause 'b_c' stands for a consecutive occurrence of the symbols 'b' and 'c'. It generates a call to 's' when it receives a message produced by the nonterminal symbol 'a'. Processes other than 'b' also produces a message to show its own occurrence as usual.

The process 'b' passes a message of its occurrence ('b(InStream)') to the right so that it will be received by a 'c' process if it ever exists. Then, the process 'c' generates a process named 'b_c' that represents a sequence of 'b' and 'c'. It then looks for a message from an 'a' process, and it successfully finds an 's'.

The above two examples show that a partial parse tree can be represented either by a data (i.e. a_b in the first example) or a process (i.e. b_c in the second example). Passing a data to the right corresponds to parsing to the right direction, and generating a process for investigating to the left corresponds to parsing to the left direction. The next section applies this idea to grammar rules locally specified in lexical entries so as to implement idiom processing in accordance with the global parsing processes.

3 Local Grammar Rules and Idiom Handling

In the previous section, we assumed that annotated grammar rules are directly compiled into a Prolog program so that the evaluation ordering agrees with the annotation.

However, when a grammar rule is written in a lexical entry as a local grammar rule, a problem arises. In practice, it is unrealistic to compile every local grammar rule in a dictionary into a Prolog program beforehand. The number of idiomatic expressions in a language is out of scope of such compilation. We cope with this problem by defining a small number of Prolog clauses that handle local grammars.

First, we show a simplified example of a local grammar specification.

Example 1: Specification of 'take care of' in the entry of 'care'

```
noun( [ cat:noun,
        lg:[ [ r([cat:p, lex:of]),
               l([cat:verb, lex:take]),
               =>, verb( ... ) ] ]
      ] ) --> [care].
```

When the word 'care' appears in a given sentence, a noun nonterminal process is invoked that has information written in this lexical entry. Lexical entries have their own information as a set of attribute-value pairs. In the example, '`lg`' stands for 'local grammar'. It says that if a preposition 'of' appears to the right and a verb 'take' appears to the left of a noun 'care' then it generates another verb, which stands for the idiom, 'take care of'. This rule represents inflected forms as well since '`lex`' indicates just the root form.

The syntax of a local grammar consists of two data structures connected by '`=>`'. The left-hand side of '`=>`' is a sequence of symbols (or conditions) which the local grammar has to satisfy, and the right-hand side specifies what is generated after the left-hand side has been satisfied. Each condition on the left-hand side is either a feature structure with a control marker or a set of logical formulae. A logical formula is either a form of feature structure that specifies a value restriction of a certain feature, or a sequence of Prolog goals. In the example, '`[cat:p,lex:of]`' specifies that the category name and the root form of the lexical entry must be '`p`' and '`of`', respectively. We have at present six control markers. They are summarized as follows[6]:

`r`(*Features*)	A grammatical category or a word that satisfies conditions '*Features*' must appear to the right.
`l`(*Features*)	A grammatical category or a word that satisfies conditions '*Features*' must appear to the left.
`c`(*Features*)	The current feature structure must satisfy conditions '*Features*'.
`ro`(*Features*)	A grammatical category or a word that satisfies conditions '*Features*' may optionally appear to the right.
`lo`(*Features*)	A grammatical category or a word that satisfies conditions '*Features*' may optionally appear to the left.
{*Goals*}	Prolog goals, *Goals*, are evaluated.

A nonterminal process that has a local grammar feature invokes a process called '`idiom_handler`' and passes the input stream and all of the grammatical information to it. The '`idiom_handler`' processes the local grammar according to its first condition.

1. If the first condition has a form '`r`(*Features*)', then the local grammar description is passed to the right.
2. If the first condition has a form '`l`(*Features*)', then it looks for a message from the input

[6] '*Features*' is of a form of a feature structure, which specifies conditions to be satisfied by the constituent

stream that is consistent with '*Features*'. Then, it continue the job with the rest of the local grammar description.

3. If the first condition has a form '`c(`*Features*`)`', then consistency of the current feature structure with '*Features*' is checked. Failure of the evaluation makes it cease the job. If the evaluation succeeds, it continues the job with the rest of the local grammar description.
4. If the first condition is of the form {*Goals*}, the Prolog goals, *Goals*, are evaluated.
5. If the first element is '`=>`', it activates the nonterminal process that follows.

A process receiving an option control marker (`ro` or `lo`) does the same job as `r` or `l` except that it also invokes a process that simply ignores the option control marker. The following Prolog clauses, although in simplified form, achieve the above operations. A nonterminal process that includes local grammar rules invokes an `idiom_handler`. The `idiom_handler` also receives the feature structure of the process. '`FS`' in the following program is assumed to be such a feature structure. (Nonterminal processes operate according to the global rules as well. When a new nonterminal process is generated by a global rule, it receives the local grammar description from the *head*[7] of the grammar rule, which is explicitly written in global rules.)

```
idiom_handler([], _, FS) --> [].
idiom_handler([LG|LGs], InStream, FS) -->
     each_idiom(LG, InStream, FS),
     idiom_handler(LGs InStream, FS)

each_idiom([r(Cond)|LG], InStream, _) -->
     [lg(Cond,InStream,LG)].
each_idiom([l(Cond)|LG], InStream, FS) -->
     search_left(InStream, Cond, LG, FS).
each_idiom([c(Cond)|LG], InStream, FS) -->
   ( { check_condition(Cond, FS) } ->
     each_idiom(LG, InStream, FS)
   ; [] ).
each_idiom([ro(Cond)|LG], InStream, FS) -->
     [lg(Cond,InStream,LG)|OutS],
     each_idiom(LG, InS, FS).
each_idiom([lo(Cond)|LG], InStream, FS) -->
     search_left(InStream, Cond, LG, FS),
     each_idiom(LG, InStream, FS).
each_idiom([{Goals}|LG], InStream, FS) :-
   ( { call(Goals) } ->
     each_idiom(LG, InStream, FS)
   ; [] ).
each_idiom([=>|Head], InS, _) :-
     { arg(1, Head, InS) },
     phrase(Head).
```

Other than this extension to the SAX program, each nonterminal process requires some additional definition for dealing with a message passed from an '`idiom_handler`. Suppose there is a grammar rule as follows:[8]

```
s(S) --> np(NP), { NP =u= S/subj }
         vp(VP), { VP =u= S }.
```

Then, the following clauses are generated by the SAX translator:[9]

[7] the same sense as GPSG and Head Grammar

[8] '=u=' stands for a unificaition operation of feature structures

[9] Output streams are implicitly represented by difference lists, which are automatically added by making use of DCG

```
np(InStream, NP) -->
     [np(InStream,NP)|OutS],
     { extract_lg(NP, LG) },
     idiom_handler(LG, InStream, NP),
     np2(InStream, NP).

vp2([np(InS,NP)|InStream], VP) -->
     { copy_term((NP,VP),(New_NP,New_VP)),
       New_NP =u= S/subj, New_VP =u= S },
     s(InS, S),
     vp2(InStream, VP).
vp2([lg(Cond,InS,LG)|InStream], VP) -->
     { copy_term((Cond,VP),(New_Cond,New_VP)),
       New_Cond =u= New_VP },
     each_idiom(LG, InS, New_VP),
     vp2(InStream, VP).
```

Every nonterminal process, when it receives a message named 'lg', checks if the received condition is satisfiable with its own feature structure. If the condition is satisfied, it invokes a new 'each_idiom' process to handle the rest of the local grammar description. Copying of the feature structure is necessary since some distinct processes may share same feature structures.

Some idioms accept modifiers to their constituents. We might say, for instance, "take good care of someone". Some other idioms may not accept any modifiers to their constituents. Such conditions are specified using 'c' marker. Suppose we do not like to have modifiers to 'care' in the above example, we can indicate it with 'c([mod:'-'])', which specifies that 'care' may not have any modifiers.

It is not essential that category names of nonterminal symbols be specified in a local grammar description. The following example defines the idiomatic expression, 'not only ··· but also ···' where 'also' is optional:

Example 2: Specification of 'not only ··· but also···' in the entry of 'only'

```
adverb([ cat:adverb,
         lg:[ [ l([cat:adverb, lex:not]),
                r([cat:Cat]),
                r([cat:conj, lex:but]),
                ro([cat:adverb, lex:also]),
                r([cat:Cat]),
                c(FS),
                { functor(Head,Cat,2),
                  arg(2,Head,FS) },
                =>, Head ],
                      ...
            ]
        ] ) --> [only].
```

This local grammar specifies that the two anonymous places in the idiom may be any category but they must have the same category name.[10] The condition marker 'c(Cond)' is used to get the current feature structure for 'Head' of the local grammar. The condition may include any Prolog calls. The first argument of the 'Head' is unified with the Input stream of the left most constituent of the idiom, which is done by the last clause of 'each_idiom'.

Figure 3 shows a sample parse tree of the system. The sentence includes idioms introduced in

notation.

[10]This is indicated by the variable 'Cat', which may take any value.

```
Enter a sentence.
|: He is at home not only with science but also with literature.
parsed
                                sentence
                                   |
 np--------------------------------vp
 |                                 |
pron be-----------------------------pred
 |   |                              |
 |   |  prep--n-------------------------pp
 |   |   |    |                         |
 |   |   |    |   adv-adv-------pp------conj-adv--------pp
 |   |   |    |    |   |         |         |   |          |
 |   |   |    |    |   |   prep---np       |   |   prep-----np
 |   |   |    |    |   |    |      |       |   |    |      |
 |   |   |    |    |   |    |      n       |   |    |      n
 |   |   |    |    |   |    |      |       |   |    |      |
 he  is  at  home not only with science but  also with literature
```

Figure 3: A parse tree with two intermingled idioms

Example 2 and Example 3. Such intermingled idioms are analyzed naturally in the system.

Example 3: Specification of 'at home with $\cdots$' in the entry of 'home'

```
noun([ cat:noun,
       lg:[ [ l([cat:prep, lex:at]),
              r([cat:pp, prep:lex:with]),
              =>, pred( ... ) ],
                    ...
          ]
      ] ) --> [home].
```

A well-known difficulty in handling idioms is an interaction with the extraposition. For instance, 'take care of' can be used in the following way:

"Good care was taken ϕ of by the man."

We are also working on integrating gap finding mechanism into the system, where an extraposition caused by a passive form, a relative clause, or an interrogative clause are analyzed by finding the gap and by restoring the extraposed constituent to the original place. The detail will be reported elsewhere. In the case of the above example, 'good care' is restored to the original place (indicated by ϕ). Then, the local grammar rule for 'take care of' is invoked at the place and the interaction of an idiom and an extraposition is settled naturally.

4 Conclusion

We have presented a concurrent parsing method in logic programming and its extension to bi-directional parsing, which is useful for handling language phenomena like idioms. This approach enables our system to deal with two kinds of grammar rules, global and local, interactively within a common mechanism. The mechanism provides a natural way of representing and analyzing idioms. Modification to an idiom, and an interaction of idioms with other idioms as well as extrapositions are handled also in a natural way.

Our bi-directionality is restricted so that users have to specify the directions of analysis explicitly. The general bi-directionality proposed by other researchers, e.g.,[12] means that the parsing system can analyze a grammar rule body not in a predetermined direction but in any direction. They propose an idea of avoiding spurious ambiguity caused by multiple usages of a single grammar rule. Our work

is more application oriented, and we do not encounter this issue since we specify the analysis ordering of grammar bodies. This may, however, lose some flexibility.

The efficiency of the system is quite satisfactory. Our basic parser produces all possible analyses of an English sentence of 20 to 30 words within one second in a compiled SICStus Prolog implementation on SUN SPARCstation1, using an English grammar with about 250 rules. Descriptions of idioms as local grammar rules do not cause heavy overhead since most of them are influential locally. We are currently developing a comprehensive English and Japanese grammars in this framework. Other than this work we are developing a disambiguation facility in the system by making use of its concurrency. The concurrency makes it possible to suspend implausible analyses to reduce ambiguity as well as to reduce the execution time. This facility, for instance, will be used to assign higher scores to the idiomatic readings obtained from local rules over the literal readings.

References

[1] Matsumoto, Y., "A Parallel Parsing System for Natural Language Analysis," New Generation Computing, Vol.5, No.1, pp.63-78, 1987.

[2] Matsumoto, Y. and R. Sugimura, "A Parsing System Based on Logic Programming," Proc. 10th International Joint Conference on Artificial Intelligence, pp.671-674, Milan, Italy, Aug. 1987.

[3] Kay, M., "Algorithm Schemata and Data Structures in Syntactic Processing," XEROX PARC, CSL-80-12, Oct. 1980.

[4] Pereira, F.C.N. and D.H.D.Warren, "Definite Clause Grammars for Language Analysis – A Survey of the Formalism and a Comparison with Augmented Transition Networks," Artificial Intelligence, 13, pp.231-278, 1980.

[5] Carlsson, M. and J. Widén, "SICStus Prolog User's Manual," Swedish Institute of Computer Science, Oct. 1988.

[6] Gross, M., "Lexicon – Grammar, The Representation of Compound Words," COLING-86, pp.1-6, 1986.

[7] Wilensky, R. and Arens, Y., "PHRAN – A Knowledge-Based Natural Language Understander," 18th ACL, pp.117-121, 1980.

[8] Wilensky, R., "A Knowledge-based Approach to Language Processing: A Progress Report," 7th IJCAI, pp.25-30, 1981.

[9] Zernik, U. and Dyer, M.G., "The Self-Extending Phrasal Lexicon," Computational Linguistics, Vol.13, No.3-4, pp.308-327, 1987.

[10] Kaplan, R.M. and Bresnan, J., "Lexical-Functional Grammar: A Formal System for Grammatical Representation," Chap.4 of 'The Mental Representation of Grammatical Relations' J. Bresnan (ed.), MIT Press, pp.173-281, 1982.

[11] Erbach, G., "Lexical Representation of Idioms," IBM Germany Science Center, Institute for Knowledge Based Systems, IWBS Report 169, April 1991.

[12] Satta, G. and O. Stock, "Formal Properties and Implementation of Bidirectional Charts," Proc. 11th International Joint Conference on Artificial Intelligence, pp.1480-1485, Detroit, Aug. 1989.

A Formalization of Metaphor Understanding in Situation Semantics

Tatsunori MORI Hiroshi NAKAGAWA

Division of Electrical and Computer Engineering
Faculty of Engineering
Yokohama National University
Tokiwadai, Hodogaya-ku, Yokohama, 240, JAPAN
mori%naklab.dnj.ynu.ac.jp@relay.cs.net nakagawa%naklab.dnj.ynu.ac.jp@relay.cs.net

Abstract

In this paper, we present a new theory which can account for understanding metaphorical expressions in a discourse. In order to deal with the information obtained from a metaphorical expression in a discourse, it is necessary that the part of semantics which deals with metaphor understanding should harmonize with the rest of semantics which deals with the ordinary discourse understanding. Accordingly we adopt the formalization based on situation semantics, in which by dealing with the interpretation of a metaphorical expression and the information in the expression separately the metaphor understanding is done according to the hearer's resource. That is, an actual meaning connoted by either a word or a phrase in a metaphorical expression may vary from its ordinary meaning. As a result, metaphorical expressions can convey some other new information. We would like to formulate this variation as a process of introducing new resources, namely, establishing correspondences between a source domain and a target domain. This theory can naturally explain metaphor understanding closely related to a discourse, such as the case that the information which the hearer obtains from a metaphorical expression depends on the context.

1 Introduction

Metaphor — one of rhetorical methods — is used not only in literary writings, but also in many conversations in our everyday life as effective means of communication[LJ80]. Therefore, metaphor understanding is one of the most important themes which we must deal with to have a sophisticated natural language understanding system. The metaphor understanding process has been discussed from various points of view by a number of researchers[TIT89, Ind87, GFS87].

It is the essence of metaphors that one thing or idea (we will call it α) is expressed by means of something else (β). For example, in a metaphorical phrase "A man like a wolf", "a man" corresponds to α and "a wolf" corresponds to β. In the rest of this paper, we will use the term *target* (or *target domain*) to refer to a thing α and *source*(*source domain*) to refer to β. According to the interaction theory of metaphors proposed by Black, an "implicative complex", which is a set of inferences that can be drawn about a domain, is associated with the domain. A metaphorical expression works by projecting an implicative complex of a source domain upon a target domain[Ind87]. There are a number of studies of the metaphor understanding based on this theory, in which such projections, or correspondences, are found by an analogical mapping[GFS87, Hol89, Ind87]. Most of these studies, however, deal only with understanding an isolated metaphorical sentence, and regard the metaphor understanding process just as establishing correspondences by some analogical mapping. Indeed the establishment of correspondences is the main part of the metaphor understanding process, but in order to deal with the information obtained from a metaphorical expression in a discourse, it is necessary that the part of semantics which deals with metaphor understanding should harmonize with the rest of semantics which deals with the ordinary discourse understanding. There are few researches based on such a viewpoint, that is, a framework which refers to the analysis of the metaphorical expression in a discourse[1].

[1]In the same viewpoint as ours, Hobbs deals with metaphor understanding in his stimulating paper[Hob90]. Especially, it should be noticed that he points out importance of the role of coherence relations in metaphor understanding. Hobbs' formalization of metaphor understanding, in which metaphor understanding process is described as axioms related to some metaphorical expression and inference based on the axioms, does not precisely deal with the difference between the

In this paper, in view of this point, we present a new theory which can account for understanding metaphorical expressions in a discourse. It is a typical case of the metaphor understanding in a discourse that the information which the hearer obtains from a metaphorical expression depends on the context in which the expression is uttered. For example, let us consider the following sentence:

With a new strategy, I attacked his position. (1)

This sentence can convey some different kind of information according to contexts in which it is embedded:

Case-1. A description about a war: The head of a winning state utters the sentence about the victory, referring to the head of the defeated state.

Case-2. A description about an argument: Someone utters the sentence about the argument in which he/she refuted his/her opponent.

A:"I argued with Mr.B about his theory last night." C:"How did it come out ?" (2)

Case-3. A description about a game: Someone utters the sentence about the result of the game (chess, go, etc.), referring to his opponent.

These connotations of the sentence (1), however, seem to share some common structure, rather than to be quite different from each other.

Generally speaking, the information in an expression may vary with hearer's circumstance, or *resource situations*, such as contexts, background knowledge, and so on. The context sensitivity like this can be explained by drawing a distinction between the interpretation of an utterance and the information in the utterance, which is one of the ideas in situation semantics[Bar89]. Accordingly we adopt the formalization based on situation semantics, in which by dealing with the interpretation of a metaphorical expression and the information in the expression separately the metaphor understanding is done according to the hearer's resource. That is, an actual meaning connoted by either a word or a phrase in a metaphorical expression may vary from its ordinary meaning. As a result, metaphorical expressions can convey some other new information. We would like to formulate this variation as a process of introducing new resources, namely, establishing correspondences between a source domain and a target domain, with which a hearer can project information about a source domain onto a target domain.

2 The Outline of Metaphor Understanding in Situation Semantics

In our theory, metaphor understanding consists of the following steps and semantic representations with which this metaphor understanding process are represented in terms of situation semantics[2]:

Step-1 Hear a fragment uf^i of an utterance and construct states of affairs corresponding to it by the convention of the language use. Then a described situation s^i_{uf} corresponding to uf^i is obtained.

Step-2 By combining s^i_{uf} with the current context s^{i-1}_d, obtain a new situation s^{new}_d. In general, s^i_{uf} is included by s^{i-1}_d, that is, $s^{new}_d = s^{i-1}_d \sqcup s^i_{uf}$. Then examine whether the situation s^{new}_d satisfies the following requirements:

1. Every relation in the situation s^{new}_d satisfies the appropriateness of its argument assignment.
2. The situation s^{new}_d is consistent.
3. All of constraints holding in the situation s^{new}_d are satisfied.

Step-3 If the requirements in Step-2 are satisfied, let the new context s^i_d be s^{new}_d and increase i by 1, then go to Step-1.

---

interpretation of a sentence including some metaphorical expression and the information conveyed by the sentence. It is important to clarify the difference for semantics of metaphorical expressions such as our example, which convey various different information according to contexts.

[2] *Situations* are parts of the world to which a *scheme of individuation* applies. The collection of all situations is partially ordered by the part-of relation $\trianglelefteq$. A situation s_1 is a part of a situation s_2 (that is, $s_1 \trianglelefteq s_2$) just in case every basic state of affairs that is a fact of s_1 is also a fact of s_2.

Step-4 Since the requirements in Step-2 are not satisfied at this point, introduce a new resource to cancel the inappropriateness and to obtain a meaningful interpretation. In metaphor understanding, as described later, the new resource is a set of constraints which represents the correspondence between the source domain and the target domain. Thus, the new situation s^i_r in which these constraints hold is regarded as a new resource situation. Now pay attention to states of affairs carried relative to the resource situation s^i_r by the described situation s^i_{uf}, which is a situation related to the source domain in this case. Since the situation $s^{i,t}_{uf}$ supporting these states of affairs corresponds to a situation related to the target domain, return to Step-2 after replacing the described situation s^i_{uf} with $s^{i,t}_{uf}$.

Not only in ordinary sentence understanding, but also in metaphor understanding, it does not seem to be plausible that the hearer begins to interpret a sentence just after he/she finished hearing all of the sentence. As described in Step-1, whenever the hearer hears/reads a fragment of an utterance such as a noun phrase, a verb (phrase), and so on, he/she must start to interpret the fragment under the current context made by the discourse.

It is Step-2 that is the process which integrates fragments of interpretation newly obtained with the current context. The way of the integration depends on the convention of the language being used. For example, the grammer of the language is one of the constraints which are effective within one sentence.

In Step-4, the described situation s^i_{uf} constitutes a part of the context about the source domain, which is going on in parallel with the main context. We call this context about the source domain the *parallel context* s^i_p. If there already exists a parallel context s^{i-1}_p, Step-2 is also applied to the parallel context and s^i_{uf}. When the requirements in Step-2 are satisfied, a new parallel context s^i_p, which consists of both s^{i-1}_p and s^i_{uf}, is made. In this case, the situation which includes both s^i_r and the resource situation which has already introduced for the parallel context s^{i-1}_p serves as a new resource situation.

3 The Details of Each Step

This section will give details of these steps roughly described in the previous section through the process of understanding the example sentence (1)

"With a new strategy, I attacked his position."

in the Case-2, in which a topic about an argument is described.

3.1 Construction of a Fragment of a Described Situation

As described in the previous section, we assume that whenever the hearer hears/reads a fragment of an utterance, he/she interprets the fragment to some extent. While how big the unit of fragments to be interpreted is depends on the language being used and his/her ability in the language use, it seems that the most primitive unit is the phrase, which is determined by the grammar of the language, such as a noun phrase and a verb phrase. Assuming that the phrase is the unit of an utterance fragment, the example sentence is regarded as the following sequence of four fragments:

[With a new strategy]$_{PP}$, [I]$_{NP}$ [attacked]$_{VT}$ [his position]$_{NP}$.

As the result of the constraint satisfaction such as the grammer of the language, the described situation s^i_{uf} corresponding to each fragment described above is obtained as follows[3]:

$$s^1_{uf} \models \langle\langle with, object{:}ST, event{:}e_1\rangle\rangle \wedge \langle\langle new, object{:}ST\rangle\rangle \wedge \langle\langle strategy, object{:}ST\rangle\rangle \quad (3)$$

$$s^2_{uf} \models \langle\langle speaker, agent{:}A\rangle\rangle \wedge \langle\langle agent, agent{:}A, event{:}e_2\rangle\rangle \quad (4)$$

$$s^3_{uf} \models \langle\langle attack, object{:}P\rangle\rangle \quad (5)$$

$$s^4_{uf} \models \langle\langle position, object{:}P\rangle\rangle \wedge \langle\langle own, agent{:}B, object{:}P\rangle\rangle \quad (6)$$

[3] A relation R, and an appropriate assignment $\mathbf{a}$ of objects, determine two *basic states of affairs*, at most one of which is factual. We denote these states of affairs by $\langle\langle R, \mathbf{a}\rangle\rangle$ and $\langle\langle R, \mathbf{a}; -\rangle\rangle$. If appropriate objects $\mathbf{a}$ stand in the relation R, then the state of affair $\langle\langle R, \mathbf{a}\rangle\rangle$ is a fact. If, on the contrary, appropriate objects $\mathbf{a}$ do not stand in the relation R, then the state of affair $\langle\langle R, \mathbf{a}; -\rangle\rangle$ is a fact. We also denote a proposition that a state of affairs σ holds in a situation s by $s \models \sigma$.

where parameters e_1 and e_2 are events, that is, some situations. They are fixed up with some anchoring by linguistic convention, such as the English grammar. In this example, we assume the following relation:

$$s^1_{uf} \trianglerighteq e_1 = s^2_{uf} \trianglerighteq e_2 = s^3_{uf} \trianglerighteq s^4_{uf} \tag{7}$$

We also assume the following situation s^0_d as the initial context corresponding to the context (2):

$$s^0_d \models \langle\!\langle argue, agent{:}A, participant{:}B, object{:}TH \rangle\!\rangle \wedge \langle\!\langle theory, object{:}TH \rangle\!\rangle \wedge \langle\!\langle own, agent{:}B, object{:}TH \rangle\!\rangle \tag{8}$$

In the following subsections, we will show how the described situation s^1_{uf} is interpreted by Step-2, Step-3 and Step-4 in the process described above. For the limitation of space, we omit the explanation of the interpretation of s^2_{uf}, s^3_{uf} and s^4_{uf}.

3.2 Examining the Appropriateness of a Described Situation

According to Step-2, let us consider the situation s^{new}_d, that is, s^0_d under the condition:

$$s^1_{uf} \trianglelefteq s^0_d (= s^{new}_d) \tag{9}$$

This situation is a candidate for the context of s^2_{uf}. If only the condition (9) is imposed on s^{new}_d, the situation satisfies the requirements in Step-2. The hearer, however, usually uses some background knowledge to fill the lack of information, whenever he/she interprets an utterance fragment. For example, the typical situation supporting a state of affairs *argue*, and the knowledge that what state of affair is involved by the state of affair *strategy*, are derived from these background knowledge. In the following part of this subsection, first we show how to represent background knowledge, then we consider how s^{new}_d is interpreted with the background knowledge which we usually have in our mind.

3.2.1 Representing Background Knowledge

In this paper, by the term *resources* we mean some information which does not appear in utterances and can be used to interpret utterances by a hearer. We will pay attention particularly to the following resource, from which the hearer will be able to obtain useful information for metaphor understanding:

> The domain knowledge of a concept such as the typical context in which the concept is used

Since the speaker assumes that the hearer has sufficient knowledge about a source domain to retrieve some information intended by the speaker from an expression related to the source domain, the speaker may inform the hearer about something relevant to the current context, that is, the target domain intentionally by comparing the two domains. Therefore, in order to formulate metaphor understanding, it should be assumed that a hearer has some knowledge both of source and target domains to a certain extent. We treat this background knowledge as a situation type, that is, $[s \mid s \models \sigma]$, which somehow the hearer has[4]. Then an instance of a typical scenes actually used as a resource becomes the situation which is of this situation type, where parameters are anchored to fit individual situations.

We also define the type hierarchy in terms of the relation *subtype* in order to express the conceptual hierarchy as follows: R_1 *subtype* R_2 is true, if all assignments of the type R_1 are of the type R_2.

3.2.2 Interpretation with Background Knowledge

We suppose that the noun "strategy" recalls the situation type T_{war} to the hearer's mind as the background knowledge of the "war"[5]. Suppose also that T_{war} is the type of the situation which supports the following constraints about a strategy[6]:

$$\langle\!\langle \Rightarrow, \langle\!\langle strategy, object{:}x \rangle\!\rangle, \langle\!\langle method, object{:}x, event\text{-}type{:}ET_{use\ m.p.} \rangle\!\rangle \rangle\!\rangle \tag{10}$$

$$ET_{use\ m.p.} = [s \mid s \models \langle\!\langle use, agent{:}a_2, object{:}mp \rangle\!\rangle \wedge \langle\!\langle mil\text{-}power, object{:}mp \rangle\!\rangle] \tag{11}$$

[4] The *situation type* $[s \mid s \models \sigma]$ is the type of situation in which the state of affair σ holds.

[5] To do this, such a mechanism as the association is necessary in implementation.

[6] A positive constraint $\Rightarrow$ is defined as follows. If $s_0 \models \langle\!\langle \Rightarrow, \sigma(\vec{x}), \tau(\vec{x}) \rangle\!\rangle$ then for every situation $s \trianglelefteq s_0$ and every anchor $f : \vec{x} \rightarrow Obj(s)$ such that $s \models \sigma(f)$, there is a situation s' such that $s' \models \tau(f)$. A negative constraint $\perp$ is also defined as follows. If $s_0 \models \langle\!\langle \perp, \sigma(\vec{x}), \tau(\vec{x}) \rangle\!\rangle$ and there is a situation $s \trianglelefteq s_0$ and an anchor f such that $s \models \sigma(f)$, then there is no situation s' such that $s' \models \tau(f)$.

For the initial context s_d^0, we also suppose that the verb "argue" recalls the type T_{argue} of the situation, in which the following states of affairs hold, as the background knowledge of the "argue":

$$\langle\langle argue, agent{:}a, participant{:}p, object{:}x\rangle\rangle \wedge \langle\langle use, agent{:}a, object{:}y\rangle\rangle \wedge \langle\langle reasoning, object{:}y\rangle\rangle \tag{12}$$

Moreover, we suppose that T_{misc} is the situation type which represents the miscellaneous background knowledge and a situation of T_{misc} supports the following constraints:

$$\langle\langle \Rightarrow, \langle\langle with, object{:}x, event{:}e\rangle\rangle \wedge \langle\langle method, object{:}x, event\text{-}type{:}et\rangle\rangle, \langle\langle et, e\rangle\rangle\rangle \tag{13}$$

$$\langle\langle \perp, \langle\langle mil\text{-}power, object{:}y\rangle\rangle, \langle\langle reasoning, object{:}y\rangle\rangle\rangle \tag{14}$$

The state of affairs (13) represents the constraint about the case that the preposition "with" occurs with a "method". The state of affairs (14) represents the constraint about the conceptual structure.

Now, let us apply the background knowledge described above to s_d^{new} such that (9), that is, assume the following conditions:

$$s_d^{new}{:}T_{argue} \tag{15}$$

$$s_d^{new}{:}T_{war} \tag{16}$$

$$s_d^{new}{:}T_{misc} \tag{17}$$

The proposition (15) leads to:

$$s_d^{new} \models \langle\langle argue, agent{:}A, participant{:}B, object{:}TH\rangle\rangle \wedge \langle\langle use, agent{:}A, object{:}y\rangle\rangle \wedge \langle\langle reasoning, object{:}y\rangle\rangle \tag{18}$$

where we assume that the state of affair *argue* supported by s_d^{new} was merged with one in T_{argue}.

On the other hand, for (16) and (17) imply that the constraints (10) and (13) hold in s_d^{new}, there exists the situation s_d' and the situation e_1 is obtained as follows:

$$s_d' \models \langle\langle method, object{:}ST, event\text{-}type{:}ET_{use\ m.p.}\rangle\rangle \tag{19}$$

$$e_1(= s_{uf}^2 \trianglelefteq s_{uf}^1) \models \langle\langle use, agent{:}a_2, object{:}mp\rangle\rangle \wedge \langle\langle mil\text{-}power, object{:}mp\rangle\rangle \tag{20}$$

In (20), it seems to be natural that we presume the state of affair *use* in (18) is merged with one in (20), because $agent{:}a_2$ should be anchored to $agent{:}A$ in (18) owing to the coherence of discourse. For this presumption leads to:

$$s_d^{new} \models \langle\langle reasoning, y\rangle\rangle \wedge \langle\langle mil\text{-}power, y\rangle\rangle \tag{21}$$

it is clear that the situation s_d^{new} violates the constraint (14). Since the requirement 3 of Step-2 is not satisfied, to assume the condition (9), that is, $s_{uf}^1 \trianglelefteq s_d^0$ has turned out to be false. Therefore s_{uf}^1 constitutes the parallel context s_p^1, that is, $s_p^1 = s_{uf}^1$.

Generaly speaking, there exists some inappropriateness about the content in metaphorical expressions. What a pair of domains is compared with in a metaphor understanding is shown by the background knowledge recalled using associations. In the example, the initial context s_d^0 and the new described situation s_{uf}^1 recall the background knowledge T_{argue} of an argument and the background knowledge T_{war} of a war respectively. This means that the target domain is the "argument", and the source domain is the "war". Finally we obtain the following conditions as the result of Step-2:

$$s_d^0 \neq s_{uf}^1 \tag{22}$$

$$s_d^0{:}T_{argue} \wedge s_d^0{:}T_{misc} \tag{23}$$

$$s_{uf}^1{:}T_{war} \wedge s_{uf}^1{:}T_{misc} \tag{24}$$

That is, the new described situation s_{uf}^1 constitutes a parallel context which is not equal to the main context s_d^0. The main context s_d^0 and the described situation s_{uf}^1 are the situations related to an argument and a war, respectively.

3.3 Introducing New Resources to Get Meaningful Information

In the previous subsection, we examined the inappropriateness which might occur when a described situation is combined with the hearer's circumstances. In metaphor understanding, however, the clue

is what kinds of information is carried by an utterance and is combined with the context to make a new context. But, as described in the previous subsection, the hearer's background knowledge of both a war and an argument is insufficient for information which contributes to the main context to be able to obtained from s^1_{uf}. Therefore, it can be presumed that the hearer should introduce some new resource to get meaningful information.

Including the interaction theory mentioned in the section 1, a number of studies suggest that the correspondence is important. As the consequence we come up with having the following supposition:

> First, the hearer introduces some correspondence between a source domain and a target domain as a new resource. Then, according to this resource, he/she converts information, that is, he/she converts some states of affairs in the source domain into the corresponding states of affairs in the target domain[7].

3.3.1 A Description of Correspondences between a Target Domain and a Source Domain

In this section, we will show what correspondences we should deal with and how to express them. Since a source domain and a target domain are generally distinct from each other, in a strict sense we should treat correspondences between the schemes of individuation for each situation in order to express correspondences between the situations. However, since it is important for metaphor understanding what kinds of information about the target domain are obtained from descriptions about the source domain, the following constraints are sufficient to express correspondences between the source domain and the target domain:

$$\langle\!\langle \Rightarrow, \langle\!\langle r_{source}, \vec{x}; p_s \rangle\!\rangle, \langle\!\langle r_{target}, \vec{y}; p_t \rangle\!\rangle \rangle\!\rangle \tag{25}$$

where s_{target} and s_{source} are the situations related to the target domain and the source domain respectively, and $r_{target} \in Rel_{s_{target}}$, $r_{source} \in Rel_{s_{source}}$.

3.3.2 How to Obtain Correspondences

How can we obtain the constraints which represent the correspondence between the source domain and the target domain ? General metaphors[LJ80] are an important concept relevant to this question. Lakoff points out that a metaphorical expression in a certain text is an instance of a general metaphor. For instance, the example sentence can be regarded as an instance of the general metaphor "ARGUMENT IS WAR". At first sight, it seems to be sufficient for metaphor understanding to prepare all of correspondences obtained from all of general metaphors. However, for metaphorical expression can be and moreover might be extended flexibly by the speaker according to the context, it seems to be difficult that we prepare universal set of correspondences. If so, how can we get correspondences, which should be set up according to the context ? In the next section, we show that some heuristics can give us some useful correspondences.

3.3.3 Heuristics for Analogical Mapping in Metaphor Understanding

Basically, there may be many possible correspondences between two domains. However, it seems that we use some heuristics to establish correspondences which can be used to retrieve some useful information from metaphorical utterances during our interpretation process. Let us enumerate some heuristics useful for metaphor understanding[8]:

H1 Coherence of information
There are no incoherent situations when constraints, representing correspondences, are applied.

H2 Similarity between corresponding states of affairs
This heuristic is based upon our assumption that it is difficult even also for human in metaphor understanding to find correspondences between two domains containing no pairs of states of affairs closely related to each other. On this assumption, we expect that some corresponding pairs of states of affairs may be found easily.

H2.1 Make correspondences between the same states of affairs in two domains as far as possible.

[7] From now on, let the terms "source domain" and "target domain" mean also situations related to a source domain and a target domain, respectively.

[8] Some measure, that is, some evaluating functions for application of heuristics may be required for the efficient search of a plausible correspondence.

H2.1.1 Especially, when a state of affairs $\sigma = \langle\langle r, \vec{x}; p\rangle\rangle$ holding in the source domain does not appear in the target domain, check whether the state of affairs can be used in the target domain, that is, checking whether the assignments to the argument roles of the state of affairs are appropriate in the target domain. If the condition is satisfied, the following constraint becomes one of the assumption: $\langle\langle\Rightarrow, \sigma, \sigma\rangle\rangle$.

H2.2 Make correspondences between *similar* states of affairs in two domains as far as possible, when the heuristic H2.1 is not applicable. Similarity among relations is defined in terms of the relation *subtype*, by which the type hierarchy is defined.

H2.2.1 If in the target domain the assignment is not appropriate for the state of affairs σ obtained by applying the heuristic H2.1.1, generalize the relation r by obtaining a super type r_{sup} such that $\sigma' = \langle\langle r_{sup}, \vec{x}; p\rangle\rangle \& r$ *subtype* r_{sup} from the type hierarchy and check whether the assignment is appropriate for this super type. If the condition is satisfied, the following constraint becomes one of the assumption: $\langle\langle\Rightarrow, \langle\langle r, \vec{x}; p\rangle\rangle, \langle\langle r_{sup}, \vec{x}; p\rangle\rangle\rangle\rangle$. This generalization, however, gives us a weaker assumption.

H2.2.2 If the heuristic H2.2.1 is applicable, examine subtypes r_{sub} of the relation r_{sup} obtained by applying H2.2.1 such that $\sigma'' = \langle\langle r_{sub}, \vec{x}; p\rangle\rangle \& r_{sub}$ *subtype* r_{sup}. If the assignment is appropriate for one of the subtypes, r_{sub}, the following constraint becomes one of the assumption: $\langle\langle\Rightarrow, \langle\langle r, \vec{x}; p\rangle\rangle, \langle\langle r_{sub}, \vec{x}; p\rangle\rangle\rangle\rangle$. This assumption is stronger than the result of H2.2.1.

H3 Coherence of objects in two domains

Each object in one domain corresponds to an object in another domain according to correspondences between relations. There must be no incoherence in these correspondences between objects.

H4 Isomorphism of constraints

Make correspondences which preserve the isomorphism of constraints in two domains. For example, suppose the following causal relations hold in the target domain and a source domain respectively.

$$\langle\langle\Rightarrow, A_{target}, B_{target}\rangle\rangle, \ \langle\langle\Rightarrow, A_{source}, B_{source}\rangle\rangle$$

It seems to be natural to assume that the consequence B_{target} corresponds with the consequence B_{source}, if it is known that the premise A_{target} corresponds to A_{source}, and vice versa.

H4.1 Strengthening the definition of constraints

Since the definition of the positive constraint, which is described in the section 3.2.2, is weaker than the implication of the predicate logic, we cannot combine several constraints into a new constraint. Therefore, the heuristic H4, which depends on the isomorphism of constraints, not always work well. In order to cope with this, we introduce the heuristic which strengthens the definition of the positive constraint only while the heuristic H4 is applied to make some correspondences. The strengthening is achieved by the following two restrictions.

- The situation s' which supports the state of affairs τ, which is carried relative to a constraint, should be a part of the situation s_0, which supports the constraint. That is, add the condition $s' \trianglelefteq s_0$ to the definition of the positive constraint.
- This definition should be bidirectional. Replace "if" in the definition of the positive constraint with "if-and-only-if".

These may correspond to the plausible restriction on the domain of consideration, with which we usually do not have to examine unrelated matters.

These restrictions contribute to the derivation of some relations as follows, which can be used in H4, rather than, to finding quite new constraints[9].

Transitivity	$s \models \langle\langle\Rightarrow, \sigma, \sigma'\rangle\rangle \wedge \langle\langle\Rightarrow, \sigma', \sigma''\rangle\rangle \rightsquigarrow s \models \langle\langle\Rightarrow, \sigma, \sigma''\rangle\rangle$
Monotonicity	$s \models \langle\langle\Rightarrow, \sigma, \sigma'\rangle\rangle \rightsquigarrow s \models \langle\langle\Rightarrow, \sigma \wedge \sigma'', \sigma'\rangle\rangle$
Conjoining of consequence	$s \models \langle\langle\Rightarrow, \sigma, \sigma'\rangle\rangle \wedge \langle\langle\Rightarrow, \sigma, \sigma''\rangle\rangle \rightsquigarrow s \models \langle\langle\Rightarrow, \sigma, \sigma' \wedge \sigma''\rangle\rangle$
Disjoining of consequence	$s \models \langle\langle\Rightarrow, \sigma, \sigma' \wedge \sigma''\rangle\rangle \rightsquigarrow s \models \langle\langle\Rightarrow, \sigma, \sigma'\rangle\rangle \wedge \langle\langle\Rightarrow, \sigma, \sigma''\rangle\rangle$
Weakening	$s \models \langle\langle\Rightarrow, \sigma, \sigma'\rangle\rangle \rightsquigarrow s \models \langle\langle\Rightarrow, \sigma \wedge \sigma'', \sigma' \wedge \sigma''\rangle\rangle$

where $\rightsquigarrow$ means that assuming the restriction described above, the right hand side is derived from the left hand side logically.

[9]The size of search spaces for finding quite new constraints based on these restrictions is often very large.

3.3.4 Computing Correspondences

In our approach, computing correspondences is equivalent to applying the heuristics described in the last section. Therefore the computational process is expressed as the order of the application of each heuristic. From the viewpoint of computing correspondences, the heuristics are classified into two groups, one group which generate hypotheses of correspondence (H2,H4) and another group which impose restrictions on generating hypotheses (H1,H3). While to clarify the process of the application of the heuristics we should study it furthermore, empirically, in many cases, the correspondences can be computed by applying the heuristics repeatedly in the following order:

1. H2.1 (H2.1.1)
2. H3 and H1
3. H4 or H2.2
4. H3 and H1

Note that from the viewpoint of the computational complexity, the generator heuristics tend to be applied in order of computational simplicity.

Gentner[GFS87, Bur86] proposed the *systematicity condition* as a general condition on the mapping process in analogy. This condition is essentially that "relations are more likely to be imported into the target domain if they belong to a system of coherent, mutually constraining relationships the others of which are mapped". This is the standpoint in which the heuristic H4 is regarded important. While H4 deals with higher order relations, that is, constraints like Gentner's method, our method differs from hers in the point that the heuristic also uses isomorphism of constraints to generate new correspondences. Indurkhya[Ind87] proposed a mapping called *T-MAP* and a model of metaphors and analogies based on it. T-MAP is a general mapping schema according to the coherence of a target domain and newly mapped relations. Since this method does not use the information about a relation like the heuristic H2 to get a clue of correspondences, there exists the problem of the computational complexity. Burstein[Bur86] also refers to non-identical mappings, which Gentner's theory cannot deal with. While this non-identical mappings corresponds to the heuristic H2, in his method, H2.2.2 is restricted to generating only "brother" of the relation in the source domain.

3.3.5 An Example of Establishing Correspondences

Let us examine correspondences for the example sentence. As described in the section 2, the described situation s^1_{uf} forms a certain part of the parallel context s^1_p in parallel with the main context s_d. Therefore correspondences we should obtain are those that connect states of affairs which hold in s^1_{uf} to states of affairs which meet the requirements of Step-2 in the context s^0_d relative to the conditions (22), (23) and (24).

First let us spell out the states of affairs which hold in s^1_{uf} under the condition (24). The states of affairs which hold in s^1_{uf} are given in (3). By applying the constraints in T_{misc} to s^1_{uf}, it is derived that there exists a situation s'_d such that (20). Now, supposing that s'_d is also used as a part of the parallel context, the condition $s'_d \trianglelefteq s^1_{uf}$ should be satisfied. Under the condition, states of affairs holding in s^1_{uf} are as follows:

$$
\begin{aligned}
s^1_{uf} \models\ & \langle\langle with, object{:}ST, event{:}e_1 \rangle\rangle \wedge & (26)\\
& \langle\langle strategy, object{:}ST \rangle\rangle \wedge & (27)\\
& \langle\langle method, object{:}ST, event\text{-}type{:}ET_{use\ m.p.} \rangle\rangle & (28)\\
e_1(\trianglelefteq s^1_{uf}) \models\ & \langle\langle use, agent{:}A, object{:}mp \rangle\rangle \wedge & (29)\\
& \langle\langle mil\text{-}power, object{:}mp \rangle\rangle & (30)
\end{aligned}
$$

Let us construct the constraints which connect the states of affairs (26),...,(30) to the states of affairs which may hold in the target domain. For the state of affairs (29), the following constraint is obtained by the heuristic H2.1, which generates correspondences related to the states of affairs which can be used also in the target domain:

$$
C^1_{w \to a} = \{\langle\langle \Rightarrow, \langle\langle use, agent{:}A, object{:}mp \rangle\rangle, \langle\langle use, agent{:}A, object{:}y \rangle\rangle \rangle\rangle\} \quad (31)
$$

To the states of affairs (26),(27), (28) and (30), however, the heuristic H2.1 cannot be applied, since they cannot be used in the target domain[10]. For the state of affair (30), the correspondence is obtained by

[10]Because (30) violates the constraint (14), (26) and (28) lead to (30) with (13), and (27) leads to (28) with (10).

comparing s^1_{uf} with the main context s^0_d under the correspondences (39) and (31). Under the condition (23) the main context s^0_d is as follows:

$$s^0_d \models \langle\langle argue, agent{:}A, participant{:}B, object{:}TH\rangle\rangle \wedge \langle\langle theory, TH\rangle\rangle \wedge \langle\langle own, agent{:}B, object{:}TH\rangle\rangle \wedge \langle\langle use, agent{:}A, object{:}y\rangle\rangle \wedge \langle\langle reasoning, object{:}y\rangle\rangle \quad (32)$$

Since the parameter mp in the source domain is connected to the parameter y in the target domain by (31), by applying the heuristic H3, the following constraint is obtained:

$$C^2_{w\to a} = \{\langle\langle \Rightarrow, \langle\langle mil\text{-}power, object{:}mp\rangle\rangle, \langle\langle reasoning, object{:}y\rangle\rangle\rangle\rangle\} \quad (33)$$

Since the relation *method* itself can be used also in the target domain, for the state of affair (28) the correspondence constraint is obtained by translating the situation type, which is in the argument of the state of affair and is related to the source domain, into the situation type related to the target domain. Since the correspondence constraints (31) and (33) are applied to the any situation s, such that $s{:}ET_{use\ m.p.}$, related to the target domain, it is derived that there exists a situation s' such that $s'{:}ET_{use\ reas}$ for every situation s, where

$$ET_{use\ reas} = [s \mid s \models \langle\langle use, agent{:}y, object{:}reas\rangle\rangle \wedge \langle\langle reasoning, object{:}reas\rangle\rangle] \quad (34)$$

We suppose the following constraint, for s' can be regarded as a situation related to the target domain:

$$C^3_{w\to a} = \{\langle\langle \Rightarrow, \langle\langle method, object{:}ST, event\text{-}type{:}ET_{use\ m.p.}\rangle\rangle, \langle\langle method, object{:}ST, event\text{-}type{:}ET_{use\ reas}\rangle\rangle\rangle\rangle\} \quad (35)$$

Lastly we examine the case of the state of affairs (27). It seems to be natural that we assume the following constraint about the "logic" holds in a situation s which is of the situation type T_{argue} about the "argument".

$$\langle\langle \Rightarrow, \langle\langle logic, object{:}x\rangle\rangle, \langle\langle method, object{:}x, event\text{-}type{:}ET_{use\ reas}\rangle\rangle\rangle\rangle \quad (36)$$

Now, applying the heuristic H4 to the constraints (10) and (36), the following correspondence constraint are obtained:

$$C^4_{w\to a} = \{\langle\langle \Rightarrow, \langle\langle strategy, object{:}ST\rangle\rangle, \langle\langle logic, object{:}ST\rangle\rangle\rangle\rangle\} \quad (37)$$

We have obtained all constraints which connect each state of affairs holding in s^1_{uf} to a state of affairs related to the target domain "argument".

3.4 Metaphor Understanding with Resources Related to Correspondences

The resource situation s^1_r in Step-4 becomes the situation which supports the constraints $C^1_{w\to a}, \ldots, C^4_{w\to a}$ in the previous subsection. In order to apply these constraints to e_1, suppose the condition $s^1_r = s^1_{uf}$. Then it is derived that there exists the following situation e^t_1:

$$e^t_1 \models \langle\langle use, agent{:}A, object{:}y\rangle\rangle \wedge \langle\langle reasoning, object{:}y\rangle\rangle \quad (38)$$

Because e_1 corresponds to e^t_1, the following constraint is obtained:

$$C^5_{w\to a} = \{\langle\langle \Rightarrow, \langle\langle with, object{:}ST, event{:}e_1\rangle\rangle, \langle\langle with, object{:}ST, event{:}e^t_1\rangle\rangle\rangle, \} \quad (39)$$

This constraint is also included by s^1_r. Then it is derived that there exists the following situation $s^{1,t}_{uf}$:

$$s^{1,t}_{uf}(\trianglerighteq e^t_1) \models \langle\langle with, agent{:}A, object{:}ST\rangle\rangle \wedge \langle\langle logic, object{:}ST\rangle\rangle \wedge \langle\langle method, object{:}ST, event\text{-}type{:}ET_{use\ reas}\rangle\rangle$$

This situation corresponds to a described situation related to the target domain. Return to Step-2 after replacing the described situation s^1_{uf} with $s^{1,t}_{uf}$. Then all requirements in Step-2 are satisfied under the condition $s^{1,t}_{uf} \trianglelefteq s^0_d$. Therefore $s^{1,t}_{uf}$ forms part of the new main context s^1_d such that $s^1_d = s^0_d \trianglerighteq s^{1,t}_{uf}$. Generally speaking, the semantic representation of metaphor understanding can be formulated as the following constituents:

- A described situation s^i_{uf} related to the source domain, that is , a part of the parallel context

- An application of the background knowledge of the source domain $s^i_{uf}{:}T_{source}$
- A resource situation s^i_r which supports constraints representing the correspondence
- A described situation $s^{i,t}_{uf}$ related to the target domain, which satisfies the conditions: $s^i_r = s^i_{uf}$, $s^i_{uf} \Vdash_{s^i_r} \sigma$, $s^{i,t}_{uf} \models \sigma$.
- The condition, which represents the construction of a new context: $s^i_d = s^{i-1}_d \trianglerighteq s^{i,t}_{uf}$.
- An application of the background knowledge of the target domain $s^i_d{:}T_{target}$

4 Conclusion

In this paper, we give a formal description of metaphor understanding in terms of situation semantics. In this formulation, we propose the model in which the hearer introduces correspondences between a target domain and a source domain as a new resource to obtain useful information about the target domain in terms of source domain. We expect that this formal description of metaphor understanding may serve as a guide to natural language understanding systems which can understand discourses with metaphors.

From the viewpoint of a computational model, our model is not perfect. Since we have concentrated upon a formal description in this paper, we have not mentioned how to search a plausible correspondence. The concrete method of how to search a plausible correspondence will be developed in our future work. (As related works of plausible search in metaphor understanding, several methods have been proposed[TIT89, Ind87].)

References

[Bar89] Jon Barwise. *The Situation in Logic*, Vol. 17 of *CSLI Lecture Notes*. CSLI, 1989.

[Bur86] Mark H. Burstein. Concept formation by incremental analogical reasoning and debugging. In Ryszard S. Michalski, Jaime G. Carbonell, and Tom M. Mitchell, editors, *Machine Learning Volume II*, chapter 13, pp. 351–369. Morgan Kaufmann Publishers, 1986.

[GFS87] Dedre Gentner, Brian Falkenhainer, and Janice Skorstad. Metaphor: The good, the bad and the ugly. In *Theoretical Issues in Natural Language Processing 3*, pp. 176–180, January 1987.

[Hob90] Jerry R. Hobbs. *Literature and Cognition*, Vol. 21 of *CSLI Lecture Notes*, chapter 4. CSLI, 1990.

[Hol89] Keith J. Holyoak. Analogical mapping by constraint satisfaction. *Cognitive Science* **13**, pp. 295–355, 1989.

[Ind87] Bipin Indurkhya. Approximate semantic transference: A computational theory of metaphors and analogies. *Cognitive Science* **11**, pp. 445–480, 1987.

[LJ80] George Lakoff and Mark Johnson. *Metaphors We Live By*. The University of Chicago Press, Chicago and London, 1980.

[TIT89] Hozumi Tanaka, Makoto Iwayama, and Takenobu Tokunaga. A computational model of understanding metaphors. In *Proceedings of Nagoya International Symposium On Knowledge Information And Intelligent Communication*, pp. 17–26, November 1989.

Metaphor vs. Anomaly: Conceptual Constraints on Verbal Metaphoric Extension

Sylvia Weber Russell

Computer Science Department
University of New Hampshire
Kingsbury Hall
Durham, NH 03824

swr@cs.unh.edu

Abstract. Making the distinction between metaphoric and "anomalous" expressions is particularly subject to variation in judgment. In the context of a program which interprets simple isolated sentences that are potential instances of metaphorical usages of verbs, with a focus on cross-modal usages, I consider some possible coherence criteria which must be satisfied for an expression to be understood metaphorically. Besides noting the preservation of "domain consistency" through extension of the verb, I consider the more problematic question of verb-noun conceivability constraints which are satisfied in a metaphor. Determining an instance of verbal metaphor is not just a matter of relaxing constraints on nominals which apply to literal interpretations. One approach is to consider how relevant constraints might be represented as metaphorically extended along with the invariant part of the verb meaning. Referring to these considerations, I suggest the role and possible limits of some potential representation components in recognizing incoherent semantic combinations and rejecting metaphoric interpretations or adjusting them accordingly.

INTRODUCTION

Why is the linguistic war horse "The ship plowed the sea" considered a metaphoric expression, but the sentence, "<Any object> plowed the spoon" anomalous? Or why can one "plow through one's memories" but not "through a spoon", even though "spoon" as a physical object may appear to satisfy more closely the selectional restrictions for the object of "plow"? The interdisciplinary literature is full of "anomalous" examples for which interpretations can (and therefore should) be found, given a particular context and a little effort. It is sometimes said that context so governs the meaning of an utterance that there is little left for the utterance itself to contribute. The "anomaly" of a linguistic expression as the term is used here, then, is a relative concept, and degrees of acceptability of an utterance are infinite. However, our initial intuitive reactions do distinguish between easily interpretable expressions and those which produce interpretations (if any) which are strained by some criterion. The determination of semantic and syntactic parsing choices, including anaphoric references, correct sense determinations, and the detection of nonmetaphoric tropes, as well as the construction of correct event sequences may depend on judgments of dependencies between concepts of such an expression. While goal planning in the sense of Perrault and Allen (1983) and Schank and Abelson (1977) can help to

disambiguate potentially metaphoric expressions, the converse is also true--the probable interpretation of an expression in isolation may help to determine a plan or a goal (Russell, 1986). At a more theoretical level, interpretability judgments based on the satisfaction of specific constraints could account for at least one sense in which an expression, while not rejected as anomalous, might be considered a "bad" metaphor. With the increasing accumulation of computational as well as noncomputational theories of metaphor, it seems timely to review constraints relevant to a theory of anomaly vs. metaphor.

What are the relationships which components of an expression must have with each other, in order for the expression to be characterized as a comprehensible metaphor as opposed to either anomaly or "somewhat incomprehensible" metaphor? How are extensions of meaning constrained to produce comprehensible, nonliteral language? How does the inadequacy of a metaphor affect its interpretation? I would like to address these questions, as they apply to verb-based metaphor, in terms of the notion of "coherence," a concept which is interpreted variously by researchers of literal and nonliteral language, and which is perhaps most systematically described by Fass (1988). The aspect of coherence which is my focus includes in its domain some classes of concepts normally neglected in computational practice, namely concepts referring to nonphysical entities and sometimes referred to as "abstract," e.g., "beauty." By considering how concepts combine comprehensibly, we may be able to arrive at better criteria for recognizing and comprehending metaphor in a computational environment.

Treatments of coherence in terms of semantic properties and constraints entail revisiting the linguistic question of "selectional restrictions." The characterization of metaphor as violating certain selectional restrictions that would be satisfied by literal language is easily criticized (Ortony, 1980) as inadequate insofar as it does not differentiate between metaphoric and anomalous language. This paper addresses this differentiation by focusing on what is preserved rather than what is violated in comprehensible metaphoric extension. In other words, both literal and metaphoric language must satisfy certain coherence criteria, i.e., selectional restrictions, in order to be termed "meaningful" rather than "anomalous." This approach, then, is an attempt to account for metaphoric comprehensibility in terms of specific lexicon representations and their applicability to variously abstract concepts--insofar as this is computationally possible.

The discussion will be confined mainly to simple sentences in which the verb is potentially used metaphorically in the immediate context of literally used nouns representing concepts dependent on the verbal concept as part of its underlying case-like structure.[1] In particular, I consider cross-modal metaphor (metaphor which crosses into and/or out of general domains corresponding to various animate faculties) and point out some similarities and differences in coherence constraints with respect to physical-domain metaphor.[2]

---

[1]Semantic tension of course often spans sentence boundaries; our interest here is on potential semantic combinations of concepts, regardless of their actual lexical/syntactic arrangement. Similarly, consideration of comprehensible elliptical language is excluded here, though the failure to find a literal or metaphoric relation can point to its detection. See Fass (1988, 1989) for a treatment of the related phenomenon of metonymy.

[2]While in most of this discussion the assumption is that the literal domain of a verb is a physical domain, extensions of verbs from other than physical domains are also possible (Russell, 1976).

BACKGROUND

The question of metaphor vs. anomaly has been addressed in the various disciplines, though less so than that of metaphor vs. literal language. This research, however, has largely neglected any general semantic criteria governing the combination of *types* of concepts represented in a linguistic expression. That is, in typically focusing on the *mechanisms* of metaphor processing or recognition rather than on *representations,* metaphor researchers who have addressed verbal metaphor have generally used physical-domain examples without indicating how concepts and constraints in cross-modal metaphor might be represented. Brugman and Lakoff (1988), although (because?) they argue for a nonsymbolic approach in their research on metaphoric and nonmetaphoric polysemy, are an exception to the lack of concern for representation, though they do not frame their work in terms of constraints or go beyond topological descriptions of concepts.

Of those computation researchers who address verbal metaphor processing, Indurkhya (1986), Fass (1989) and Fass and Wilks (1983) include search of the source and target domains for a match (at some level) of nominals and of verb components. None, however, has devoted much consideration to cross-modal metaphor. Indurkhya's formal approach provides no rationale for his constraints. Fass uses hierarchies to establish that "The ship plowed the waves" satisfies a "tool plowing a medium". However, explicit categories such as "tool" and "medium" (if in fact they can be located and matched in the hierarchy) are not easily applicable to nonphysical domains. We might ask what it is about tools and media that *would* make them extensible. Carbonell (1982) and Carbonell and Minton (1983) present approaches for interpreting various kinds of metaphor, including some cross-modal metaphor. However, they do not attempt to judge metaphoric comprehensibility *per se*. Martin (1988) also ventures into a variety of types of metaphor, but, given his focus on a system which builds on and extends conventional metaphoric knowledge, does not concern himself with any specific representations governing semantic relations between concepts. MAP (Russell, 1976, 1985, 1986), which provides rough "literal" interpretations of isolated expressions containing a metaphorically used verb, is based on a model in which certain simple coherence criteria must be satisfied in order to interpret the expression as metaphoric (or literal) rather than as anomalous. Relevant parts of this program will next be outlined as a context for observations on coherence.

MAP

MAP as discussed here focuses on cross-modal verbal metaphor, though an early pilot program processed within-physical-domain metaphor, and further research has provided a limited treatment of nominal metaphor in context (Russell, 1986). The program is based on a semantic model in which the lexicon assigns to verbs and nouns a "conceptual domain" in which the verb or noun is considered to operate literally. Conceptual domains with subcategories as used in MAP (see Russell, 1986 for elaboration) are:

MENTAL: intellect, attitude, volition
SENSORY: sight, sound, other (unimplemented) senses

CONTROL: intrinsic (talent), extrinsic (rights/duties, possession)
PHYSICAL: animate, inanimate

SPATIAL (for nominals only, e.g., "space," "room")
TEMPORAL (for nominals only, e.g., "time," "year")

In addition, verbs are deterministically assigned "invariant" abstract descriptors in an extensible representation somewhat close in level to LNR representations (Norman & Rumelhart, 1975). Basic structures specify a (unary) attributive or (binary) relational STATE, any external agent that CAUSEs the state, and CHANGE of state (enter, leave, or pass-through the state); these provide the most general and minimal level of interpretation of an expression. Further descriptors relating to these structures specify whether the state is symmetric, attempted (vs. accomplished) and potential (vs. real). Other abstract descriptors characterize the action (if any) represented by the verb (speed, repetition, continuity), the actor (whether he/she intends the action) and any subjective response to the verbal concept (evaluation, intensity, emotional values). While the basic state structure preserves the existential content of the source and target meanings of the verb, a potential *reason* for the metaphor is represented in the verb definition by connotations expressed as a simple effect, e.g., "effort" for the verb construct "plow through,"[3] or as a secondary structure. An optional "purpose" structure is allowed for but not considered salient or extended unless there is a match with this purpose in the discursive context or a failure of constraints on concepts in the basic structure.

In defining verbs in all conceptual domains in terms of similar structures, the assumption is that conceptual categories get shifted around in cross-modal metaphor extension in a way which is interdependent with syntactic form. In other words, conceptual attributes, actions and relationships are reified and treated syntactically as objects which themselves have attributes or enter into relationships. This semantic model therefore does not represent real-world actions literally in the sense attempted through systems such as Schank's (1975) Conceptual Dependency representations, but rather assigns extensible structures to concepts according to an *ABSTRACT-CONCEPT-AS-OBJECT* metaphor. In the interpretation process of a sentence satisfying coherence, the abstract structure of the verb, together with the invariant descriptors, including subjective effects, are retained, with optional translation of the verb to a verb in the target domain, i.e., the domain of the conceptual (affected) object of the expression. Any physical instruments of the process are not carried over into the target domain. While the paraphrases produced often seem too general, they appear to be close in content to the types of responses Gentner and France (1988) observed empirically, which they classify as "minimal subtraction" (of meaning from the verbal concept in its original sense).

---

[3]In MAP, verbs are considered together with any subsequent preposition because this enables the model to recognize "X plow through Y" abstractly as a relation which can metaphorically characterize the possession of abstract concept Y by animate concept X, bringing with it connotations which might not be salient to either the verb or preposition taken separately. See Brugman and Lakoff (1988) for an alternative, systematic, purely topological approach to this type of expression.

METAPHORIC COHERENCE

Domain Consistency and Metaphoric Extension

A necessary though not sufficient principle governing recognition of cross-modal metaphor in MAP is that of "domain consistency". Since each domain is intended to represent a conceptually different modality, we can speak of "conceptual domain consistency". The assumption is that a verbal concept and its possible associated objects are semantically interdependent--they share each other's properties. A verb in one conceptual domain with an object in another is thus incoherent; physical concepts do not mix with mental ones. A verbal extension to a new domain is "conceptually coherent" if its constituents combine in a "conceivable" way, rather than necessarily a "usual" way, according to identifiable properties of the verbal concept and its dependent nominals. In linguistic terms, we could speak of domain consistency as a fundamental "selectional restriction" that is a minimum necessary for coherence. For a simple expression in which one or more nouns provide the immediate context for a potentially metaphoric verb, this principle can be a means of recognizing nonliteral language; if the verb in its literal sense and the conceptual object of the sentence are inconsistent in domain, as in "siphon off the idea" or "his pride broke", the expression is not a literal one.[4] If, further, the verb can be "comprehensibly" extended to the domain of the object, as determined by any selectional constraints, the expression is a metaphoric one in that domain; otherwise it is anomalous, i.e., the object cannot serve as the "real" conceptual object of the verb. The extended verb can be seen as analogous to the verb in its literal sense, the (abstract) structure being the same but the domain different.

Domain inconsistency as a possible indicator of metaphoric extension of a given verb as modeled by MAP invites the question of what happens to the meaning of verbs and nouns in metaphoric extension, and therefore to their combinatorial semantics. The empirical research conducted by Gentner and France (1988) addresses this question. They discuss subjects' interpretations of semantically strained sentences in isolation in terms of "verb mutability," i.e., the potential of the verb to change in meaning to adapt to the domain of the object. They note that mutability may account for polysemy (at least some of which, it would seem, is due to frozen metaphor). This appears reasonable. In terms of models such as MAP, however, as well as in terms of Gentner's and France's observations that abstract relations are extended to interpretations, whereas specific objects are not, there may be a co-existing reason for both verb mutability and polysemy. It could be that a verb can be understood in so many different contexts, though changed in meaning, precisely because a core component of its basic meaning is *im*mutable and is therefore recognized in whatever conceptual domain it is used. The fact that the verb submits to the noun in adapting its meaning thus in no way contradicts the centrality of the verb as the authors suppose; structurally, it is still the verb which determines the expression of the thought to be conveyed.

[4]Conditions of course also apply to the actor of the verb. Broadly, an actor in a MENTAL, SENSORY, CONTROL or PHYSICAL-animate domain must be animate for the expression to be literal. Otherwise we potentially have a personification, as in "The trees began to think about dropping their leaves" or as in *one* interpretation of "My car drinks gasoline," an example pursued at length in earlier work by Fass and Wilks (1983).

Rather than downplaying the verb, Gentner's and France's studies, as well as elements of MAP and other programs, appear simply to reflect both *dissimilarity* and *similarity* views of metaphor, namely that the domain changes but some part of the structure must remain the same. It is the "similarity" aspect which determines the extension of constraints as well as of structure.

Conceptual Constraints

Domain consistency alone does not guarantee a metaphorically coherent expression. For within-(physical-) domain verbal metaphor, descriptors of a dependent nominal can serve as constraints which determine a literal, metaphoric or anomalous reading of an expression, or select from various metaphoric interpretations.[5] These descriptors can take the form of features describing static characteristics of objects, a structure describing the use or function of the object, or a category or categories in which these features and functions are implicit. One application of constraints is illustrated by Fass (1989), who, as mentioned earlier, shows how "The ship plows (or 'plows through') the waves" can be interpreted metaphorically as satisfying a "tool plowing a medium." However, explicit categories on the level of "tool" and "medium" do not help in comprehending extensions to nonphysical domains, such as "The cashier plowed (or 'plowed through') her memories." The difficulty of application is at least partly due to the fact that such categories include domain-specific content, which is not extensible. Extensibility requires some degree of abstraction, suggesting a "factoring" of semantic components, most easily discussed in terms of features.[6] In terms of the previous section, some constraints, expressible as features, appear to be extended along with immutable components of the verbal concept, i.e., those which are perceived as similar in the source and target usages of the verb. The purpose of abstract features of this kind is to determine semantic relations which are *conceivable* rather than ordinary or expected. In the physical domain, descriptors at this level allow interpretation of examples such as "plowed through the trash" (even though "trash" would not be expected to be defined as a "medium"), because "trash," in being a collection of parts as opposed to a solid unit, is +COMPLEX and +FLUID.

For cross-modal metaphor, such features, since they apply to concepts which are "objects" only in a metaphoric sense, are themselves metaphoric. For example, "plowing" an abstract concept implies thinking of that concept as metaphorically +COMPLEX, i.e., not elementary or simple. As a constraint, this feature value would be satisfied by a noun representing a composite event, e.g., a "concert," but not a simple action, e.g., a "throw" or attribute, e.g., "truth". This means that in specifying abstract features which will "work" for cross-modal metaphor, knowledge base editors have to do some metaphoric thinking. As Tourangeau and Sternberg (1982) note, features cannot be identical for

[5]This approach does not imply that metaphor is to be thought of as "deviant" with respect to literal language, or sought only after literal interpretations fail, but rather as a different kind of predication with its own set of constraints (Russell, 1986).

[6]The term "features" is used loosely to mean some formalized property which may apply in combination with other properties. While some features in MAP simply have "+/-/variable" values, the CONTAIN property associates with a "+" value a further simple specification as to what *type* of concepts can conceivably be contained by the nominal. Others are or could be 3-state variables. Of greater importance is that the set of semantic descriptors be kept small and not open-ended.

concepts in different domains; at best they are analogous.

How do the members of a feature set extensible to abstract concepts compare with those of a physical-domain set from which it is drawn? It would be absurd to say that either set in MAP is "valid" or even well-developed. However, we can say about the relationship between the two types of sets that, just as physical aspects disappear from a concept extended to a nonphysical domain, physical features merge (are mutually redundant) in nonphysical domains, because of the irrelevance of distinct topological characteristics. The effective abstract feature set can therefore be expected to be smaller than a literal feature set. Relevant basic nominal features used in MAP so far are: SHAPE (vs. "amorphous"), 1-DIMENSIONAL (linear-like), 2-DIMENSIONAL (area-like), PART (of something), FIXED (attached to something), CONTAIN (surround), COMPLEX, FLUID, ANIMATE ("dynamic"). In addition, a FUNCTION of the nominal, if any, is expressed as a verbal concept in terms of abstract components as described above. In the extensible feature set, literal features such as PART and FIXED merge into a vague kind of "connectedness." Similarly, COMPLEX and FLUID are not easily distinguished. A 2-DIMENSIONAL, flat or area-like quality does not mean much in nonphysical domains and thus is irrelevant as a feature.

When used in processing potential PHYSICAL-domain metaphor, a basic set of this small size is liberal in allowing computational interpretation of novel metaphor prevalent in various kinds of discourse. A disadvantage of course is that such features may underconstrain the interpretation. This disadvantage is somewhat offset by the earlier observation that conceivability judgments carry only a part of the burden of interpretation. Applied to concepts in potential cross-modal metaphor, such a feature set is subject to the problem of human consensus on what a feature means.

Besides the problem of determining the relevance of extensible features with respect to literal features, there is a potential problem of using as a constraint for a PHYSICAL-domain usage a feature value which is redundant and therefore "harmless" for PHYSICAL domains, but which would impose a false constraint by the verb in an extension to a -PHYSICAL domain. For example, an early version of MAP specified +FIXED for the object of "plow." While a physical object of plowing will generally be +FIXED (due to size constraints alone, since free-floating, plowable areas are difficult to imagine), this constraint is not dependent on the nature of "plowing" and thus should not be specified as an extensible constraint. If it is, then even if +FIXED maps to a +CONNECTED feature in nonPHYSICAL domains, examples such as "plowing memories" or other conceptual objects will be wrongly hypothesized as anomalous. It seems that extensible representations and their use are particularly susceptible, not only to idiosyncratic understandings of word meanings, but also to undesired consequences resulting from inadequately considered representations.

Given the necessarily vague nature of judgments on the applicability of such features to a concept represented by a noun, are they of any value in distinguishing metaphor from anomaly? As an example, the verb construct "plow through" has a TRANS-STATE (pass-through a state) structure, which implies a beginning and an end and therefore calls for an object which is not thought of as a point, i.e., a +COMPLEX concept. A "plural" concept will satisfy this. This constraint would appear to account for the observation that

"She plowed through her memories" is less strained than "She plowed through the (musical) note." Similarly, even though "plowing through memories" extends a physical action to a mental domain, it is less strained and more conceivable in isolation from discursive context than "plowing through a (-COMPLEX) spoon," because the immutable part of the meaning of COMPLEX applies to "memories," but not to "spoon." While "plow" has a somewhat different interpretation than "plow through," constraints on the object are similar; therefore "plowed the waves" and "plowed her memories" (arranged them in some way) are both more comprehensible than "plowed the spoon."

While such explanations are of semantic interest, we might ask whether constraints on abstract nominal concepts associated with a verbal concept play a significant role in the AI goal of text comprehension. Cases in which humans could not resolve a metaphorically ambiguous sentence through extra-sentential context are probably infrequent. In practice, however, successful computational use of such context is not easy, and the above constraints do serve as an aid. As another example, for "The farmer plowed through the book," the features could be used to select the MENTAL aspect of "book" rather than the PHYSICAL aspect, despite the strong literal association of "plow" with "farmer.[7]" Disambiguation at this intra-sentential level is due in part to the fact that the MENTAL sense of "book" (which is made up of many MENTAL items) is defined with a +COMPLEX feature value, satisfying the conceptual selectional restrictions of "plow through", whereas the PHYSICAL sense of "book" is arguably -COMPLEX and certainly -FLUID, and is therefore relatively anomalous.

INTERPRETATIONS

Most of this discussion has been concerned with descriptors which focus on static/topological features of nominals. A brief consideration of descriptors which refer to an object's use,[8] as mentioned earlier, offers an opportunity to comment on the variety of ways an object nominal can enter into a metaphor interpretation, depending on what types of selectional constraints are satisfied, since potential interpretations are automatically selected by expressions determined not to be anomalous. One type of difference in interpretation is given by differences in salient components of the same verb. Examples for comparison are "His efforts sharpened his mind (or 'memory' in the sense of 'faculty')" vs. "His efforts sharpened his memories (in the sense of 'thoughts of the past')." While the verb "sharpen" (transitive sense) requires its object to have a point or an edge for a literal interpretation, no feature corresponding to this is extensible to -PHYSICAL domains. However, "sharpen" is defined with the PURPOSE, either that the object cuts or pierces better (if the object possesses some FUNCTION), or that one can distinguish the object better (with no FUNCTION constraint on the object). "Mind" and "memory" as mental faculties are mental FUNCTIONs and thus satisfy general requirements for an interpretation in terms of thinking or recalling better. "Memories" have no such function, so that

[7]This example could also be viewed as a "frozen metonymy" (see Fass, 1983, for alternative resolution).

[8]or what an object "affords" in the sense of Gibson (1977)

"He sharpened his memories" does not yield the interpretation that memories were functioning better, but rather satisfies the interpretation that he could distinguish his memories better, since for this interpretation there are no constraints.

Another example of interpretation differences is given by different structurings of the object nominal by different verbs, e.g., "He boarded up his mind" vs. "He sharpened his mind." The first requires the object to be able to CONTAIN, for the interpretation that he did not let any new concepts into his mind, and the second requires the object to have a FUNCTION as above. In the approach of Lakoff (1986) and Lakoff and Johnson (1980), these are equivalent to *MIND-AS-CONTAINER* and *MIND-AS-TOOL* metaphors respectively. However, here we are letting the verb connect selectional restrictions with interpretations in terms of metaphoric extension, rather than checking for pre-existing metaphorically related nominal pairings (which can only be reliably found if all possible metaphor themes have been thought of in advance).

What if constraints are not satisfied, but an interpretation is required due to other text processing considerations? One approach is to provide an interpretation in which only the most abstract, i.e. the existential parts of the verb structure, enter into the interpretation, under the assumption that this is simply an unrecognized idiom or a "bad" metaphor due to the lack of interconnections which would have been justified by satisfaction of all constraints. If the only available parse is one which represents "plowing through a spoon" (corresponding to the situation in which humans would find an interpretation if forced), a possible, minimal interpretation is that the "plower" was in motion, made contact with the spoon and ended up somewhere else. "Plowed the spoon," also initially judged as anomalous, would receive a default interpretation saying that something was done to the spoon to change its character in some way.

CONCLUSION

A method of determining metaphoric coherence in computational approaches requires a consideration of how we as language users might extend properties comprehensibly across conceptually different domains. A principled analysis of extension in turn requires attention to semantic components. The above observations on constraint criteria for determining cross-modal and other metaphoric relations may give some indication of what we can expect and not expect representations of such components to do. While we can disclaim the need for robust judgments of coherence of expressions in the light of contextual contribution, a certain limit is imposed by the relative uncertainty in defining extensible properties. This is perhaps a reflection of our inability to completely explain our perception of similarity (Russell, 1988). A further observation is that specifications of selectional constraints are circular; only "plowable" things can be "plowed." This, however, is a natural consequence of the way our language splits an event into verb and object in the first place; in a sense coherence means the recognition that a verbal concept and its dependent nominals function as a unit.

Computational experience with a more developed system of abstract descriptors, using examples in context, as well as extensions of experiments of the type run by Gentner and

France, may show whether models can (or have to) rely on such descriptors and what form they should take. It could be, for example, that reifications of attributes and of other concepts which are not literally "things" would be more effectively described, and their use constrained, in ways other than by features borrowed from literal language about physical concepts. For example, rather than representing the noun "hope" as a -COMPLEX (etc.) nominal, it might be more practical to simply allow its general categorization as a conceptual attribute in the MENTAL-attitude domain to itself serve as a constraint. Here there is a need to further explore and specify how extensions to different domains are *not* strictly analogous. Regardless of such choices, an exploration of how coherent metaphoric extension works in terms of constraints requires further attention to lexicon representations which take into account the ways in which language allows coherent metaphoric predications about concepts other than ships and seas.

REFERENCES

Brugman, C. and Lakoff, G. (1988).
Cognitive topology and lexical networks. In S. Small, G. Cottrell & M. Tanenhaus (Eds.), *Lexical Ambiguity Resolution.* San Mateo, CA: Morgan Kaufmann Publishers, Inc.

Carbonell, J. (1982).
Metaphor: An inescapable phenomenon in natural-language comprehension. In W. Lehnert & M. Ringle (Eds.) *Strategies for Natural Language Processing.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Carbonell, J. and Minton, S. (1983).
Metaphor and common-sense reasoning (Rep. No. CMU-CS-83-110). Pittsburgh: Carnegie-Mellon University.

Fass, D. (1988).
An account of coherence, semantic relations, metonymy, and lexical ambiguity resolution. In S. Small, G. Cottrell & M.Tanenhaus (Eds.), *Lexical Ambiguity Resolution.* San Mateo, CA: Morgan Kaufmann Publishers, Inc.

Fass, D. (1989).
Met: A method for discriminating metonymy and metaphor by computer* (Report CSS/LCCR TR 89-15). Burnaby, BC: Simon Fraser University.

Fass, D. and Wilks, Y. (1983).
Preference semantics, ill-formedness, and metaphor. *American Journal of Computational Linguistics, 9,* 178-187.

Gentner, D. and France, I. (1988).
The verb mutability effect: Studies of the combinatorial semantics of nouns and verbs. In S. Small, G. Cottrell & M. Tanenhaus (Eds.), *Lexical Ambiguity Resolution.* San Mateo, CA: Morgan Kaufmann Publishers, Inc.

Gibson, J. (1977).
The theory of affordances. In R. Shaw & J. Bransford (Eds.), *Perceiving, Acting and Knowing.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Indurkhya, B. (1986).
Constrained semantic transference: A formal theory of metaphors. *Synthese, 68,* 515-551.

Lakoff, G. (1986).
A figure of thought. *Metaphor and Symbolic Activity, 1,* 215-225.

Lakoff, G. and Johnson, M. (1980).
Metaphors We Live By. Chicago: Chicago University Press.

Norman, D. and Rumelhart, D. (Eds.) (1975).
Explorations in Cognition. San Francisco, CA: Freeman.

Ortony, A. (1980).
Some psycholinguistic aspects of metaphor. In R. Honeck & R. Hoffman (Eds.), *Cognition and Figurative Language.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

REFERENCES

Perrault, C. and Allen, J. (1980).
A plan-based analysis of indirect speech acts. *American Journal of Computational Linguistics, 6,* 167-182.

Russell, S. Weber (1976).
Computer understanding of metaphorically used verbs. *American Journal of Computation Linguistics. Microfiche 44.*

Russell, S. Weber (1985).
Conceptual analysis of partial metaphor. In L. Steels & J. Campbell (Eds.), *Progress in Artificial Intelligence.* Chichester, England: Ellis Horwood.

Russell, S. Weber (1986).
Information and experience in metaphor: A perspective from computer analysis. *Metaphor and Symbolic Activity, 6* 227-270.

Russell, S. Weber (1988, April).
Metaphor and computational limits. Presented at the Symposium on the Human Dimension of Artificial Intelligence, sponsored by the University of Kentucky and Asbury Theological Seminary (Computer Science Dept. Report 89-57). Durham, NH: University of New Hampshire (1989).

Russell, S. Weber (1989).
Verbal concepts as abstract structures: The most basic conceptual metaphor? *Metaphor and Symbolic Activity, 4,* 55-60.

Schank, R. (1975).
Conceptual Information Processing. Amsterdam: North Holland.

Schank, R. and Abelson, R. (1977).
Scripts, Plans, Goals and Understanding. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Tourangeau, R. and Sternberg,R. (1982).
Understanding and appreciating metaphors. *Cognition, 11,* 203-244.

Register and Speech Act Theory in Text Generation

Cameron Shelley and Neil Randall and Chrysanne DiMarco
Departments of Computer Science, English, and Computer Science
University of Waterloo, Waterloo, Ontario, Canada N2L 3G1
{cpshelle@logos|nrandall@watserv1|cdimarco@logos}.waterloo.edu

Abstract

The generation of effective text is becoming a greater concern for artificial intelligence as the need for sophistication and complexity rises. The information required to generate appropriate text by humans and programs is not only literal, but situational. The linguistic study of *register theory* addresses the systematic relation of situation to utterance, and should be integrated into current, computational text generation paradigms. This paper will advance a perspective of register theory which will allow its incorporation into such a text generation model. The role of register in the model will then be argued for, followed up by the outline of further relevant directions of research in the fields of computer science, linguistics, and rhetoric.

1 Incorporating register into text generation

Register, in any linguistic sense, is difficult to characterize precisely. Theories of register attempt to describe how the topics, settings, and participants of discourse systematically affect the forms of utterance. Few would dispute that certain kinds of language are appropriate to a given situation, while others are not. However, it is not immediately obvious how such an observation can be turned into a computational model for generating appropriate text.

A notable model of basic text planning is that based on speech act theory [Cohen and Perrault, 1979]. Speech act theory classifies utterances according to use or intended goal, rather than structure. The theory underlies the planning model of a number of computer text generators and interfaces, and provides a promising avenue for further development.

This paper will develop an approach for the integration of register into text generation by applying knowledge of register to constrain the text planning process implied by speech act theory. Specifically, we will review speech act theory and register theory, also touching on a computational model of each, and proceed to propose a text-generating structure incorporating both speech act and register theories. General and implementational aspects of the model will then be addressed.

2 Introducing speech act theory

2.1 Speech acts

People use speech to attain practical effects. While this notion is not new, and in fact is the subject of all classical rhetorical theory [Bizzell and Herzberg, 1990], the first comprehensive attempt at characterizing the achievement of pragmatic effects through speech acts was by the philosopher J. L. Austin [1962]. Austin proposed a classification of language use [Austin, 1962] based on a hierarchy of speech acts that allows the description of the intention of an utterance.

His classification is as follows:

- **Locutionary act**: the act of uttering a proposition;
- **Illocutionary act**: an action intended by a locutionary act;
- **Perlocutionary act**: an effect intended by means of illocutionary acts.

The *locutionary act* is directly related to the utterance of speech—it is the production of a grammatically correct proposition. The *illocutionary act* is once removed from utterance—it is named *by* a proposition (though not always explicitly, as we will see). The *perlocutionary act* is twice removed from utterance—it is a goal accomplished by some string of illocutionary acts.

For example, if someone says to you *You stink!* he has probably attempted: the perlocutionary act of offending you, the illocutionary act of informing you of his belief about your odour, and the locutionary act of communicating a proposition.

Other ways of "pigeonholing" language use by functional level are possible (*i.e.*, [Searle, 1969] and [Searle, 1976]), but Austin's system has remained prevalent[1] and will be adhered to in this paper. Note

[1] See [Saddock, 1988] and [McCawley, 1981] for cur-

that Austin used *speech act* to refer to any of the three acts just described, while it has since largely come to refer only to *illocutionary act* [Searle and Vanderveken, 1985]. The reason for this change will become apparent later. Throughout the rest of this paper, the terms *speech act* and *illocutionary act* will be used interchangeably, but *speech act theory* will still stand for the entire system of perlocutionary, illocutionary, and locutionary acts.

2.2 Performative verbs

The centerpoint of Austin's speech act theory [1962] is the performative verb. Such a verb names a corresponding illocutionary act. Examples of performative verbs include *promise*, *order*, *inform*, *ask*, *tell*, and so on. Instances of their performative uses are:

(1) I promise to be home by 9:00.

(2) I order you to feed the dog.

(3) I inform you of the new deadline.

(4) I ask you, was that necessary?

(5) I tell you, it was awful!

In each case, the main verb *names* the illocutionary act being carried out. Since each speech act above overtly contains the verb naming it, the usages above are said to be *explicit*. These can be converted to *implicit* speech acts, *i.e.*, with no overt performative verb, as follows:

(6) I will be home by 9:00.

(7) Feed the dog.

(8) There is a new deadline.

(9) Was that necessary?

(10) It was awful!

The absense of a performative verb can introduce ambiguity into the inference of which act is being performed. For instance, (6) can be interpreted as a promise (as in (1)), or a threat, or a prediction, *etc.* Implicit or explicit, performative verbs may be regarded as a means of partitioning the perlocutionary from the locutionary level.

2.3 Summary of speech acts

Although speech act theory is not uncontroversial, (*e.g.*, the proposals of [Ballmer and Brennenstuhl, 1981]), it provides a means for describing the functions that language fulfills and has proven useful in the computer modelling of text generation.

3 Planning communication by speech acts

An important contribution to the translation of speech act theory into a computational paradigm was made by Cohen and Perrault [1979]. The premises and results of their work relevant to our proposal will be presented in this section. The model they describe will then become the platform for our approach to text generation.

3.1 A plan-based model of speech acts

Cohen and Perrault (C&P) submit that perlocutionary acts can profitably be viewed as goals that a speaker wishes to achieve. Having a goal usually implies having a plan to attain it. Therefore, C&P model a human speaker as someone "executing a plan that prespecifies the sequence of actions to be taken" [Grosz *et al.*, 1986, 423]. In this framework, speech acts can be regarded as operators designed to change what the speaker believes and what the speaker thinks the hearer understands. In other words, speech acts are the actions that attempt to carry out the planned goals of some agent.

Each operator defined by C&P has the following components:

- a **name**—*e.g.*, CAUSE-TO-WANT or REQUEST;
- a set of **preconditions**—which specify the physical and mental (CANDO.PR and WANT.PR) circumstances necessary to invoke the operator;
- a set of **effects**—(labeled EFFECT) that dictate which conditions will result from executing the operator.

A goal such as CAUSE-TO-WANT represents an effect an agent intends to have on someone, and so is a perlocutionary goal. The goal REQUEST, however, refers to the act of informing someone of an agent's desire, and is therefore an illocutionary goal in terms of speech act theory.

Cohen and Perrault's system may carry out a perlocutionary goal (produce its EFFECT) by planning to invoke the set of illocutionary acts that create its required preconditions (by means of *their* EFFECTs). For instance, for a speaker to incite a hearer to some action (CAUSE-TO-WANT), he must convince the hearer that he desires the action. This precondition can be brought about by planning a REQUEST for the hearer to do the action. When the hearer receives the request, she will (in the system) then believe that the speaker desires the action. Thus, the speaker has acomplished the communication of a perlocutionary goal by means of an illocutionary act. In general, a wide range of perlocutionary goals can be accomplished by the same chaining process with illocutionary acts as described above.

This process of planning to chain a certain sequence of operators having preconditions and postconditions (EFFECTS) is computationally equivalent to deriving a plan-tree from a set of semantically constrained planning rules, *i.e.*, an *attribute grammar* [Knuth, 1973]. In this way, we may use formal, grammatical terminology in discussing planning systems like C&P's. Much of our proposal will be phrased in terms of grammar rules and attributes.

rent discussion.

3.2 Summary of the computational model

Cohen and Perrault's formalism successfully showed that speech act theory can be translated into a computational model for utterance planning. Their model will be the starting point of our own. Cohen and Perrault also make a number of pragmatic simplifications in their work, one of which is that situational factors are not considered. Our model will partially relax this simplification.

4 Considering register and illocutionary force

Many computational text generators have either ignored or fixed the situations for which their text is intended. Such systems treat all users and topics as identical. To alleviate this, several current researchers in text generation are seeking to apply results of *register theory*. We will briefly present the register concepts of *features*, *domains*, and *roles*[2], which we will later translate into components of our model. Since these components will be introduced at the perlocutionary level, but must affect the locutionary (utterance) level, we will discuss *illocutionary force* [Searle, 1969] as a means through which register can help govern utterance production.

4.1 Features, domains, and roles

Although a precise definition of register is elusive[3], the theory we propose to adopt characterizes the systematic variation of language with socio-linguistic setting in terms of features, domains, and roles.

Features are discrete contextual factors that act as parameters to the process of forming an utterance. Common examples of features are: the 'formality' of the situation, the 'sensitivity' of the topic under discussion, and the 'importance' of the topic. The canonical illustration of how formality can affect the form of a proposition is shown by two sentences requesting a salt shaker:

(11) Give me the salt!

(12) Could you pass me the salt, please?

the first of which is very informal, while the second is more formal. In fact, the second sentence does not literally express the desire for a salt shaker. The hearer must infer the true goal of the request from his or her own knowledge of formal requests.

The *key*, or the *level of planning* appropriate to producing an utterance in a given situation, may also be regarded as a feature. Gleason (excerpted in [Clark *et al.*, 1981], based on Joos [Joos, 1962]), proposes five keys of interaction, in order of increasing planning effort necessary: *intimate*, *casual*, *consultative*, *deliberative*, and *oratorical*. So, as the hearer becomes more critical, the number of constraints which must be simultaneously obeyed increases, thus also increasing the key level.

Domains label generic categories of text that exhibit particular, sometimes unique, rules of composition. A classic example of a domain is the 'recipe' domain, in which utterances such as the following occur (in [Attardo, 1990], due to [Haegeman, 1987]):

(13) Skin and bone chicken, and cut 0 into slices.

The "0" marks the place where the pronoun *it* would be in normal speech. Other examples of domains could include highly formal poetry (sonnets, villanelles, and haiku), specifically formatted technical documentation (*e.g.*, reference manuals and guides), and formalized academic essays. Indeed, domains such as poetry must be analyzed in terms of subdomains, such as ballads, limericks and the three mentioned above[4].

Role characterizes the relationship the speaker has with the hearer. Typical examples of the sublanguages produced by roles are 'motherese' and 'journalese', from the *mother* and *journalist* roles. Many of us have seen examples of journalese such as the following while in grocery store lineups:

(14) Woman has Alien Baby!

Both *Woman* and *Alien Baby* lack the usual article before them.

It is appropriate to note here that the rhetorical theorist Kenneth Burke, in his seminal work [Burke, 1945], identifies *role* as the most essential component of human interaction. In his lengthy analysis of motivation (which continues in [Burke, 1950]), Burke argues that all communicative acts (hence, all motivations) operate according to a five-part structure (*pentad*) consisting of *act*, *scene*, *agent*, *agency*, and *purpose*. Together, these five elements constitute a "drama" of interaction in which *roles* are (usually unconsciously) assumed.

Burke's pentad has informed a great deal of rhetorical study in the past two decades. Like Aristotle, Burke is concerned primarily with considering the circumstances of an action [Golden *et al.*, 1976], but Burke carries the idea further than Aristotle by assessing specific ratios among the individual elements of the pentad. For Burke, the goal of all language is the communication of motivation, and only through the assumption of roles can this communication occur.

Drawing on Burke's dramatistic theory of rhetoric, Ernest Bormann [1980] has established a dramatistic theory of all communication. Like Burke, Bormann establishes *role* as the dominant determiner of what can be communicated. It is beyond the scope of this paper to apply Burke's pentad or Bormann's theory of communicative processes to the

[2] Adapted from [Attardo, 1990].

[3] See [Halliday, 1978] for more discussion.

[4] The direct effect of the subject-matter in a domain, *e.g.*, the frequent occurrence of "chicken" in the recipe domain, does not concern register and may constitute a 'subject-dialect'. The choice between "capon" and "chicken" however, should be classed as a register variation caused by the need for technical language.

notion of *role* in register theory presented here, but ultimately such study will help considerably in determining the means by which *role* establishes communicative interaction.

One objection to the account of register advanced here is the difficulty of justifying any particular enumeration of features, domains and roles. Why is one quality a feature, and another not? Why is there a motherese, but no nephewese? This difficulty suggests that some instances of register components (*e.g.*, roles) are produced by what Searle calls *constitutive rules* ([Searle, 1967]). These are rules 'made up' dynamically to create behaviour appropriate to a novel situation, and formed from pragmatic knowledge. This consideration lies, however, outside our present proposal.

A last fact to be considered about register is its relation to each participant in a situation. Any participant may have her own register (or *point-of-view*) with regard to a single setting. Thus, a drill instructor addressing a recruit on a parade square may use very informal language such as:

(15) Get a grip, you bone idle man!

The recruit so addressed will usually be confined to a formal register and reply as follows:

(16) Yes sir!

This implies that register mechanisms must be independantly available to every speaker or speaker model.

Thus, we can see that register represents part of each speaker's systematic attempts to adjust his language to its setting.

4.2 The function of register

Although it is easy to say that language varies with setting, it is also important to see why such variation is necessary for both human and computer text generators. Grice's maxims of conversation [Cole and Morgan, 1975] can be viewed as ways a speaker can maximize the chances of the hearer understanding him correctly. Likewise, the effects of register can be seen as considerations that maximize the potential success of speech act plans when dealing with different social environments.

Since, as we noted previously, speech act theory can be interpreted as a grammar of utterance-planning operators, and since the components of register theory discussed describe systematic variations of language production with setting, we propose to view those components as a set of constraints on the selection of text planning rules. The application of such constraints essentially defines a subgrammar of appropriate language.

4.3 Illocutionary force

As mentioned previously, we will argue that register constrains the text planning process explicitly at the perlocutionary level. This leaves us to explain a method by which such perlocutionary considerations affect the locutionary planning component. In other words, why is a structure like *Could you pass me the salt, please* considered more formal than *Give me the salt?* One approach is to apply the concept of *illocutionary force.* Illocutionary force relates broad categories of illocutionary acts to the speaker's attitudes they reflect.

Searle [1969] describes[5] a taxonomy of illocutionary force with five primary categories as follows: (S means Speaker, H means Hearer, P means the Proposition expressed; each entry in the list gives the category label followed by its definition with an example afterwards.)

- **Representative**: Commits S to the truth of P *e.g., The Emperor has no clothes on.*;
- **Directive**: S attempts to get H to do something *e.g., Begin the parade!*;
- **Commissive**: Commits S to a future course of action *e.g., I will march through the city.*;
- **Expressive**: Expresses psychological state of S regarding P *e.g., Thank you for your clever work.*;
- **Declarative**: Creates the condition expressed in P *e.g., You are sentenced to one year in jail.*.

The above statements correspond roughly to the illocutionary acts of saying, ordering, promising, thanking, and sentencing, respectively. Thus, the speaker's varying intentions can be communicated effectively by syntactic and lexical selections at the locutionary or utterance level. If our proposed register-constrained perlocutionary planning component can decide which illocutionary force category is most appropriate, then illocutionary planning rules will be able to map those intentions onto particular deep syntactic or other structures for the locutionary component to realize.

It is not clear, however, that illocutionary force alone will be sufficient to generate the variation of language already discussed under register. But we may look to goal-directed computational models of style [DiMarco, 1990], goal-directed computational models of rhetoric [Hovy, 1987], and contemporary rhetorical theories for further ideas and details of implementation at the illocutionary level.

In fact, the study of rhetoric is precisely the study of a speaker's intention. The classical rhetoricians, from Isocrates and Gorgias through Cicero and Quintilian, were concerned with educating speakers to reveal their intentions in the most effective possible way, and contemporary rhetoric attempts to determine means by which audiences can determine those intentions. In other words, rhetoric studies illocutionary force.

Illocutionary force need not be conscious. Burke ([1945], [1950]) insists that motivation is at the heart

[5] See [Searle and Vanderveken, 1985] for a longer and more technical exposition of illocutionary force.

of the communicator's intention, and motivation unquestionably guides illocutionary force (you get what you want by saying what you must). Perelman [1969] discusses the relationship between practical (i.e., logical) reasoning and the twin rhetorical focusses of *argumentation* and *demonstration*, arguing that effective argumentation demonstrates the speaker's will to be heard and understood.

Furthermore, the relationship between speaker and hearer, which is at the centre of Searle's taxonomy, is also the focal point for the study of *ethos* in rhetoric. Aristotle and Cicero linked *ethos* to the *character* of the speaker, while Burke suggests that the relationship might best be called *identification* [Burke, 1945], and Perelman [1969] argues for the study of the *presense* of the speaker in the discourse. Because the study of *ethos* is essentially the study of perceived intentions (and these intentions range from a speaker's credibility through such complexities as moral stance), it is also the study of illocutionary force. Once again, further exploration of the relationship of rhetoric and illocutionary force should contribute to the current theory.

4.4 Summary of register and illocutionary force

The linguistic concepts of *register* and *illocutionary force* provide a framework for understanding how the setting of language use is translated into particular utterances. Such register constituents as *key* demonstrate that these theories can be used to constrain a text generator in which text planning is based on speech act theory. Approaches for implementing register computationally will be brought forward in a later section.

5 Bringing register and planning together

Before proceeding to describe our model, we will briefly review ongoing work at the Information Sciences Institute (ISI) which has taken a similar direction to that which we propose. Specifically, it is the development of a sophisticated text planning component to explain the reasoning of an expert system—the Explainable Expert System (EES) [Moore and Paris, 1989].

The EES produces text by planning it in a manner similar to that suggested by Cohen and Perrault, but adapts its plan language from the *Rhetorical Structure Theory* described by Mann and Thompson [1987]. As laid out by Moore and Paris (M&P) [1989], the plan language uses operators consisting of:

- **Effects**—which state the perlocutionary goal(s) the operator intends to achieve;
- **Constraints**—which state the preconditions for executing the operator;
- **Nuclei**—which specify the illocutionary act(s) to be expressed by the operator;
- **Satellites**—which give additional goals and/or acts that may be needed to achieve the effect.

In two recent papers, Bateman and Paris ([1990], [1991]) discuss register theory and how it might be used to enhance the EES framework. Specifically, they wish to tailor the form of generated text to the category of user who is interacting with the system. They divide users into three different roles depending on domain-knowledge:

- **System developers**—who wish to make sure the expert system knowledge base is correct and working properly;
- **End-users**—who want to follow the system's reasoning but do not know about expert systems;
- **Students**—who are naive about both expert systems and the domain.

After some experimentation ([Bateman and Paris, 1990]), they arrived at several distinct kinds of linguistic variation that characterized interactions with these different kinds of hearer. These include:

- The lexical choice between using *exist* rather than something like *be*;
- Selecting between possessive constructions *X's Y* and *Y of X*;
- Pronominalizing a noun or leaving it expanded.

The roles of the system and the user form only a part of linguistic register, but this augmentation of EES is a useful start and demonstrates that register theory can be applied to produce effective variation in the text generated during user-interaction.

6 Unveiling the model

This discussion of our model will consist of three parts: the first will examine and justify that same text-planning structure depicted in figure 1, the second will project methods for the implementation of our model, and the third will give a walk-through example of the proposed model, for clarification.

6.1 The big picture

In justifying the placements of components in figure 1, we will assume the placement of *perlocutionary*, *illocutionary*, and *locutionary* on the basis of our discussion of speech act theory. Also, the incorporation of register into every individual planning agent will rest on the point-of-view principle discussed previously in section 4.1.

Firstly, as Cohen and Perrault noted, a generalized planning agent could have access to both physical and linguistic means to achieve goals. There must therefore be a stage in the planner in which the planning process is undifferentiated, and stages in which planning for the two types of action has diverged. Perlocutionary goals such as offending

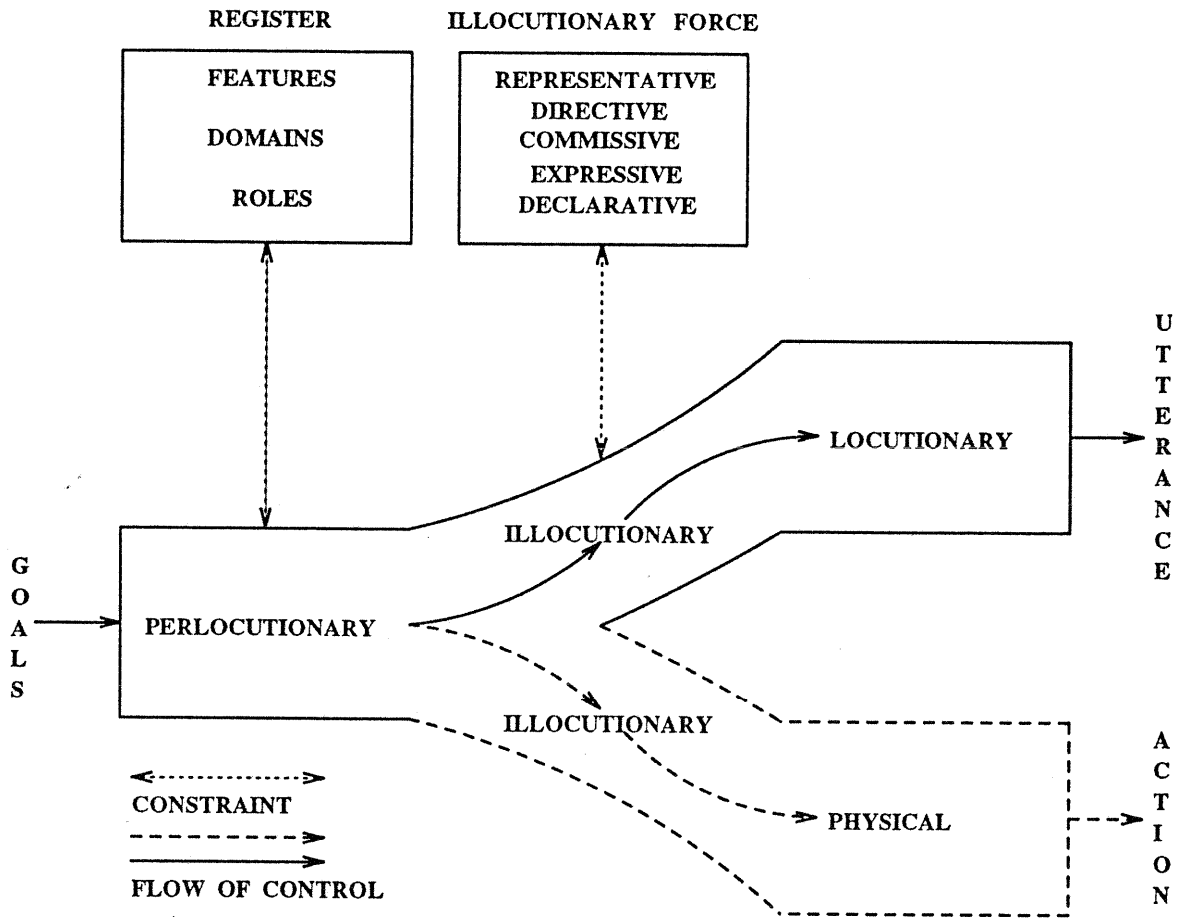


Figure 1: Speech acts, register, and illocutionary force

someone can clearly be accomplished by an appropriate gesture *or* insult (or a combination), so that we follow [Appelt and Konolige, 1988] in not distinguishing perlocutionary acts by physical or linguistic properties.

However, the planning process must be dedicated to utterance production if it reaches the locutionary stage, indicating that the planning structure must split by that point. Therefore, the division of the structure must occur at the illocutionary stage. Whether we label both paths at that stage "illocutionary" depends on how further studies in multimedia presentation progress[6]. Since our proposal does not deal further with plans involving physical acts, we will not attempt to resolve the issue here.

Secondly, the placement of illocutionary force away from the initial planning goals reflects the fact that it resides further from conscious control. For instance, it is easier to imagine someone setting a goal of changing the register of a situation (*e.g.,* a *pick-up line* in a singles bar that attempts to change formal to informal) than setting a goal to change his concept of what sentence structure corresponds to which force. Also, we have shown that illocutionary force lends itself to being a transitional process between the perlocutionary and locutionary levels.

Finally, register is not placed at the illocutionary stage since its information is needed at the perlocutionary stage. The eventual control of register of illocutionary force emphasizes its function as *look-ahead* information for constraining the planning search space in a heuristic system.

6.2 Implementation of the register model

We have shown that both speech act and register theories can be incorporated into a single model of text generation. Here, we will further substantiate this claim by suggesting computational mechanisms that implement speech acts and register together in a single framework.

We propose to model the roles described by register with a set of perlocutionary planning operators. This is justified by the view that roles exist to maximize communicative effectiveness ([Burke, 1945], and [Bormann, 1980]) discussed earlier, and that assuming a role represents the best way of achieving a relevant goal. As a perlocutionary operator is also intended to be the best rule for achieving an effect, it is natural to import roles into the planning system as instances of those operators. For example, since the *mother* role is effective in *put-the-kids-to-bed*, the planning agent could invoke the *mother rule* which would know to plan an illocutionary *order-kids-to-the-bathroom* first.

Situational features would then act as parameters to the perlocutionary operators. Thus, different utterances can be consistent with a particular role (and other goals), but vary with the formality, sensitivity, *etc.*, of the situation. For instance, if there are guests in the house, the *mother* might plan a *put-the-kids-to-bed* by invoking a polite request as opposed to an order, due to the formal situation. These parameters could then be inherited by any further planning operators invoked as the higher role operator sees fit.

This also suggests a treatment of domains as roles or operators in which the relevant features are fixed. For example, recipes are not written to sound abusive, but always formal. (We will ignore infelicitous uses of register for our present purposes.)

Another feature of register previously discussed, key, appears to be a function of a number of other features, such as sensitivity, importance, and so forth. We propose that an effective model would be to interpret each key as an upper bound on the amount of planning effort and assign each planning operator a value describing the effort required to complete it. This value would vary with the features involved. By inheriting and adding up the effort expended, the system can avoid exceeding the amount of processing appropriate to the current key[7]. We leave the determination function for key to further empirical study.

6.3 An enterprising example

Let us suppose that Captian Kirk (of Star Trek fame) is seated at the dinner table in the officer's mess eating a synthetic steak. Since the taste is not to his liking, he forms the desire to sprinkle some salt on it. Looking about, he spots a salt shaker nearby on the table—at the other side of Mr. Spock's place setting. In order to fulfill his desire, Kirk must first obtain the salt shaker. Please refer to figure 1 starting at "GOALS" to follow the flow of control.

The usual situation in the officer's mess is formal, but relaxed. However, Kirk knows that, in his role as captain, he must maintain a high standard of decorum and authority. Therefore, his planner passes the goal *obtain-salt-shaker* along with such features as *formal*, *relaxed*, and *authority* to the domain planner for *table-manners*.

Kirk's *table-manners* planner is faced with several choices. Since the salt-shaker is nearby, Kirk could simply reach over Spock's setting and grab it. He could also just indicate the salt shaker to Spock with a gesture, or he could ask for Spock to hand it to him. But, due to the formality of the situation, Kirk decides reaching over Spock's setting would be inappropriate, and Vulcans have been known to apply the death grip to those who reach over their plates at mealtime. Just pointing to the salt shaker for Spock to see and interpret is however too imperious and violates the relaxed atmosphere of the setting. So, the goal *request-salt-shaker* is passed to a final perlocutionary planner, along with any appropriate features.

[6] See [McKeown *et al.*, 1990], [Hovy and Arens, 1990], and [Marks and Reiter, 1990].

[7] See [Gibson, 1991] for a similar paradigm applied to syntactic processing.

At this point, Kirk's planner must decide what series of speech acts will get him the salt shaker. Since the event is not unusual, and there are no other complications, Kirk will simply determine to inform Spock of his desire for the shaker. The goal of *inform-of-desire-for-salt-shaker* is then passed on to the illocutionary planning component.

The captain's illocutionary planner has to decide what specific kinds of utterance will meet the goal. Kirk's position of authority dictates that he use the *directive* illocutionary force, and also that he specify literally what he wishes done. However, the relaxed setting obviates the need for an explicit performative, and also calls for the use of a polite lexical item such as *please*.

So, the appropriate goal is passed to the locutionary component which realizes it with the single utterance "Pass the salt please, Spock."

By placing register directly into the planning mechanism, we can obtain the constraint of the planning grammar mentioned previously to produce the desired subgrammar appropriate to a situation. The possible selection of a subset of other perlocutionary rules by *our* proposed rules will narrow the planning search space and focus the search itself on the operators relevant to a given goal. This, in turn, helps to provide the communicative effectiveness that is a primary concern of non-trivial text generation systems.

Planning from perlocutionary acts to locutionary acts can be accomplished by applying rules governed by illocutionary force. Illocutionary force can be augmented by current computational models of style and/or rhetorical structure to help produce the range of variation that register considerations make possible.

7 Concluding remarks

We have presented an overview of speech act theory and register theory and shown that both are relevant to a model of appropriate text generation. Further, we have proposed a structure in which aspects of both theories are directly active in the defining of a subgrammar most effective for planning utterances that accomplish goals. We have also given computational, linguistic, and rhetorical justifications for the correctness of the proposal, which is vital when considering further augmentations of text generation systems. Accounting for register phenomena promises to allow the production of text which is more natural for its intended audience, which in turn maximizes its usefulness.

The appropriateness and effectiveness of text will become a greater concern for artificial intelligence as the complexity of text generators (such as machine translation systems) increases and as demand for sophistication rises. The information required for text production by humans is not only literal, but also situational, and a complete understanding of natuaral language utterances will not be accomplished unless this information is investigated. Effective text will not be realized serendipitously—amplification of current linguistic and computational models is indispensible to further progress in useful text generation.

Acknowledgements

We would like to express our gratitude to Graeme Hirst for his erudite comments on a draft of this paper and Salvatore Attardo for his insight on the problems of register.

We acknowledge the financial support of the Department of Computer Science, University of Waterloo, the Natural Sciences and Engineering Research Council of Canada, the Information Technology Research Centre, and the Ontario Government for its Graduate Scholarship.

References

[Appelt and Konolige, 1988] D. Appelt and K. Konolige. A practical nonmonotonic theory for reasoning about speech acts. In *Proceedings of the 25th Annual Meeting of the ACL*, pages 170–178. SUNY at Buffalo, Buffalo, New York, 1988.

[Attardo, 1990] S. Attardo. A polythetic theory of register and its relevance to esl/efl. Presented at the Conference on Pragmatics and Language Learning, 1990.

[Austin, 1962] J. L. Austin. *How to do things with words.* Oxford University Press, 1962.

[Ballmer and Brennenstuhl, 1981] T. Ballmer and W. Brennenstuhl. *Speech act classification.* Springer-Verlag, 1981.

[Bateman and Paris, 1990] J. A. Bateman and C. L. Paris. User modeling and register theory: a congruence of concerns. Presented at the 2nd International Workshop on User Modeling, 1990.

[Bateman and Paris, 1991] J. A. Bateman and C. L. Paris. Constraining the deployment of lexicogrammatical resourses during text generation: towards a computational instantiation of register theory. In *Proceedings of the 16th International Systemic Congress*, 1991.

[Bizzell and Herzberg, 1990] P. Bizzell and B. Herzberg. *The rhetorical tradition.* Bedford Books, 1990.

[Bormann, 1980] E. G. Bormann. *Communication theory.* Holt, Rinehart and Winston, 1980.

[Burke, 1945] K. Burke. *A grammar of motives.* Prentice Hall, 1945.

[Burke, 1950] K. Burke. *A rhetoric of motives.* Prentice Hall, 1950.

[Clark *et al.*, 1981] V. P. Clark, P. A. Eschholz, and A. E. Rosa. *Language: introductory readings.* St. Martin's Press, 1981.

[Cohen and Perrault, 1979] P. R. Cohen and C. R. Perrault. Elements of a plan-based theory of speech acts. *Cognitive Science*, **3**(3):177–212, 1979.

[Cole and Morgan, 1975] P. Cole and J. L. Morgan, editors. *Syntax and semantics*, chapter Volume 3. Academic Press, 1975.

[DiMarco, 1990] C. DiMarco. Computational stylistics for natural language translation. Technical Report CSRI-2239, Dept. of Computer Science, University of Toronto, 1990.

[Gibson, 1991] E. Gibson. A computational treatment of processing overload. Presented at the fourth annual CUNY sentence processing conference, 1991.

[Golden *et al.*, 1976] J. L. Golden, F. B. Goodwin, and W. E. Coleman. *The rhetoric of western thought.* Kendall/Hunt, 1976.

[Grosz *et al.*, 1986] B. J. Grosz, K. S. Jones, and B. L. Webber, editors. *Readings in natural language processing*, chapter Discourse Interpretation. Morgan Kaufmann, 1986.

[Haegeman, 1987] L. Haegeman. Register variation in english: some theoretical observations. *Journal of english linguistics*, **20**(2):230–248, 1987.

[Halliday, 1978] M. A. K. Halliday. *Language as social semiotic.* University Park Press, 1978.

[Hovy and Arens, 1990] E. Hovy and Y. Arens. How to describe what? towards a theory of modality utilization. In *Proceedings of the 12th Annual Conference of the Cognitive Science Society*, 1990.

[Hovy, 1987] E. Hovy. Generating natural language under pragmatic constraints. *Journal of pragmatics*, **11**(6), 1987.

[Joos, 1962] M. Joos. *The five clocks.* Harcourt and Brace, 1962.

[Knuth, 1973] D. E. Knuth. *Sorting and searching.* Addison Wesley, 1973.

[Mann and Thompson, 1987] W. C. Mann and S. A. Thompson. *The structure of discourse*, chapter Rhetorical Structure Theory: a Theory of Text Organization. Ablex, 1987.

[Marks and Reiter, 1990] J. Marks and E. Reiter. Avoiding unwanted conversational implicatures in text and graphics. In *Proceedings of the National Conference on AI.* Boston, MA, 1990.

[McCawley, 1981] J. D. McCawley. *Everything that linguists have always wanted to know about logic*, chapter Speech Acts and Implicature. University of Chicago Press, 1981.

[McKeown *et al.*, 1990] K. McKeown, R. Kathleen, and S. Feiner. Coordinating text and graphics in explanation generation. In *Proceedings of the National Conference on AI.* Boston, MA, 1990.

[Moore and Paris, 1989] J. D. Moore and C. L. Paris. Planning text for advisory dialogues. *Journal of the ACL*, 1989.

[Perelman and Olbrechts-Tyteca, 1969] C. Perelman and L. Olbrechts-Tyteca. *The new rhetoric: a treatise on argumentation.* University of Notre Dame Press, 1969.

[Saddock, 1988] J. M. Saddock. Speech act distinctions in grammar. In F. J. Newmeyer, editor, *Linguistics, the cambridge survey.* University of Cambridge Press, 1988.

[Searle and Vanderveken, 1985] J. R. Searle and D. Vanderveken. *Foundations of illocutionary logic.* Cambridge University Press, 1985.

[Searle, 1967] J. R. Searle. Human communication theory and the philosophy of language. In F. E. X. Dance, editor, *Human communication theory: original essays.* Holt, Rinehart and Winston, 1967.

[Searle, 1969] J. R. Searle. *Speech acts: an essay in the philosophy of language.* Cambridge University Press, 1969.

[Searle, 1976] J. R. Searle. A taxonomy of illocutionary acts. In K. Gunderson, editor, *Language, mind and knowledge.* University of Minnesota Press, 1976.

Learning Metaphorical Relationships Between Concepts Based on Semantic Representation Using Abstract Primitives

Masaki Suwa **Hiroshi Motoda**
Advanced Research Laboratory, Hitachi Ltd.
Hatoyama, Hiki-gun, Saitama Pref. 350-03, Japan.
suwa@harl.hitachi.co.jp, motoda@harl.hitachi.co.jp

Abstract: The goal of this paper is to elicit metaphorical relationships between concepts from metaphors we use in everyday life. In the past literatures, there have been two major approaches to interpreting metaphors; natural language processing approach and analogical approach. We follow the latter. We provide a set of abstract primitives by use of which to represent the meanings of words. Based on that semantic representation, metaphor interpretation is attempted in a structure-based way with a semantic constraint of "primitive matching" at the initial stage of finding the entire correspondence the metaphor stands on. Out of the acquired mappings, a mapping between abstract primitives is elicited as knowledge about "metaphorical relationships between concepts" underlying the metaphor. The use of the abstract primitives is essential in two ways. First, the semantic constraint of "primitive matching" contributes to reducing the calculation cost of determining the entire correspondence. Secondly, the generality of the acquired knowledge comes from the abstractedness of the primitives.

1 Introduction

Metaphor is important because the act of viewing a concept as another concept for the purpose of highlighting a feature of the former is central to our thinking or expressing ideas [4] [14]. In AI community, there have been two major computational approaches to interpreting the metaphorical meaning of a non-literal sentence [16], natural language processing approach and analogical approach, although both share the view that metaphor stands on a similarity between a targe concept and a source one. In natural language processing approaches, a metaphorical interpretation is attempted in a unified framework of processing both literal and non-literal sentences, by replacing the source concept with some target concept which semantically matches the source best [7] [22] [19], when the source violates a selection restriction. On the other hand, in analogical approaches, new metaphorical meanings are created by finding a target concept whose structure corresponds to that of the source in terms of analogical mapping and then transferring relations into the target domain from the source domain[8] [4], in which the advantage of the unified interpretation of literal and non-literal sentences is discarded.

Lakoff [14] says that we human beings have "general metaphor", prototypical structures of metaphorical expressions, as commonsense knowledge and that we communicate to each other using this knowledge. One of the attempts of codifying information about "general metaphor" as metaphorical knowledge is Carbonell's work [4], in which he specifies that a general metaphor should consist of four types of knowledge; *a recognition network*, *basic mappings*, *implicit intention component*, and *transfer mappings*. He implies that metaphorical interpretation consists of two tasks. One is to find an appropriate target concept, using a recognition network and basic mappings. The other is to discover the spearker's (or writer's) intention about why he dares to use the source concept instead of using the target concept, *i.e.* to highlight the features which would not usually be included or salient in the target domain, for which an implicit intention component and transfer mappings are used. Especially, the latter task is important, as Ortony asserts that *salience imbalance* between the source and the target is a source of metaphoricity [18].

Comparing the above two approaches from this viewpoint, analogical approaches are more advantageous than natural language processing approaches because an analogically directed transfer of relations from the source to target can naturally implement the process of highlighting non-salient features of the target or creating new features in the target. In this paper, we follow the analogical approach for interpreting metaphor.

Most works in analogical approaches have focused mainly on what kind of constraints need to be imposed on in determining analogical mappings of two structures. According to Holyoak's analysis [12], there are three kinds of constraints; *structural consistency*, *semantic similarity*, and *pragmatic centrality*. Those supporting *pragmatic centrality* insist that similarity cannot be determined without goals or contexts in using the analogy. On the contrary, those supporting the former two constraints insist that we human beings have already had some heuristics for determining similarity out of contexts, although *structural consistency* constraint and *se-*

mantic similarity constraint are different types of heuristics. The former is a purely syntactic heuristic, *e.g.*, Gentner's *systematicity principle* [8], and the latter is a heuristic based on semantic representation of concepts. These three constraints are analysed in detail in the second section, which motivates us to take the following approach in this paper.

Our approach, which belongs to the heuristics approaches, employs both semantic and syntactic heuristics in determining analogical correspondences between a target and a source concept. In this paper, the following metaphor will be used for discussions as an example (Example 1 [14]).

He shot down my opinion in the argument.

The verb '*shoot down*' is used metaphorically and the word which has the closest meaning may be '*criticize*'. We treat only those metaphors in which verbs are used metaphorically. In the third section, we introduce the method of semantic representation of verbs as the basis for using semantic heuristics. In the fourth section, we propose an algorithm of using both semantic and syntactic heuristics.

Here, we have to make clear our motivation of processing metaphors computationally. Lakoff asserts that metaphor is pervasive in everyday life, not just in language but in thought and action [14]. Barnden [1] claims that people cannot describe or explicate the state of their minds without using metaphors in which objects in mental worlds are viewed as those in the actual world we live in. In other words, they suggest that our feeling about similarities between words or concepts is essential to our thinking and acting. According to what they say, metaphors are treasure houses of human beings' "metaphorical relationships between concepts".

Our goal in this paper is to propose a method of eliciting knowledge about "metaphorical relationships between concepts" from metaphors we use in everyday life. We mean by "understanding a metaphorical sentence" not only finding correspondences between the source and the target concepts in the sentence but also learning this kind of knowledge. In the fifth section, we show an example of how those metaphorical relatioships between concepts are elicited from metaphors, which suggests that the knowledge representation in our approach is essential to the elicitation.

2 Constraints in Making Analogical Mapping

2.1 Structural consistency

Many researchers, particularly Gentner [8] [9], have stressed the importance of the constraint of structural consistency. This constraint requires that two analogues should have any partial correspondence in syntactic structure. In Gentner's structure-mapping theory, correspondence must be one-to-one. In addition to this, he proposed another syntactic heuristics, *systematicity principle*. The systematicity priciple states that a base predicate which belongs to a mappable system of mutually interconnecting relations is more likely to be imported into the target than an isolated predicate.

Structural consistency constraints can be formulated in terms of "morphism". Let T denotes a target concept. T can be represented as $< O_t, R_t >$, where O_t is a set of objects and R_t is a set of relations on O_t. $o_i R_k o_j$, $(o_i, o_j \in O_t)$ means a theorem which holds in T. Then, let S denotes a source concept and if the following condition,

$T \models o_i R_k o_j$ implies
$\exists f,\ S \models f(o_i) f(R_k) f(o_j)$

holds, we say that there is a structural correspondence between T and S.

If f is required to be one-to-one morphism and also if the above condition is required to hold for all the theorems in T, this would be extremely strong constraints [12]. It is called isomorphism. This strict formal definition of isomorphism is clearly inappropriate as a characterization of the kinds of analogy we human beings feel. Rather, two analogues may have correspondences for only some theorems in T, or some correspondences may be many-to-one. These can be viewed as approximations to isomorphism.

However, discovering similarities based on structural consistency causes us to encounter a problem inherent in "open" domains if we want to process metaphors computationally; in case huge amount of relations constitute the structures of both the source and the target concepts, the calculation cost needed to determining the most plausible correspondence becomes extremely large [12], which would be indesirable as a computational modeling on understanding metaphors.

2.2 Semantic similarity

Some researchers insist that the formal definition of isomorphism makes no reference to the similarity of the objects and the relations involved in two analogues. Halford [10] claimed the necessity of a constraint, *semantic similarity*, by mentioning the following example;

(1) John is taller than Bill.
(2) Mary is heavier than Susan.
(3) Communism is more radical than socialism.

Even though a human being would feel that the sentences (1) and (2) are more similar to each other than sentences (1) and (3), the degree of structural

consistency is the same in both pairs of analogues.

The major disadvantage of this approach is that defining *semantic similarities* of objects or relations itself is very difficult; similarity of objects can potentially be reduced to similarity of relations by specifying that two objects are similar if they serve as arguments in the same location of similar relations. Similarity of relations must in turn be analysed in terms of some feature overlap [21]. However, defining what "the feature of a relation" should be is the very question we are asking.

The method commonly used in the past studies is to add a strong restriction to the constraint of structural consistency which would reflect a human being's feeling about *semantic similarity, i.e.* the corresponding relations must be identical [5] [11], or share common superordinate in an abstraction hierarchy of features [3]. The former constraint does not match the characteristics of human beings that we would find similarities even between two non-identical concepts [3]. The latter cannot explain the fact that making correspondence between such relations that have no common superordinate would rather yield more vivid and acute metaphors [4] [14].

2.3 Pragmatic centrality

Researchers in this approach share the idea that similarity cannot be determined without predetermined focuses or goals or contexts. Some researchers have insisted that the mappings for some "specific relations" should work as a constraint and influence the entire correspondence between two analogues. *Cause*, *purpose*, *motivation* and *reason* etc. are selected as "specific relations" [23], which are called "causal relations". This constraint can apparently be an answer to the problem of large calculation cost in using only *structural consistency* constraint. These models have the common implication that they reflect an insight that human beings tend to focus on causal relations mainly.

However, in reality there are possibly lots of causal relations involved in describing a concept, which requires us to provide another criterion for determining which relation is the most relevant to the current situation of using the analogy. Some researchers have stressed the significant roles of high-level purpose/goal; they have the view that in order to determine the most relevant "causal relation" people use explicit or implicit knowledge about the purpose the analogy is intended to accomplish. In Burstein's model [2], an explicit description is given as an index into the most relevant causal relation. Kedar-Cabelli's model [13] is provided with a goal concept suggesting the aspect of the target concept which people would want to focus on using the analogy and she insists that during the explanation process of the goal concept only the relevant causal relations come to appear dynamically. Although these models have some differences, they all share the idea that high-level purposes/goals provided implicitly or explicitly contribute to focusing on the relevant causal relations. Holyoak [12] referred to the knowledge for making access to specific "causal relations"or "high-level purposes/goals" as pragmatic knowledge.

Getting back to the metaphorical sentences we treat in this paper, "He shot down my opinion in the argument", is it possible for us to provide pragmatic knowledge that is central to understanding the sentence ? The answer may be no. The speaker's intention of this sentence is to highlight a non-salient aspect of the target concept "criticize" by using "shoot down" as the source concept. In the domain of metaphors, it is a speaker's intention itself that corresponds to the pragmatic knowledge. It implies that pragmatic knowledge should not be provided explicitly in advance, but should be a new discovery as a result of understanding the metaphor. Therefore, we cannot depend on the constraint of pragmatic centrality for the purpose of this paper.

Gentner's statement in [9] supports our claim. She makes a clear distinction between soundness and relevance of an analogy; an analogy can be judged sound if sufficient structural overlap to support importing relations from source to target can be found as a result of structure-based bottom-up inference, while it can be judged relevant if there are such goal-oriented top-down inferences as can be transferred from source to target and satisfy the goal in the target domain. She points out that the purely relevance-based approach cannot account for the fact that people can comprehend or spontaneously generate a metaphorical sentence even outside of a goal-context.

From her point of view, the most appropriate constraint is dependent on the situation where the analogy is used or the metaphorical sentence is spoken. In case a metaphor appears in a context among other sentences[1], information about what topic is being talked of influences the mapping processes in the metaphorical sentence as a main constraint. And yet, note that we can comprehend some metaphorical sentences independently in a context-free situation, for example in isolation [2]. From the above discussions, we need to provide a mechanism of interpreting metaphors in a bottom-up way without pragmatic information.

[1] Gentner calls it "analogy in problem solving".
[2] Gentner calls it "analogy in isolation".

2.4 The overview of our approach

We provide a hierarchy of finite number of abstract primitive concepts, by use of which to represent the sematic structure of verbs. The characteristics of our approach is to first find an exact matching of primitive relations between the semantic representation of both concepts and then to determine in a structure-based way the entire correspondence which is consistent with the mappings derived from the "primitive matching". We expect that introducing the semantic constraint at the initial stage contributes to reducing the calculation cost of the whole structurally-based interpretation of metaphor.

3 Semantic Representation

3.1 Describing the meanings of verbs

"Causal relations" which are important in the constraint of pragmatic centrality could be regarded as higher-order primitive relations for describing the meanings of words. In this paper, we provide a set of first-order primitive relations and functions as well, by use of which to describe the meanings of verbs.

The meaning of "*A shoots down B*", for example, is described as in Fig.1. It is read as,

> *A 'makes' C 'transfer'* toward *B*, which *'causes' C* to *'touch' B*, which *'causes' 'power to operate from' C* to *B*, which further *'causes' 'the function of' B* to *'disappear'*, where *A 'possesses' C*.

The words written in slant style are examples of the primitives.

The characteristics of this representation is that the meaning is described as an assembly of abstract primitive relations, in which lower order primitives are connected to each other through higher order primitives. We also provide taxonomy of objects in which each of the objectives or the subjects of the verbs, 'A' through 'C', belongs to a certain class.

3.2 Classification of primitives

In logic programming [15], a world is represented by terms, functions and predicates. The primitives we have provided are the ones concerning functions and predicates. Primitives of predicates are classified into four; the ones representing a state of an object, a relation among several objects, an action or operation concerning an object and a higher-order relation of relations.

The set of primitives provided in this paper is shown in Appendix A. We do not claim that this is the result of enumerating all the possible primitives. Rather, we understand that this set is enough only for use in a tool to construct, as a first trial, a system that understands metaphors. We have already described the meanings of about two hundred verbs using only these primitives.

4 Finding a Target Concept

We argue the case of Example 1. Those verbs which are normally accompanied by *'human beings'* as a subject and *'opinion'* as an objective are specified as candidates of the target concept corresponding to *'shoot down'*. Out of the candidates, the one which yields the maximum number of mappings is selected as the target concept.

4.1 Mapping algorithm

The definition of mapping ϕ is a set of one-to-one correspondence between terms or functions or predicates. We use the following rules in finding mappings.

1. For atoms $\alpha = p(t_1, \ldots, t_n)$ and $\alpha' = p'(t'_1, \ldots, t'_n)$, and an already existing mapping ϕ,
   if $p = p'$, then revise the mapping ϕ as follows; $\phi = \phi \cup \{< p, p' >\} \cup \{< t_i, t'_i > | i = 1, \ldots, n\}$.
   Here, we say that α and α' can be identified through ϕ and we write $\alpha\phi\alpha'$.

2. For atoms $\alpha = p(t_1, \ldots, t_n)$ and $\alpha' = p'(t'_1, \ldots, t'_n)$, and an already existing mapping ϕ,
   if $\forall i < t_i, t'_i >\in \phi$, then revise the mapping ϕ as follows; $\phi = \phi \cup \{< p, p' >\}$.
   And also we write $\alpha\phi\alpha'$.

3. Let f, f' denote n-ary function symbols. For terms $t = f(t_1, \ldots, t_n)$ and $t' = f'(t'_1, \ldots, t'_n)$, $< t, t' >\in \phi$ is equivalent to $< f, f' >\in \phi \wedge \forall i < t_i, t'_i >\in \phi$.

4. For terms $t = f(t_1, \ldots, t_n)$ and $t' = f'(t'_1, \ldots, t'_n)$, and an already existing mapping ϕ,
   if $f = f'$, then revise the mapping ϕ as follows; $\phi = \phi \cup \{< f, f' >\} \cup \{< t_i, t'_i > | i = 1, \ldots, n\}$.

5. For terms $t = f(t_1, \ldots, t_n)$ and $t' = f'(t'_1, \ldots, t'_n)$, and an already existing mapping ϕ,
   if $\forall i < t_i, t'_i >\in \phi$, then revise the mapping ϕ as follows; $\phi = \phi \cup \{< f, f' >\}$.

6. For higher-order relations $hr = h(\alpha_1, \ldots, \alpha_n)$ and $hr' = h'(\alpha'_1, \ldots, \alpha'_n)$, and an already existing mapping ϕ,
   if $h = h'$ and $\exists i, \phi_i \subseteq \phi$, $\alpha_i\phi_i\alpha'_i$ holds and
   for all $j (1 \leq j \leq n)$ where $i \neq j$, $\exists \phi_j$ such that $\alpha_j\phi_j\alpha'_j$,
   and finally $\bigcup_j \phi_j \cup \phi_i$ is one-to-one mapping,

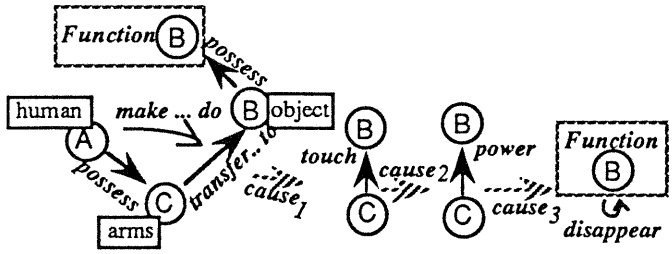


Fig.1 : Semantic representation of 'shoot down'

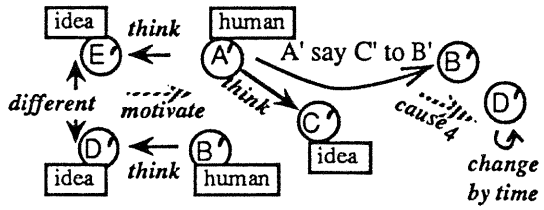


Fig.2 : Semantic representation of 'criticize'

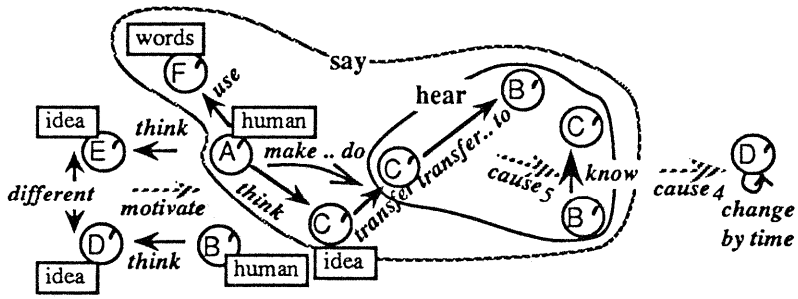


Fig.3 : Semantic representation of 'criticize' purely by primitives

then revise the mapping ϕ as follows;
$\phi = \phi \cup \{< h, h' >\} \cup \bigcup_j \phi_j$.
And we write $hr\phi hr'$.

In order to determine the whole mapping between the semantic representations of two verbs, we take three steps as follows;

1. **(Primitive Matching)**
   First, Rules 1, 3 and 4 are used. We obtain a mapping ϕ_1 by finding exact matchings of primitives of first-order predicate symbols p or function symbols f. If there exist several exact matchings which yield an inconsistent mapping as a whole in terms of the principle of one-to-one mapping, then we select only the matching which yields the maximum number of mapped elements.

2. **(Higher-Order Structure Mapping)**
   Secondly, Rule 6 is used. We identify those pairs of higher-order relations which give birth to a mapping that does not contradict ϕ_1. We denote the mapping obtained by this procedure as ϕ_2.

3. **(The Remaining Mappings)**
   Rules 2 and 5 are used for obtaining a mapping ϕ_3 of first-order predicate symbols and function symbols, based on the mapping $\phi_1 \cup \phi_2$.

The whole mapping between the two semantic structures is $\phi_{total} = \phi_1 \cup \phi_2 \cup \phi_3$.
The point is that we look for primitive matchings to obtain initial mappings, which will constraint the whole structure-based mapping process.

4.2 Example

Here, we show that the verb *criticize* forms a sufficient mapping as a counterpart of *shoot down*. The semantic representation of "*A' criticizes D'*" is shown in Fig.2. It is read as

> *A' 'thinks' E'* which is *'different'* from *D'* that *B' 'thinks'*, and this *'motivates' A'* to **say** *C'* to *B'*, where *C' 'is equivalent to'* that *D'* is *'wrong'*. This further *'causes' D'* to *'change by time'*.

We use the primitive *'change by time'*, superordinate to *'disappear'*, *'appear'*, *'change'* and *'remain still'*, because we do not know whether *B'* will come to think differently after being criticized. Note that we use the verb 'say' which does not belong to the set of primitives in our classification. We describe the meaning of "A says C to B" as follows;

> *A 'thinks' C* and *'makes' C 'transfer' 'using'* words *F* and this *'causes' B* to **hear** *C*.

And we describe the meaning of "B hears C" which is referred to in the above description as follows;

> *C 'transfers'* to *B* through *B*'s ear and this *'causes' B* to *'know' C*.

Even if we admit that the meaning of a verb is described referring to other non-primitive verbs, we can reduce it to a representation structured by only primitives after a finite number of substitutions. Figure 3 is a description of the primitive level-representation of *A' criticizes D'*.

Between the semantic representations of *shoot down* (Fig.1) and *criticize* (Fig.3), we obtain in the first step, primitive matching,

$\phi_1 = \{ <C,C'>, <B,B'>, <transfer, transfer> \}$.

The pairs of higher-order relations we can identify in the second step are

make(*A*, *transfer*(*C*,*B*)) vs.
make(*A'*, *transfer*(*C'*,*B'*))

$cause_3$(*make*(*A*,*transfer*(*C*,*B*)),
disappear(*function*(*B*))) vs.
$cause_4$(*make*(*A'*,*transfer*(*C'*,*B'*)),
change-by-time(*D'*))

and we obtain a mapping

ϕ_2 = { <*A*,*A'*>, <*make*,*make*>,
<$cause_3$, $cause_4$>, <*function*(*B*), *D'*>,
<*disappear*, *change-by-time*> }.

Further we obtain in the last step a mapping

ϕ_3 = { <*possess*, *think*> }.

5 Understanding and Learning

5.1 Deriving metaphorical relationships

For every correspondence $< a, a' > \in \phi$ (a, a' are terms or function symbols or predicate symbols), if $a \neq a'$ (In case a, a' are terms, if they belong to different entries in the taxonomy), a' is viewed as a.

The pairs in ϕ_{total} which meet the above condition are <*C*,*C'*>, <*B*,*B'*>, <*function*(*B*),*D'*>, <*disappear*, *change-by-time* >, <*possess*, *think*>. For each correspondence, we learn metaphorical implications included in the sentence as follows;

1. The idea *C'* can be viewed as arms *C*.
2. A human being *B'* can be viewed as object *B*.
3. When *B'* *'thinks'* *D'*, *D'* can be viewed as *'a function of' B'*.
4. *D'* can be rather viewed as having *'disappeared'*, instead of having *'changed by time'*.
5. *'Thinking'* can be viewed as *'possessing'*.

When we human beings say that we have understood metaphorical sentences, it normally means that we have gone through these implications. The viewings No. 1 and 2 constitute the basic views between the objects which are specific to this sentence. The viewing No.4 is an addition of implication resulting from using "shoot down" instead of using "criticize"; in general if a person has been criticized he/she may sometimes change his/her thinking or may remain thinking the same way. In case of Example 1, however, it is implied that the person may have changed his/her ideas.

The viewings No.3 and 5. are the very examples of mappings between abstract primitives in case of this metaphor. We claim that this type of mappings can be regarded as knowledge about "metaphorical relationships between concepts" which we human beings certainly have but cannot articulate when asked independently of the context of understanding metaphors. We also claim that this kind of knowledge may be a key to accessing to the "feeling about similarities" which is central to our thinking activity. Our belief is that metaphors are good evidences from which to elicit "feeling about similarities" efficiently.

5.2 Creating new meanings

The intention of a speaker's using metaphors may be to add to the target concept some implications which the target is not supposed to have in normal cases, by viewing the target as the source concept [4]. In metaphor interpretation by analogical approach, the addition of new meanings is done through analogically directed transfer from a source concept. In this paper, only higher-order relations are transfered, following Gentner's systematicity principle.

Let ϕ denotes the mappings between the target and the source. We define a new mapping ϕ_m as $\phi \cup \{< X, X > \mid X$ is an arbitray symbol $\}$, where a variable X will be used for transfering terms or functions or predicates which are not included in ϕ. The procedure of transfering is to create a new relation hr_t in the representation of the target concept such that $hr_s \phi_m hr_t$, where hr_s denotes a higher-order relation in the source.

In case of the example, the sets of hr_s is

$cause_1$(*transfer*(*C*,*B*), *touch*(*C*, *B*)),
$cause_2$(*touch*(*C*, *B*), *operation-power*(*C*, *B*)),
$cause_3$(*operation-power*(*C*, *B*),
disappear(*function*(*B*))).

For these relations, new hr_t's,

$cause_1$(*transfer*(*C'*,*B'*), *touch*(*C'*, *B'*)),
$cause_2$(*touch*(*C'*, *B'*), *operation-power*(*C'*, *B'*)),
$cause_4$(*operation-power*(*C'*, *B'*),
change-by-time(*D'*)),

are produced. These relations express as a whole the following knowledge;

C', the idea possessed by *A'*, transfers toward *B'*, *i.e.* *A'* says *C'* to *B'*, and as a result of this, *C'* touches *B'*, which causes a power to operate from *C'* to *B'* and also changes *B'*'s way of thinking.

This process corresponds to "new discoveries accompanied by understanding metaphors" [14]. The obtained knowledge can be regarded as commonsense concerning concepts, corresponding to Carbonell's *implicit intention component*[4]. In case of

the example, it expresses a commonsense that words sometimes have a strong power enough to put an impact to people's mind and make them change their ideas.

6 Related Works

The main difference from the past studies supporting *semantic similarity* constraint [5] [11] [3] is that we employ exact matchings of pre-provided primitives only at the first stage of finding the whole mapping.

In natural language processing approaches, several methods are proposed for finding a target which matches best semantically to the source. Fass [7] provides an abstraction hierarchy of concepts and asserts that the target should be a sibling of the source concept in the hierarchy. Obviously it may have the same disadvantage as Burstein's method [3] (*see* the second section). Russell [19] uses Conceptual Dependency primitives [20] as her underlying knowledge representation. Our approach is different from this representation method in two ways. One is about how to use primitives as a tool of representation. In our method a verb is represented as an assembly of abstract primitives, while in their method only one CD primitive is written in the "Action" slot as the major action of the verb. The other, a larger difference, is about the very semantic constraint in making mappings. We use the abstract primitives only for detecting their exact matchings between the analogues, while she uses pre-determined knowledge about semantic similarities between CD primitives for finding a target similar to a source. For example, the most plausible candidate of a target concept to a source concept "offer" must involve a MTRANS primitive, because the verb "offer" involves PTRANS and PTRANS is supposed to be semantically similar to MTRANS. This kind of knowledge is the very target knowledge we try to acquire in this paper, although what we should acquire is not similarities between primitives but metaphorical relationships.

Schank provides only the primitives about actions, which are approximately equivalent to our primitives about "Action or Operation" as a whole. On the other hand, we need to provide a comprehensive collection of primitives concerning not only actions but also functions, states, relations and higher-order relations, because we do not fill "Action" slot with one primitive, but represent verbs as an assembly of primitives.

The main difference from the Structure-Mapping Engine of Falkenheiner et.al. [6] is that we incorporate a semantic constraint, "primitive matching", at the initial stage of making correspondence. The constraint saves repetitions of backtracking in finding the candidates of mappable higher-order relations, because we can restrict our attention only on those higher-order relations having in their arguments at least one pair of first-order relations which are found to be identifiable by the already existing mapping ϕ resulting from the primitive matchings (see Rule 6 in the fourth section).

Deriving metaphorical relationships between concepts from metaphors is quite different from the kind of learning Martin does in his book [16]. He tries to learn new metaphors through the systematic extension, elaboration, and combination of already well-understood metaphors, which will in turn be applied directly as metaphorical knowledge to processing novel metaphors in future. Our trial is an elicitation of more basic metaphorical relationships between concepts underlying metaphorical sentences. They may be similar to the *view-links* Norvig [17] provides to his system as knowledge base for understanding texts. Norvig does not mention how *view-links* should be acquired.

Lastly, we discuss that providing a finite set of abstract primitives is essential to learning metaphorical relationships. First, the requirement that the number of primitives must be small enough is critical. If the meaning of a word is, as seen in an ordinary dictionary, constructed by referring to other words included in an entry of the dictionary whose meanings are also structured by referring to other words, it would cause the following disadvantages in finding appropriate mappings between a source and a target. For example, the meaning of *shoot down* would be represented as follows;

> The subject S fires an object A which S possesses toward something like a machine O, and this causes A to collide with O, which further causes A to destroy O.

The actions "fire", "collide with" and "destroy" are referred to. On the other hand, the meaning of *criticize* would be

> The subject S thinks that the idea of the other person O is wrong, which motivates S to say the opinion to O, and this influences O's idea.

The actions "say" and "influence" are referred to. The problem we encounter in this method of representation is that we cannot determine how far we have to trace down the referred verbs in making correspondence between the two analogues. The deeper we trace down to, the larger the whole stuctures of the analogues to be mapped with each other becomes, which implies that the calculation cost in making correspondence becomes tremendously huge.

Secondly, the abstractedness of primitives as a basis of knowledge representation is essential because it directly influences the generality of the ac-

quired knowledge about "metaphorical relationships between concepts". If we employ the above method of representation, we would obtain the following result of learning;

"Saying can be viewed as firing."

"Influencing can be viewed as destroying."

We can say that the result is too specific to be regarded as useful metaphorical knowledge.

7 Conclusion

We provided a set of abstract primitives by use of which to represent the meanings of verbs. Based on that semantic representation, metaphorical sentences are interpreted basically by making mappings in a structure-based way with a semantic constraint of "primitive matching" at the initial stage, and as a result of that, knowledge about metaphorical relatioships between concepts is acquired. The use of abstract primitives is essential in two ways. First, exact matchings of primitives in the semantic structures of the target and source concepts work as a strong constraint in determining the whole mapping of both concepts. Secondly, the generality of the acquired metaphorical knowledge comes from the abstractedness of primitives. We expect that this method will provide a way to accessing to "feeling about similarities" which we human beings already have but cannot easily articulate.

References

[1] J. A. Barnden. Belief, metaphorically speaking. In *Proceedings of 1st international conference on principles of knowledge representation and reasoning*, pp. 21–32, 1989.

[2] M. H. Burstein. A model of learning by incremental analogical reasoning and debugging. In *Proceedings of AAAI83*, 1983.

[3] M. H. Burstein. A model of learning by incremental analogical reasoning and debugging. In *Machine Learning: An Artifitial Intelligencd Approach (Vol.2)*. Morgan Kaufmann, 1986.

[4] J. G. Carbonell. Metaphor - A key to extensible semantic analysis. In *Proceedings of the Third Annual Conference of the Cognitive Science Society*, pp. 292–295, 1980.

[5] B. Falkenhainer, K. D. Forbus, and D. Gentner. The structure-mapping engine. In *Proceedings of AAAI86*, 1986.

[6] B. Falkenheiner, K. D. Forbus, and Gentner D. The structure-mapping engine: Algorithm and examples. *Artificial Intelligence*, Vol. 41, pp. 1–63, 1989.

[7] D. Fass. *Collative Semantics: A Semantics for Natural Language*. PhD thesis, New Mexico State University, New Mexico, 1988.

[8] D. Gentner. Structure mapping: A theoretical framework for analogy. *Cognitive Science*, Vol. 7, No. 2, pp. 155–170, 1983.

[9] D. Gentner. Analogical inference and analogical access. In *Analogica*. Morgan Kaufmann, 1988.

[10] G. S. Halford. A structural-mapping approach to cognitive development. *International Journal of Psychology*, Vol. 22, pp. 609–642, 1987.

[11] M. Haraguchi and S. Arikawa. A foundation of analogical reasoning and its realization. *Journal of Japanese Society for Artificial Intelligence*, Vol. 1, pp. 132–139, 1986.

[12] K. J. Holyoak and P. Thagard. Analogical mapping by constraint satisfaction. *Cognitive Science*, Vol. 13, pp. 295–355, 1989.

[13] S. Kedar-Cabelli. Toward a computational model of purpose-directed analogy. In *Analogica*. Morgan Kaufmann, 1988.

[14] G. Lakoff and M. Johnson. *Metaphors we live by*. University of Chicago Press, 1980.

[15] J. W. Lloyd. *Foundations of Logic Programming*. Springer-Verlag, 1984.

[16] J. H. Martin. *A Computational Model of Metaphor Interpretation*. Academic Press (Perspectives in Artificial Intelligence series), 1990.

[17] P. Norvig. *A unified theory of inference for text understanding*. PhD thesis, University of California, Berkeley, CA, 1986.

[18] A. Ortony. Beyond literal similarity. *Psychological Review*, Vol. 86, No. 3, pp. 161–180, 1979.

[19] S. W. Russell. Computer understanding of metaphorically used verbs. *American Journal of Computational Linguistics*, Vol. 44,, 1976.

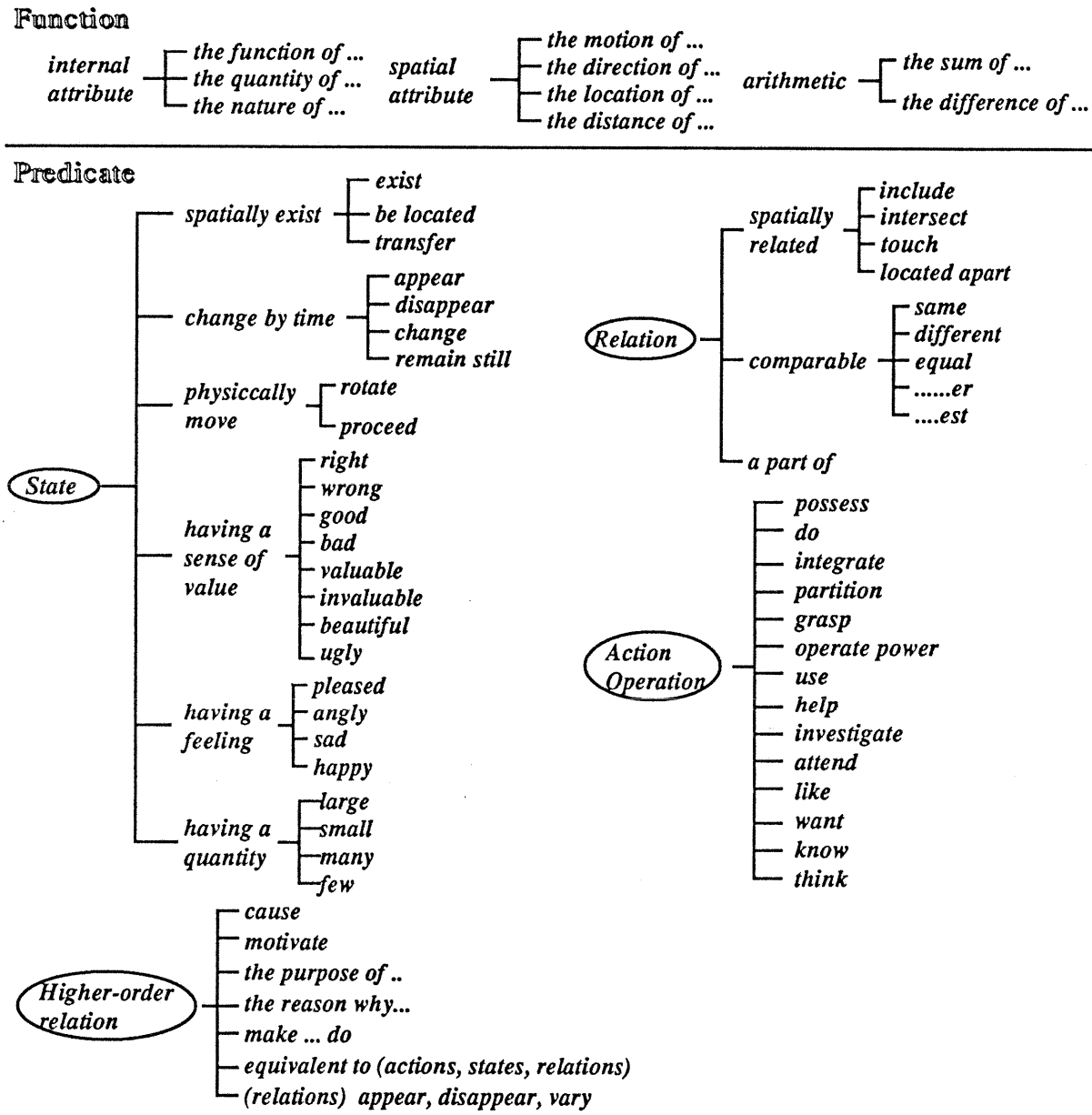
[20] R. C. Schank. Conceptual dependency: A theory of natural language understanding. *Cognitive Psychology*, Vol. 3, pp. 552–631, 1972.

[21] A. Tversky. Features of similarity. *Psychological Review*, Vol. 84, pp. 327–352, 1977.

[22] Y. Wilks. Making preferences more active. *Artificial Intelligence*, Vol. 11, pp. 197–223, 1978.

[23] P. Winston. Learning and reasoning by analogy. *Communications of ACM*, Vol. 23, No. 12, pp. 689–703, 1980.

Appendix A : A collection of primitives



Beyond Literal Meaning : Conversation Acts

David R. Traum*
Computer Science Department
University of Rochester
Rochester, New York 14627 USA
(716) 275-7230
traum@cs.rochester.edu

Elizabeth A. Hinkelman†
Center for Information and Language Studies
University of Chicago, 1100 East 57th St.
Chicago, IL. 60637 USA
(312) 702-8887
eliz@tira.uchicago.edu

July 3, 1991

Abstract

Theories of speech acts view dialogue as a series of *actions* which attempt to change the mental states of agents participating in a conversation. It is increasingly recognized that these actions serve to coordinate both the conversation itself and joint activities of the participants. This paper presents an approach to discourse modelling that focuses on these *conversation acts*, on their collaborative nature, and on the richness of their relationship to surface features. A speaker may, for instance, signal acceptance of a proposal with a word ("ok"), a proposition ("Sounds like a good idea"), or implicitly in a response that presupposes that acceptance. Conversation acts do not have literal meaning, since they involve speaker intentions, and yet they are necessary to the understanding of various linguistic structures. As such they downplay the role of the literal/nonliteral distinction, in favor of a much broader conception of meaning.

1 Introduction

The "literal meaning" of a sentence is a truth condition composed from the meanings of its constituent words; composed, on most accounts, using simple functions determined by sentence syntax without benefit of extrasentential information. Following Strawson, [Strawson, 1969] we hold that there can be no adequate truth-functional account of meaning in human language, unless it includes an analysis of the attitude or intention of the speaker. For Strawson, the argument is that in order to state the conditions under which a speaker could truthfully assert a proposition, it is necessary to know what it is to assert in the first place. We carry the argument much further. We take as our subject of study informal spoken communication, the core form of human language from which other forms of language are derived or abstracted. Having done so, we find that several problems afflict the notion of literal meaning, and reduce the value of the literal/nonliteral distinction.

The first is the difficulty of assigning literal meaning to certain frequent forms. These include lexical items such as "ok", "huh", "please", and "hello", and constructs such as sentence mood and formulaic social expressions. We approach the problem by relating these items to speaker intentions, as shown in Sect. 2.

The second problem is that the truth conditions of asserted (or otherwise presented) propositional material vary with the purposes of the conversation. For instance, the assertion that "I have a VCR" is valid when volunteering for a party, even if the volunteer has two VCR's. The assertion would be invalid

*This author was supported in part by the NSF under research grant no. IRI-9003841 and by ONR under research grant no. N00014-90-J-1811

†This author is an Ameritech Fellow

as a reply to a survey question on consumer electronics, where the exact count is desired. It will not do to address each such example by proliferating word senses. Nor will it do to restrict our attention to the so-called null context, which is in practice filled with unarticulated assumptions. We regard this problem as one of determining *grounding criteria* [Clark and Schaefer, 1989]. Grounding of belief is achieved when the hearer acknowledges the adequacy of the material for conversational purposes.

The third problem is that the process of establishing grounding is a collaborative, interactive one that requires negotiation and error recovery. We examine these processes in detail, since they are crucial to understanding what makes any conversation act successful. In this paper, we describe a model of conversation which views all conversation acts as collaborative, and which incorporates algorithms for recognizing speaker intentions. This leads to a new view of what it is to assert, and hence, of the role of the literal meaning distinction.

2 Linguistic Cues of Speaker Intentions

Hinkelman [Hinkelman, 1990] gives a method of speech act recognition that draws hypotheses about speaker intentions from patterns of features present at several levels of linguistic structure. The hypotheses are then verified against the larger semantic context of utterance. For instance, "Would you take a memo for me, please?" is marked as a directive act by the lexical "please", syntactic modal verb + aux-inversion construction, and the presence of the thematic role of beneficiary. The action being requested is a subconstituent of the utterance's logical form. By contrast, "It's cold in here" is recognized as a directive only by tone of voice. Its requested action is indicated, if at all, by inference using both the entire logical form and contextual factors. Still another act, the greeting "hello!", is recognized from lexical and contextual factors, with no logical form whatsoever.

The recognition process uses fine-grained production rules, whose left hand side specifies a pattern of linguistic features, and whose right hand side specifies a range of possible interpretations. The process is:

1. attempt to unify each LHS with utterance structure as this is built, and if it succeeds, build the corresponding partial speech act interpretations.
2. conjoin all such sets of interpretations, and find the most general unifier.
3. filter disjuncts for consistency with context, by checking agent beliefs related to their structure as plans. Here we leave the linguistic notion of compositionality for a more computational sense of the term.
4. apply extended reasoning, if it is necessary for planning purposes to further resolve the interpretation by further using contextual information.

The process thus uses a select subset of utterance features, which determine when elements of logical form enter into the interpretation. The role of logical form is reduced to building pieces of structure that may or may not be used. Logical form itself must already incorporate nonliteral constructs like metaphor. Thus, interpretations are given to certain lexical and syntactic constructs.

The model presented below fills several major remaining gaps. It provides interpretations for additional lexical items, as *grounding* and *turn-taking* acts (described in Section 3.3, below) which coordinate the dialogue itself. For example, a person who says "Oh, OK" communicates first comprehension ("Oh") and then acceptance ("OK") of another's presentation. It shows how grounding occurs as a collaborative activity across multiple utterances. For example, the above progression might also be communicated propositionally by a paragraph-length turn in the dialogue, or might be conveyed only as a presupposition, by elaborating that suggestion. Conversation acts, as we will see, emerge as speaker meaning in a much richer framework than that presumed by a notion of literal vs. nonliteral meaning.

3 Conversation Acts

Most prior computational speech act work has included the following simplifying assumptions:

1. Utterances are heard and understood correctly by the listener as they are uttered, and it is expected that they will be so understood.
2. Speech acts are single agent plans executed by the speaker. The listener is only passively present.
3. Each utterance encodes a single speech act.

In fact all of these assumptions are too strong to be able to handle many of the types of conversations people actually have:

1. Not only are utterances often misunderstood, conversation is structured in such a way as to take account of this phenomenon. Rather than just assuming that an utterance has been understood as soon as it has been said, this assumption is not made until some positive evidence is given by the listener (an acknowledgement) that he has understood. Some acknowledgements are made with explicit utterances (so called *backchannel responses* such as "okay", "right", "uh huh"), some by continuing with a next relevant response (e.g. a second part of an adjacency pair such as an answer to a question), and some by visual cues, such as head nodding, or continued eye contact. If some sort of evidence is not given, however, the speaker will assume that he has not made himself clear, and either try to clarify, or request some kind of acknowledgement (e.g. "did you get that?")
2. Since the traditional speech acts require at least an initial presentation by one agent and an acknowledgement of some form by another agent, they are inherently multi-agent actions. Rather than being formalized in a single agent logic, they must be part of a framework which includes multiple agents.
3. Each utterance can encode parts of several different acts. It can be a presentation part of one act as well as the acknowledgement part of another act. It can also contain turn-taking acts, and be a part of other relationships relating to larger scale discourse structures.

These problems are particularly acute in the study of task-oriented conversations. As part of the TRAINS Project [Allen and Schubert, 1991] a small corpus of spoken language task-oriented conversations have been collected and studied [Nakajima and Allen, 1991]. This corpus has been used to develop a speech act classification scheme based on intentions of the speaker. We have tentatively identified the following different levels of conversation acts, summarized in Table 1. Acts at each of these levels may be signalled by direct surface cues in the discourse.

3.1 The Core Speech Acts: DU Acts

We would like to keep as much of the previous analysis and work on speech acts as possible, while still relaxing the unwarranted assumptions described above. We maintain most of the traditional speech acts, such as **Inform**, **Request** and **Promise**, calling them *Core Speech Acts*. Instead of the traditional, indefensible assumption that these acts correspond to a single utterance, we posit a level of discourse structure which we call a **Discourse Unit** (DU), which is composed of the initial presentation and as many subsequent utterances by each party as are needed to make the act mutually understood (grounded). Typically, a DU will contain an initial presentation and an acknowledgement (which may be implicit in the next presentation), but it may also include any repairs which are needed. A DU corresponds more or less to a top level *Contribution*, in the terminology of [Clark and Schaefer, 1989].

Discourse Level	Act Type	Acts
Sub UU	Turn-taking	release-turn keep-turn assign-turn take-turn
UU	Grounding	Initiate Continue Ack Confirm Clarify ReqClar ReqAck
DU	Core Speech Acts	inform WHQ YNQ Acc Req Den Sug eval ReqPerm offer promise
Multiple DUs	Argumentation	Convince Summarize Find-Plan Elaborate

Table 1: The Conversation Acts

3.2 Argumentation Acts

We may build higher level conversation acts out of combinations of DU acts. We may, for instance, use an **inform** act in order to summarize or elaborate a prior position. We may use a combination of informs, and questions to convince another agent of something. We may even use a whole series of acts in order to build a plan, such as the top-level goal for the task oriented conversations in the TRAINS domain. The kinds of actions generally referred to as *Rhetorical Relations* take place at this level, as do many of the actions signalled by cue phrases.

3.3 Grounding Acts: UU Acts

An *Utterance Unit* (UU) is roughly defined as more or less continuous speech by the same speaker, punctuated by clause boundaries or prosodic cues. Each utterance corresponds to one *Grounding act* for each DU it is a part of. An Utterance Unit may also contain one or more turn-taking acts. Grounding acts include,

Initiate(DU-type) An initial utterance component of a Discourse Unit - traditionally this utterance alone has been considered sufficient to accomplish the core speech act. This also corresponds roughly to the first utterance in Clark and Schaeffer's presentation phase of a contribution [Clark and Schaefer, 1989].

Continue A continuation of the act begun with the **initiate** act. Part of a separate phonetic phrase, but syntactically and conceptually part of the same act.

Acknowledge Shows understanding of a prior utterance. May be either an explicit backchannel response (e.g. "okay", "right"), or implicit signalling of understanding, such as by proceeding with a second part of an adjacency pair.

Confirm Like an acknowledge, but actually repeating or paraphrasing part of a prior utterance.

Clarify Includes corrections, adds new material or changes material in previous utterances. It can change either the content or type of the current DU (e.g. a Tag question can change an inform into a question).

ReqClar Asks for a clarification by the other party. This is roughly equivalent to a *Next Turn Repair Initiator* [Schegloff *et al.*, 1977].

ReqAck Tries to get the other person to acknowledge the previous utterance.

3.4 Turn-taking Acts: Sub UU Acts

We hypothesize a series of low level acts to model the turn taking process [Sacks *et al.*, 1974]. The basic acts are **keep-turn**, **release-turn** (with a subvariant, **assign-turn**) and **take-turn**. Turn-taking acts are discussed further in Section 4. There may be several turn-taking acts in a single utterance. The start of an utterance may be a take-turn action (if another party initially had the turn), the main part of the utterance may be keeping the turn, and the end might release it.

3.5 Building a Discourse Unit from Utterance Units

A completed Discourse Unit is one in which the intent of the Initiator becomes mutually understood (or *grounded*) by the conversants. While there may be some confusion among the parties as to what role a particular utterance plays in a unit, whether a DU has been finished, or just what it would take to finish one, only certain patterns of actions are allowed. For instance, a speaker cannot acknowledge his own immediately prior utterance. He may utter something (e.g. "ok") which is often used to convey an acknowledgement, but this cannot be seen as an acknowledgement in this case. Often it will be seen as a request for acknowledgement by the other party.

Meanings of States

State	Relevant Context	Preferred Next
S		Initiate_I
1	Initiate_I	Ack_R
2	$\text{ReqClarify}_\text{R}$	Clarify_I
3	Clarify_R	Ack_I
4	$\text{ReqClarify}_\text{I}$	Clarify_R
F	Done	next DU

Table 2: Preferred Nexts of Discourse Unit States

We can identify at least six different possible states for a Discourse Unit to be in. These can be distinguished by their relevant context and what is preferred to follow, as shown in Table 2. State S represents a DU that has not been initiated yet, state F represents one that has been grounded, though we can always add on more, as in a further acknowledgement or some sort of repair. The other states represent DUs which still need one or more utterance acts to be grounded. State 1 represents the state in which all that is needed is an acknowledgement by the Responder, this is also the state the results immediately after an initiation. However, the Responder may also request a clarification, in which case we need a clarification by the Initiator before the responder acknowledges; this is State 2. The Responder may also clarify directly (state 3), in which case the Initiator needs to acknowledge this clarification. Similarly the Initiator may have problems with the Responder's clarification, and may request that the Responder clarify further; this would be state 4. [Traum, 1991] presents a finite automaton to track DUs through these states in a conversation, as well as ideas on how to implement a conversation system which can perform and recognize grounding acts.

Speaker	U#	Utterance	UU Act
		<long pause>	
M	087	system, why don't we uhh take uhh engine E-two TT KT KT	Initiate(1)
M	088	and go get tanker T-one KT	continue(1)
M	089	and bring it back to city D KT AT	continue(1)
S	090	okay TT RT	Ack(1)
		<short pause>	
M	091	and why don't we . use engine E-three .. to uhh TT KT	Initiate(2)
M	092	go to city I to get..get boxcar B-eight,	Continue(2)
M	093	go to city B to get tanker T-two KT	Continue(2)
M	094	go to city B to get tanker B-seven KT	Continue(2)
S	095	sorry, those are boxcars, you mean TT AT	ReqClar(2)
M	096	aaah I'm sorry, yes TT RT	Clarify(2)
M	097	I wanna get boxcar seven and eight and tanker T-two KT	Initiate(3)
		<short pause> RT	
S	098	okay TT	–
S	099	and tanker T-two at B KT RT	Confirm(3)
M	100*	yes TT	Ack(3)
S	101	yes TT RT	Ack(3)
M	102	and I would like to . bring TT RT	Initiate(4)
S	103*	use E-three for that TT RT	Confirm(2)
M	104	yes TT	Ack(2)
M	105	and then I would like to take those to uhhh city F KT KT	Initiate(5)
		<short pause> RT	
S	106	okay TT	Ack(5)

Table 3: Dialogue Fragment with Conversation Acts

DU#	DU Act	Initial U#	Final U#
1	Suggestion	087	090
2	Suggestion	091	104
3	Inform	097	101
4	Suggestion	102	–
5	Suggestion	105	106

Table 4: DU Acts from Dialogue Fragment

4 Examples of Conversation Acts

Table 3 presents a small conversation fragment from the TRAINS domain, annotated with examples of conversation acts. The goal of the TRAINS Project [Allen and Schubert, 1991] is to build an intelligent planning assistant that can communicate with a human manager in natural language to cooperatively construct and execute a plan to meet the manager's goal. The domain is transportation and manufacturing, with the execution being carried out by remote agents such as train engineers and factory operators. As a guide to the types of interactions such a system should be able to handle, a corpus of (spoken) task oriented conversations in this domain has been collected with a person playing the part of the system. Table 3 is a small excerpt taken from the TRAINS corpus (experiment 8, utterance units 87-104). This experiment requires the manager to get 100 tankerloads of beer to a particular city within three weeks time. The manager and the system are trying to form a plan to accomplish this. The transcription breaks the discourse into utterance units, numbered consecutively from the beginning of the dialogue. The entire problem takes 451 utterances (about 17 minutes), so this fragment is taken from near the beginning. After querying the system as to the available resources (beer already in warehouses, locations of beer factories, train cars, engines, and raw materials), the manager is now in the middle of formulating a plan to collect some of the train cars together.

The table shows the dialogue as well as some of the conversation acts which are performed. The table can be read as follows: the first column shows the speaker: *M* for manager, or *S* for system. The second column gives the number of the utterance, the third column the transcription of the utterance, and the last column the utterance act which is performed, with the number of the Discourse Unit of which it is a part (numbered in order of initiation from the beginning of this fragment). Utterance numbers appended with an asterisk indicate utterances which overlap temporally with the previous utterance, with the text lined up directly under the point in the previous utterance at which the overlap begins. Turn-taking acts are shown directly under the part of the utterance which signals this attempt. Turn-taking acts are labelled TT, for take-turn, KT for keep-turn, RT for release-turn, and AT for assign-turn. Table 4 shows the *core speech acts* which correspond to the DUs numbered in Table 3.

DU#5 exemplifies the fewest possible number of Grounding acts needed to complete a Discourse Unit: an initiation followed by an acknowledgement. On the other hand, DU#2 shows a moderately complicated one, with several continuers, a clarification request, and even an embedded inform act which further serves to clarify the suggestion. DU#4 is interrupted and never acknowledged, it is as if the suggestion has never been made. This forces the manager to start a new suggestion with DU#5.

The DUs in this fragment are also part of higher level conversation acts, though these are not shown in the table. The whole thing is part of a large action of finding a plan to satisfy the domain goal. At a smaller level, all of these suggestions are part of an action of formulating a plan to put this large train together which will later be used to ferry beer along. On a still smaller scale, DU#3, an inform act, is used to summarize the intentions of the suggestion in DU#2. Topic switching markers, such as the name address "System" in utterance 087, signal the start of a higher level conversation act, in this case consisting of the suggestions shown in Table 3 and the confirmational rechecking and acceptance which immediately follows the presented fragment.

Conversants can attempt turn-taking acts by any of several common speech patterns, but it will be a matter of negotiation as to whether the attempt succeeds. Other participants may also use plan recognition on seeing certain kinds of behavior to determine that the other party is attempting to perform a particular act, and may then facilitate it. For example, in utterance 102 in Table 3 the manager is speaking, and hears the system interrupt. The manager can deduce that the system is attempting a **take-turn** action, and stops talking, handing over the turn to the system.

Any instance of starting to talk can be seen as a take-turn attempt. We say that this attempt has succeeded when no one else talks at the same time (and attention is given to the speaker). It may be the case that someone else has the turn when the take-turn attempt is made. In this case, if the other party stops speaking, the attempt has been successful. If the new speaker stops shortly after starting, while the other party continues, we say that the take-turn action has failed, and a keep-turn action by the other party has succeeded. If both parties continue to talk, then neither has the turn, and both actions fail.

Similarly, any instance of continuing to talk can be seen as a keep-turn action. Certain sound patterns, such as "uhh", seem to carry no semantic content beyond keeping the turn (e.g. 087, 091).

Pauses generally release the turn. Certain pauses (for example the one between utterances 86 and 87 which begins the dialogue fragment in table 3) are marked by context as to who has the turn. Even here, an excessive pause can open up the possibility of a take-turn action by another conversant. Other release-turn actions can be signalled by intonation. Assign-turn actions are a subclass of release-turn in which a particular other agent is directed to speak next. A common form of this is a question directed at a particular individual.

5 Recognizing Conversation Acts

We now consider the recognition of discourse structure in this framework. Each utterance as it occurs is subjected to Hinkelman-style speech act recognition. This will identify many UU and DU acts directly, including signalled Ack, Confirm, ReqClar, ReqAck, and the initiations and continuations of core speech acts. At this point, a simple turn-taking calculation and UU tracking (as mentioned in Section 3.5) using the finite-state automaton occur. We discuss below the use of predictions from it.

Core speech acts ascend to the status of mutually believed events when their initiations are acknowledged. Therefore when the system performs such an initiation, it leaves the initiation's objective on its list of goals until such an acknowledgement is detected. The acknowledgement may be explicit, or preceeded by clarification dialogue. The hardest case to distinguish is when the acknowledgement is implicit in an elaboration; we must rely on a notion of semantic coherence to determine whether this is not a change in topic. When the system recognizes an initiation, it prefers to treat subsequent speaker utterances as part of that act until it has identified what is being said or a problem with understanding. In the former case it generates an acknowledgement, and in the latter a request for clarification.

Integration of predictions about DU and UU acts can proceed consistently with the [Hinkelman, 1990] model. It eliminates proposed interpretations if they conflicted with existing beliefs about speaker intentions, as a part of checking consistency with context. It would only be necessary to add the prediction to the hearer's beliefs before the recognition process began, to eliminate interpretations that would lead to inconsistency with the prediction. Such a model would be improved by the use of weighted evidence combination.

6 The Literal/Nonliteral Distinction

The theory of conversation acts provides interpretations for lexical and larger forms that had not been treated successfully as literal meaning. It provides for the grounding of mutual understanding according to the criteria of the particular conversation, although much more work remains to be done here. Finally, it provides an approach to the core speech acts that takes into account their collaborative nature.

The notion of meaning discussed here is broader than the traditional one of literal meaning, since it concerns the purposes of conversation and utterance. While it makes use of the classical logical form, it is so tightly interwoven with that form as to render the literal/nonliteral distinction irrelevant. We suggest that much more intellectual leverage is to be had by focus on the general notions of computability and speaker intentions than on the notions of literal and nonliteral meaning.

References

[Allen and Schubert, 1991] James F. Allen and Lenhart K. Schubert, "The TRAINS Project," TRAINS Technical Note 91-1, Computer Science Dept. University of Rochester, 1991.

[Clark and Schaefer, 1989] Herbert H. Clark and Edward F. Schaefer, "Contributing to Discourse," *Cognitive Science*, 13:259 - 94, 1989.

[Hinkelman, 1990] Elizabeth Hinkelman, *Linguisic and Pragmatic Constraints on Utterance Interpretation*, PhD thesis, University of Rochester, 1990.

[Nakajima and Allen, 1991] Shin-Ya Nakajima and James F. Allen, "A Study of Pragmatic roles of prosody in the TRAINS Dialogs," TRAINS Technical Note, Computer Science Dept. University of Rochester, forthcoming 1991.

[Sacks *et al.*, 1974] H. Sacks, E. A. Schegloff, and G. Jefferson, "A Simplest Systematics For the organization of Turn-Taking for Conversation," *Language*, 50:696–735, 1974.

[Schegloff *et al.*, 1977] E. A. Schegloff, G. Jefferson, and H. Sacks, "The Preference for Self Correction in the Organization of Repair in Conversation," *Language*, 53:361–382, 1977.

[Strawson, 1969] P F Strawson, "Meaning and Truth," In A. P. Martinich, editor, *The Philosophy of Language*. Oxford University Press, 1985.

[Traum, 1991] David R. Traum, "A Computational Theory of Grounding in Natural Language Conversation," Unpublished Thesis proposal, 1991.

Idioms, non-literal language and knowledge representation [1]

Erik-Jan van der Linden

Institute for Language Technology and AI
Tilburg University
PO Box 90153
5000 LE Tilburg
The Netherlands
E-mail: vdlinden@kub.nl

Abstract

This paper has two aims. Firstly, idioms are defined and located in the space of non-literal expressions. Secondly, the application of knowledge representation techniques in three different models for the representation and processing of idioms is discussed. The first, a symbolic procedural model extends the *two-level model* which was originally developed in computational morphology. The second is a simple *localist connectionist* model. The third, a symbolic hierarchical model, represents idioms as part of a lexicon conceived as an *inheritance hierarchy*. A comparison between the models is made in which the focus lies on the resolution of the ambiguity of idioms, the relation between the literal and non-literal interpretation and the syntactic flexibility of idiomatic expressions.

1 Introduction

Two issues are of importance in a computational theory of idioms. Firstly, a description of idioms should be provided (section 2). Definitions of idioms in the linguistic literature are not adequate, as will be argued here, since they define what idioms are *not*: a positive definition should be supplied, that defines idiomaticity as a property. Furthermore, idioms should be located in the space of non-literal expressions in order to understand why these expressions are non-literal. Secondly, models for the representation and processing of idioms should be provided. In section (3), three different models for the representation and processing of idioms will be presented, which use different KR techniques. The first extends the two-level model which was originally developed in computational morphology. The second is a simple localist connectionist model. [2] The third represents idioms in a lexicon that is modelled as an inheritance hierarchy. The focus in comparing the three models will be on the resolution of the ambiguity between the idiomatic and non-idiomatic interpretation of an idiom.
The present paper will concentrate upon aspects of idioms that relate to the division between literal and non-literal language, and syntactic flexibility. For a more elaborate discussion of other aspects like syntactic-semantic processing and elaboration on motivation and isomorphism, see van der Linden (in prep.).

[1] Thanks to Harry Bunt, Walter Daelemans, Koenraad De Smedt, Martin Everaert, Dirk Geeraerts, Wessel Kraaij, Michael Moortgat, Wietske Sijtsma, Ton van der Wouden, and anonymous reviewers for discussion and comments.

[2] The first two models are reported on in van der Linden and Kraaij (1990).

2 Idiomatic expressions and non-literal language

2.1 Definition

Within the linguistic literature 'Traditional wisdom dictates that an idiomatic expression is by definition a constituent or series of constituents where interpretation is not a compositional function of the interpretation of its parts.' (Gazdar et al. 1985, p. 327). [3] Three aspects of this definition deserve consideration.

non-literal expressions The sceptical tone of the quote is due to the fact that Gazdar et al. try to account for the meaning of idiomatic expressions under the principle of compositionality (see also Gibbs and Nayak 1989). Such analyses run, however, into considerable problems (van der Linden 1989; in prep.): idiomatic expressions should be regarded as expressions with non-compositional meaning. Here, non-compositional meanings will be considered to be non-literal meanings; although other opinions exist (Dascal 1987, and see section 2.2.2): idiomatic expressions are non-literal expressions.

positive definition The definition describes what idiomatic expressions are *not*: [4] their meaning can be subject to any other principle that describes in what way the meaning of an expression should be derived (contextuality, meaning postulates...). A definition that states what the meaning *is*, is preferable because, firstly, it makes stronger claims about the meaning of idiomatic expressions and, secondly, fits better in the general methodology of linguistics which is not to state negative facts, but supply positive definitions. [5]

Idioms or idiomaticity Idiomatic expressions do not form a homogeneous class (Wood 1986; Napoli 1988). Also, expressions that are no idioms *proper* may be partly idiomatic. A first example are collocations, that are idiomatic with respect to generation, but not with respect to analysis (Fillmore et al. 1988). If a language user merely knows the meaning of the words *school* and *whales*, he will be able to arrive at the interpretation of a group of fishes when encountering the expression *a school of whales* without knowledge of the collocation. However generating such an expression without this knowledge is not possible. A second example are constructions like *it is raining cats and dogs* in which *it is raining* can be assigned a literal interpretation, although the expression as a whole is idiomatic. Therefore it seems more fruitful to define a notion of *idiomaticity* as a property, and to apply this notion to separate expressions, parts of these expressions, or parts of the meanings of expressions, than to claim that a certain class should be described as idioms with an all-or-none property that distinguishes them from all other classes of expressions. *Idiomaticity*, then, **is a property of aspects of the meaning of complex (multi-lexemic) expressions which states that these aspects are exclusively a property of the whole expression. An *idiomatic expression* is a complex expression some aspect(s) of the meaning of which are subject to idiomaticity. An *idiom* is a complex expression all aspects of the meaning of which are subject to idiomaticity.**

2.2 Metaphorical properties of idioms

Idiomaticity does not imply *arbitrariness* of meaning. The metaphorical properties of idioms are important in this respect, but have mistakingly been taken as an argument in favour of the compositionality of the meaning of idioms.

---

[3] For this 'Traditional wisdom', see Hocket (1958); Fraser (1970); Heringer (1976); Chomsky (1980); Wood (1986); Di Sciullo and Williams (1987); Abeillé and Schabes (1989); Erbach (1991). Also: various papers in Everaert and van der Linden (1989).

[4] The same holds for the circumscription of non-literal language, as applied to idioms, in the call for papers of the present workshop: "devices (...) whose meaning cannot be obtained by direct composition of their constituent words."

[5] Wasow et al. 1983 give a comparable 'positive' circumscription "the idiomatic meaning is assigned to the whole phrase" (p. 110). See also Fillmore et al. 1988, p. 501, and compare Wilensky and Arens (1980): "... these constructs are phrasal in that the language user must know the meaning of the construct as a whole to use it correctly" (p. 117)

2.2.1 Motivation and isomorphism

Metaphors are general principles that link some domain to some target. [6] An example might be ANGER IS THE HEAT OF A FLUID IN A CONTAINER. Metaphors like this may be applied in several metaphorical expressions (1, taken from Lakoff (1987), p. 380-381) Metaphors may underlie basic (1c) and complex expressions (1a;b) .

(1) a. You make my *blood boil.*
    b. He's just *letting off steam.*
    c. He *exploded.*

Metaphorical properties of idioms can be described with the notions *motivation* and *isomorphism.* [7] Motivation and isomorphism are important notions because among other things the syntactic flexibility of idioms depends upon them (4).

Motivation Idioms are to a large extent frozen, conventionalized metaphorical expressions (or considered as such), and (part of) the conventional image underlying an idiom may result in the possibility of establishing a relation, a *motivating* link between the idiomatic interpretation of the idiom, and the non-idiomatic interpretation of the idiom (Lakoff 1987). [8]. For instance *blow the fuse* offers an image for loss of temper; *spill the beans* offers an image for making secret information public; *saw logs*, meaning *to be sound asleep*, can also be interpreted on the basis of a conventional metaphor. The relationship between the two is *motivated* just in case there are independently existing elements of the conceptual system that link the idiomatic and non-idiomatic meaning (Lakoff 1987, p. 451-2).

Isomorphism Not only may a relation exist between the non-idiomatic interpretation as a whole, and the idiomatic interpretation as a whole, it may also be the case that parts of the idiomatic and non-idiomatic interpretation maintain relations: an *isomorphism* may exist between the parts of the idiomatic and the non-idiomatic interpretation. For instance, in *blow the fuse* it is possible to find a part-to-part-correlation. *The beans* in *spill the beans* may refer to the information that is supposed to be kept secret. *Spill* refers to making that information public.
Idioms may be both motivated and isomorphic, motivated or isomorphic, or may be neither motivated nor isomorphic.

2.2.2 Motivation, isomorphism and literal meaning

Just because of the fact that parts of idioms may have metaphorical referents, it has been claimed that they should be considered literal expressions (Gazdar et al. 1985; Gibbs and Nayak 1989). Although parts of some idioms have identifiable meaning, this does *not* imply that the property of having this meaning is a property of the lexeme outside the idiom, and that the meaning of the idiom can be composed on the basis of these parts. The crucial point is that it is a property of the idiom as a whole whether the parts can be assigned metaphorical referents (and whether the idiom as a whole can be motivated). Thus although this might seem paradoxical at first sight, it is possible to enable distribution of meaning while adhering to the definition of idiomaticity.
Looking at dictionaries, one observes the same: idioms are listed in the entry of one (or more) of the content-words in the idiom as a *lexical unit.* The dictionary user does not find 'idiomatic meaning' of every content word leaving him to find out the meaning of the whole himself. The observation that idioms should be stored and accessed as lexical items and not from some special list that is distinct from the lexicon[9] (Swinney and Cutler 1979) arises from psycholinguistic research as well. Furthermore, idioms are stored as holistic entries in the mental lexicon (Swinney and Cutler 1979; Lancker

[6] Lakoff and Johnson (1980); for computational models of metaphorical language see for instance Martin (1989).

[7] The notions *isomorphism* and *motivation* stem from a lecture by Dirk Geeraerts (Leiden, November 7, 1990). The interpretation of these notions here differs somewhat from Geeraerts interpretation. In van der Linden (in prep.) motivation and isomorphism are compared with other notions and approaches throughout the literature (for instance Zernik 1987).

[8] More specific, this may be the case for idioms in which the referential meaning is subject to idiomaticity as opposed to idioms in which for instance pragmatic aspects of meaning are idiomatic: *how do you do?*

[9] As has been defended by Bobrow and Bell (1973).

and Canter 1980; Lancker, Canter and Terbeek 1981; cf. the notion 'configuration' in Cacciari and Tabossi 1988).

Given the definition of idiomaticity presented here, idioms differ from other non-literal expressions. In the case of idioms, non-literal meaning is represented within lexical entries, whereas in the case of other non-literal expressions (like metaphorical expressions, irony), meaning is derived on the basis of other information sources (like metaphorical principles (Martin 1989), Grice's maxims respectively). This is a second reason why idioms have been included under literal language ('moderate literalism'; Dascal 1987), a view that is not agreed with here since the meaning of idioms and other non-complex lexical elements are unpredictable on different grounds. Besides, the inclusion makes literal interpretation too important a design principle for natural language, whereas it is the conviction here that the existence of other principles is warranted.

3 Representation and processing of idioms

The issue concerning the representation and processing of idioms that will be concentrated upon in the models to come, will be the resolution of the ambiguity of idioms. The general approach to NLP here, is that the NL processor operates efficiently if it adopts an incremental mode of interpretation, and interprets input as immediate as possible (Thibadeau et al. 1982). Ambiguities are resolved on the basis of a best-first strategy. The question, then, is which possibility is the best one, and on the basis of what knowledge choices should be made.

3.1 Conventionality

A choice between the literal and non-literal reading of an idiom can be made using various kinds of linguistic information, but the claim here is that the mere fact that one of the analyses is idiomatic suffices. Besides, this choice does not have to be stipulated explicitely. Rather it follows naturally from the architecture of the lexicon and the retrieval process, provided an appropriate model of the lexicon is used.

Phrases consisting of idioms *can* in most cases be interpreted non-idiomatically as well. Very rarely, however, an idiomatic phrase *should* in fact be interpreted non-idiomatically (Koller 1977, p. 13; Chafe 1968, p. 123; Gross 1984, p. 278; Swinney 1981, p. 208). Psycholinguistic research indicates that there is clear preference for the idiomatic reading (Gibbs 1980; Schweigert and Moates 1988). We will refer to the fact that phrases should be interpreted according to the lexical, non-literal meaning, as the 'conventionality' principle. If this principle could be modeled in an appropriate way, this would provide a heuristic that would render the interpretation process more efficient since other than lexical knowledge is not nessecary for the resolution of ambiguities. So, the resolution of the ambiguity occurs as soon as the idiom has been encountered in the input.

When can and does an incremental processor *start* looking for idioms? Psycholinguistic research indicates that idioms are not activated when the 'first' (content) word is encountered (Swinney and Cutler 1979). There is, from the computational point of view, no need to start 'looking' for idioms, when only the first word has been found since this would only result in increase of the processing load at higher levels. In Stock's (1989) approach to ambiguity resolution the idiomatic and the non-idiomatic analysis are processed in parallel. An external scheduling function gives priority to one of these analyses. Also, the disambiguation process already starts when the 'first' word has been encountered. As we have stated, this increases the load on higher processes.

3.2 An extension of the notion *continuation class*

Lexical representation Lexical entries in two-level morphology are represented in a trie structure, which enables incremental lookup of strings. A lexical entry consists of a lexical representation, linguistic information, and a so-called continuation class, which is a list of sublexicons "the members of which may follow" (Koskenniemi 1983, p. 29) the lexical entry. In the continuation class of an adjective, one could, for instance, find a reference to a sublexicon containing comparative endings (ibid. p. 57). An

obvious extension is to apply this notion beyond the boundaries of the word. A continuation class of an entry **A** could contain references to the entries that form an idiom with **A**. An example is (2a).

Algorithm A simple algorithm is used to retrieve idioms (in (2b) the relevant fragment of the algorithm is represented in pseudocode). The result of the application of the algorithm is that linguistic information associated with the idioms is supplied to the syntactic/semantic processor. The linguistic information includes the precise form of the idiom, the possibilities for modification etc. Note that conventionality is modeled explicitely. [10]

(2)

(a)

```
k-i-c-k*---b-u-c-k-e-t*
          h-a-b-i-t*
           \e-e-l-s*
```

(b)

```
DO read a letter
   IF word has been found THEN
     IF this word forms a lexical item
        with previous word(s)
     THEN make its information
          available to syn/sem process
     ELSE make word information
          available to syn/sem process
UNTIL no more letters in input.
```

3.3 A connectionist model

The second model we present here is an extension of Cottrell's (1988) localist connectionist model for the resolution of lexical ambiguity. The model consists of four levels. Units at the lowest level represent the smallest units of form. These units activate units on the level that represents syntactic discriminations, which in turn activate units on the semantic level. The semantic features activate relational nodes in the semantic network. *Within* levels, inhibitory links may occur; *between* levels excitatory links may exist, however, there are no inhibitory links within the semantic network. The meaning of idioms is represented as all other relational nodes in the semantic network. On the level of semantic features, the idiom is represented by a unit that has a *gate* function similar to so-called *SIGMA-PI units* (Rumelhart and McClelland 1986, p. 73): in order for such a unit (**A**) to receive activation, all units activating **A** bottom-up should be active. If one of the units connected to a unit **A** is not active, **A** does not receive activation. Thus when the first word of an idiom is encountered, the idiom is not activated, because the other word(s) is (are) *not* activated. However, once *all* relevant lexemes have been encountered in the input, it becomes active. Note that an external syntactic module activates one of the nodes in case of syntactic ambiguity. Since there is more than one syntactic unit activating the idiom, the overall activation of the idiom becomes higher than competing nodes representing non-idiomatic meanings. The idiom is the strongest competitor, and inhibits the non-idiomatic readings. The conventionality principle is thus modeled as a natural consequence of the architecture of the model. [11]

3.4 Idioms in an inheritance hierarchy

Inheritance mechanisms are becoming increasingly important in the study of natural language processing. [12] A lexicon modeled as an inheritance hierarchy allows for the stipulation of general principles on high and abstract levels of representation, and therefore avoids the stipulation of redundant information. An idiom and its verbal head (*kick* in the case of *kick the bucket*) can be said to maintain an inheritance relation: the idiom inherits part of its properties from its head.

[10] A toy implementation of the lexicon structure and the algorithm has been made in C.

[11] A more detailed description of this model can be found in van der Linden and Kraaij (1990). The model has been implemented in C with the use of the Rochester Connectionist Simulator (Goddard et al. 1989) by Wessel Kraaij.

[12] See Daelemans and Gazdar (1990) for recent research and references.

Syntactic information Idioms can be represented as functor-argument structures [13] and have the same format as the verbs that are their heads. It is therefore possible to relate the syntactic category of the idiom to that of its head. The information that the object argument is specified for a certain string, can be added monotonically. The verb itself does not specify a string value for the argument; the idiom is a specialization of the verb because it does specify a string value for its argument. So, the relation between verb and idiom could be specified as KICK $\succ$ KICK_THE_BUCKET, where KICK and KICK_THE_BUCKET are represented as in (3): $\succ$ denotes an inheritance relation.

(3) KICK: $< (np\backslash s)/np >$
KICK_THE_BUCKET: KICK $\cup$
$prosody(argument(syntax(KICK\_THE\_BUCKET))) \approx [THE, BUCKET]$

Semantics It follows from the definition of idioms that the meaning of the idiom cannot be inherited from the verb that is its head, but should be added non-monotonically. Bouma (1990) defines *default unification* in order to model non-monotonic inheritance for feature-value structures. In (4) the representation of the semantics of *kick the bucket* is presented. Note that semantics is represented as non-default information (denoted with !).

(4) b. KICK: $< (np\backslash s)/np, kick(x)(y) >$
c. KICK_THE_BUCKET:
KICK $\cup$
$prosody(argument(syntax(KICK\_THE\_BUCKET))) \approx [THE, BUCKET] \wedge$
$!semantics(KICK\_THE\_BUCKET) \approx die(y)$

As in the model of the lexicon proposed by Zernik and Dyer (1987), the model proposed here features a lexicalist theory (categorial grammar) and therefore puts the syntactic and semantic burden on the lexicon. Also, Zernik and Dyer relate idioms to their heads. Flickinger (1987) presents a hierarchical structure of the lexicon, but does not include idiomatic expressions.

3.4.1 Conventionality and blocking

Ambiguous lexical items are linked in the lexicon by means of an inheritance relation ($\succ$), or a Boolean relation ($\wedge$). The inheritance relation can be exploited to model conventionality with the application of the notion *blocking*. Blocking is "the nonoccurrence of one form due to the simple existence of another" (Aronoff 1976, p. 41). In morphology it is an absolute principle. For instance the existence of the nominal derivation * *graciosity* of *gracious* is blocked by *grace*. Daelemans (1987) and De Smedt (1990) show that in a hierarchical lexicon structure, blocking is equivalent to the prevalence of more specific information lower on in the hierarchy over more general information.
Blocking can be applied as a preference in the case of ambiguity if the different interpretations of the item are ordered. For instance, KICK and KICK_THE_BUCKET are ordered by means of inheritance: the idiom frame is more specific than the frame without the argument. When (and only when) all the relevant lexical material of the idiom has been encountered in the input, the expression as a whole is preferably considered an idiom. Conventionality is thus a natural property of the use of the specialization relations in the hierarchy.
It is of course possible to order the different frames a verb has in some linear order, and to stipulate that the first should always be preferred. However, this order must be stipulated, whereas in the case of the hierarchical lexicon structure presented here, the relation between the categories is linguistically motivated. Frequency of occurence, that is, giving forms with higher frequency prevalence over those with lower frequency, is not an alternative: more specific forms do not necessarily appear more frequently than the forms they inherit from.

[13] See van der Linden (in prep.) and similar representations in Abeillé (1990), Abeillé and Schabes (1989) and Erbach (1991).

3.5 Comparison of the models

The three models presented here are all able to model the conventionality principle. There are, however, a number of differences between them, that can be used to evaluate them.

- In the two-level model and the connectionist model, in the case of ambiguity the simplest hypothesis that covers the largest part of the input is preferred, and it is assumed that the largest part also constitutes the conventional interpretation. Although this is mostly the case, it does not necessarily have to be so. In the hierarchical model, conventionality is modeled by means of the specialization relation. Specialization seems to be more closely related to conventionality. In PHRAN (Wilensky and Arens 1980), specificity only plays a role in suggesting patterns that match the input, but evaluation takes place on the basis of *length*, and *order* of the patterns. Zernik and Dyer (1987) do not discuss ambiguity.
- In the two-level model conventionality has to be modeled explicitely, whereas in the other two models, it follows naturally from the architecture of the lexicon and the retrieval process.
- The hierarchical model is linguistically motivated, whereas the other models are merely models of the lexical retrieval process.
- The hierarchical model gives a less redundant representation of linguistic information. The other two models, however, could be extended with a hierarchical structure for the representation of syntactic and semantic information.
- A disadvantage of the connectionist model is the necessity for parallel processing: in the hierarchical model most processing takes place in serial order, and it therefore demands smaller processing capacity.

On the basis of these considerations, the hierarchical model seems to be the better of the three.

4 Syntactic flexibility

Idioms seem to deviate from their literal counterparts, with respect to the syntactic constructions idioms can occur in. For instance (5) does not have an idiomatic interpretation.

(5) # The bucket was kicked by John.

Most research on the flexibility of idioms has been devoted to explations for this deviation, without firstly assessing the extent to which idioms differ from non-idiomatic expressions. The point to be made here is that for a considerable part idioms do *not* deviate from their literal counterparts: the syntactic flexibility of idioms can for a considerable part be explained in terms of properties of its verbal head, and this behavior can best be explained if the idiom is said to inherit these properties from its head. This thus supplies an argument in favour of a hierarchical model of the lexicon. In order to illustrate this, the *passive* will be considered in detail here: non-passivizability of a large group of idioms can be explained in terms of properties of its verbal head.

4.1 Passive

Only transitive verbs occur in passive constructions (Bach 1980). [14] Bach mentions a number of classes in which verbs occur that seem to be transitive, but that are in fact complex intransitives, and therefore do not passivize. This classification seems to apply as well to idioms and explains why these do not passivize. A first rather trivial class are idioms which are already in passive form.

(6) Van de aarde weggenomen worden.
From the earth away_taken to_be.
To be dying.

[14] For passivization of Dutch intransitives, see van der Linden (in prep.)

If the object of an idioms is a lexical reflexive, passivization is not possible. Reflexivity includes reflexive pronouns and inalienable objects.

(7) Zijn beste beentje voorzetten.
His best leg_[dim] in_front_to_put.
Put one's best foot forward.

If the object of a verb is a lexically stipulated expletive pronouns, passivization is not possible.

(8) Hij zal het niet lang meer maken.
He will it not longer again make.
He will soon die.

The same applies to subjects.

(9) Het loopt af met hem.
It comes to_an_end with him.
He is dying.

Bach mentions a group of verbs that have objects that are no 'true' object NP's. Examples are predicative or copulative verbs, or verbs like *wegen* (*to weigh*) or *spelen* (*to act*).

(10) Hij speelt stommetje.
He plays dumb_[dim].
He keeps his mouth shut.

Verbs of possession are not transitive either

(11) Een bord voor de kop hebben.
A sign in_front_of the head to_have.
To be thick-skinned.

Although for a number of idioms it can thus be argued that the verb it comprises disables passivization, there is still a group of idioms for which this explanation does not do. In van der Linden (in prep.) it is shown that if these idioms do not passivize, this should still be attributed to them being intransitive, but that intransitivity is not inherited from the verbal head, but is a consequence of lack of underlying metaphorical properties: idioms that are neither isomorphic, nor motivated loose their transitivity and cannot be passivized.

Concluding remarks There is a large group of idioms, the non-passivizability of which should be accounted for in terms of the non-transitivity of the verb that is the head of the expression. The most natural way to represent this, is by means of inheritance: the idioms inherits certain properties from its verbal head, which determine its syntactic flexibility.

5 Concluding remarks

Idioms have a non-literal interpretation that is lexically represented as a property of the expression as a whole, where parts of the expression may have metaphorical referents. As a model of the representation and processing of these expressions, a lexicon structure that is considered as an inheritance hierarchy seems the most viable, at least when the resolution of ambiguity and syntactic flexibility are concerned. When issues outside the scope of this paper are taken into consideration, the comparison becomes slightly different.

- Subsymbolic approaches can model more easily the *interactive* nature of natural language processing.

- With respect to *learning*, here learning idioms, it is clear from recent work in AI and cognitive psychology that distributed subsymbolic representations are promising. Algorithms for learning hierarchical structures exists. An underlying principle of inheritance, structure sharing, goes well with such distributed representations: inheritance hierarchies could be considered a linguistically sufficient generalization of an underlying subsymbolic representation. The symbolic model proposed by Zernik and Dyer (1987) for learning idioms only works in case of detection of a gap in lexical knowledge: bootstrapping in case of an empty lexicon is not possible.
- Upon failure of the principle of conventionality (in the end it is a heuristic) the hierarchical model provides an easy way to model *backtracking* by means of the choice of a different node in the hierarchical structure.

The fact that it is easy to model a principle of conventionality, could render the interpretation process of other forms of non-literal language efficient, and it is therefore worth to examine the scope of the principle.

References

Abeillé, A. (1990), 'Lexical and syntactic rules in a tree adjoining grammar', in: *Proceedings of the ACL 1990*, pp. 292-298.

Abeillé, A. and Y. Schabes (1989), 'Parsing idioms in Lexicalized TAGs', in: *Proceedings of EACL 1989*, pp. 1-9.

Aronoff, M. (1976), *Word formation in generative grammar*, The MIT Press, Cambridge, Massechusetts.

Bobrow, S. and S.Bell, (1973), 'On catching on to idiomatic expressions', *Memory and Cognition* 3, pp. 343-346. Bouma, G. (1990), 'Defaults in unification grammar', in: *Proceedings of ACL 1990*, pp. 165-172.

Cacciari, C. and P. Tabossi (1988), 'The comprehension of idioms', *Journal of Memory and Language*27, pp. 668-683.

Cottrell, G. (1989), ' A model of lexical access of ambiguous words'. In: Small, S., G. Cottrel, G., and M. Tanenhaus, M., 1988, p. 179-194.

Daelemans, W. (1987) *Studies in language technology: an object-oriented model of morphophonological aspects of Dutch*, PhD-thesis, Univesity of Leuven.

Daelemans, W. and G. Gazdar (1990) *Inheritance in natural language processing*, workshop proceedings, Tilburg, ITK.

Dascal, M. (1987), 'Defending literal meaning', in: *Cognitive Science* 11, pp. 259-281.

De Smedt, K. (1990), *Incremental sentence generation*, PhD-thesis, University of Nijmegen.

Di Sciullo A. and Williams (1987), *On the definition of word*, MIT press, Cambridge, Massachusetts.

Erbach, G. (1991), 'Lexical representation of idioms', IWBS report 169, IBM TR-80.91-023, IBM, Germany.

Everaert, M. and van der Linden, E. (Eds.) (1989) *Proceedings of the First Tilburg Workshop on Idioms*, Institute for Language Technology and AI.

Fillmore Ch. , P. Kay and M. O'Connor (1988), 'Regularity and idiomaticity in grammatical constructions: the case of *let alone*', *Language* 64, pp. 510-538.

Flickinger, D. (1987) *Lexical rules in the hierarchical lexicon*, PhD-thesis, Stanford University.

Fraser, B. (1970), 'Idioms within a transformational grammar'. *Foundations of language* 6, pp. 22-42.

Gazdar, G., E.Klein, G.Pullum and I.Sag (1985) *Generalized Phrase Structure Grammar*. Basil Blackwell, Oxford.

Gibbs, R. (1980), 'Spilling the beans on understanding and memory for idioms in conversation'. *Memory and Cognition* 8, pp. 149-156.

Gibbs, R., and N. Nayak (1989), 'Psycholinguistic studies on the syntactic behavior of idioms', *Cognitive Psychology* 21, pp. 100-138.

Goddard, N, K. Lynne, T. Mintz, and L. Bukys (1989), 'Rochester connectionist simulator'. Technical Report. University of Rochester.

Gross, M. (1984), 'Lexicon-grammar and the syntactic analysis of French', In: Proceedings COLING '84, pp. 275-282.

Heringer, J. (1976), 'Idioms and lexicalization in English', In: M. Shibatani (ed.), *Syntax and Semantics 6: The grammar of causative constructions*, Academic Press, New York, pp. 250-216.
Hocket, Ch. (1958), *A course in modern linguistics*, Macmillan, New York.
Koller, W. (1977), *Redensarten: linguistische Aspekte, Vorkommensanalysen, Sprachspiel.* Tübingen, Niemeyer.
Koskenniemi, K. (1983), *Two-level morphology*, PhD-thesis, University of Helsinki.
Lakoff, G. (1987), *Women, fire and dangerous things*, Chicago University Press, Chicago.
Lakoff, G., and M. Johnson (1980), *Metaphors we live by.* Chicago, University Press, Chicago.
Lancker, D. van and G.Canter (1981), 'Disambiguation of ditropic sentences: acoustic and phonetic cues'. *Journal of Speech and Hearing Research* 24, pp. 64-69.
Lancker, D. van, G.Canter and D.Terbeek (1981), 'Disambiguation of ditropic sentences: acoustic and phonetic cues'. *Journal of Speech and Hearing Research* 24, pp. 330-335.
Van der Linden, E. (1989) 'Idioms and flexible categorial grammar', in: Everaert and van der Linden (1989), pp. 127-143.
Van der Linden, E. (in prep.) 'A categorial, computational theory of idioms', PhD-thesis, Tilburg University.
Van der Linden, E. and Kraaij, W. 1990, Ambiguity resolution and the retrieval of idioms: two approaches. In: *Proceedings of COLING 1990*, vol 2 , pp. 245-251.
Martin, (1989), 'Representing and acquiring metaphor-based polysemy', In: U. Zernik (Ed.) *Proceedings of the First International Lexical Acquisition Workshop*, August 21, 1989, Detroit, Michigan.
Napoli, D. (1988), 'Subjects and external arguments, clauses and non-clauses'. *Linguistics and Philosophy* 11, pp. 323-345.
Rummelhart, D. and J. McClelland (1986), *Parallel Distributed processing. Explorations in the microstucture of cognition.* Volume 1: *Foundations*; Volume 2: *psychological and biological models*, MIT press, Cambridge, Massachusetts.
Schweigert, W. (1986), 'The comprehension of familiar and less familiar idioms', *Journal of Psycholinguistic Research* 15, pp. 33-45.
Stock, O. (1989) 'Parsing with flexibility, dynamic strategies, and idioms in mind'. *Computational Linguistics* 15, 1, pp. 1- 19.
Swinney, D. (1981), 'Lexical processing during sentence comprehension: effects of higher order constraints and implications for representation'. In: T. Meyers, J. Laver and J. Anderson (eds.) *The cognitive representation of speech*, North-Holland.
Swiney, D. and A.Cutler (1979), 'The access and processing of idiomatic expressions'. *Journal of Verbal Learning and Verbal Behavior* 18, pp. 523-534.
Thibadeau, R., M. Just and P. Carpenter (1982), 'A model of the time course and content of reading', *Cognitive Science* 6, pp. 157-203.
Wasow, T, I.Sag and G.Nunberg (1983), 'Idioms: an interim report'. In: Shiro Hattori and Kazuko Inoue (Eds.), *Proceedings of the XIIIth international congress of linguists.* Tokyo, CIPL, pp. 102-115.
Wilensky, R., and Y. Arens, (1980), PHRAN, A knowledge-based natural language understander, In: *Proceedings of the ACL 1980*, University of Pennsylvania, Philedelphia, Pennsylvania.
Wood, M. McGee (1986), *A definition of idiom.* Masters thesis, University of Manchester (1981). Reproduced by the Indiana University Linguistics Club.
Zernik, U. (1987), *Strategies in language acquisitions: learning phrases from examples in context*, PhD-thesis, UCLA, Los Angeles.
Zernik, U. and M. Dyer, (1987), *The self-extending phrasal lexicon*, *Computational Linguistics*, 13, p. 308-327.

A connectionist model of literal and figurative adjective noun combinations

Susan H. Weber
International Computer Science Institute,
1947 Center St,. Ste. 600, Berkeley CA 94704
tel: 415-643-9153 fax: 415-643-7684
weber@icsi.berkeley.edu

July 2, 1991

Abstract

This paper describes a connectionist model of metaphor in adjective noun combinations that features context sensitive lexical semantics, direct inferences, property abstraction and scalar value transference as interpretive mechanisms for novel metaphoric usages, and a frequency based word sense acquisition mechanism common to both literal and figurative usages.

1 Introduction

Despite the traditional treatment of metaphor as an anomalous semantic form [Richards 1937, Black 1979], models that place metaphor on the same footing as 'literal' semantics are gaining widespread acceptance [Lakoff 1987, Martin 1988, Gildea & Glucksberg 1986]. In the domain of adjective-noun combinations, examples abound of metaphoric word senses with comprehension times comparable to literal senses. A case in point is the word sense of 'light' meaning low calorie. Traditional models hold that the phrase "light beer" is initially interpreted as a low weight (or pale colored) beer and only upon rejection of this initial 'literal' interpretation is the correct word sense arrived at. The problem with this account is that the 'literal' meaning quite often makes sense (eg. light-colored beer) so no incentive apparently exists to retrieve the contextually appropriate but more fanciful word sense. A more robust model of metaphor processing uses context sensitive adjective semantics [Barsalou 1982] so that the meaning of 'light' when used to describe typically calorific food items is immediately assumed to mean low calorie, with no intervening 'literal' interpretive steps.

The distinction between 'literal' and 'metaphoric' word senses, it is argued here, should be based not on processing but on context sensitivity. The word senses that come to mind on hearing the term in isolation (eg. when asked to define the term) are the 'literal' meanings. The word senses that either come to mind only in specific contexts or carry with them some analogical mapping baggage (eg. light beer, green recruit) are the 'figurative' meanings. There is psycholinguistic evidence [Tabossi 1988] that since both literal and figurative meanings display the same comprehension timings, they are processed with the same mechanisms. Since selecting the more appropriate of two literal word senses (eg. light color vs. light weight) is also a context sensitive process, distinguishing between literal and metaphoric usages requires considering either the default (context free) semantics, or the historical derivation of the term, or both. The term 'light', for example, is not only used far more commonly to refer to color and weight than to calorie content (if only because foods comprise only a small fraction of all objects with color and weight) but the calorie word sense is also a relatively recent extension to the word's meaning. The two points are related if word senses are acquired by compiling frequency statistics on apparent semantics of use, since more common usages would thus be acquired first.

So what determines the "apparent semantics of use"? If the context sensitive semantics of a term is a result of its observed meanings in various contexts, how does one interpret the semantics of an unfamilar or novel expression? The context sensitive versus context free distinction applies only to known meanings of a term. If an adjective is used in an unfamiliar and apparently inappropriate context (eg. green idea), pragmatics demand that there be some reasonable interpretation of the phrase's meaning. When the usual context sensitive semantic retrieval process fails to supply any interpretations, there must be a supplemental metaphoric meaning interpretation mechanism available to suggest some possibilities. The proposal is to exploit the term's known meanings in other contexts to suggest meanings in the novel context. For adjective noun combinations, the adjective can be considered to propose feature values for the salient features of the object denoted by the noun.

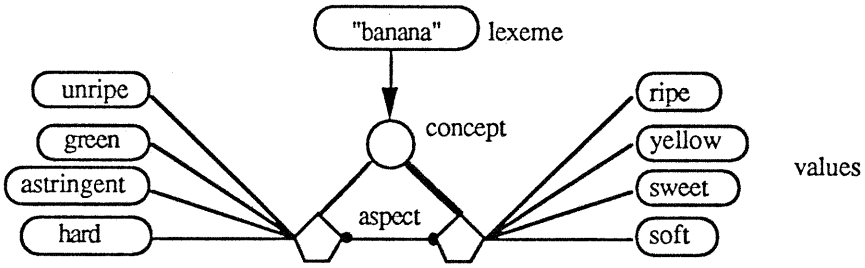


Figure 1: Approximate structure of the "banana" concept, showing two relevant aspects, one for ripe bananas (the default) and one for unripe. Heavier lines are used to denote preferentially weighted links.

The process of semantic acquisition and interpretation is bootstrapped in experience: the literal meanings of perceptual property terms are learned first, thus supplying a basis for developing abstract features and metaphoric meanings. Each subsequent novel use is interpreted in the light of all previously encountered expressions, so that the semantics of one metaphor can be based on another. For example, the weight and color word senses of 'light' would be acquired first, by observing the correlation between the lexeme 'light' and certain relative weights and degrees of color saturation. The meaning in the context of beer can be derived by a simple scalar mapping process, from low weight and low saturation to low calorie. With sufficient repetition this semantic interpretation, initially derived by a simple analogical mapping from previously acquired word senses, is itself catalogued in the dictionary and becomes a figurative word sense in its own right. An even more elaborate example can be devised for the semantics of 'green': the color word sense of 'green' is acquired first, by observing the correlation between the lexeme and a certain color hue. Later, the tendency of unripe fruit to be green in color may (with sufficient reinforcement, such as consistent metonymic references to unripeness) become a word sense proper. The interpretation of conventional metaphoric uses, such as the phrase 'green recruit', is derived from 1. the observed correlation between green and unripe and 2. the abstraction linking unripeness with inexperience, a salient property of recruits.

2 A context sensitive model of category structure

'Literal' semantics: direct inferences and category structure

Since the interpretation of novel figurative adjective noun phrases is based on the semantics of literal adjective noun phrases, any discussion of metaphor interpretation mechanisms must be prefaced with an analysis of literal interpretive mechanisms. While a limited form of analogical mapping can be established directly between two scalar property values (eg. low weight/saturation mapping to low calorie), the really interesting cases arise when observed correlations between property values are brought into play. When an adjective modifies a noun in a literal context, the category denoted by the noun is cast in a new light. Sometimes the shift in perspective is a minor one: the phrase "green car" carries little additional information over the selection of a specific color. The phrase "green banana", however, entails a significant modification to the default values of bananas: in addition to the color changing from yellow to green, one also infers that a green banana is unripe, difficult to peel, astringent tasting, and so on (see Figure 1). Since it seems untenable that this (completely different) view of the property values of bananas constitutes a proper subcategory, this information must exist within the category itself.

This habit of rapidly and automatically inferring changes in property value settings from the knowledge of one property value is called *direct inferencing*. Immediate inferences are the direct inferences available at the level of the category under consideration. They are performed quickly, in a few hundred milliseconds, and without conscious thought. These immediate inferences must reflect the structure of stored knowledge, as they are available too quickly and effortlessly to involve any complex form of information retrieval. The argument is that the patterns of immediate inferences reflect the structure of connections in the underlying spreading activation model, implemented here as a structured connectionist network.

With the observation of direct inferences as a starting point, a model of the internal structure of categories can be developed. Given that physical objects possess certain properties with characteristic values, the question is how to represent the correlations between values. The answer suggested by [Weber 1989] is that functional properties of the object supply the necessary organizational structure, as each value of a functional property participates in a distinct *aspect*, or informal coalition of property values, of the category. For example, the functional property *ripeness* of fruit motivates three distinct aspects of bananas (unripe, ripe and overripe), each with partially overlapping but distinctive characteristic property values.

Aspects, by controlling the priming of correlated property values, support not only straightforward 'literal' reference but also metonymy, when one property value is used to refer to another, and unusual or unfamiliar choice

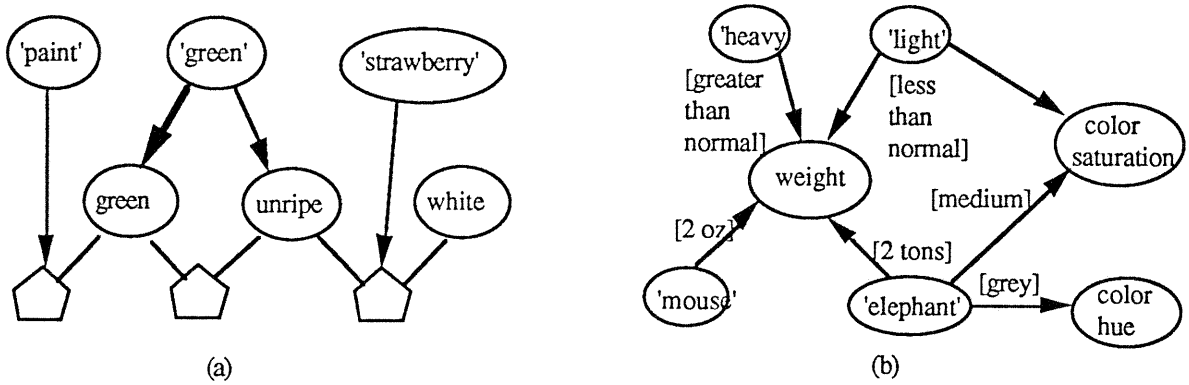


Figure 2: the component of network structure capturing lexical semantics. (a) polysemy: lexemes can have multiple word senses; spreading activation supplies the contextual cues to select the appropriate one. (b) relativity: the adjective is treated as a scaling factor on the normative property value associated with the noun.

of adjectives, when the individual words are known but the meaning of the combination is not. When faced with the need to suggest a semantics for an adjective noun phrase, a source domain must be established for the analogical mappings of potential meanings. This domain consists of the adjective's connotations in all previously encountered contexts. The connotations considered for the purposes of figurative interpretation are the immediate inferences arising from the modification of a category commonly associated with the given adjective.

Figurative interpretation: property value transference

The central component of the model is the use of direct inferences and scalar value transference in interpreting both familiar and fanciful adjective-noun combinations. What is a "green peach"? A "thirsty fern"? A "happy box"? In a literal adjective-noun phrase, the descriptive adjective names a property value of the category denoted by the noun, that is, it indexes a feature value unit associated with the noun. As correlations often exist between feature values [Malt and Smith 1984], indexing one will tend to excite others, thus supplying the adjective with the characteristic connotations known as direct inferences. For example, while a (ripe) peach is normally pink on the outside and juicy and sweet on the inside, a green peach is unripe, dry and sour in addition to being green in color. These direct inferences suggest themselves so strongly that perceptual property values are often used metonymically to stand for certain (non-observable) functional or constitutive property values. For example, unripe is an extended word sense of the adjective 'green'.

The interesting thing about the phrase "green peach" is that the literal word sense of the adjective (meaning color), while certainly applicable, is not the most appropriate one in the context. Yet it seems difficult to argue that the *unripe* word sense of 'green' is a literal one; hence the suggestion that the metaphoric versus literal distinction should not be made on semantic processing grounds but rather be considered a matter of context sensitivity.

When the modifying adjective indexes a feature value *not* possessed by the noun, mappings from feature values associated with the adjective in literal noun contexts are used to indirectly index feature values of the noun. A "thirsty fern" is readily understood to be a fern in need of watering, despite the fact that ferns lack the cognitive facilities needed to subjectively experience thirst as humans do. This is an example of what Keil [1979] calls "category error", a situation in which a predicate (i.e. adjective) is applied to a category (i.e. noun) that does not normally allow it. Keil suggests that natural categories form a "predicability hierarchy", where predicability is defined to be the knowledge of which predicates can be combined with which terms in a natural language. A predicate is said to 'span' a term when the predicate can be meaningfully applied to the term and the resulting phrase can be assigned a truth value, be it true or false. A category error occurs when a predicate is used in conjunction with a term it does not span. For example, "green idea" is a category error, since only physical objects can have color as a property. Thus it is neither true that the idea is green nor that it is not green, since the latter statement implies that the idea has a color, the value of which is something other than green.

When faced with a category error, needless to say, the phrase is far from uninterpretable. Typically, the superficially inapplicable adjective is understood to refer elliptically to a legitimate property of the noun, in this case, hydration, as described in Section 3.

These property transference mechanisms can be used to supply suggestions for even the most fanciful of expressions such as "happy box". The context surrounding the phrase will usually supply some clues as to the properties of the noun that are being alluded to, but all else being equal the more salient properties of the noun will

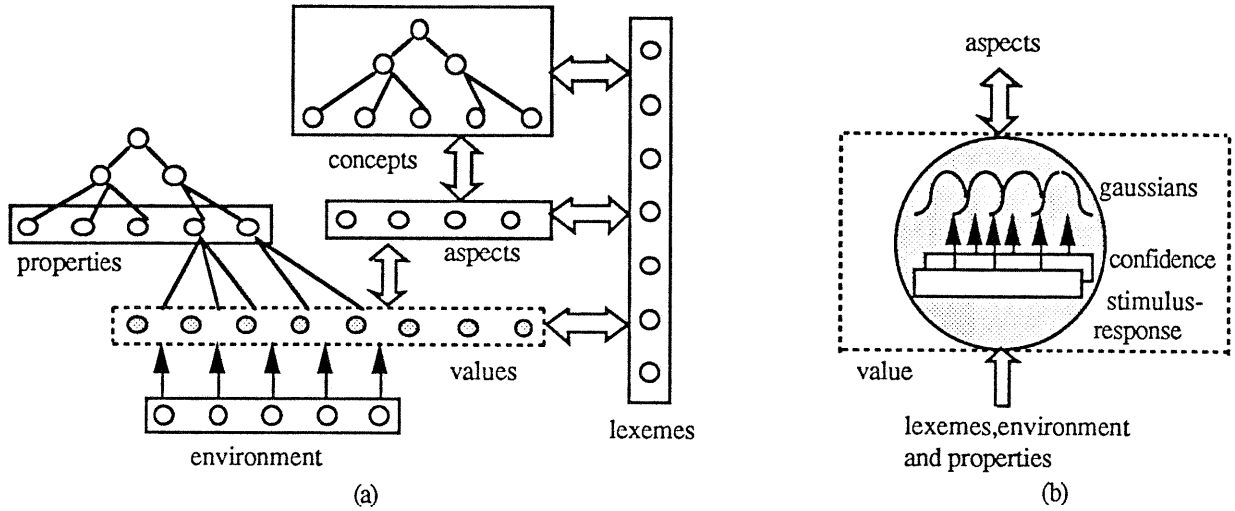


Figure 3: (a) Network configuration prior to training. Plain arrows denote one way links, plain lines two way links and open arrows full connectivity. (b) The implementation of the virtual "value" nodes shown in (a).

be favored. The degree of confidence in any such interpretive suggestions fall off with the semantic distance in the property abstraction hierarchy between the mapping's source and target properties.

Acquiring lexical semantics

Another crucial component of the model is context sensitive lexical semantics. Descriptive adjectives correspond to feature values and nouns to concepts. An adjective noun phrase is understood by the spread of activation from the corresponding lexemes through the semantic network, where the spread of activation is modulated by the relevance of each intermediate node to the concept under consideration.

The model captures two forms of context sensitivity in descriptive adjectives. In one form, the appropriate word sense of a polysemous lexeme is chosen by spreading activation [Cottrell 1982]. For example, a "green strawberry" tends in fact to be more white than green in color; the adjective is used to refer indirectly to unripeness, a feature of strawberries that is strongly correlated with the color green in the context of other fruits and vegetables (see Figure 2(a)).

The other form of context sensitive semantics involves terms such as 'light' and 'large', terms whose semantics is relative to some normative value associated with the noun. A heavy mouse weights considerably less than even a light elephant, and green leaves are greener than green sea water. One mechanism that could account for this semantic pattern is to have adjectives supply a scaling factor which is applied to the value fo the [implicitly] named feature. For example, if mice normally weigh two ounces then a light mouse may only weight 1.6 oz., or 80% of the normal weight (see Figure 2(b)). One flaw with this idea is that the variance is fixed. The difference in weight between a styrofoam and lead box is much greater than the difference in weight between a scrawny and an overweight house cat. Nonetheless, for natural kind terms, the focus of this study, the mechanism seems reasonably appropriate.

There are two situations in which lexical semantics can be learned. In one, an object is presented by the teacher and explicitly named for the student, eg. "Mommy", "cookie". In the other, far commoner situation, the exact referent of the lexeme is left up to the student to infer, as when a child is exposed to adult conversation. A crude estimate of the meaning of a term can be made by compiling frequency statistics of use for each lexeme. The meaning hypothesis can be iteratively refined with each new training example, eventually resulting in associations being formed between the lexeme and its referent.

2 The connectionist implementation

This model of adjective noun semantics has been implemented in an adaptive structured connectionist network on the Rochester Connectionist Simulator [Goddard *et al.*, 1988]. The network operates by allowing activation from the noun to spread through the aspects relevant to the adjective, thus priming all contextually relevant direct inferences, retrieving any relevant stored meanings, and dynamically suggesting figurative interpretations based on the available direct inferences, all in parallel. Since the dynamic interpretations are proffered with a relatively low confidence, if a catalogued word sense is available it will be favored over the figurative meaning suggestions. Thus

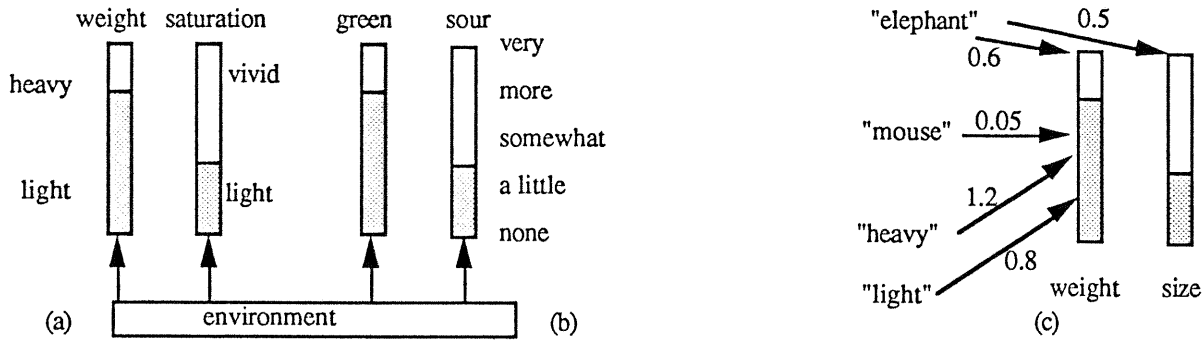


Figure 4: Implementing (a) scalar and (b) non-scalar values. Adjective semantics (c) are implemented as weights on the links to the corresponding feature unit, so that a heavy elephant weighs more than a (normal) elephant, which is considerably more than a heavy mouse.

while no explicit attempt is made to catalogue adjective noun combinations according to their degree of apparent semantic anomaly, the confidence level of interpretation implicitly reflects the distance from a 'normal' semantics for the constituent terms.

The network blends prestructured elements with adaptive components. The layered architecture reflects certain *a priori* assumptions about the internal structure of cognitive categories. The links between layers, however, are initially uniform and must develop in response to training input.

There are two forms of input to the system, 'environmental' stimuli, corresponding to the feature values 'perceived' as occurring in the outside world, and lexical items, the adjectives and nouns used to describe the environment. While environmental inputs are strictly feedforward, recurrent links connect lexemes to the net. Both sources of input are initially fully connected to an internal semantic representation of feature values, and lexemes are additionally connected to the conceptual level.

Figure 3 shows the initial configuration of the network layers. At the lowest layer are the units representing the environmentally supplied stimuli, corresponding to perceptual features. These units are designed to be turned on and off by the environment (i.e. by hand). They feed forward in a one-to-one mapping to the feature value layer.

The output from the feature value layer feeds into a layer of aspect binder units, whose task is to store and retrieve characteristic correlation patterns observed in the gaussian layer. One shot learning is used to recruit aspect binders one at a time as the environmentally supplied stimuli change. Each aspect thus recruited allies itself with the currently active unit in the next layer up, the concept layer.

Representing feature values

All layers but the feature value layer adhere to the unit value principle, which holds that the degree of activation of a unit reflects the confidence in the proposition associated with that unit. The feature value layer, on the other hand, incorporates a symbolic element [Lange and Dyer 1989] in that activation levels of units representing scalar features reflect actual feature values.

While some features such as weight, size and saturation are scalar quantities whose feature values correspond to positions in the feature scale, others such as taste and hue seem to be collections of scalar feature values. In the former case (see Figure 4(a)) the feature values are mutually exclusive, since the feature scale can only take on one value at a time. In the latter case (see Figure 4(b)), adjectives map not on to the feature values but to the features themselves, and since features can co-occur, the mutual exclusion property is missing for these lexemes. This method of implementing the mutual exclusion properties of scalar features violates the unit value principle; thus each feature unit is paired with a unit representing the confidence in the proposition that the feature in question is either present in the environment or being referred to linguistically. That is, the feature units can either be activated from the "external world" or from linguistic input. In either case the confidence in the featural proposition is high; when activation arrives from other channels, however, confidence is considerably lower, as in the case of scalar value transference, an activation propagation mechanism employed in novel metaphor interpretation.

Each 'unit' in this layer is actually a group of several units (see Figure 3(b)), a stimulus response unit, a confidence unit, and a collection of gaussian response units to quantify the joint response of the stimulus response and confidence units. The stimulus response unit simply transmits the value received on its inputs; when more than one value is seen, the one with the highest associated confidence is chosen. The confidence unit paired with each stimulus response unit places highest confidence in signals received from the environment and considerably lower confidence on values from any other source, such as those suggested by the figurative interpretation mechanisms. The stimulus response-confidence signals are fed into a layer of gaussian response units whose

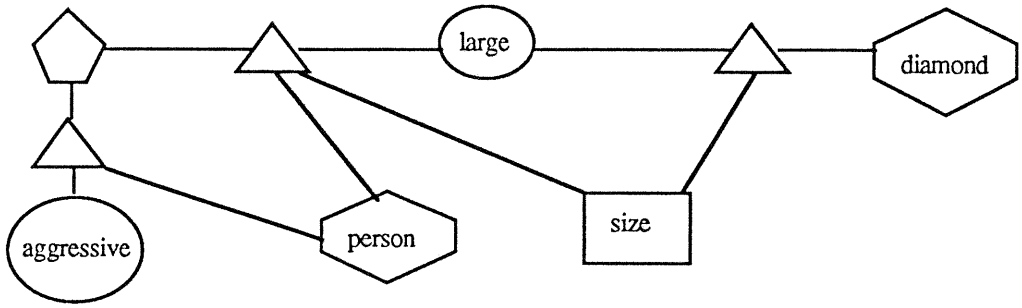


Figure 5: Direct value transference: the phrase 'aggressive diamond' is interpreted to mean a large diamond, due to the correlation of aggressive and large within an aspect of the person category.

function is to transform the response value, itself highly symbolic, into a quantized unit-value form more consonant with connectionist principles. For example, a given response unit may be represented at the gaussian level by three gaussian units, one tuned to respond maximally to low input values, one to median values, and one to high. Then a low stimulus value, while producing a different pattern of activation over the gaussian layer from a high value of similar confidence, will be represented with the same intensity of response. Needless to say, the response of the gaussians is modulated by the corresponding confidence input, so that a lower confidence signal results in lower activation levels of the activation pattern characteristic of the given value.

This architecture provides the integrated feature value-confidence response that is used in forming conceptual aspects. The weighted input from the currently active noun supplies the default value for each feature associated with that noun, and the adjectival input supplies modifying signals. The two input values are simply multiplied together, so an adjectival input less than one will lessen the default value while an input greater than one will increase it (see Figure 4(c)). This scheme, while suffering from the failing of forcing a semantics of uniform and fixed variance on adjectives, offers the advantage of saving the space required by the combinatoric number of nodes that would be needed to represent the cross product of nouns with adjectives.

Conceptual aspects and direct inferences

One of the preconfigured components of the network has each noun known to the system mapped to a single node in the concept layer. Units in the concept layer use an activation function which simply reflects the input from its corresponding lexeme (a logistic output function is used). After training each concept unit is linked with varying degrees of confidence to a small number of relevant aspect units which are in turn linked to their characteristic semantic feature value units. Activation from a noun preferentially excites all context free aspects of the concept. Inter-aspect competition is implemented not as an all-or-nothing inhibition but rather as a signal attenuation factor: if the unit's current activation is less than the maximum, it is reduced by a constant (non-cumulative) amount. Aspects output the logistic function of the weighted sum of all inputs from concepts or values, so aspects can receive activation both from nouns (concepts) and adjectives (feature values). Aspect activation percolates down one more layer to the feature values, producing a characteristic pattern of activation within the feature value layer. The full lexical semantics of a noun with or without adjectival modification is taken to be the pattern of activity over the aspect and feature value layers of the network.

All three layers, concept, aspect and feature value, are considered candidates to supply the semantics for the lexical layer. A modified form of Hebbian reinforcement learning is used to successively refine the lexical semantics hypotheses.

Proposing novel interpretations

In addition to the straightforward retrieval mechanisms for catalogued meanings described above, there are some simple analogical mapping mechanisms in place to supply semantic suggestions when the stored meaning retrieval mechanisms fail to supply any semantics other than the default or context free ones. These mechanisms are based on the premise that mappings can be established between two feature values at analogous scalar positions on their respective feature scales. The process of figuratively interpreting an adjective noun phrase involves setting up mappings from the values primed by the literal connotations of the adjective in various contexts (the source domain) to values of the category denoted by the noun (the target domain). The set of all values activated in this manner form the interpretive basis for understanding a figurative usage. Semantic correspondences must somehow be established between the property value denoted by the adjective and property values belonging to the category denoted by the noun.

There are two methods used to establish semantic correspondences: (direct) value transference, and (indirect) scalar correspondence. The most straightforward interpretations arise when a property value of the target category is made available through an immediate inference associated with another category. For example, suppose it was 'known' to the system that aggressive people are also large in size, i.e. large and aggressive both participate in the same aspect of person, then the unfamiliar figure of speech 'aggressive diamond' would be interpreted as denoting a large diamond (see Figure 5). This form of mapping value-to-value mapping, known as property value transference [Aarts and Calbert 1978], is applicable when a property value (and hence its associated property) is common to both the source and target fields.

To establish correspondences in the non-overlapping areas of the source and target fields, however, more indirect mappings must be resorted to. One possible indirect mapping exploits the scalar nature of many properties, particularly properties of physical objects. Properties that are quantitative in nature, such as size, weight, density, malleability and so on, tend to impose a natural scalar ordering on their values [Kittay 1987]. Thus assuming that two such properties have somehow been placed in correspondence, the value mappings become obvious. For instance, if size in the source field corresponds to weight in the target field, then small maps to lightweight, large to heavy, and so on. This method of value mapping is called scalar correspondence.

Scalar correspondence is handled by the property abstraction hierarchy. Each property node is in fact a pair of value-confidence units like the ones used in feature value nodes. As activation propagates through the abstraction hierarchy, the confidence signal is steadily decremented but the value signal is maintained at full strength. Thus the confidence measure of a value input from the property abstraction hierarchy directly reflects the semantic distance between the source and target.

3 Acquiring network structure from data

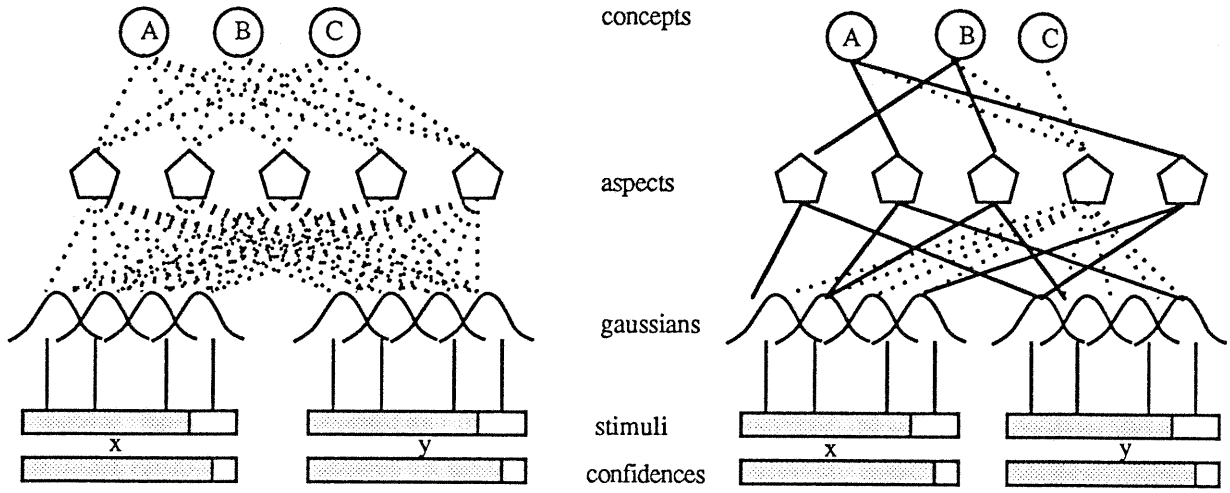
The network has two distinct learning modes, perceptual feature learning and lexical semantic learning. The behavior of lexical layer units is affected by the lexical learning signal and all other units, concept, aspect, gaussian and stimulus-response, react to the perceptual learning signal. Both signals are implemented with single dedicated control units which can be clamped on or off independently of each other, but in practice the perceptual learning is run before the lexical learning, in order to furnish a coherent lexical semantics as a learning target.

In perceptual learning mode, the network responds to the presentation of an exemplar, which is defined to be a pattern of co-occurring feature values associated with a single concept. Before running the network in this mode, the user clamps on one concept unit and any (meaningful/appropriate) subset of the environmental stimulus units, where the semantics of each stimulus unit is known to the user but not to the system. That is, even though a given input node may be labelled as representing the feature value green, since it is initially fully connected to the aspect layer (albeit indirectly), an initially uniform semantics for all input units exists. The labels merely provide the teacher with a map to the input unit's supposed relationship with the outside world (see Figure 6).

When learning from an exemplar, stimulus response units simply transmit their environmental input. Gaussian units weight their response to input from feature units by the input from the corresponding confidence unit. They will also recruit themselves to an aspect unit that is emitting the distinguished signal of new recruitment, by setting the weight on each link to the value currently being transmitted. Aspects respond with the weighted sum of the gaussian layer inputs. If an unrecruited unit's response is over a certain threshold and greater than the current maximum of all other aspect units, then it will recruit itself to represent the current pattern of gaussian activity by setting the weights on the incoming links to reflect the values they transmit. This has the effect of severing all links to currently inactive gaussians. If the unit's response equals the maximum of the other aspect units, then there is contention for recruitment, so the weights on the gaussian layer links are re-randomized to (eventually) break the response deadlock. Once a unit has been recruited it will respond as before to gaussian input but will not be eligible to represent a new pattern. The effect of the threshold on the recruitment scheme is to widen the radius of sensitivity of aspect units to include not only the precise pattern it was recruited on, but also reasonably similar patterns, including sub-patterns.

The response of concept units is tied to the corresponding behavior of aspect units. Immediately after recruitment an aspect unit emits a distinguished value (1.0) as a signal to the concept layer, which is monitoring the responses of aspects to the given stimuli. Only when an aspect has been newly recruited at the gaussian level will a concept node recruit it. Concept layer recruitment involves cutting the links to all but the maximally active aspect, the link to which has its weight boosted to the maximum. This procedure means that while a concept can acquire many aspects, each aspect will be associated with one and only one concept, just as required by the model.

When not running an exemplar in perceptual learning mode, activation functions are generally simpler. Stimulus response units are the one exception: if no environmental input is available, the unit will respond to suggestions as to its value coming in from alternative internal sources, specifically, from the property inheritance hierarchy used by the figurative interpretation mechansisms. Gaussians again weigh their response by the confidence input; if no inputs are being received from the feature layer, the weighted sum of inputs from any active



(a) (b)

Figure 6 (a) network connectivity before training and (b) after learning two aspects of two concepts. Dotted lines indicate small random link weights and solid lines correspond to large positive weights. After learning, concept A's first aspect associates the values "somewhat x" with "very y" and its second aspect groups the values "very x" with "not y" (see Figure 4(b)).

hubs effects the retrieval of the activation pattern stored on recruitment. Aspects simply respond with the weighted sum of gaussian and concept layer inputs. Finally, only the concept unit associated with the currently active noun is activated; all others remain quiescent.

Lexical units are initially fully connected to both the feature layer and the concept layer, since these two layers contain compact representations of the semantics of adjectives and nouns respectively. Lexical semantic acquisition is achieved with simple Hebbian weight reinforcement. When a lexical unit is coactivated with either a feature value or a concept node, the strength of the connection between them is augmented, and when the lexeme is active it slightly decrements the weight on all links to inactive values and concepts.

An Example

Some components of the system's initial structure, as well as the data used during perceptual training, are obtained from the user supplied input file. A small example of such an input file appears below:

```
sfeature ( ripeness: unripe, ripe, overripe, rotten )
sfeature ( experience: inexperienced, experienced )
feature  ( taste: sweet, sour, astringent )
feature ( color: green, yellow, red )
concept ( apple, banana, recruit )
abstracts ( development: ripeness, experience )
exemplar ( apple: red, sweet, ripe )
phrase: pomme
exemplar ( apple: green, sour, unripe )
phrase: pomme verte
exemplar ( banana: yellow, sweet, ripe )
phrase: banane
exemplar ( banana: green, astringent, unripe )
phrase: banane verte
exemplar ( recruit: inexperienced )
```

There are six statements known to the system, *feature*, *sfeature*, *abstracts*, *concept*, *exemplar* and *phrase*. The *feature* statement is used for non-scalar features such as hue (color), saturation, taste, and so on; features for which multiple values are permissible. Separate feature nodes are created for each value associated with the property (see Figure 4(b)). The *sfeature* statement is used for scalar features such as temperature, weight, size and so on; features whose values are mutually exclusive and arranged along an implicit scale. The values named in the statement

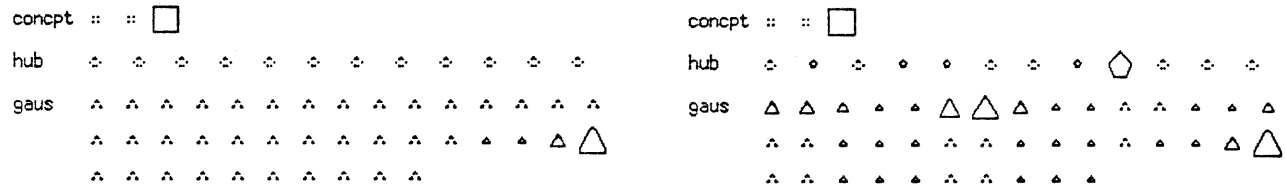


Figure 7: Results of running the example on the system developed on the Rochester Simulator. On the left is the result of running the phrase "green recruit" before training on the exemplars; on the right is the result after training.

should appear in order of increasing scalar magnitude. A single feature node is created for the property, and the given values are allocated scalar positions within the feature node (see Figure 4(a)). The *abstracts* statement is used to build the property abstraction hierarchy. The property abstraction hierarchy plays such a crucial role in the novel metaphor interpretation process that it would be better to acquire the structure of the hierarchy during training rather having the user specify it beforehand, but until a solution is found to the problem of automatic abstraction the hand-coding is needed. The *concept* statement is somewhat redundant, since concepts appear in exemplars, but was included for implementational simplicity. The statement has no effect on the network's initial structure; its only use is in initializing the lexical dictionary used to devise training sequences. The *exemplar* statement specifies the form of the perceptual mode training sequences. For example, the first set of co-occurring property values would be red, ripe and sweet. These values are run with the lexeme "pomme", which is the French word for apple (*phrases* apply to the previous exemplar). Lexical training with the given set of exemplars results in "pomme" mapping to the apple concept and "banane" to the banana concept, but the best guess as to the semantics of "verte" is that is could equally well refer to either unripe or green, as seems appropriate.

The network produced by these statements consists of three concept nodes (apple, banana, recruit), 12 aspects (the total number of feature values and as good a guess as any), 45 gausssian units (five for each stimulus), 8 stimulus response--confidence pairs (ripeness, experience, sweet, sour, astringent, green, yellow, red), 15 internal environment (i.e. training input) inputs, one for each concept and feature value and 15 lexical level units, all initially arranged as shown in Figure 3. Figure 7 shows a pair of before and after pictures of the network's response to the stimulus "green recruit", both before and after training. Concept units appear as square icons, whose size is proportional to the unit's level of activation. The pentagons are the aspect units (or "hubs"), and the triangles the gaussian units. Before training no aspects have been recruited to concepts and features, so activation does not spread from the input stimuli 'green' and 'recruit'. After training, however, activation is permitted to spread from the features representing green color to its correlated feature unripeness, and, via the property abstraction hierarchy, from unripeness to inexperience, represented by the cluster of gaussians near the center of the first row.

4 Summary and Future Work

One possible avenue for future work is to investigate learning property abstractions, as opposed to having them built in according to the user's hand designed specifications. The difficulty with learning property abstractions under the current scheme is that properties and their attendant values are not represented at a fine enough grain to support abstraction. Some form of microfeature representation would be the obvious first place to start in remedying this deficiency.

Another intriguing problem raised by the work to this point is the difficulty in integrating dynamic gaussian distribution with the other forms of training supported by the network. Preliminary work on training the gaussian layer to conform to the representational demands of the underlying property values seemed encouraging, but so far all attempts to integrate these techniques with the exemplar training routine have failed. The problem is one of a moving target: the gaussian layer needs a clear indication of the representational capacity of each feature value unit, but at the same time the exemplar training procedure requires a clear signal from the gaussian layer to be effective.

One final suggestion for future work on the system is that it be extended to allow mediated inferences to play a role in the metaphor semantic suggestion mechansisms. Only direct inferences are currently considered when devising dynamic figurative interpretations.

To summarize, the model lexical semantics for adjective-noun phrases suggested by the observation of direct inferences has three components, functionally organized internal structuring of categories into aspects, context sensitive lexical semantic retrieval for both literal and figurative usages, and dynamic interpretation mechanisms to suggest figurative meanings of novel usages. The dynamic mechanisms base their interpretive suggestions on catalogued meanings in alternative contexts. These suggestions can, with sufficient repetition or situational reinforcement, be added to the catalogue of known meanings, thus enriching the semantic basis for future figurative interpretations.

References

[Aarts and Calbert, 1979] Jan M. G. Aarts and Joseph P. Calbert, *Metaphor and Non-Metaphor: the semantics of adjective-noun combinations*, Max Niemeyer Verlag, 1979.

[Barsalou, 1982] Laswrence W. Barsalou, "Context-Independent and context-dependent information in concepts", *Memory and Cognition*, 10(1), pp. 82-93, 1982.

[Black, 1979] Max Black, "More about metaphor", in Andrew Ortony, editor, *Metaphor and Thought*, Cambridge University Press, 1979.

[Cottrell 1982] Garrison W. Cottrell, Computer Science Department, University of Rochester, 1982.

[Gildea and Glucksberg, 1982] Patricia Gildea and Sam Glucksberg, "On Understanding Metaphor: The Role of Context", *Journal of Verbal Learning and Verbal Behavior*, 21, pp. 85-98, 1982.

[Goddard et al., 1988] Nigel Goddard, Kenton Lynne and Toby Mintz, "The Rochester Connectionist Simulator User Manual", Technical Report 233, Computer Science Department, University of Rochester, May 1988.

[Keil, 1979] Frank C. Keil, *Semantic and Conceptual Development: an Ontological Perspective*, Harvard University Press, Cambridge Mass., 1979.

[Kittay, 1987] Eva Feder Kittay, *Metaphor, its Cognitive Force and Linguistic Structure*, Clarendon Press, 1987.

[Lakoff, 1987] George Lakoff, *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*, University of Chicago Press, 1987.

[Lange and Dyer, 1989] Trent E. Lange and Michael G. Dyer, "High Level Inferencing in a Connectionist Network", Technical Report UCLA-AI-89-12, University of California, Los Angeles, October 1989.

[Malt and Smith, 1984] B. C. Malt and E. E. Smith, "Correlated properties in natural categories", *Journal of Verbal Learning & Verbal Behaviour*, 23:250-269, 1984.

[Martin, 1988] James H. Martin, "A Computational Theory of Metaphor", Report no. UCB/CSD 88/465, Computer Science Division, University of California, Berkeley, 1988.

[Richards, 1937] I. A. Richards, *The Philosophy of Rhetoric*, London: Oxford University Press, 1937.

[Tabossi 1986] Patrizia Tabossi, "Effects of Context on the Immediate Interpretation of Unambiguous Nouns", *Journal of Experimental Psychology: Language, Memory and Cognition*, 14(1), 153-162, 1988.

[Weber 1989] Susan Hollbach Weber, "A Structured Connectionist Approach to Direct Inferences and Figurative Adjective-noun Combinations", University of Rochester Computer Science Department TR 289, June 1989.

Extending the Lexicon by Exploiting Subregularities*

Robert Wilensky
Division of Computer Science
University of California, Berkeley 94720

(415) 642-7034
wilensky@teak.berkeley.edu

Abstract

The term "literal meaning" conflates a number of distinct notions. In particular, it is often assumed that the literal meanings of words are conventionalized in the lexicon, while non-literal meanings arise from some more generative process of interpretation. This dichotomy is already known to be false. However, the degree to which the lexicon embodies non-literal meanings is not fully appreciated. I have proposed a theory of lexical relations, of relations between word senses, which are divided into three classes, grammatical relations, core relations, and transforming relations. Relations of the last case often involve conventionalized but non-literal word senses. Many types of these have been uncovered, but it is conjectured that they are not in principle limited in number. These lexical relations are in general useful for lexical acquisition. However, they also demonstrate the degree to which the notion of literality reflects an intuition about the centrality of a word sense rather than about its conventionality.

*The research reported here is the product of the Berkeley Artificial Intelligence and Natural Language Processing seminar; contributers include Michael Braverman, Narcisco Jaramillo, Dan Jurafsky, Eric Karlson, Marc Luria, James Martin, Peter Norvig, Michael Schiff, Nigel Ward, and Dekai Wu. This work also benefitted from discussions with George Lakoff, Eve Sweetser and Paul Kay. James Martin pointed the use of metaphor knowledge in learning; Peter Norvig's lexical network work demonstrated the existence of other, useful lexical relations. This research was sponsored by the Defense Advanced Research Projects Agency (DoD), monitored by Space and Naval Warfare Systems Command under Contract N00039-88-C-0292 and by the Office of Naval Research, under contract N00014-89-J-3205.

1. Introduction

In a previous paper (Wilensky 1987), I argued that important aspects of the "classical" notion of literal meaning were mistaken. In particular, it has generally been assumed that, since the term "literal" distinguishes central-sense uses of words from idiomatic uses, non-metaphoric from metaphoric, and direct from indirect, that literal meanings must be the same as sentence meanings, i.e., that the meaning of utterances intented literally could be computed from knowledge of words and core grammar rules of the language. However, the intuitions justifying this position seem to confuse *sentence meanings* with *sentence interpretations*. Literal interpretations of a sentence, even out of context, generally make recourse to extra-linguistic knowledge. In other words, even literal interpretations of an utterance generally are non-compositional.

Another aspect of this argument involves the relation between the non-literal and the conventional. In particular, some non-literal interpretations are purely linguistic in nature; thus, at least linguistically, their meanings are as compositional as those involving so-called literal meanings. This point was made forcefully by Lakoff and Johnson (1980) and Martin (1988), who underscore the prevalence of conventionalized but metaphoric word senses. Norvig and Lakoff (1987) extend this argument to several other kinds of non-literal word senses.

In this paper, I wish to clarify the notion of word sense, and examine the ways words senses can themselves be "non-literal". My belief is that virtually all interesting "non-literal" and "creative" uses of language have conventionalized reflexes in the lexicon; moreover, there is no small number of ways in which literal word senses are extended to non-literal ones. The language coheres not so much by virtue of compositionality and a small number of types of non-literal phenomena, but by virtue of rampant subregularities, i.e., motivated but non-predictable conventionalizations.

The term "literal", incidentally, is still useful, but not as something that strictly distinguishes conventional from non-conventional language. Rather, our intuitions about literality encode intuitions about the centrality of words senses, rather than intuitions about their conventionality.

2. The Nature of Lexical Entries

It seems generally agreed upon that words often have multiple senses, can serve as different parts of speech, and manifest variable valence. Moreover, these facts about a given word go together. To use a trite example, the word "bank" can be a verb whose meaning has to do with using a financial institution, and when it does, it subcategorizes for a subject expressing the customer and a prepositional complement beginning with "at" that expresses the financial institution. Alternatively, when "bank" is a verb meaning to bang a billiard ball off a cushion, it subcategorizes for a subject expressing the agent and a direct object expressing the ball. Of course, the word has at least two meanings as a noun, none of which subcategorizes for anything.

I assume that each lexical entry contains this kind of information. I call each basic unit of lexical information a *lexical grammatical construction* (or *lex-con* for short). Each lex-con specifies a phonological form (i.e., a "word"), grammatical information such as part of speech and valence, and a meaning.

The meaning in a lex-con is simply a concept that the word denotes. These concepts are supposed to be wholy non-linguistic, that is, they are objects in some internal representation language, and everything we know about these concepts (other than associated linguistic information) is spelled out in the conceptual domain. I call a concept associated with a word via a lex-con a *sense*: A sense is just a concept that happens to be lexicalized, i.e., that has gotten attached to a word.

I will use the term *lexical relation* to mean any systematic relation between lex-cons. I divide relations between lex-cons into three classes:

1) *Grammatical relations* – These are meant to capture the trivial case of grammatical variations of the same stem.

2) *Core relations* – Lex-cons are *core-related* when their senses share some component of meaning.

3) *Transforming relations* – These are relations between lex-cons one of whose senses is thought of as derived from the other through some process of meaning transformation.

Each class of relations is now described in greater detail.

(1) Grammatically-related Lex-cons

Lex-cons as described so far require a rather fine level of granularity. In particular, I require separate lex-cons for the cross-product of all grammatical, semantic, and phonological distinctions. So, for example, there is a distinct lex-con for the word 'bank' corresponding to the first person singular verb meaning to use a financial institution, another for the same word used as the second person plural verb with the same meaning, and others for all combinations of tenses, person, number and senses of this word and its variants.

However, the semantic relations between these lex-cons are in some sense trivial. Some of these lex-cons differ only in grammatical features like person, so their senses are identical; others, such as the tenses or number have different but completely generative meanings. Thus, even though the phonological relation between the words might vary, I do want to recognize these lex-cons as not having an independent status. I call two lex-cons *grammatically related* if their word senses are in one of a set of "grammatical meaning relations". Grammatical meaning relations are just semantic relations that are grammatically realized, such as singular-multiple, present-past, and especially, the identity relation.

To facilitate certain generalizations, we will allow *abstract lex-cons*, that is, lex-cons that are lacking in information necessary for their participation in grammatical constructions. The most abstract lex-con of a hierarchy of lex-cons is called a *basic lex-con*, and corresponds intuitively to the root or stem of a word.

(2) Core-related Lex-cons

One way in which the relation between two lex-cons may be more interesting is when their meanings intersect in ways that are not completely predictable. I call this shared meaning a *core meaning* (or *core concept*) of the senses, and say that these senses (and thereby, the lex-cons which provide them) are *core-related*. Such relations are assumed to hold between basic lex-cons of the same word, and between those of different words. When I have occasion to refer to all the senses that extend a core meaning, I will use the term *meaning complex*.

As an example, there are lex-cons of "have", "give", "take", "get" and "receive" whose senses are core-related. The core meaning is the notion of possession, which happens to be a sense of the word "have". The senses of (the lex-cons of) the other terms extend this core by specifying a change in possession, along with different relations between their roles; in addition, these lex-cons have different valences. For example, "give" has at least two lex-cons relevant to this possession meaning complex; let us temporarily call the sense specified by both of these lex-cons Give. The Give concept has a Giver, a Givee, and a Given, say; the meaning of these terms is established in the conceptual domain, i.e., by representing facts predicating their relation to other concepts in the system, e.g., that each Give is an Action by the Giver resulting in the Givee possessing the Given. One of the two lex-cons specifying this sense would have the valence information that the Given is specified by the direct object and the Givee by a prepositional complement beginning with "to", the other that the Givee and the Given are specified by the two consecutive noun phrases (e.g., a ditransitive), in that order.

Similarly, "get" may also have two lex-cons, in this case with different but related senses: One sense specifies that the recipient is also the active agent of the transfer, the other that the recipient plays a passive role. Indeed, the first of these senses may be identical to a sense of "take", and the second identical to the sense of "give" just mentioned. However, the lex-cons are quite different. For example, the lex-con of "get" that would have the same sense as one of "give" would specify that the syntactic subject is the Givee, rather than the Giver, which can be specified by an optional "from" complement.

Of course, if a single concept is shared by a lex-con of "give" and a lex-con of "get", and another shared by a lex-con of "give" and one of "take", they should be named more neutrally. For example, instead of Give, I should have named this concept something like Agent-causes-recipient-to-possess; instead of Take, I should have Agent-causes-self-to-possess. I should similarly rename the various thematic roles. In any case, the semantics of these terms is established on the conceptual level, e.g., by stating in one's knowledge representation language that an Agent-causes-other-to-possess event is a kind of Cause-to-possess event, etc.

I illustrate a few of these lex-cons below. For example, here are the lex-cons for "give" and "get" that share the same sense, and the ones for "get" and "take" that share the same sense:

Lex:	root-give
Sense:	Agent-cause-recipient-to-possess
POS:	V
Val:	[Subj: Agent, DO1: Recipient, DO2: Patient]

Lex:	root-give
Sense:	Agent-cause-recipient-to-possess
POS:	V
Val:	[Subj: Agent, DO: Patient, P[to]: Recipient]

Lex:	root-get
Sense:	Agent-cause-recipient-to-possess
POS:	V
Val:	[Subj: Recipient, DO: Patient, (P[from]: Agent)]

Lex:	root-take
Sense:	Agent-cause-self-to-possess
POS:	V
Val:	[Subj: Agent, DO: Patient, (P[from]: Donor)]

Lex:	root-get
Sense:	Agent-cause-self-to-possess
POS:	V
Val:	[Subj: Agent, DO: Patient, (P[from]: Donor)]

These are basic lex-cons, so the lexical entries contain roots rather than word forms per se. The valence structure is given as a list of syntactic functions and associated thematic roles, with parentheses indicating optionality, and "P[*word*]" indicating a prepositional phrase headed by *word*.* On the conceptual level,

*I am appealing rather casually in this discussion to some notion of thematic role, or conceptual case frame. Actually, the

facts are needed representing the relation of the concepts in these lex-cons to other concepts. For example, assuming that there is a predicate "AIO", representing the relation between an individual and a category, the following might be expressed:

```
V g E a,p,d,c,i,t
  AIO(g,Agent-cause-self-to-possess)
   -> Agent(a,g) & Patient(p,g) & Donor(d,g)
        & AIO(c,Causing)
           & Cause(i,c)
           & Effect(t,c)
        & AIO(i,Action)
           & Agent(a,i)
        & AIO(t,Transfer-of-possession)
           & Donor(d,t)
           & Recipient(a,t)
           & Patient(p,t)
```

That is, every instance of a Agent-cause-self-to-possess event, there is an Agent, a Patient and a Donor, and, moreover, the Agent does some action that causes the Transfer-of-possession of the Patient from the Donor to that Agent. (This particular style of representation, in which there are objects corresponding to categories of events, is discussed further in Wilensky (1990), although nothing in the present discussion hinges of these details.)

(3) Transforming Lexical relations

We now come to those basic lex-cons of a word that can be related to other basic lex-cons through some non-core relation. I give a few brief examples here to illustrate the idea. For example, the basic lex-cons of "have", "give", and "get" mentioned above are each metaphorically extended to the domain of being infected with a disease. Thus, "have a cold" means being infected with a cold; "give Lynn a cold" means to infect Lynn with a cold, etc. (This example of metaphoric extension is discussed in Martin (1988).) In our terminology, for each of the initial set of lex-cons, there is another lex-con which is metaphorically based on the first.

I will call such relations *transforming* lexical relations. The intuition is that transformating lexical relations extend one sense to another that does not contain the original. Furthermore, let us say that one of the senses is *based* on the other, and that such a sense is an *extended sense* of another. A meaning of a word that is not based on another will be called a *central sense*. Thus, the possession meaning of "give" and the knot meaning of "tie" are central senses, while the infect lex-con of "give" and, say, the join-in-marriage lex-con of "tie" appeal to extensions of these senses via metaphoric relations.

A transforming relation is a synchronically logical relation, that is, it pertains when a speaker cognizes one sense as being based on another in some fashion. It does not imply that one sense is derived from another during ordinary language use (although the theory of learning alluded to below presumes that central senses would ordinarily be learned prior to extended ones). In cases in which a synchronic priority of one word sense relative to another cannot be established, I will consider both senses central.

---

thematic roles are specified too generally in these examples. In the spirit of this analysis, I might have specialized thematic roles for each distinct concept, e.g., an Agent-of-Agent-cause-self-to-possess, rather than simply Agent. That every such role is also an Agent role would be captured on the conceptual level. I generalized to more common case roles primarily for expositional purposes, but shall return to related issues below. See Wilensky (1990) for a more detailed analysis of the nature and status of such objects.

While metaphorical relations are quite common, and perhaps worthy of special attention, many other kinds of relations can be discerned. For example, I allege that the verb "tie" has a central sense of "make a knot" (as in "tie a knot"). This sense is core-related to the meaning "secure/connect by making a knot" (as in "tie Jan to the chair"). However, it is a transforming relation that extends it to the sense of physically connect (as in "tie the beam to the joist"), since the latter does not involve tying in the first sense. (These senses are related to nominal senses in an interesting fashion. The metaphorical sense of the noun "tie", meaning the abstract connection (as in "family ties"), seems to be most readily available; the sense of physical connection is rather restricted (as in "railroad tie"); and there appears to be no general nominal senses related to the central sense of the verb (i.e., a nominal lex-con "tie" meaning the rope used to connect two things).)

As a less obvious example, consider first the central sense of "gate", which I presume to be a movable structure controlling entrance through a barrier. While there are certainly metaphorical extensions of this concept (e.g., in electronics), consider the sense involved in a sporting arena, airport, and some university campuses, in which the gate refers to a passageway, and not to a movable structure controlling access. This relation appears to be some kind of "frame-complementation", in which the term used for central component is extended to a nearby component. Since this relation seems somewhat curious, I make the point of arguing that it is not an isolated case. Consider for example, the use of the term "hole" in the sense of "doughnut hole", which refers centrally to the empty space, and via frame-complementation to the dough ball formed creating this empty space; "window" can refer to the space in the wall as well as the object that fills it; "bed" can refer to the frame or the mattress and box spring that it supports.

We elaborate the nature of such extensions below.

A Test for Centrality

Speakers seem to judge central senses as *literal* or *actual* uses of a word. This observation is in sharp contrast with the view that non-literal language is non-lexicalized language. It also provides us with a convenient, but limited, tool for judging whether a sense is central or not. For example, consider the following:

(1a) You didn't literally cut your class.
(1b) ?You didn't literally cut your hand.

(2a) That's not really a doughnut hole; it's the part they removed to make the hole.
(2b) ?That's not really a doughnut hole; it's the space created from removing the hole.

(3a) You didn't buy an actual bed; you just bought a frame.
(3b) ?You didn't buy an actual bed; you just bought a frame, box spring and mattress.

In each case, the adverb or adjective selects the central sense; there is no question that the incompatible usages are conventionalized lexical entries.

While this test is useful, it is limited in several ways. First, it seems to apply to only polysemous senses of the same parts of speech; second, it seems difficult to apply to some parts of speech, for example, adverbs and connectives; third, it is unclear what the test says for the most closely related hypothesized senses, which of course are the most problematic. For example, it is unclear that there are strong judgments about cases like the following

(4) You didn't literally take Lynn to dinner.

Nevertheless, the test is both a useful tool for analyzing word senses, and a good demonstration that our intuitions about literality embody distinctions involving the centralize of word senses rather than their

lexicalization.

2.1. Kinds of Lexical Subregularities

I view the various lexical relations just discussed as linguistic subregularities. By subregularity, I mean any phenomenon that is systematic but not predictable. To qualify as a useful lexical relation, the relation must hold between a number of different examples; at the same time, most of these relations do not predictably hold wherever they might. For example, the systematic metaphoric structuring alluded to above, and studied extensively by Lakoff and Johnson (1980), comprise one important class of subregularities. Thus, it is worth recognizing the metaphoric lexical relation between central sense "open" and the "open" of "The play opens Thursday" because it appears in other places, such as "He opened with P-K4" and "The play closes Thursday". But the use of the relation is not predictable: "?He closed with QxB" is awkward at best.

The question arises as to just what kinds of subregularities there might be. One set of candidates is approached in the work of Brugman (1981, 1984) and Norvig and Lakoff (1987). In particular, Norvig and Lakoff (1987) offer six types of links between polysemous word senses in what they call *lexical network theory*. However, there appear to be many, many more links than these, or at least, many subcases which are necessary to distinguish. Indeed, I have no reason to believe that the number of such subregularities is bounded in principle, and thus do not offer a theory that pithily summarizes them.

I have developed a partial list of about 30 subregularities I have encountered that appear to have some predictive value. These may be found in Wilensky (1991). Here I list, in an informal rule format, some of these subregularities that I have found that extend word senses to a "non-literal" but conventionalized meaning:

(1) function-object-noun –> primary-activity-"determinerless"-noun
("the bed" –> "in bed, go to bed"; "a school –> at school"; "my lunch –> at lunch"; "the conference –> in conference")

(2) noun –> "involve-concretion"-verb
("a tree" –> "to tree (a cat)"; "a knife" –> "to knife (someone)")

(3) verb –> verb-w-role-splitting
("take a book" –> "take a book to Mary", "John shaved" –> "The barber shaved Bill")

(4) verb –> frame-imposition-verb
("take a book" –> "take someone to dinner", "go" –> "go dancing")

(5) noun –> resembling-in-appearance-noun
("tree" –> "(rose) tree"; "tree" –> "(shoe) tree"); "tiger" –> "(stuffed) tiger", "pencil" –> "pencil (of light)")

(6) noun –> having-the-same-function-noun
("bed" –> "bed (of leaves)")

(7) verb –> frame-imposition-verb
("take a book" –> "take someone to dinner", "go" –> "go dancing")

(8) activity-verb –> purpose-of-activity-verb
("tie a knot" –> "tie the beam to the joist"; "open the jar" –> "open the road"; "open the jar" –> "open the meeting")

(9) activity-verb –> enable-purpose-of-activity-verb
("open the door" –> "open the lock")

(10) object-noun –> central-component-of-object
("a bed" –> "bought a bed" [=frame with no mattress]; "an apple" –> "eat an apple [=without the core]"))

(11) central-object-noun –> frame-complement-noun
("gate" –> arena gate, campus gate, "hole" –> doughnut hole)

Consider the first rule. This rule is intended to capture the idea that, for some noun whose central meaning is a functional object, there is another nominal lex-con of that word that occurs without determination, and means something like the primary activity associated with the central sense of the term. For example, the word "bed" has as a central sense a functional object used for sleeping. However, the word can also be used in utterances like the following: "go to bed", "before bed", and "in bed", (but not, say, "*during bed"). In these cases, the noun is determinerless, and seems to mean something akin to being in bed for the purpose of sleeping for a significant period of time (i.e., to retire). Note, for example that it would be infelicitous to say I went to bed if I merely went over to a bed and sat down on it for a few minutes, or even if I took a short afternoon nap in one.

Other examples include "jail", "conference", "school" and virtually all the scheduled meal terms of English, e.g., "lunch", "tea", "dinner". For example, it would be misleading to say that I was "in jail" yesterday if I were visiting a relative, but acceptable to say that I was "in *the* jail" under such a circumstance. Note further than I can "send my children to school", but not "*to school down the street", while "to the school down the street" is acceptable. (Indeed, non-referential use of the noun presumably motivates its determinerless nature.) Also, which words conform to this subregularity is apparently a function of dialect. British English allows "in hospital" and "in university", while American English does not.

I belabor this example not because it in itself is a particularly important generalization about English, but precisely because it is not. That is, there appear to be many such facts of limited scope, and each of them may be useful for learning analogous cases.

For another, rather different example, consider the second rule, which relates function nouns to verbs. Examples of this are "tree" as in "The dog treed the cat" and "knife" as in "The murderer knifed his victim". The applicable rule states that the verb means some specific activity involving the central sense (e.g., "knife" and "tree", respectively). I.e., the verbs are treated as a sort of conventionalized denominalization. Note that the activity is presumed to be specific, and that the way in which it must be "concreted" is assumed to be pragmatically determined. Thus, the rule can only tell us that "treeing" involves a tree, but only our world knowledge might suggest to us that it involves cornering; similarly, the rule can tell us that "knifing" involves the use of a knife, but cannot tell us that it means stabbing a person, and not say, just cutting.

I leave to reader to peruse the remaining examples. However, it is important to reiterate that it is not each of these individual cases that is deemed to be important. Rather, it is the diversity and pervasiveness of conventionalzed non-literal word senses that they represent.

3. Lexical Acquisition

The import of this analysis for systems build lies in exploiting subregularities for language acquisition. Previously, it has been shown how it is possible to acquire metaphoric word sense extensions (Martin (1988)). Recently, I have analyzed extending this procedure to a general one of analogical word sense extension. A proto-algorithm for such a procedure can be found in Wilensky (1991).

4. Conclusions

I have distinguished three kinds of relations between lexical entries: core relations, grammatical relations, and transforming relations. Core-relations capture the commonality of meanings within a meaning complex; grammatical relations hold between different forms of a word whose semantic relation is completely productive; transforming relations map meanings into new meanings. Meanings not "produced" by transforming relations are central senses. The amount of knowledge to be contained in the lexicon is large, but we can take solace in its systematicity, and hope to exploit this systematicity in acquisition.

Many of the extended senses are intuitively non-literal in nature. Indeed, the use of the adverb "literally" is compatible with a central sense of a word; non-literal conventionalized word senses abound. Thus the lexicon is a repository for a large amount of diverse, conventionalized non-literal language.

5. References

(1) Brugman, Claudia. The Story of *Over*. University of California, Berkeley M.A. thesis, unpublished. Distributed by the Indiana University Linguistics Club. 1981.

(2) Brugman, Claudia. The *Very* Idea: A Case-Study in Polysemy and Cross-Lexical Generalization. In *Papers from the Twentieth Regional Meeting of the Chicago Linguistics Society*. pp. 21-38. 1984.

(3) Lakoff, G. and Johnson, M. *Metaphors We Live By*. University of Chicago Press, 1980.

(4) Martin, James. A Computational Theory of Metaphor. Berkeley Computer Science Technical Report no. UCB/CSD 88/465. November 1988.

(5) Norvig, Peter and Lakoff, George. Taking: A Study in Lexical Network Theory. In the *Proceedings of the Thirteenth Annual Meeting of the Berkeley Linguistics Society*. Berkeley, CA. February 1987.

(6) Nunberg, G. *The Pragmatics of Reference*. Bloomington, In.: Indiana University Linguistics Club. 1978.

(7) Wilensky, R. Primal Content and Actual Content: An Antidote to Literal Meaning. In the *Journal of Pragmatics*, Vol. 13, pp.163-186. Elsevier Science Publishers B.V. (North-Holland), Amsterdam, The Netherlands. 1989.

(8) Wilensky, R. Sentences, Situations and Propositions. In J. Sowa (ed.) *Formal Aspects of Semantic Networks*. Morgan Kaufman Publishers, Los Altos, California, 1990.