An Alternative to Deep Case for Representing Relational Information

Nigel WARD[1]
University of Tokyo

Abstract

No one has come up with a completely satisfactory set of deep cases relations (or thematic relations). The underlying reason is that any finite set of case relations can capture only some of the generalizations desired. I propose instead a feature-space representation of relational information, where the axes are such things as degree of responsibility, degree of activity, and degree of affectedness. The role of a participant in an event can then be described as a point in this space, allowing more accurate representation of relational information. The domain of validity of each relevant linguistic generalizations corresponds to a prototype-centered region in the space. This proposal is easy to implement.

1. Background

There are several things that a representation of relational information should do, and case does them all, but not very well.

1.1 The Problems with Case

The continued failure to come up with a satisfactory set of cases is a symptom of the impossibility of fixing a single set of cases that has all the desired properties.

A system of cases should work for the description of more than a few syntactic generalizations. Yet, for example, the set of things which can be passive subjects is not the same as the set of things that can be direct objects, and so any definition of patient can account for at most one of these two.

A system of cases should reflect similarities of form. Yet there are many dimensions of similarity, and any set of cases will account for only some. This can be seen by considering the fact that many prepositions, for example *"with"*, have meanings which span several cases (Tsujii & Yamanashi 1985).

Moreover, similarities of meaning do not always line up with similarities of meaning. For example, when assigning a case to *"wind"* in *"the wind closed the door"*, syntax suggests agent, as does the semantic feature 'no-covert-controller', but the the semantic feature 'not-animate' suggests that wind is an instrument.

For two languages the problems get worse; choosing a set of cases to capture the generalizations of one language tends to obscure the generalizations of another. To use another old example, a definition of agent that works well for English will not suffice to rule out inanimate subjects in Japanese.

A representation for relational information should be good not only for capturing similarities (generalizations) but also for precise representation. Here too, case often comes up short. For example, in both *"load the wagon with hay"* and *"load hay onto the wagon"*, *"wagon"* is traditionally assigned the same case, but this obscures the difference that in the first sentence the wagon is more affected -- it is more likely to be fully loaded. In general, the goal of precise representation suggests many specific cases with narrow meanings, but the goal of capturing generalizations suggests broader cases.

1.2 The State of the Art

To summarize, the problem with case is not that 'we haven't found the right set of cases yet' but that it is impossible to find a set of cases which does everything. The goals of representing various types of similarity conflict with each other, and these goals conflict with the goal of being able to precisely represent relational information. While there are refinements which help somewhat (sub-cases allow more precision, and multiple inheritance from supercases increases the number of generalizations capturable) the problems remain. (For further discussion of past work on case see Ward (to appear) and the references cited therein.)

Of course it is always possible to cope -- to make do with a set of cases which satisfies only some of the desiderata. For one thing, it is possible make do with limited expressive power. For example, many machine translation researchers appear satisfied if their case system is just detailed enough to account for choices among target language prepositions. It is also possible to make do with a system of cases that misses generalizations. Designers of machine translation systems, again, presumably make rough trade-offs as to the relative value of simplifying the parser (by choosing a set of cases convenient for the source language) or simplifying the generator (by choosing the cases to suit the target language).

The problems with a system of cases are not always identified as such. No one has ever written a paper saying 'I can't make case work for my application' -- shortcomings can always be compensated for by complicating the rules that refer to cases. That is, any proposal for a set of cases is unfalsifiable ... but it is possible to do better.

2. Proposal

2.1 Participatory Profiles

I propose to represent in detail the 'participatory properties' of objects. For example, in the scene involving Judas, Jesus, and a kissing, Judas can be described as active, volitional, very responsible, basically unaffected, a direct-cause, and so on. I will refer to the set of these properties as the 'participatory profile' of that object. A participatory profile is implemented as a vector of values over 'case features'.

[1] thanks Dan Jurafsky, Jane Edwards, Toshiaki Hisada, and Mitsubishi Heavy Industries.

kiss

Judas: agent

Jesus: patient

Figure 1:
A traditional representation

kiss

Judas:
- active .7
- volitional .7
- responsible .7
- affected -.2
- direct-cause .7

Jesus:
- active -.7
- volitional -.7
- responsible -.2
- affected .5
- direct-cause .2

Figure 2:
A 'participatory profile' representation

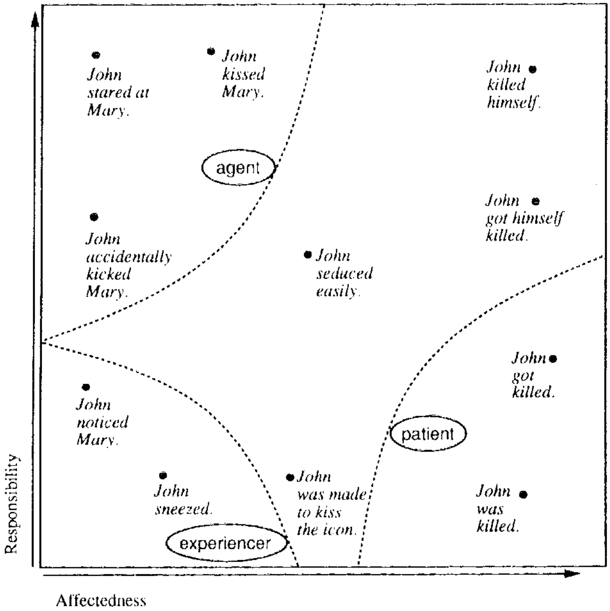

Figure 3: A slice of case space, chosen to focus on causal/inchoative events

For example, Judas as a kisser may be (−.2 affected), (+.7 responsible), and so on, as shown in Figure 2 and as contrasted to the traditional representation shown in Figure 1.

A participatory profile is a precise description. To illustrate this with a spatial metaphor, a participatory profile can be identified with a point in an n-dimensional space, which I will call 'case space', where the axes are the case features. Figure 3 shows an impressionistic projection of this space onto two dimensions, populated with sentences about John, positioned appropriately for his role in them. Superimposed on this with curved lines is a suggestion of the way that a traditional case account might divide up this space. This illustrates how case allows only a relatively coarse description, providing only the opportunity to describe a participant's role as being in a certain region of the space.

This proposal also makes it easy to explain similarities. For example, comparing the roles of *"yeast"* in *"yeast makes bread rise"* and *"spoon"* in *"eat with a spoon"*, they are similar in that both are concrete and directly acting, but different in that the yeast is not manipulable, nor is it identifiable as a separate entity afterwards. Profile representations of the roles of yeast and spoon can show that they are similar on specific shared dimensions, while not obscuring the differences on other dimensions. Profile representations also make it easy to quantitatively describe similarity on a single dimension. For example, it is possible to describe John as active in both *"John speculated in commodities"* and in *"John watched the ducks"*, but somewhat less active in the latter; there is no forced choice between assigning John to a case where he is active and one where he is not.

Case is traditionally considered to be a classification of the semantic relations between predicates and their arguments, but the proposal replaces it with an account of the roles of participants in events. In some languages things like individuation or definiteness, which would seem to have nothing to do with the verb, affect choice of case markers and constructions (Fillmore 1968; Hopper & Thompson 1980). Thus it seems that meaning relations should relate to the situation, not just to the predicate. (Here 'situation' is meant in a narrow sense (DeLancey 1991), where *"John asked Mary to leave"* involves two situations.)

2.2 Profiles and Language

Language refers to regions of case space. This is true, in particular, of 'case markers', constructions, and grammatical roles.

Consider for example the family of uses of *"of"* exemplified in *"John died of cancer"*. *"Of"* is used for causes which are direct causes, invisible, immaterial, of unknown origin, and at most only slightly controlled (DeLancey 1984). If direct-cause, visible, and so on are treated as case features, this use of *"of"* can be described as appropriate for participants in a certain region of case space. In general, the meanings of 'case markers', that is, words conveying relational information, can be identified with regions of case space

The meanings of some constructions also can be identified with regions. ('Construction' here is meant in the sense of Fillmore, Kay, and O'Connor (1988).) For a given participant, the extent to which its profile leads to selection of function words or to mobilization of constructions (affecting word order), or to both, depends entirely on the language.

Regions in case space can also be used to describe grammatical roles. For example, consider the set of things which can be subjects of passive sentences. Rather than saying that this includes themes, patients, and recipients, provided they meet certain conditions, we can describe this as the set of things which are highly topicalized, not very active, and more or less affected; this of course describes a region of case space. The set of things which can be direct objects is another region, overlapping that for passive subjects, but also including the region of highly affected things even if they are not at all topics, and excluding all highly topicalized things, and also mildly topicalized things unless they are highly affected. The set of things that can be passive subjects in Japanese is yet another region, again overlapping but slightly different.

To summarize the ways in which this proposal solves the problems raised in Section 1: it allows precise representation because instances are represented as points, and this does not conflict with the need to capture generalizations, because generalizations are represented as regions; and it can capture all generalizations because there is no assumption of correspondence between the regions required for different generalizations.

2.3 Examples and Details

To define the regions for various case markers by precisely specifying their boundaries would be onerous at best. Instead we can define these implicitly by reference to their prototypical meanings. For example, the prototypical use of *"of"* in *"die of cancer"* can be described as a point in case space. By computing the proximity of a participant's profiles to such prototypes for various case markers it is possible to determine the most suitable case marker for that participant.

Similarly for constructions; they are used when a participant's profile is sufficiently close to the construction's prototype. (Polysemous constructions can probably be analyzed as having several prototypes.) For example, one can analyze the Passive Construction as being relevant if a participant expressed in subject position has a profile is 'closer than 1.2' to the prototype (affected +1., volitional −1., responsible −1.), as shown in Figure 4.

Unlike prepositions, constructions' meanings do not form a partition of case space; thus a single point can fall into the regions of several constructions. It is sometimes necessary to employ more than one construction to adequately specify the profile of a participant. For example, to describe a participant who is active and possibly affected, but not responsible nor directly affected, the Passive and Causative Constructions must be used together, as in *"John was made to kiss the statue"*; each construction expressing some dimensions of the participant's profile. The idea of additive contributions from several constructions can also be applied to, for example, *"John was kissed"*, where *"John"* is a perfectly good subject, and also a perfectly good passive subject. This style of analysis means factoring out information, which of course makes for simple constructions.

Use of constructions provides a way to account for the 'subcategorization' properties of verbs. To explain why *"John broke the dish"* is English but *"*the magician vanished the rabbit"* is not, one can say that the verb *"break"* can participate in the Lexical Causative Construction but *"vanish"* can only participate in the Periphrastic Causative Construction. Thus it is not necessary to directly describe the allowable cases of a verb and their mappings to prepositions and grammatical roles; that information can be factored out into constructions. That is, the case frame (valence) of a verb can be explained in terms of the constructions the verb can participate in.

Grammatical roles can also be analyzed in terms

The Passive Construction
example: *"Mary was given a fork"*
condition for relevance:
expression of a participant closer than 1.2 to the prototype, using the weights below

	prototype	weights
affected	+1.	1
volitional	−1.	.5
responsible	−1.	.5

The Periphrastic Causative Construction
example: *"John made Mary go to Chicago"*
condition for relevance: expression of a participant closer than 3.5 to the prototype below

	prototype	weights
volitional	+1	1.
responsible	−1	1.
active	+1	1.
affected	−1	1.
direct-cause	−1	1.

The State Change Construction
example: *"John died"*
comment: rival to the Passive Construction; prevents *"John was died"*
condition for relevance: expression of a participant closer than 2.5 to the prototype below; also the availability of a state-change-verb

	prototype	weights
affected	+1	1.
volitional	−1	1.
responsible	−1	1.
object-of-force	−1	1.

Figure 4: Some constructions whose relevance depends on profiles

feature	prototype location	weights
topic	+1	.6
volitional	+1	.4
active	+1	.4
responsible	+1	.2
individuated	+1	.2
partial-cause	+1	.1
affected	−1	.2

Figure 5: The first constituent of the Subject-Predicate Construction

of prototypes — for example it has long been said that the prototypical direct object is probably that of *"kill"* — and these prototypes can be mapped into case space. Proximity to prototypes can then be computed. This allows, for example, the simple rule: 'for subject, select the participant which is closest in case space to the prototypical subject' (to slightly modify a proposal by Dowty (1991)). As some factors are more important than others, it is appropriate to assign weights to the various case features, to bias the computation of proximity. For example, the weights for subject shown in Figure 5 account for subject selection (in the context of the system described in Section 4), explaining:

1a) *John kissed Mary*
1b) *Mary made the boy eat a peach*
1c) *Mary was kissed by John* (if she is the topic in the larger context)
1d) *the wind broke a dish*
1e) *Mary was killed* and *Mary died*

This account of subject is more parsimonious than a subject hierarchy, that is, a list of cases in order of preference for which can become the subject (Fillmore 1968), plus rules for overriding it for the sake of topics. This description also obviates the need for explicit statements that topicness is more important than agentivity or that volition is more important than activity; such facts are simply encoded in the weights.

In the current implementation of case space, the range of values for each feature go from −1 to +1. Whereas participants can be located at any point in the space, it seems appropriate to site prototypes at the corners or edges of the space. A few more examples of profiles are shown in Figure 4, and many more in Ward (to appear).

3. Related Work

Although the synthesis is novel, many of the major components of the proposal have been previously proposed, if in somewhat different guises and for different purposes For example, Cruse (1973) and Delancey (1984) studied the components of various meaning relations, Labov (1973) and Miikkulainen and Dyer (1991) pioneered the use of vector spaces for describing meaning, Hopper and Thompson (1980) showed how to relate grammatical reflexes to lists of scalar-valued parameters (features), Hinton (1981) noted the possibility of using a 'distributed representation of roles', Tsujii and Yamanashi (1985) viewed cases in terms of prototypes and their extensions, Fukuda et al (1986) and Pederson (1991) introduced the spatial metaphor for meaning relations, and Dowty (1991) explained how to relate grammatical relations to prototype-structured clusters of meaning relations.

4. Implementation

I have built a parser (Ward 1992) and a generator (Ward to appear) which use participatory profiles. This section discusses the generator, not as a presentation of the best or only way to use profiles, but merely as a demonstration that case profiles are workable.

FIG, a 'Flexible Incremental Generator', produces English and Japanese sentences starting from a meaning representation, using spreading activation in a knowledge network. One task of a generator is, given an input including some items with case profiles, to build a sentence whose syntactic form and function words reflect those case profiles.

In FIG case features are implemented as nodes in the associative network. They are linked to constructions and words, with appropriate weights. For example, the node $\mathbf{responsible}_r$ has a link to the node $\mathbf{by}_w$, representing the word *"by"*, and this link has weight +1.

The participatory profiles of concepts in the input are represented by links to nodes for case features, appropriately weighted. For example, the node for Mary may have a link with weight .5 to $\mathbf{responsible}_r$ to represent a given input.

For such an input, when $\mathbf{mary}_n$ becomes activated, case features will become activated to the degree appropriate for her profile. In turn $\mathbf{by}_w$ and other prepositions will receive activation from these case features. The net effect is that the profile for a participant activates prepositions proportionally to their proximity in case space to that profile. (The measure of proximity computed is, to be precise, the dot product of the vector for the participant and the vector for the prototype.) The preposition whose prototype is closest will receive the most activation, and hence appear in the output. Like case markers, constructions receive activation from the profiles of participants, via case features. They thus become mobilized to the extent that there is a participant with a profile matching that of the construction. (Some case markers appear before the word they flag, others after, and so FIG has a distinction between activation from the profiles of concepts which remain to be expressed and activation from the profile of the concept just expressed.)

Constituents which involve profiles also are linked to nodes for case features; from these activation flows to concepts, and so the concept whose participatory profile is closest to that activated by a constituent will receive the most activation. (Actually the case feature nodes used for activation flow from constituents to concepts are distinct from those used for activation flow from concepts to case markers and constructions. That is, each case feature is implemented as a pair of nodes; this is for technical reasons.) There are multiple profiles in any non-trivial conceptualization, and it would seem that crosstalk might be a problem, but this has not been the case in FIG, primarily because generally there is one construction and one concept with enough activation to dominate.

FIG originally expected deep case relations in its inputs, and its grammar and lexicon referred to those cases. One problem was that, as I extended FIG's coverage of the two languages, the number of cases kept growing and the grammar got uglier and uglier. In particular, there were lengthening lists of possible cases for constituents, for example there was a list of four possible cases to use for subject. Switching to profiles solved these problems. Conversion was relatively easy; other than the new references to profiles, the grammar did not need to be changed. The additional computation required is negligible.

FIG currently uses 10 case features: volitional, responsible, active, affected, direct-cause, partial-cause, individuated, topic, object-of-force, and touched; these replace the cases agent, instrument, patient, experiencer, cause and percept. At this point the meanings of the case features derive less from their names than from the way they are related to the constructions of Japanese and English. This is because the numeric values for the profiles, although originally chosen according to common sense and with reference to the literature, have had to be tuned in the course of making FIG able to generate sentences in both languages for a largish number of inputs. I ascribe no special significance to the particular set of profiles currently in use: they are specific to FIG's current grammar and implementation details.

5. Summary and Hopes

To summarize the advantages of the proposal: Participatory profiles are a representation mechanism that allows both precision and generalization-capturing. Precision is important for being able to represent accurately the information that people can get from language, and it will probably also be useful for artificial intelligence systems in the near future. Better generalization-capturing allows simpler and better grammars. This is important for linguistics, and also for computational linguistic, where the cash value will be improved manageability and performance for natural language systems. One example is machine translation. If the parser/understander arrives at a narrow enough case profile for a participant, then it is possible to directly find the relevant target language constructions by computing in which regions the point lies. It should thus be possible to eliminate the need for contrastive knowledge relating the regions of the various constructions and words of the two languages.

Judging from my experience converting FIG to profiles, these advantages may be easy to achieve in practice. Of course, to come up with a general theory of relational information will require a great deal more work, both on the mechanism and on the analysis of language.

This proposal is in some ways a logical continuation of Fillmore's (1968) research program. Fillmore wanted to capture linguistic generalizations in terms of meaning, not syntactic structures. In Ward (forthcoming) I suggest that a processing model can dispense with surface syntax structures also; doing without case eliminates yet another type of intermediate structure typically interposed between thought and language, allowing an even more direct account of linguistic generalizations in terms of meaning.

References

Cruse, D. A. (1973). Some Thoughts on Agentivity. *Journal of Linguistics*, 9:11–23.

DeLancey, Scott (1984). Notes on Agentivity and Causation. *Studies in Language*, 8(2):181–213.

DeLancey, Scott (1991). Event Construal and Case Role Assignment. In *Berkeley Linguistics Society, Proceedings of the Seventeenth Annual Meeting*.

Dowty, David R. (1991). Thematic Proto-Roles and Argument Selection. *Language*, 67:547–619.

Fillmore, Charles J. (1968). The Case for Case. In E. Bach & R. Harms, editors, *Universals in Linguistic Theory*, pp. 1–88. Holt, Rinehart, New York.

Fillmore, Charles J., Paul Kay, & M. C. O'Connor (1988). Regularity and Idiomaticity in Grammatical Constructions: The Case of Let Alone. *Language*, 64(3).

Fukuda, Kanae, Jun Yamaguchi, Jun ichi Tsujii, & Masaaki Yamanashi (1986). Kaku Kaishaku to Ninchi Kikoo, Sono 2 (Case Interpretation and Cognitive Structure, Part 2). In *Proceedings 3rd Japanese Cognitive Science Society*, p. 66.

Hinton, Geoffrey E. (1981). Implementing Semantic Networks in Parallel Hardware. In Geoffrey. E. Hinton & J. A. Anderson, editors, *Parallel Models of Associative Memory*, pp. 161–188. Lawrence Erlbaum Associates.

Hopper, Paul J. & Sandra A. Thompson (1980). Transitivity in Grammar and Discourse. *Language*, 56:251–299.

Labov, William (1973). The Boundaries of Words and their Meanings. In Joshua Fishman, editor, *New Ways of Analyzing Variation in English*, pp. 340–373. Georgetown University Press.

Miikkulainen, Risto & Michael Dyer (1991). Natural Language Processing with Modular PDP Networks and Distributed Lexicon. *Cognitive Science*, 15:343–400.

Pederson, Eric (1991). The Ecology of a Semantic Space. In *Berkeley Linguistics Society, Proceedings of the Seventeenth Annual Meeting*.

Tsujii, Jun-ichi & Masaaki Yamanashi (1985). Kaku to sono Nintei Kijun (Cases and Criteria for their Identification). Technical Report 52-3, Information Processing Society of Japan, Natural Language Working Group, Tokyo.

Ward, Nigel (1992). An Evidential Model of Syntax for Understanding. Technical Report 88-3, Information Processing Society of Japan, Natural Language Working Group, Tokyo.

Ward, Nigel (forthcoming). A Parallel Approach to Syntax for Generation. *Artificial Intelligence*.

Ward, Nigel (to appear). *A Connectionist Language Generator*. Ablex. revised and extended version of A Flexible, Parallel Model of Natural Language Generation, Ph. D. thesis and Technical Report UCB-CSD 91/629, Computer Science Division, University of California at Berkeley.