Supporting rule-based representations with corpus-derived lexical information.

Annie Zaenen
Cleo Condoravdi
Daniel G. Bobrow
PARC
3333, Coyote Hill Road
Palo Alto, CA, 94304, USA
{zaenen,condorav,bobrow}@parc.com

Raphael Hoffmann
University of Washington
Box 352350
Seattle, WA, 98195, USA
raphaelh@cs.washington.edu

Abstract

The pervasive ambiguity of language allows sentences that differ in just one lexical item to have rather different inference patterns. This would be no problem if the different lexical items fell into clearly definable and easy to represent classes. But this is not the case. To draw the correct inferences we need to look how the referents of the lexical items in the sentence (or broader context) interact in the described situation. Given that the knowledge our systems have of the represented situation will typically be incomplete, the classifications we come up with can only be probabilistic. We illustrate this problem with an investigation of various inference patterns associated with predications of the form 'Verb from X to Y', especially 'go from X to Y'. We characterize the various readings and make an initial proposal about how to create the lexical classes that will allow us to draw the correct inferences in the different cases.

1 Introduction

Machine Reading requires a level of Natural Language Processing that allows direct inferences to be drawn from the processed texts. Most heavy duty inferencing will be done by a reasoning engine working on the output of the linguistic analysis (with possible loops between the two) but for this to be possible, the linguistic analysis should deliver representations where a certain level of disambiguation and content specification has been done. For instance, a human will draw different conclusions from the following two sentences about the position of the referent of the subject: 'Eric went from Paris to Lyon' and 'The road went from Paris to Lyon'. The first sentence implies that a person named Eric was in Paris at some time and in Lyon at a later time, whereas the second sentence implies that a part of the road was in Paris and a part of it was in Lyon at the same time. For the reasoner to draw such conclusions, the linguistic analysis should assign appropriate roles to the subject argument and the *from-to* adjunct or argument phrases of the verbal predicate *go* so as to convey that the first sentence involves movement, while the second involves spatial extent.

In this paper we look at a range of such inferences associated with *from-to* phrases. We limit ourselves to rather simple cases of the use of *from-to* phrases: those that describe no change or gradual changes in the physical world. We show that beyond inferences about time-dependent locations and spatial extent of particular entities, *from-to* phrases give rise to inferences about change of an entity in some dimension (e.g. temperature or width) either through time or through space. We first discuss the inferences we would like to be able to draw, and describe features of a representation that captures enough distinctions to enable these inferences to be drawn. This allows us to isolate the factors leading to such inferences. Finally, we give a preliminary sketch of a corpus analysis that would help make the required distinctions

and characterize appropriate lexical classes.

2 Some simple inferences

Consider the following sentences:

1. Eric went from Paris to Lyon.
2. The road went from Paris to Lyon.
3. The meeting went from 3 p.m. to 5 p.m.
4. The temperature in the room went from 20 degrees to 30 degrees from 10 to 11 a.m.
5. The temperature went from 20 to 30 degrees from the front to the back of the room
6. The temperature went from 20 degrees to 30 degrees.
7. The room went from 20 to 30 degrees.

As indicated above, we would like the system to be able to conclude from (1) that Eric was in Paris before being in Lyon, and from (2) that one part of the road is in Paris whereas another part is in Lyon at the same time. From (3) the system should infer that the mentioned event, the meeting, started at 3 p.m. (or no later than 3 p.m.) and ended at 5 p.m. (or no earlier than 5 p.m.). From (4) the system should infer that the value of the function *temperature* as it applies to the room increases over the given temporal span. It is worth noting at this point that the two sets of *from-to* phrases in (4) play different roles. The temporal *from-to* phrases specify the relevant domain of the temporal argument of the function, while the measure *from-to* phrases specify the range of the function on the given domain. (5) has a similar implication to that of (4), that the temperature changes, but this time over a spatial dimension: the temperature is implied to vary in different parts of the room, being 20 degrees in the front of the room and 30 degrees in the back. Again the two sets of *from-to* phrases in (5) play different roles. The spatial *from-to* phrases specify the relevant domain of the spatial argument of the function and the measure *from-to* phrases specify the range of the function on the given domain. (6) and (7) have similar implications to those of (4) and, in the right context, to those of (5) but they present challenges of their own. In (6) the temporal (or spatial) dimension is implicit and needs to be inferred. (7) requires the inference that a change of the values of the function *temperature* is involved.[1]

These examples show that sentences that have substantially the same syntax and even use the same main verb can exhibit very different relations between their parts. The first question we want to address is how to explicate these differences and the second question is how to get from the words used in these sentences to the information needed about their type of referent to ensure the right interpretation in each case.

The verb 'to go' is, of course, not the only one that exhibits this behavior. The difference in interpretation between examples (1) and (2) can also be found with manner-of-motion verbs such as 'run' and 'zigzag'. Some verbs do lexically encode a particular functional dimension, such as temperature or width. These are known as degree achievements (Dowty, 1979; Abusch, 1986).[2] Examples of degree achievements include 'widen', 'lengthen', 'shorten', 'cool', 'age'. They exhibit similar patterns of modification with *from-to* phrases as we saw above:

8. The road widens from Palo Alto to Menlo Park.
9. The road widens from 12 to 24 feet.

Here 'widen' is interpreted statively, like 'go' in (2), and the two sentences imply spatial change in width, over subparts of the road. The two *from-to* phrases, however, have a different function giving rise to different implications. (8) implies that the road is wider in Menlo Park than it is in Palo Alto. (9) specifies the relation between the measures of width at two different subparts of the road. The *from-to* phrases in (8) specify

[1] It is not clear that the change has to be in one directional in all cases:

> This summer, the temperature went from 20 degrees to 30 degrees.

In this example, it seems that the temperature varied from 20 to 30 degrees, not necessarily that 20 degrees was a starting point or 30 degrees an end point. See section 4.1 for some further discussion.

[2] In English most degree achievements are derived from gradable adjectives. When this is the case, the meaning of degree achievements and underlying adjectives is systematically related, as argued in (Hay et al., 1999).

the domain of the spatial argument of the function *width* as it applies to the referent of 'the road'. Those in (9) specify the range of the values of the function *width* as it applies to different parts of the referent of 'the road'.

In what follows we will distinguish between *extent readings* and *change readings*. Extent readings specify, in full or in part, the temporal or spatial extent of a temporal or spatial entity, as seen in (3) and (2). Change readings specify the values of a function as applied to a given entity through a temporal or spatial span. The function is either determined directly by the verb, as in (8) and (9), or by the verb in combination with one of its arguments, as in (4) – (6), or it has to be inferred, as in (1) and (7).

3 Representing the different readings

For the sake of concreteness, in this section we show how the distinctions discussed above are represented and implemented in AKR, an abstract knowledge representation language into which sentences are mapped after they are parsed in the NL system developed at PARC (Bobrow et al., 2007). The idea behind AKR is to canonicalize many variations of an input text with the same underlying meaning into a more uniform representation. This ought to make the task of interfacing with reasoners easier.

The AKR of a sentence consists of a list of assertions. Terms are generated for each of the content words of a sentence, such as verbs and nouns, and are associated with assertions about the types of events and objects their corresponding words refer to. Predicates and their arguments or modifiers are related via role relations. The inventory of roles we use extends the set of semantic or thematic roles often assumed in linguistic analyses and found in resources such VerbNet or FrameNet. It includes among other things temporal or spatial relations of inclusion, precedence, etc.

We assume that sentences with *from-to* phrases imply the existence of a path and that the further information about the path specified is about the "location" of its initial and final points. In representing such sentences a term is created to represent a path and the path term is linked by a role *initial* to the term for the complement of *from*, and by a role *final* to the term for the complement of *to*. On our analysis then the *from-to* phrases are used to specify restrictions on the path term and do not translate into thematic roles relating the verbal predicate and the complement NP, such SOURCE or GOAL. The path term is related to the verbal term via different roles, depending on the type of interpretation. Below is an example that shows the role relations in AKR for sentence (1).

role(theme, go:13, Eric:7)
role(mpath, go:13, path:23)
role(initial,path:23,loc(-at-,Paris:4))
role(final,path:23,loc(-at-,Lyon:6))
role(dimension,path:23,loc)

3.1 Extent interpretations

In extent readings the subject argument denotes an entity extended in space, as seen in (2), or a non-punctual event, as seen in (3). The verb itself does little work other than to signal that the *from-to* phrases give information about the spatial or temporal extent of its subject argument. The way they do that is by saying that the given path is a spatial or temporal part of the entity that is the referent of the subject argument. Let us start with the representation of (3), as the representation of its meaning in our terms is quite intuitive. Temporal paths, such as from-to-span:11, correspond to time periods.

role(initial,time-span:11,timepoint(-at-,3pm))
role(final,time-span:11,timepoint(-at-,5pm))
role(temporalWithin,time-span:11,meeting:1)

It should now be clear that the representation for the spatial extent reading would differ minimally from that of the temporal extent reading: the relation between the path and the road terms would be that of spatial inclusion and the dimension of the path is locational.

role(initial,path:23,loc(-at-,Paris:4))
role(final,path:23,loc(-at-,Lyon:6))
role(spatialWithin,path:23,road:10)

3.2 Change interpretations

As discussed in section 2, change interpretations establish a dependency between two paths which should be represented explicitly. The paths themselves may be specified overtly by *from-to* phrases or they may be implicit. Functionally relating two paths of this type was first discussed, to our knowledge, in (Jackendoff, 1996) and further developed in (Gawron, 2005) and (Gawron, 2009).

Let us consider first example (4), where the two paths are given explicitly. (4) implies a change in the temperature of the room over time so the function *temperature* should be construed as time-dependent. The temporal path specifies the time period over which the given change in temperature takes place; the scalar path *partially* specifies the range of the function over the given temporal domain. What we can conclude for certain from (4) is that the temperature in the room was 20 degrees at 10 a.m. and 30 degrees at 11 a.m. The sentence gives no specific information about the temperature of the room in between 10 and 11 a.m. though in this case, given that change in temperature is continuous, we can conclude that every degree between 20 and 30 was the temperature of the room at some point within the relevant time period.

In order to represent the dependency between the two paths we use a higher order predicate *path-map* that specifies a function, that varies over a range (in this case the scalar path from 20 degrees to 30 degrees) with a domain (in this case the temporal path from 10 a.m. to 11 a.m.). More generally: the higher-order predicate, **path-map(F,D,R)**, relates a function F and two posets D and R. The path-map relation expresses that the image of D under F is equal to R.[3] For (4) we end up with the following representation.

role(scale,go:5,path:4)
role(dimension, path:4,temperature)
role(initial,path:4,temperature(-at-,20 deg))
role(final,path:4,temperature(-at-,30 deg))
role(initial,time-span:11,timepoint(-at-,10am))
role(final,time-span:11,timepoint(-at-,11am))
path-map(function(temperature,room:2),
time-span:11,path:4)

The fact that path:4 is a scalar path is marked by relating it to the verbal term via the role *scale*.

The other examples discussed in section 2 receive representations based on this model. (5) implies a change in the temperature of the room over its spatial extent oriented from the front to the back, so the function *temperature* should be construed as location-dependent. Below we give the assertions for the representation of (5) that differ from those of (4). Note the additional assertion relating the spatial path term to the room term.

role(initial,path:11,loc(-at-,front:10))
role(final,path:11,loc(-at-,back:12))
role(spatialWithin,,path:11,room:2)
path-map(function(temperature,room:2),
path:11,path:4)

The representation of sentences with degree achievements, such as *The road widens from 12 to 24 feet from Palo Alto to Menlo Park*, would the same in all relevant respects except that the dimension of the scalar path would be determined by the verb, in this case being *width*.

To derive full representations for (6) and (7) we need to be able to infer the second and the first argument of *function*, respectively. Moreover, we need to fix the dimension of the implicit path. Generally, when only one path is specified overtly, as in (6), (7) and (8) and (9) the existence of the other type of path is understood. When only the range path is given, the understood domain path can be either temporal or locational.

We come now to the prototypical use of a *from-to* phrase with verbs like 'go' to describe movement whose origin is specified by the *from* phrase and whose destination is specified by the *to* phrase. We gave a preliminary representation for (1) at the beginning of section 3. Missing from that representation is the explicit link between the location of the theme argument during

[3]Depending on what F, D and R are, this mapping may also be order preserving, i.e. for all elements x, y in D, if x precedes y then F(x) precedes F(y).

the time of the movement. This link, of course, can now be given in terms of the following path-map assertion:

path-map(function(location,Eric:7),
time(go:13),path:23)

4 Which elements in the sentence guide interpretation?

In our system roles and dimensions are introduced by rules that take the output of the syntactic parse of the sentence as input. The exact form of these rules need not to concern us here. But an important question for NLP is where the information comes from that allows us to determine which role and dimension a path has. As the examples show, the verb is not necessarily the place to look: most of the examples use the verb 'to go'.

In fact, the information can come from various places in the sentence (or the broader textual context: ellipsis and anaphoric relations play their usual roles here). Moreover in some cases information about, say, the dimension can come from the arguments of *from* and *to* whereas in other cases this information can come from the verb. 'Widen' for instance imposes the width-dimension but if we use the verb 'to go' to describe a widening event, the information about the dimension has to come from the arguments of *from* and *to* and the subject.

Similar problems arise with respect to the determination of the roles. Example 1 and 2 seem to have straightforward interpretations where the path role in the first case is clearly a movement path whereas in the second case we have to do with a stative interpretation. At first blush, it seems that this information could be straightforwardly lexically encoded: people move and roads don't. But further reflection shows that this will not do. Take the following example:

10. The train went from one end of the station to the other.

In this case we can have two interpretations: either the length of the train is such that it covers that of the whole station or the train moved from one end of the station to the other. What is important is not an intrinsic characteristic of the lexical item but whether it is appropriate for the extent (length) of its referent to be measured by the *from-to* phrase.

Some more or less stable relations between syntax and semantics can help us determine which analysis to give. For instance, the starting and end points of movement paths and stative locational paths are referential (in contradistinction to those of scalar paths). As such, they tend to be expressed by proper names or by a noun phrase with a determiner.[4]

Manner of motion verbs are surprisingly uninformative: many of them can have a moving object or a stationary object or a function such as the temperature as their subject. The combinations summarized in the following are all possible:

11. Liz/the road/the temperature
went/crawled/moved/meandered
from X to Y.

With verbs of inherent directed motion, the verb contributes a polarity for the direction but very little else, as example 12 illustrates:

12. Liz/the road/the temperature
descended/climbed/ascended/fell/tumbled
from X to Y.

Again whatever information there is about the type of path or the dimension it has to come from the subject or from the *from-to* arguments. *From-to* arguments can give the necessary information about the dimension (locations, money, time, degrees) but when they are scalar or temporal, the measurement units will often be omitted and the theme will indicate the dimension.

Degree achievements tend to be more specialized. They indicate the dimension (width, temperature). Lexicons can contain many of the function names but will not help with the cases of metonymy (where an argument is given instead of the name of the function itself).

[4]There are, however, exceptions:

He ran from where Bill was to where the field ends.
His tattoo goes from head to toe.
The path meanders from mountain to mountain.

4.1 Characterizing components of the representations

In the previous subsection we have discussed different types of *from-to* phrases, and the roles that link the elements of the representations of these types. The question we address now is how we can provide our system with the necessary information to make these distinctions. This is a preliminary investigation as yet without implementation.

Ideally, we would have ontologies to give us the right characteristics of the entities underlying our lexical items and we would have adequate mappings from the lexical items to these ontologies. These ontologies and these mappings are currently not available. Natural language processing applications, however, have taught us that even if humans can do surprising things and language can express surprising thoughts, most of the time, the reality that human language expresses is rather predictable, so that the mapping to ontologies can up to a certain point be mimicked by probabilistic feature assignments to lexical items. For 'Eric' we can assume that with a high probability it will be the theme of a movement path and whereas for 'the road' a high probability assigns it as the theme of a stative path. In other cases, however, we need concrete co-occurrence statistics to assign the right representations. Next, we sketch a preliminary investigation of some Wikipedia data that can be brought to bear on this issue. We indicate how the data might help and point out some of the new problems it brings up.

A first question that arises is of how much practical relevance the different types that we have discussed are. We looked at the first 100 'went from X to Y' sentences pulled out of Wikipedia parsed with the Stanford dependency parser, that had the required syntactic pattern and found that 61 fell into the categories described in the previous sections (gradual change or no change in the physical domain) whereas about 39 are clearly transformational *from-to*'s (for instance 'The SU-152 went from design concept to field trials in a record twenty-five days'). Of these 61, 4 had temporal *from-to* modifiers, 19 had various scales or numeric *from-to* modifiers and 38 were locational. Of the locational ones, 11 had a stationary reading and 17 had a movement reading. So all the cases under discussion are well represented in naturally occurring text.

A second question is how we can obtain the relevant features from the data. We see four potential methods: (1) the characterization of words within existing ontologies like WordNet (Miller, 1995), (2) the combination of stated facts through reasoning, (3) co-occurrence statistics of words in text, and (4) solicitation of novel features from human annotators. We illustrate these methods based on Wikipedia examples.

A first idea might be that there is at least a straightforward ontological characterization for difference between the movement and the stative reading: for the movement reading we require living beings and for the stative reading we require long stationary entities. These impressions are, of course, not completely wrong but in the first case, we have to include in the living beings not only groups such as brigades but also ships (as in 'She went from the Red Sea to the Mediterranean to relieve USS Coral Sea ...'), flights (as in 'This flight went from Spitsbergen (Svalbard) to Alaska nonstop, so there is little doubt that they went over the North Pole.') and messages (as in 'The message went from the Palace in Stockholm to the King at Drottningholm.'). And in the second categories we have not only roads and various transportation lines but also borders (as in 'The boundary of Manila province went from northeast to southwest, ...') and trade routes and things such as (rifle) suppressors as in 'The suppressor, 2 inches in diameter, went all the way from the back of the barrel to well beyond the muzzle ...'). A quick inspection of WordNet shows that there is no interesting ancestor node that covers all the movement cases but it also suggests that a great number of the cases can be covered with 'conveyance, transport' together with 'motion, movement, move' as well as 'organism, being'. But 'organism, being' also covers 'plants' and 'sitter' and 'stander' and other subclasses that

don't seem to be plausible candidates for the movement analysis. There is no interesting hypernym for both 'road' and 'border' before we get to the useless level of 'object, physical object' and no already existing ontology will help with the suppressor case. Thus we might get some data by using the first method but most likely not everything we want.

As far as the arguments of the *from-to* phrases themselves, locations can be indicated by place names, institution names, nouns referring to locations, but also nouns referring to spatial located entities that we do not think of as locations, such as parts of pieces of equipment. The very limited inspection of data we have done up to now does not lead us to expect that the nature of the *from-to* arguments occurring with movement readings is very different from that found with stationary readings. In the current state of affairs, many of the arguments of the *from-to* phrases can be found either in gazetteers or through the analysis of a reasonably well-circumscribed spatial vocabulary.[5]

Some cases, however, fall outside of these resources. The most interesting problem is presented by the reference to spatial entities that are not clearly flagged as locations in ontologies, such as those found in the suppressor-sentence ('The suppressor, 2 inches in diameter, went all the way from the back of the barrel to well beyond the muzzle ...') above. We admit that his type of sentence seems to be rather rare in the Wikipedia corpus but it is problematic because detailed ontological representations of even common objects are not readily available. Wikipedia, however, has some information that might help one to formulate reasonable hypotheses about parts. For instance, the article that contains the suppressor-sentence, also contains a structured specification of the carbine under description mentioning the barrel and the muzzle. Here we need to use the second method, reasoning. The question then becomes whether we can find reasoning patterns that are general enough to give interesting results.

[5]Whereas it is possible to enumerate an extensive part of the relevant vocabulary, there is no extensive description of meaning contribution of these elements.

The third method, already demonstrated in the context of semantic parsing (Poon and Domingos, 2009), seems also to be promising. For instance, even staying within the class of movement verbs, different verbs have different signatures that might help us with the classification of their subjects and their *from-to* arguments. While 'go' has indeed the wide range of meanings that we expected, 'run' is rather different: apart from three examples where 'run' refers to the movement of living beings and three referring to vehicles moving, the other examples of the combination of 'run' with *from-to* fall in two classes: indications of the spatial extent of roads, railways and the like (27) and temporal extensions of shows, games or strips running (16). The nature of the corpus has certainly an influence here (Wikipedia does not contain narrative texts) but this type of information might be valuable to disambiguate parses: if we can distinguish the cases where 'run' occurs with spatial extent readings and the cases where it occurs with temporal extent meanings, we can harvest a set of possible subjects that are also possible subjects for the spatial extent meaning of 'go'. The distinction between the two readings of 'run' is not very difficult to make as most of the temporal extent readings of 'run' have a temporal *from-to* phrase.[6]

A different way in which the characteristics of specific verbs or verb argument combinations might at least probabilistically disambiguate possible readings is illustrated with a difference between 'go' and 'range' with scalars. In section 3.2, we observed that scalar 'go' does not always imply that there is a steady increase or decrease over time or space. However in all the numerical or scalar examples except for one in our first sample, the interpretation implies such

[6]But those readings themselves bring up a new classificatory problem: most of the time the subject is an event, a show, or a game. However, in most cases the meaning is not that one performance of the show ran for several months or year but that several successive performances ran. Moreover, the construction cannot only be used with event-referring expressions but also with entities such as 'strips'. Here we get into problems of regular polysemy. The treatment we have given above needs to be complicated to take these into account.

a steady increase or decrease. We also examined the sentences with 'price ranged' and 'price went' in the whole of Wikipedia. Unfortunately there are very few examples but for these, the difference in interpretation for 'range' and 'go' seems to hold up: all 4 examples with 'go' had the interpretation of steady increase or decrease. So 'the price ranged ...' and 'the price went ...' statistically might get a different interpretation even if in some cases 'go' can be synonymous with 'range'.

Finally, there is a possibility that due to sparseness some required features can neither be derived from existing ontologies nor from natural language text itself. For example, in 'The 2006 Trek the Trail event was organised on the Railway Reserve Heritage Trail and went from Mundaring to Darlington' we assume an extent interpretation, and may thus be inclined to classify all events that way. However, in 'The case Arklow vs MacLean went all the way from the New Zealand High Court to the Privy Council in London.' we assume a change interpretation (movement), although WordNet sees 'event' as a hypernym of 'case'. Interestingly, it is not the arguments that determine the right interpretation here, but rather our distinction between different kinds of events: those for which spatial extent is important (street festivals) and those for which not (lawsuits). More generally, in cases where we are unable to make such fine distinctions based on features derived from available corpora, we can use our fourth method, soliciting additional features from human annotators, to group concepts in novel ways.

5 Conclusion

In this paper we first described the distinctions that need to be made to allow a correct interpretation of a subclass of *from-to* sentences. We then looked at the resources that are available to help us guide to the correct interpretation. We distinguished four different ways to obtain the information needed: features in an existing ontology, features statistically derived for the relations used with a concept, features computed through reasoning and features obtained through human annotation. We saw that a small, very preliminary examination of the data suggests that the three first methods will allow us to make the right distinctions in an important number of cases but that there will be cases in which the fourth method, human annotation, will be necessary.

Acknowledgments

This material is based in part upon work supported by the Air Force Research Laboratory (AFRL) under prime contract no. FA8750-09-C-0181. Any opinions, findings, and conclusion or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the view of the Air Force Research Laboratory (AFRL).

References

Dorit Abusch. 1986. *Verbs of Change, Causation, and Time*. Report CSLI, Stanford University.

Daniel Bobrow, Robert Cheslow, Cleo Condoravdi, Lauri Karttunen, Tracy, Rowan Nairn, Valeria de Paiva, Lotti Price, and Annie Zaenen. 2007. PARC's Bridge question answering system. In *Proceedings of the GEAF (Grammar Engineering Across Frameworks) 2007 Workshop*. Stanford, CA.

David Dowty. 1979. *Word Meaning and Montague Grammar: The Semantics of Verbs and Times in Generative Semantics and in Montague's PTQ*. Springer.

Jean Mark Gawron. 2005. Generalized Paths. *SALT 17*.

Jean Mark Gawron. 2009. The Lexical Semantics of Extent Verbs.

Jennifer Hay, Christopher Kennedy, and Beth Levin. 1999. Scale structure underlies telicity in 'degree achievements'. pages 127–144.

Ray Jackendoff. 1996. The Proper Treatment of Measuring Out, Telicity, and Perhaps Even Quantification in English. *Natural Language and Linguistic Theory 14*, pages 305–354.

George A. Miller. 1995. Wordnet: A lexical database for english. *Communications of the ACM*, 38(11):39–41.

Hoifung Poon and Pedro Domingos. 2009. Unsupervised semantic parsing. In *EMNLP*.