Formal Ontology in Information Systems
M. Donnelly and G. Guizzardi (Eds.)
IOS Press, 2012

doi:10.3233/978-1-61499-084-0-221

Axiomatizing Change-of-State Words

Niloofar Montazeri[1] and Jerry R. Hobbs
Information Sciences Institute
University of Southern California
Marina del Rey, California

Abstract. We describe our effort in defining change-of-state words from Core WordNet by constructing axioms anchored in core theories that are crucial in characterizing event words, including change of state, composite entity, scales and event structure. Our methodology consists of three steps: Analyzing the structure of a word's WordNet senses, writing axioms for the most general senses, and testing the axioms on textual entailment pairs. We also describe some common issues we faced and decisions we had to make during axiomatization. We look at two specific textual entailment examples in detail to illustrate the power of this method.

Keywords. Ontology of Event Structure, Change-of-State, Deep Lexical Semantics, Commonsense Knowledge, Textual Entailment

Introduction

Words describe the world, so if we are going to draw the appropriate inferences in understanding a text, we must have underlying theories or ontologies of aspects of the world, and we must have axioms that link these to words. This includes domain-dependent knowledge, of course, but 70-80% of the words in most texts, even technical texts, are words in ordinary English used with their ordinary meanings. We are engaged in an enterprise we call "deep lexical semantics", in which we develop various core theories of fundamental commonsense phenomena and define English word senses by means of axioms using predicates explicated in these theories. Among the core theories are cognition, microsociology, and the structure of events. The last of these is the focus of this paper (more specifically change-of-state words). We use textual entailment pairs to test out subsets of related axioms. This process enforces a uniformity in the way axioms are constructed, and also exposes missing inferences in the core theories. In our previous experimental work [1], we focused on a systematic way to find holes in the core theories by axiomatizing a number of event words and testing them on a small set of textual entailment pairs. In this paper we address some common issues in axiomatizing change-of-state words and in particular show how our decision to axiomatize general senses influences the entailment power of our axioms.

In Section 1 we describe three aspects of the framework we are working in—the logical form we use, abductive interpretation and defeasibility, and the relevant core theories. In Section 2 we describe the methodology we use for constructing axioms,

[1] Corresponding Author; E-mail: niloofar@isi.edu.

deriving from WordNet senses a smaller set of general or abstract "supersenses" and encoding axioms for these. In section 3 we describe our effort in axiomatizing change-of-state words from Core WordNet and describe some common issues and possible solutions. In section 4 we introduce our evaluation method which consists of collecting textual entailment pairs related to change-of-state and testing our axioms on these pairs. Finally we present a detailed analysis of two textual entailment examples.

1. Framework

We use a logical notation in which states and events (eventualities) are reified. Specifically, if the expression `p(x)` says that `p` is true of `x`, then `p'(e,x)` says that `e` is the eventuality of `p` being true of `x`. Eventuality `e` may exist in the real world (`Rexist`), in which case `p(x)` holds, or it may only exist in some modal context, in which case that is expressed simply as another property of the possible individual `e`. The logical form of a sentence is a flat conjunction of existentially quantified positive literals, with about one literal per morpheme. (For example, logical words like "not" and "or" are treated as expressing predications about possible eventualities.) We have developed software[2] to translate Penn TreeBank-style trees (as well as other syntactic formalisms) into this notation. Boxer parser [2] now produces this notation and we used that as well. The underlying core theories are expressed as axioms in this notation [3].

The interpretation of a text is taken to be the lowest-cost abductive proof of the logical form of the text, given the knowledge base [4]. That is, to interpret a text we prove the logical form, allowing for assumptions at cost, and pick the lowest-cost proof. Factors involved in computing costs include, besides the number of assumptions, the salience of axioms, the plausibility of axioms expressing defeasible knowledge, and consilience or the degree to which the pervasive implicit redundancy of natural language texts is exploited. We have demonstrated that many interpretation problems are solved as a by-product of finding the lowest-cost proof. This method has been implemented in an abductive theorem-prover called Mini-Tacitus[3] that has been used in a number of applications [5], and is used in the textual entailment problems described here.

Most commonsense knowledge is defeasible, i.e., it can be defeated. This is represented in our framework by having a unique "et cetera" proposition in the antecedent of Horn clauses that cannot be proved but can be assumed at a cost corresponding to the likelihood that the conclusion is true. For example, the axiom

```
bird(x) & etc-i(x) → fly(x)
```

would say that if `x` is a bird and other unspecified conditions hold, `etc-i`, then `x` flies. No other axioms enable proving `etc-i(x)`, but it can be assumed, and hence participate in the lowest cost proof. The index `i` is unique to this axiom. (This approach to defeasibility is similar to circumscription [6].)[4]

[2] http://www.rutumulkar.com/nl-pipeline.html.

[3] http://www.rutumulkar.com/tacitus.html.

[4] Since most of the axioms presented in this paper are defeasible, we omit the "etc" for sake of simplicity.

We find that reifying states and events as eventualities and treating them as first-class individuals is preferable to employing the event calculus [7; 8]; which makes a sharp distinction between the two, because language makes no distinction in where they can appear and we can give them a uniform treatment.

We have articulated a number of core theories[5]. The most relevant to this paper are the theories of change of state, composite entity, scales and event structure. In the following, for each relevant core theory, we give a brief overview of the properties and predicates that are used in this paper.

The predication `change′(e,e1,e2)` says that `e` is a change-of-state whose initial state is `e1` and whose final state is `e2`. The chief properties of change are that there is some entity whose state is undergoing change, that change is defeasibly transitive, that `e1` and `e2` cannot be the same unless there has been an intermediate state that is different, and that change is consistent with the *before* relation from our core theory of time [9]. Since many lexical items focus only on the initial or the final state of a change, we introduce for convenience the predications `changeFrom′(e,e1)` and `changeTo′(e,e2)`, defined in terms of `change`.

Because we reify states and events, their temporal properties are simply relations between them and temporal entities like instants and intervals of time. States generally occupy intervals of time.

A composite entity is a thing composed of other things. The concept is general enough to include complex physical objects (e.g., a telephone), complex events (e.g., the process of erosion) and complex information structures (e.g., a theory). A composite entity is characterized by a set of components, a set of properties, and a set of relations. In particular, the predication `componentOf(y, x)` says that `y` is a component of `x`. An important part of the theory of composite entities is the figure-ground relationship which is of fundamental importance in language and cognition. We encode this with the predication `at(x,y,s)`, saying that a figure `x` is at a point `y` in the ground `s`, where `y` is a component of `s` and `x` is external to `s`. A ground is a composite entity whose parts are all uniform in that they all share some property. Examples using the preposition “at” include “Nuance closed at 57 3/8”, where the ground is scale of numbers/prices; “John is now at a competing company”, where the ground is set of organizations; and “At that moment John stood up”, where the ground is the time scale.

Our core theory of scales provides axioms involving predicates such as `scale`, `lts`, `subscale`, `top`, `bottom`, and `at`. These are abstract notions that apply to partial orderings as diverse as heights, money, and degrees of happiness. A scale is a composite entity where the components are the members of the set and one of the relations among them is the partial ordering. The standard properties of partial orderings can be defined in terms of the predicate `lts`. In particular, the predication `lts'(e1,x1,y1,s)` says that `e1` is the relation of `x1` being less than `y1` on scale `s`; where `x1` and `y1` can be either points on (components of) or entities that are "at" points on the scale. The "less than" relation is defined to be antireflexive, antisymmetric and transitive.

The theory of event structure is about how changes of state and causality compose into more complex events, processes and scenarios. It includes definitions of conditional, iterative, cyclic, and periodic events, and is linked with several well-

[5] http://www.isi.edu/~hobbs/csk.html.

developed ontologies for event structure, e.g., PSL [10]. In this theory, the aggregate, or conjunction, of two events is an event. Thus we can define a composite eventuality such as "carry" as being the conjunction of the eventualities "hold" and "move". An event sequence is the reified conjunction of two events that are in sequence. More specifically, the predication eventSequence'(e,e1,e2) says that e is an event sequence of e1 followed by e2. To account for event sequences of arbitrary length, we let e1 and/or e2 be an event sequences themselves.

Eventualities are very finely individualized. Thus Pat's running and Pat's going are distinct though related eventualities. The expression arg*(x,e) says that entity x is an argument or participant in eventuality e, or more precisely, x is an argument of e or an arg* of an eventuality argument of e. Substitution is explicitly axiomatized. The expression subst(x,e1,y,e2) says that x plays the same role in e1 that y plays in e2. The expression subst2(x,y,e1,z,w,e2) says that x and y play the same role in e1 that z and w play in e2 respectively.

2. Methodology for Axiomatization

Our methodology consists of three steps: Analyzing the structure of a word's WordNet senses, writing axioms for the most general senses, and testing the axioms on textual entailment pairs.

Our focus in this paper is on change-of-state words such as "enter", "cut", "rise", "return", "carry", etc. For each word, we analyze the structure of its WordNet senses. Typically, there will be pairs that differ only in, for example, constraints on their arguments or in that one is inchoative and the other causative. This analysis generally leads to a radial structure indicating how one sense leads by increments, logically and perhaps chronologically, to another word sense [11]. The analysis also leads us to posit "supersenses" that cover two or more WordNet senses. (Frequently, these supersenses correspond to senses in FrameNet [12] or VerbNet [13], which tend to be coarser grained; sometimes the desired senses are in WordNet itself.)

For example, for the verb "enter", three WordNet senses involve a change into a state:

V2: enter, participate (become a participant; be involved in) "enter a race"; …

V4: figure, enter (be or play a part of or in) "Elections figure prominently in every government program"; "How do the elections figure in the current pattern of internal politics?"

V9: embark, enter (set out on (an enterprise or subject of study)) "she embarked upon a new career"

We group these three senses into supersense S1 [6]. Two more senses specialize supersense S1 by restricting the target state to be "in a location":

V1: enter, come in, get into, get in, go into, go in, move into (to come or go into) "the boat entered an area of shallow marshes"

V6: enter (come on stage)

Two other senses add a causal role to this:

V5: record, enter, put down (make a record of; set down in permanent form)

V8: insert, infix, enter, introduce (put or introduce into something) "insert a picture into the text"

[6] In our framework, it is possible to assign a WordNet sense to two supersenses.

One other sense adds a causal role to S1 and restricts the target state to be membership in a group:

V3: enroll, inscribe, enter, enroll, recruit (register formally as a participant or member) "The party recruited many new members"

Figure 1 shows the radial structure of the senses for the word "enter"[7], together with the axioms that characterize each sense. A link between two word senses means an incremental change in the axiom for one gives the axiom for the other. For example, the axioms for S2 and S2.1 are obtained by adding a causal role to the axioms for S1 and S1.1 respectively. Thus S2 is linked to S1 and S2.1 is linked to S1.1. More specifically, the expanded axiom for S1 says that if x1 enters eventuality e1, then there is a change to e1 where x1 is an argument of e1; and the axiom for S2 says that if x1 enters x2 in eventuality e1, then x1 causes a change into the state e1 where x2 is an argument of e1. So S2 and S1 are closely related and linked together.

Abstraction is a particularly important kind of incremental change; one sense S1 is an abstraction of another sense S1.1 when S1.1 specializes S1 either by adding more predications to or specializing some of the predications in S1's axiom. We represent abstractions via arrows pointing from the subsenses to the supersenses. In Figure 1, S1.1 specializes S1 by adding an extra predication describing e1 as an "in" eventuality and V3 specializes S2 by specializing e1 to "membership in group".

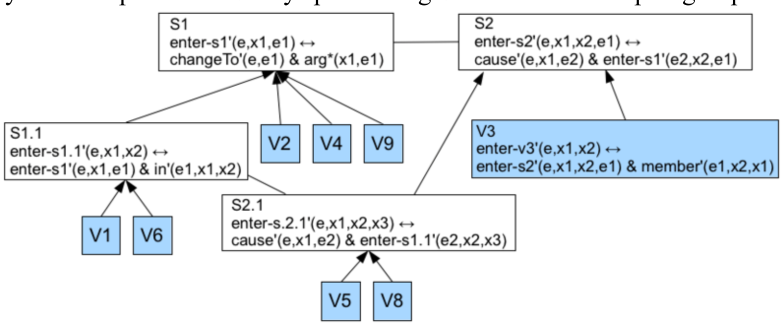

Figure 1: Senses and axioms for the verb "enter"

Knowing this radial structure of the senses helps enforce uniformity in the construction of the axioms. If the senses are close, their axioms should be almost the same. We are currently only constructing axioms for the most general or abstract senses or supersenses. In this way, although we are missing some of the implications of the more specialized senses, we are capturing the most basic topological structure in the meanings of the words. This decision has a number of advantages which are listed below:

- Axiomatizing the general senses is much more feasible than axiomatizing all the senses. (e.g., axiomatizing 4 supersenses for the verb "give" rather than all 44 specific senses).
- Axiomatizing general senses allows us to make correct (although not precise) inferences (ambiguous inference) when the disambiguation into fine-grained senses is not possible, either because no context is available or because the context doesn't help in resolving the ambiguity. For example, in the sentence "she

[7] Throughout this paper, we refer to "WordNet sense" simply as "sense". We refer to the nth sense of a particular part of speech (POS) of a word as POSi (e.g., V3 or ADV2).

had a baby”, it is not clear which sense of “had” is being referring to. Is it referring to the “have/possess” sense or to the “giving birth” sense? In any case we can be sure there is some relation between “she” and the “baby”. Of course it would be always good to have axioms for the specific senses too, so that we get more accurate inferences when context is strong enough to do lexical disambiguation. In fact, one can view the choice between fine-grained and coarse-grained axioms as the decision of where to place the unreliability. If we only employ fine-grained senses, we have less reliable lexical disambiguation, but more precise inferences after the right sense is determined. If we axiomatize coarse-grained senses, we get more reliable lexical disambiguation (since we have a smaller branching factor); but introduce uncertainty (and impreciseness) in the axioms. In the latter case, it is the task of our abduction engine to find more precise inferences by searching for the lowest cost proof of the text.

- Extracting the basic topological structure of words gives us a framework for investigating synonymy, near synonymy and other types of similarity between words [14].
- In metaphors, it is often the topological properties captured in a supersense that is transferred from source to the target [15]. For example, we can view “entering a course of study” as resting on a metaphor of “entering a room”. As shown in Figure 1, the supersense S1 unpacks and makes explicit the “change of state” property from the source domain senses (V1 and V6) of “enter” that is transferred to the target domain senses (V2-V5, V8, V9).
- WordNet makes a distinction between senses based on argument structure, in addition to semantic differences, differences in sets of syntactic frames and/or differences in selectional restrictions (cf. [16]). We can make the distinction between senses more semantic-oriented by grouping such non-semantically - differentiated senses under one supersense, assigning them multiple argument realization patterns and explicating only one predicate axiomatically. For example, we group senses 1 and 8 of the verb “begin” (which correspond to “We began working at dawn” and “begin a cigar” respectively), as one concept (supersense) that is realized by two different argument realization patterns.

In constructing the axioms in the event domain, we are very much informed by the long tradition of work on lexical decomposition in linguistics (e.g., [17; 18]). Our work differs from this in that our decompositions are done as logical inferences and not as tree transformations as in the earliest linguistic work, they are not obligatory but only inferences that may or may not be part of the lowest-cost abductive proof, and the “primitives” into which we decompose the words are explicated in theories that enable reasoning about the concepts.

3. Axiomatizing change-of-state words

WordNet [19] contains tens of thousands of synsets referring to highly specific animals, plants, chemical compounds, French mathematicians, and so on. Most of these are rarely relevant to any particular natural language understanding application. To focus on the more central words in English, the Princeton WordNet group has compiled a Core WordNet[8], consisting of 4,979 synsets that express frequent and salient concepts. We classified these word senses manually into sixteen broad categories, including composite entities, scales, events, space, time, etc. Each of these categories was then given a finer-grained structure. The internal structure of the category of event words consists of state, change, cause, instrumentality, process, opposition, force and functionality. In this effort, we focus on change category which we have further classified into the following sub-categories:

- Abstractly: incident, happen
- A change of real or metaphorical position: enter, return, take, leave, rise
- A change in real or metaphorical size or quantity: increase, fall, . . .
- A change in property: change, become, transition, . . .
- A change in existence: develop, revival, decay, break, . . .
- A change in real or metaphorical possession: accumulation, fill, recovery,..
- The beginning of a change: source, start, origin, . . .
- The end of a change: end, target, conclusion, stop, . . .
- Things happening in the middle of a change: path, variation, [take a] break, . . .
- Participant in a change: participant, player, . . .

Our categories of course have fuzzy boundaries and overlaps, but their purpose is only for grouping together concepts that need to be axiomatized together for coherent theories. There are 134 word senses in the change-of-state category. Since in our methodology we axiomatize coarse-grained senses, we determined the radial structure of all WordNet senses of the head words and axiomatized the most general senses.

In the rest of this section we describe some common problems we faced and decisions we had to make during axiomatization effort.

3.1. Specificity of Details

An important question is: How much information should be encoded in an axiom? For example, one possible axiom for a generalization of senses 1, 16 and 32 of the verb "carry" is[9]:

```
carry-s0'(e,x,y,p0,p1) ↔ and'(e,e1,e2) & hold'(e1,x,y) & move'(e2,x,p0,p1)
                         & cause(e,e3) & move'(e3,y,p0,p1)
```

which means the eventuality `e` of `x`'s carrying `y` from `p0` to `p1` is an eventuality e of both `x`'s holding `y` (`e1`) and `x`'s moving from `p0` to `p1` (`e2`) where e causes the eventuality `e3` of `y`'s moving from `p0` to `p1`. Another possibility for defining "carry" is:

```
carry-s0'(e,x,y,p0,p1) ↔ and'(e,e1,e2) & hold-v2'(e1,x,y) &
                         move'(e2,x,p0,p1)
```

where the following axiom would be in a core theory of attachment and causality:

```
and'(e,e1,e2) & hold'(e1,x,y) & move'(e2,x,p0,p1) ↔
cause(e,e3) & move'(e3,y,p0,p1)
```

which means that `x`'s moving while holding `y` causes `y`'s moving. In fact, this is a fundamental property of attachment and hence should be part of the core theory of attachment. In general, we prefer to factor out all the information that can be inferred

[8] CoreWordNet is downloadable from http://wordnet.cs.princeton.edu/downloads.html

[9] Variables occurring in antecedent of implications are universally quantified and variables occurring in the consequent are existentially quantified.

from world knowledge and explicate it in our core theories. In this way, the knowledge becomes available for examples not involving the word "carry".

3.2. Metonymy vs. Lexical Disambiguation

WordNet is not quite consistent in handling metonymy. For example consider the following senses of the verb "begin":

V1: get down, begin, get, start out, start, set about, set out, commence (take the first step or steps in carrying out an action) "We began working at dawn"; "Who will start?";...

V8: begin, start (begin an event that is implied and limited by the nature or inherent function of the direct object) "begin a cigar"; "She started the soup while it was still hot"; …

WordNet has assigned a separate sense (V8) to the use of the verb "begin" with a metonymycal argument. However, despite different argument realization patterns, V1 and V8 are semantically equivalent: both mean starting an eventuality e. The only difference is that e is missing in V8, since it can be determined by the arguments. Now consider another case where metonymy is not handled in separate senses. One sense of "cut" is:

V36: cut, cut off (cease, stop) "cut the noise"; "We had to cut short the conversation"

This sense semantically means "stopping an eventuality". While "conversation" is an eventuality, "noise" is not. What we really mean with "cut the noise" is "cut making noise"; where "making" is omitted due to its recoverability.

One way to handle this case is to split the sense V36 into two sub-senses V36-a and V-36b and axiomatize them differently:

cut-v36-a′(e,x,e0) ↔ cause′(e,x,e1) & changeFrom′(e1,e0)
cut-v36-b′(e,x,z) ↔ cut-v36-a′(e,x,e0) & arg(z,e0)*

The first axiom states that x's cutting eventuality e0 means that x causes a change from e0. The second axiom states that x's cutting of z means x causes a change from some eventuality e0, where z is an argument of e0. A better approach is to handle metonymy separately using axioms such as:

cut-v36′(e,x,e0) → cut-v′(e,x,e0) & eventuality(e0)
cut-v36′*(e,x,e0)* → cut-v′(e,x,y) & arg*(y,e0)

and have only one predicate cut-v36 that we explicate as:

cut-v36′(e,x,e0) ↔ cause′(e,x,e1) & changeFrom′(e1,e0)

The logic behind this is that our predicates explicate situations; therefore if two situations are the same, their predicates should also be the same. We also merge those senses of WordNet that differ only in argument realization patterns and assign them a unique axiom. e.g., for the "begin" case above, we generate a supersense that unifies V1 and V8:

begin-v1-8′(e,x,e0) ↔ changeTo′(e,e1)& eventSequence*(e0,e1,e2) & arg*(x,e1)*

and add the following axioms to handle metonymy:

begin-v1-8′(e,x,e0) → begin-v′(e,x,e0) & eventuality(e0)
begin-v1-*8′(e,x,e0) → begin-v′(e,x,y) & arg*(y,e0)*

Which say that the eventuality e of x beginning an event sequence e0 is a change to an eventuality e1, where e1 has x as its argument and is the first eventuality in e0. Ideally we would like to have a general mechanism to handle metonymy, but since such a mechanism is not implemented yet, we can handle some cases of metonymy using the above technique. Besides solving metonymy, this method reflects the fact that the same concept can be realized differently in text.

3.3. Metaphors

Sometimes a WordNet sense has a clear metaphorical origin and that metaphorical interpretation has become conventionalized as one of the senses of the word. Thus it is not inconsistent to say that an instance of a word is both a metaphor and an example of a conventional sense of a polysemous word.
Consider sense 2 of the verb “descend”:

V2: derive, come, descend (come from; be connected by a relationship of blood, for example) "She was descended from an old Italian noble family"; "he comes from humble origins"

Other senses of “descend” mean a change from being at a higher region to being at a lower region on a vertical scale[10]:

V1: descend, fall, go down, come down (move downward and lower, but not necessarily all the way) "The temperature is going down"; "The barometer is falling"; "The curtain fell on the diva"; ...

V3: condescend, deign, descend (do something that one considers to be below one's dignity)

Sense 2 doesn’t seem to indicate a change on a scale; but it would if we consider a tree of life having its root in top and the children deriving from it downwards. In such cases, axiomatizing a word sense with metaphorical interpretation allows us to fit those senses within our radial structure, making the structure more coherent. However we won't do so if such abstract axioms are not practical for reasoning. In this example, we axiomatize sense 2 as

```
descend-v2'(e,d,o) ↔ ancestorOf'(e,o,d)
```

which falls into kinship domain. Note however that often times we can account for many conceptual metaphors [15] by using abstract predicates. For example, consider the following senses of the verb “cut”:

V32: cut (shorten as if by severing the edges or ends of) "cut my hair"

V37: abridge, foreshorten, abbreviate, shorten, cut, contract, reduce (reduce in scope while retaining essential elements) "The manuscript must be shortened"

A generalization of both these senses can be captured by the following axiom

```
cut-s1'(e,x,y) ↔ cause'(e,x,e10) & changeFrom'(e10,e0) & connect'(e0,w,z) &
                 componentOf(w,y) & componentOf(z,y)
```

which says, x’s cutting y is x’s causing a change out of a state in which w and z, two components of y, were connected[11]. Although “cutting a manuscript” is very different from “cutting hair”, our abstract definition of composite entity allows us to handle them similarly.

3.4. Scales:

We have axiomatized many words in terms of change on a scale. Consider for example these senses of the word “rise”:

V1: rise, lift, arise, move up, go up, come up, uprise (move upward) "The fog lifted";…

V7: ascend, move up, rise (move to a better position in life or to a better job) "She ascended from a life of poverty to one of great renown”

V11: rise, jump, climb up (rise in rank or status) "Her new novel jumped high on the bestseller list"

---

[10] A scale can be stipulated to be vertical for a variety of reasons.

[11] As we will see in the evaluation section, we can infer y’s getting shorter by combining this axiom with some axioms from the core theory of changes to composite entities.

We can axiomatize all these senses via this very general axiom:

```
rise-s0'(e,x,p1,p2) ↔ change'(e,e1,e2) & at'(e1,x,p1) & at'(e2,x,p2) &
                      lts(p1,p2,s) & verticalScale(s)
```

which states that x's rising from p1 to p2 is a change from x being at point p1 to x being at point p2, where p1 is less than p2 on a vertical scale s. This, at the first glance, may seem to be too general to be useful; especially because of ignoring the specification of the scales in the sub-senses. However, we have observed that in many cases the scale can be recovered from knowledge about the arguments. If we construct a knowledge base containing information such as

- different scales that an entity is associated with (e.g., "fog" is associated with "altitude" scale)
- components of each scale (e.g., "poverty" is a fairly low component of a complex "status in life" scale that includes factors such as wealth and fame)
- entities referred to as scales (e.g., "best seller list" is a "rank" scale)

then we may be able to determine that the scale s is "altitude" in V1, "position in life" in V7 and "rank" in V11.

3.5. Event Types:

Sometimes we need to state that the type of two events are similar although their arguments or time are different. For example, consider the following senses of the adverb "back" which refer to an eventuality that is the same as a previously occurring eventuality, with the difference that the agent and the patient have switched roles:

ADV5: back (in reply) "he wrote back three days later"

ADV6: back (in repayment or retaliation) "we paid back everything we had borrowed"; …

We can axiomatize both these senses via the following axiom:

```
back-adv5-6'(e,e2) ↔ subst2(x,y,e2,y,x,e1) & before(e1,e2)
```

The above axiom means that eventuality e2's having the property "back" means there has been an eventuality e1 that has occurred before e2 and e1 and e2 are the same except that their first and second arguments are flipped.

Now consider the word "repeat". We want to say that when x repeats an action e, then x is doing the same type of action again.

V1: repeat, reiterate, ingeminate, iterate, restate, retell (to say, state, or perform again) "She kept reiterating her request"

We can capture this meaning with the following axiom:

```
repeat-v1'(e1,x,e0) ↔ before(e0,e1) & subst(x,e0,x,e1)
```

So the above axiom states that e1 is x's repetition of e0 if and only if e0 has occurred before e1, both e0 and e1 are the same type of eventualities and x plays the same role in both e0 and e1. (Another sense of repeat changes the argument as in "repeat after me")

4. Evaluation Method

For each set of inferentially related words we construct textual entailment pairs, where the hypothesis (H) intuitively follows from text (T), and use these for testing and evaluation. The person writing the axioms does not know what the pairs are, and the person constructing the pairs does not know what the axioms look like. The ideal test then is whether given a knowledge base K consisting of all the axioms, H cannot be proven from K alone, but H can be proven from the union of K and the best interpretation of T. This is often too stringent a condition, since H may contain irrelevant material that doesn't follow from T, so an alternative is to determine whether the cost of the lowest cost abductive proof of H given K plus T is substantially lower than the lowest cost abductive proof of H given K alone, where "substantially lower" is defined by a threshold that can be trained [20].

A preliminary study on 50 entailment pairs shows that we get 37 correct inferences without adding any axioms to the core theories. This number increases to 43 if we add the required axioms.

Ideally we would like to test the axioms on standard data sets like RTE data sets[12], but such data sets were designed with other problems in mind and they contain very small number of text and hypothesis pairs for which the inference depends on our change-of-state word senses. This is true even for large sets such as WIKI corpus [21]. So we decided to build our own specialized dataset of pairs. We are currently in the process of gathering a larger set of textual entailment pairs made by several non-experts using the following procedure: For each sense Si of a given word W (Currently we only consider verbs), we select several sentences that contain W in sense Si. A few sentences can be obtained from WordNet's examples for each synset and more can be extracted from sense-tagged corpora or large web-page collections such as ClueWeb[13]. These sentences are then given to several subjects to make inferences focusing mainly on W. Table 1 shows some of the inferences that we got for the verbs "cut" and "return". We have categorized them into relevant and non-relevant inferences. We can put more restrictions in our guidelines to get more relevant results, however we prefer not to restrict the annotators too much and let them think of as many inferences as they can. We can always filter the inferences later, but a large variety of inferences gives us insight about different aspects of individual as well as compositional word meanings.

Table 1. Sample inferences made by annotators

Text	Relevant Hypothesis	Id
He cut the paper along the dotted line.	The paper was divided into smaller pieces.	1
He cut the rope.	The length of the rope decreased.	2
He cut his daily fat intake.	He reduced the amount of fat that he consumed.	3
The coach cut two players from the team.	Two players on the team were asked to leave.	4
The telephone call was immediately returned.	He was called right back.	5
He returned to his work on the script.	He took a break from writing the script.	6
Text	**Irrelevant Hypothesis**	**Id**
The glass from the shattered windshield cut into her forehead.	Blood was dripping from her forehead.	7
The Vietnamese cut a lot of timber while they occupied Cambodia.	The Vietnamese were forced to work a lot during the occupation.	8
The telephone call was immediately returned.	The caller left a message.	9
Results from the FBI fingerprint check would be returned to the State agency for evaluation.	Fingerprints were discovered at the scene.	10

As we can see, irrelevant inferences are those that require extra knowledge (e.g., cutting into forehead results in bleeding in inference #7); or not related to the verb under consideration (e.g., inference #8 and #10); or making assumptions that are not necessarily true (e.g., inferences #7-9).

[12] http://pascallin.ecs.soton.ac.uk/Challenges/RTE/

[13] http://lemurproject.org/clueweb09.php/

Of course, some of the inferences shown in Table1 can be handled by axioms that capture such shallow inference types as synonymy, frame structure, paraphrasing, etc. (cf. [20]). However, the focus of this work is on the general case where deeper reasoning is needed.

Here we work through two examples to illustrate how textual entailment problems are handled in our framework and what kind of missing knowledge they reveal. In this example we do not show the costs, although they are used by our system.

First we start with inference #1. Here are the logical forms for text (T) and hypothesis (H)[14]:

```
T: Male'(e14,x0) & cut-v'(e13,x0,x1,u9) & paper-n'(e5,x1) & along-
   p'(e18,e13,x12) & dotted-a'(e17,x12) & line-n'(e16,x12)
H: paper-n'(e3,x1) & divide-v'(e1,u5,x1,u6) & into-p'(e8,e1,x2) & smaller-
   a'(e7,x2) & piece-n'(e4,x2)
```

We have the following general axioms for "cut" and "divide":

```
AX1: cut-s0'(e,x,y)   ↔
     cause'(e,x,e10) & changeFrom'(e10,e0) & connect'(e0,w,z) &
     componentOf(w,y) & componentOf(z,y)

AX2: divide-v1-5'(e,x,y,s)↔
     cause'(e,x,e0) & changeFrom'(e0,e1) & connect'(e1,w,z) & member(w,s) &
     member(z,s) & componentsOf(s,y)
```

The first axioms says that x's cutting y means x causes a change from w and z to be connected, where w and z are components of the composite entity y. The second axiom says that x's dividing y into a set s, is x's causing a change from w and z to be connected, where w and z are members of s and s is the set of components of the composite entity y.

In order to map from the logical form to the predicate divide-v1-5', we need the following axiom:

```
divide-v1-5'(e1,x,y,s) → divide-v'(e1,x,y,u) & into-p'(e8,e1,s)
```

We also need the following axiom from the core theory:

```
componentOf(c,x) ↔ componentsOf(s,x) & member(c,s)
```

which means if c is a member of the set of components of x, then c is a component of x. With these set of axioms, we are able to infer H from T as shown in figure 2.

As the second example, we consider inference #2: We would like to infer "The length of the rope decreased" from "He cut the rope" using axiom AX1 for cut. This means that we need to infer a decrease in length of the rope from the disconnection of two pieces of the rope. This inference may seem intuitive, but in fact, it is defeasible: what if we cut the rope along the length instead of the width? Then we will get a narrower rope with the same length. In fact, in this case, T is taken from WordNet's example for sense V1 of "cut" which is defined as "separate with or as if with an instrument". One can argue that in this case "cut" should be considered in sense V32 defined as "shorten as if by severing the edges or ends of". In any case, we believe that the general axiom AX1 is sufficient and the type of arguments should help to determine the most probable inference. However there are some complementary axioms that are necessary to help our general axiom work. For example, the fact that the parts of a physical composite entity are smaller than the whole in particular dimensions:

[14] These logical forms are generated by the Boxer tool: http://svn.ask.it.usyd.edu.au/trac/candc/wiki/boxer

```
componentOf(c,y) & dimensionOf(p,y) & dimensionOf(p,c) & valueOf'(v0,p,y) &
valueOf'(v1,p,c) → lts(v1,v0,s) & scaleFor(s,p)
```

This axiom states that if c is a component of y; and y and c both have a physical dimension p; and their values for this dimension are v0 and v1 respectively, then v1 (e.g., the length value of c) is less that v0 (e.g., the length value of y) on the scale associated with p.

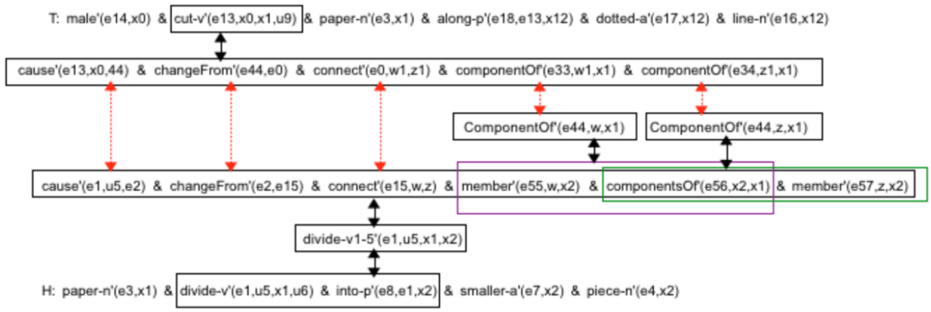

Figure 2: Abduction results on the pair T: "He cut the paper along the dotted line", H:"The paper was divided into smaller pieces". Solid arrows show abductive inference and dashed arrows show merging (which results in substantial cost reduction).

We also need to account for the fact that after separation, we refer to the parts of a rope as the rope itself (In contrast to parts of a car): One implication of this is that the properties of either of the components can be assigned to the original entity[15]:

```
changeFrom'(e10,e0) & connect'(e0,c1,c2) & componentOf(c1,y) & componentOf(c2,y) &
dimensionOf(p,y) & dimensionOf(p,c1) & dimensionOf(p,c2) & valueOf'(e5,v0,p,y) &
valueOf(v1,p,c1) & valueOf(v2,p,c2) ↔
change'(e10,e5,e2) & or'(e2,e3,e4) & valueOf'(e3,v1,p,y) & valueOf'(e4,v2,p,y)
```

This axiom means if c1 and c2 are components of y and y, c1 and c2 all have a dimension p and their values for this dimension are v0, v1 and v2 respectively, then a change from c1 and c2 being connected, means a change from y having the value v0 for dimension p (e.g., rope having length v0) to y having either v1 or v2 for dimension p (e.g., rope having length v1 or v2). Our general axiom for decrease is:

```
decrease-v1'(e,p,x,v1,v2) ↔ change'(e,e1,e2) & valueOf'(e1,v1,p,x) &
                            valueOf'(e2,v2,p,x) & lts(v2,v1,s) & scaleFor(s,p)
```

Which says a decrease of x's value for its dimension p from v1 to v2 means a change from x's value for dimension p being v1 to being v2, where v2 is less than v1 on scale s associated with p. With all these axioms we are now able to infer that when a rope is cut into two pieces c1 and c2, the size of rope will change to the size of either c1 or c2 which is less than the size of the original rope. A change (along some scale) from a higher value to a lower value of a property is a "decrease" of that property. The only thing that remains is inferring that the property under discussion is "length". This would require axiomatizing the rough shapes of common objects.

As this example shows, not embedding the "shortening" aspect of "cut" in our axiom can be recovered by appropriate, although complex, knowledge about composite entities and their properties.

[15] We won't need this axiom if we defined "cut" as "removal of a component from a composite entity": When a component is removed from a composite entity, the entity becomes smaller (shorter).

5. Conclusion and Future Work

We are continuing the evaluation of our change-of-state axioms on a larger set of textual entailment pairs. So far we have considered only positive textual entailment examples. We are aware that our evaluation should also consider non-entailment pairs. This however remains as a future work. Finally, once we have completed testing of change-of-state words, our next effort will be to axiomatize words from Core WordNet that involve causality.

References:

[1] N. Montazeri, J. Hobbs, *"Elaborating a Knowledge Base for Deep Lexical Semantics"*, in J. Bos and S. Pulman (eds.), Proceedings of the Ninth International Conference on Computational Semantics (IWCS 2011), pp. 195-204.

[2] J. Bos. *"Wide-Coverage Semantic Analysis with Boxer"*. In J. Bos and R. Delmonte (Eds.), Semantics in Text Processing. STEP 2008 Conference Proceedings, Research in Computational Semantics, pp. 277–286. College Publications. (2008)

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

[4] J.R. Hobbs, M. Stickel, D. Appelt, and P. Martin. *Interpretation as abduction*. Artificial Intelligence 63, 69–142. (1993)

[5] R. Mulkar, J.R. Hobbs, and E. Hovy, *"Learning from Reading Syntactically Complex Biology Texts",* Proceedings of the AAAI Spring Symposium Commonsense'07. Stanford University, CA, 2007.

[6] J. McCarthy, *"Circumscription: A Form of Nonmonotonic Reasoning"*, Artificial Intelligence, 13: 27-39. (1980)

[7] M. Gruninger, C. Menzel, *"The Process Specification Language (PSL) Theory and Applications"*, AI Magazine, Vol 24, No 3.(2010)

[8] E.T. Mueller, *Commonsense Reasoning*, Morgan Kaufmann Publishers, Inc., San Mateo, California. (2006)

[9] J.R. Hobbs and F. Pan. *"An Ontology of Time for the Semantic Web"*. ACM Transactions on Asian Language Processing (TALIP): Special issue on Temporal Information Processing, Vol. 3, No. 1, pp. 66-85. (2004)

[10] C. Bock, M. Gruninger, *PSL: A Semantic Domain for Flow Models*, Software and Systems Modeling Journal, 4:2, 209 - 231.12. (2005)

[11] G. Lakoff, *"Women, Fire, and Dangerous Things: What Categories Reveal About the Mind"*, University of Chicago Press, Chicago. (1987)

[12] C. Baker, C. Fillmore, B. Cronin, *The Structure of the Framenet Database*, International Journal of Lexicography,Volume 16.3: 281-296.14. (2003)

[13] K. Kipper, A. Korhonen, N. Ryant, and M. Palmer, *"Extensive Classifications of English verbs"*, Proceedings, 12th EURALEX International Congress. Turin, Italy. (2006)

[14] N. Montazeri, J. R. Hobbs, *"Synonymy and Near-Synonymy in Deep Lexical Semantics"*, Proceedings of Workshop on Computational Approaches to Synonymy, Helsinki, Finland, (2010)

[15] G. Lakoff, M. Johnson, *Metaphors We Live By,* University of Chicago Press. (1980)

[16] M. Palmer, *"Are Wordnet sense distinctions appropriate for computational lexicons?"* In Advanced Papers of the SENSEVAL Workshop, Sus-sex, UK. (1998)

[17] J.C. Gruber, *Studies in Lexical Relations*, unpublished Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, Massachusetts. (1965)

[18] R.S. Jackendoff, *Semantic interpretation in generative grammar*. Cambridge, MA: The MIT Press. (1972)

[19] G. Miller, *"WordNet: a lexical database for English"*. In Communications of the ACM 38 (11), pp. 39-41.22. (1995)

[20] E. Ovchinnikova, N. Montazeri, T. Alexandrov, J. Hobbs, M. McCord, and R. Mulkar-Mehta, *"Abductive Reasoning with a Large Knowledge Base for Discourse Processing"*, Proceedings of the 9th International Conference on Computational Semantics, Oxford, United Kingdom. (2011)

[21] F.M. Zanzotto, M. Pennacchiotti. *"Expanding textual entailment corpora from Wikipedia using co-training."* Proceedings of the 2nd Workshop on Collaboratively Constructed Semantic Resources, COLING 2010. (2010)