Veridicity annotation in the lexicon? A look at factive adjectives

Annie Zaenen
CSLI / Stanford University
azaenen@stanford.edu

Lauri Karttunen
CSLI / Stanford University
laurik@stanford.edu

Abstract

In this note, we look at the factors that influence veridicity judgments with factive predicates. We show that more context factors play a role than is generally assumed. We propose to use crowd sourcing techniques to understand these factors better and briefly discuss the consequences for the association of lexical signatures with items in the lexicon.

1 Veridicity: what and why

Recognizing the inferential properties of constructions and of lexical items is important for NLU (Natural Language Understanding) systems. In this paper we look at FACTUAL INFERENCES, inferences that allow the reader to conclude that an event has happened or will happen or that a state of affairs pertains or will pertain. We will refer to events and states together as SOAs. Factuality is in the world and outside of the text. In cases where the reader has no direct perceptual knowledge about the SOAs, she has to evaluate the factuality of a SOA referred to in a text based on her decoding of the author's representation of the factuality of the SOA and on her knowledge about the world and about the author's reliability. Authors have a plethora of means to signal whether they want to present SOAs as factual, as having happened or going to happen or as being more or less probable, possible, unlikely or not factual at all. We will call this presentation of a SOA the VERIDICITY of a SOA. We will call the reader's interpretation of the author's intention, the RIV (READER INFERRED VERIDICITY) and the

reader judgment about the factuality of a SOA, RIF (READER INFERRED FACTUALITY).

Annotation can, at its best, only provide us with RIVs as the author is typically not available for consultation. This leads to a methodological problem. A reader will in his interpretation of a sentence be sensitive, not only to the way an author signals her intentions but also to what he knows about the world. To circumvent this problem as much as possible, corpus annotation for veridicity is typically done by trained annotators with extensive guidelines (see e.g. (Saurí, 2008), (Saurí and Pustejovsky, 2012)) but corpus annotation by trained annotators is an expensive enterprise, hence looks at a limited number of cases. For instance, to anticipate on a case we will discuss later in the paper, *lucky* occurs only once in the FactBank ((Saurí and Pustejovsky, 2009). Given that annotation is done on running text, it is also difficult to avoid that the reader's evaluation of the wider extralinguistic context might still play a role. We propose to supplement corpus annotation with crowd sourcing experiments. In these, sentences are presented to Mechanical Turk workers in limited contexts, very similar to the contexts in which linguists judge the effect of the contribution of a lexical item or a construction. But contrary to linguistic practice, we derive our examples from really occurring ones culled from the web and, more importantly, present them to many native speakers (typically 100) and in different variations to explore factors that can influence the interpretation. This kind of variation is very difficult to find in naturally occurring corpora of the type that are used for annotations (e.g. FactBank). This type of study comple-

ments the corpus studies in controlling the variation in the environment and in minimizing the external factors. With these experiments we intend to isolate a lexical signature for the lexical items we are interested in, in contradistinction to the interpretation in context that is provided by corpus annotation. It is, however, not intended to replace corpus studies because it has the drawback of not being able to take into account the influence of a wider linguistic environment.

2 Subclasses of veridical phenomena

The means an author uses to signal the factuality status of a SOA can be syntactic and/or lexical. Examples of syntactic means that have been exploited in textual inferencing tasks are appositives. They typically contain presupposed material and are in general factive. For a theoretical discussion of syntactically presupposed material see (Potts, 2005). Here we look at lexical sources of veridicity. They can be subdivided in IMPLICATIVE, FACTIVE and EPISTEMIC MODAL predicates. In all cases a lexical item occurs in the matrix clause of a syntactic frame where the embedded clause refers to a SOA. The veridicity status of the embedded clause is considered to be triggered by the lexical item, in the case of implicatives because there is an entailment-type relation, in the case of factives because there is a presupposition and in the case of epistemic modals because the embedded clause is under the scope of the modal.

(Karttunen, 1971; Karttunen, 2012) has studied several classes of IMPLICATIVE verbs and verb-noun collocations. Implicative constructions yield entailments about the veridicity of a complement clause. The entailment may be positive (+1) or negative (-1) depending on the polarity of the containing clause. Examples are:

(1) a. John managed to get the job done. (implies that the job got done)

     b. John didn't manage to get the job done. (implies that the job did not get done)

     c. John forgot to do the job. (implies that the job did not get done)

     d. John didn't forget to do the job. (implies that the job got done)

There are several different inference patterns described in detail in the references given above. The polarity computation must take into account the many ways of expressing negation by particles (*not*), adverbs (*never, almost*), quantifiers (*no one*) and counterfactual mood as in (2).

(2) a. Rand Paul would have fired Clinton.

     b. I wish I had been there.

FACTIVES were first studied (Kiparsky and Kiparsky, 1970). Their use indicates that the author considers the material in the embedded clause as presupposed (see e.g. (Beaver, 2010) for a discussion of relevant aspects of theories of presupposition). For the purpose of NLU, their most important characteristic is that their veridicity status does not change under negation or questioning ((Karttunen, 1971).

(3) a. It is annoying that people post stuff that no one cares about.

     b. It isn't annoying that people post stuff that no one cares about.

     c. Is it annoying that people post stuff that no one cares about?

Many implicative verbs are also presupposition triggers. For example, (1c) and (1d) both presuppose that John intended to do the job but carry opposite implications about whether the job got done. job The class of lexical items that express EPISTEMIC MODALITY includes verbs such as *must, have to, ought to, should, may, might*, adjectives such as *certain, likely, possible* and adverbs *certainly, likely, possibly*. There is a rich literature on this topic (Palmer, 2001; Kratzer, 2012).

With respect to veridicality, the most striking aspect of modal assertions is that even the *necessity* modals such as *must* and *have to* involve a weaker author commitment than the corresponding statements without the modal. An author who says

(4) It must be raining.

indicates that she has reasons to conclude that it is raining although she is not herself a witness to the event. A man who sees drops of water falling from the sky and recognizes it as rain would say *It is raining*; it would be odd for him to say (4). Direct evidence trumps reasoning.

The *possibility* modals such as *may* also indicate an inference or a guess that is made in the absence of direct evidence. The author of

(5) It may be raining.

indicates that she has no direct knowledge of whether it is raining but that conclusion is consistent with the evidence she has, but so is *it might not be raining*.

As epistemic modals show, author commitment to the veridicality of a SOA is a matter of degree ranging from *definitely true* to *definitely false* through a scale of weaker stances: *must, have to – probably, likely – possibly, perhaps, may – possibly not, perhaps not – probably not, most likely not – must not.*

Epistemic weakening applies to implications but not to presuppositions.

(6) John may have forgotten to do the job.

implies that the it is possible that John did not do the job but commits the author to the view, just as strongly as (1c) and (1d), that John had the intention to do the job. Presuppositions tend to "project" out of the embedded clauses that express them.

3 Annotating veridicity in the lexicon

Given the description above, one might come away with the idea that the only thing that needs to be done is to mark the veridical predicates in the lexicon and then have the system transmit a veridicity mark to the embedded clause. The mark would be different for the three classes as it would need to be sensitive to negation in different ways and the commitment might be absolute (negative or positive) or relative but the calculation would only have to look at one level of embedding. As the implementations discussed in (Nairn et al., 2006) and (MacCartney and Manning, 2009) show, the situation is quite a bit more complex. For factives specifically, we need to take into account what is known as the projection problem (Langendoen and Savin, 1971; Karttunen,

1974). But even if one assumes the projection problem solved, the picture is quite a bit more complicated than the short description in the previous section would let us to assume. We look in more detail at the complications that one finds with factive adjectives.

4 Factive adjectives

A great number of adjectives have been classified as factive in one or more of the following syntactic environments (see (Norrick, 1978) for the most extensive study that we are aware of):

(7) a. it be ADJ that S: It is annoying that he left early.

b. it be ADJ (for NP or of NP) to VP: It was daring for John to climb on the roof.

c. NP be ADJ that S: John is happy that the work got done.

d. NP be ADJ to VP: John was happy to get his paycheck.[1]

(7a) and (7b) are extraposition constructions, so there are also non-extraposed variants but as they are rare we leave them out of consideration here. We counted about 800 adjectives taking the (7a) construction, a slightly smaller number is supposed to occur in the (7b) one. Note that the syntactic frames themselves are not specific to factive adjectives. We can find non-factive adjectives in exactly the same syntactic environments:

(8) a. It is probable that he left early.

b. It is unlikely for John to come early.

c. He is certain that it will rain.

d. He is likely to come early.

4.1 Problem 1: variation

The first problem that arises is that when one looks at the data available on the web: several of these adjectives are used as non-factive implicatives in the construction in (7d), as we can see from the following examples (simplified from web examples):

---

[1]Constructions with -ing forms are also possible. We leave them out of the picture here because they have not been studied systematically.

(9) a. This is my first trip to Italy, so I was not brave to venture out alone.

    b. Luckily, she was not stupid to send them any money.

    c. He was not stupid to think she would remain the same weak little girl.

    d. It was raining and snowing like crazy in March here, so I was not stupid to risk the customer car, my license and my life.

    e. She still was not brave to approach the cars, even the couple of cars right in front of her.

    f. I was not lucky to have a good view.

The intended meaning of these sentences is clear but are the implicative interpretations of these sentences available to all speakers or are they just the creation of netizens whose command of English is weak or are they part of a bona fide unrecognized variant of English?

4.2 Problem 2: context[2]

As explained above, the received wisdom is that the factors that determine the inferential properties of a lexical item are the lexical item itself and its syntactic frame. The syntactic frame is in general meant to refer to a loose notion of subcategorization[3]. It consists of the environment of the item expressed in terms of syntactic categories, be it Phrase Structure categories or Dependency Grammar ones. This is the approach taken in VerbNet (Kipper-Schuler, 2005), where some semantic frames are associated with Phrase Structure syntactic frames, giving it a potential basis to make some inferential properties explicit. We work here too with Phrase Structure categories because most of the preceding literature on adjectives is in Phrase Structure terms. Under such an approach, we would list the adjectives that can be found in the frames given above in (7) and associate the factivity marker with them in case they

---

[2]The problem we discuss here for adjectives is discussed in a more theoretical setting for verbs by (Beaver, 2010)

[3]It is a loose notion because some elements that are recognized as part of the frame might reconsidered adjuncts rather than arguments in a strict syntactic sense. We do not go into this debate here as the distinction between arguments and adjuncts is often rather difficult to draw.

are factive and different marker in case they are not. But both introspection and experimental studies will tell us that this is not sufficient. Consider the following pair:

(10) a. It was fool hearted of John to go on a trip around the world.

    b. It is fool hearted to go on a trip around the world.

(10a) will indeed get a factive interpretation but (10b) will not.

4.3 Crowd sourcing for RIVs

The two problems above convinced us that, before proposing an veridicity annotation scheme for lexical items, we should study variation and context in more detail. To do this, we set up several Mechanical Turk experiments. In one, we presented Mechanical Turk workers with sentences like those in (9) (not including 9f) asking them both about how they understood them and whether they would use them to express the interpretation they had given. The preliminary results show that most speakers indeed interpret the sentences as implicating that the embedded clause is false but, more importantly, this non-factive interpretation was considered unobjectionable by 20% of native users of English (we controlled for this by asking the MT workers explicitly about their command of English ("Was English the primary language you used in ...") and by asking them to judge sentences that any native speaker of English would get right). 20% seems to be a large minority to ignore.

For the one adjective that we have studied in detail, *lucky*, the native speakers that consider examples such as (9f) as ill-formed and would require an *enough* to get the intended interpretation are in the minority according to an informal survey we did in parallel with the MT study. This suggests that the split between users accepting the implicative interpretation and those that don't might not be the same for each adjective.

With respect to the examples in (9), we presented the MT workers with several variants: tense variation (past/present) and three different subject conditions (specific subject, non specific but explicit sub-

ject and no subject), as well as the difference between *of* and *for* PPs as illustrated below:

(11) a. It was fool hearted of John to go on a trip around the world.

b. It was fool hearted of old people to go on a trip around the world.

c. It was fool hearted for John to go on a trip around the world.

d. It was fool hearted for old people to go on a trip around the world.

e. It was fool hearted to go on a trip around the world.

f. It is fool hearted of John to go on a trip around the world.

g. It is fool hearted of old people to go on a trip around the world.

h. It is fool hearted for John to go on a trip around the world.

i. It is fool hearted for old people to go on a trip around the world.

j. It is fool hearted to go on a trip around the world.

We haven't yet analyzed the results in detail but only 4 Turkers out of 10 rated (11j) as having happened whereas 9 out of 10 found (11a) to be factual. There is no study of what is going on here but theoretical linguists wouldn't be too upset about the facts observed and invoke something like generic readings to account for the difference. From our more practical point of view, we observe that having an explicit subject and being in the past tense makes a difference. We need further studies to determine what the importance of various factors is.

The insufficiency of the syntactic frame information is illustrated even more dramatically with *lucky*. Here the use of the future tense changes the interpretation dramatically. Whereas in the past tense, *lucky* behaves as a factive or implicative adjective (see above), in the future it can have an idiomatic meaning illustrated in

(12) Wong Kwan will be lucky to break even. (from theFactBank (Saurí and Pustejovsky, 2009))

Here the speaker expresses the opinion that it is unlikely that Wong Kwan will break even. This idiomatic meaning seems to be the meaning that is predominant with the future tense but, unfortunately for annotation purposes, it is not the only possible one (see (Karttunen, 2013) for more details on *lucky*):

(13) Sooner or later, a drug company will be lucky to find such a molecule.

It is clear then that there are several factors beyond the syntactic frame as generally understood that play a role in determining the inferences of lexical items.

This situation is not specific to adjectives. Factive verbs have been studied in some detail and it has been noticed that, for some of them, the veridicity status depends on factors such as the person of the matrix clause. The most recent study that we are aware of is (Beaver, 2010) from which we the following examples.

(14) a. He is not aware that Morris saw the letter.

b. I am not aware that he [Morris] saw the Daschle letter. (CNN, November 2001, taken from (Beaver, 2010))

Whereas in the a-example, the embedded clause seems factual, this is not the case in the b-example. (Beaver, 2010), however, also gives examples of third person use where the factive presupposition is cancelled:

(15) Mrs London is not AWARE that there have ever been signs erected to stop use of the route, nor that there has ever been any obstruction to stop use of the route. (County Environment Director, Definitive Map Review 1996/2000, Public Rights of Way Committee, Parish of Aveton Gifford, 2000)

Another well-known environment that influences the status of factives is the antecedent of a conditional. The case of first person cancellation has been known for a long time. The third person case in mainly documented in (Beaver, 2010).

(16) a. If I REALIZE later that I have not told the truth, I will confess it to everyone.

 b. If anyone DISCOVERS that one of our volunteers is charging money for being a volunteer, please notify me ASAP.(Tom Elliott, GenWeb, Waldo County, Maine, 30 Nov., 2000, taken from (Beaver, 2010))

4.4 What can be done with lexical signatures?

Lexicon annotation practice tends to take the lexical item into account and the syntactic frame. The data above suggests that much more needs to be taken into account, even in experimental settings that mimic that of linguistic introspection. The existence of variation shows that we have to allow for ambiguities in inference patterns, even when there are no detected meaning differences and the syntactic frames, as usually understood, are the same. The data in section 4.2 suggests two possible approaches: we could try to encode more specific patterns or we could base the attribution of a feature such as +factive on a 'prototypical environment', the kind of environment linguists have assumed tacitly. The first approach would most likely lead to an unmanageable explosion of frames. The second approach makes features such as +factive conditional; contrary to linguistic practice it is important to spell out the exact conditions in which they are supposed to hold. Further study of IMPLICATIVES and EPISTEMIC MODALS will most likely lead to similar conclusions.

Is it, however, possible to spell out these conditions? In what precedes we have talked as if the variation that we observe is due to morpho-syntactic factors such as tense. But providing more context, it is of course also perfectly possible to have generic interpretations, not implying factuality, with factive adjectives in the past as the following example illustrates:

(17) In the Middle Ages it was daring to express anything except orthodox opinions.

And again, although the idiomatic meaning of *lucky* occurs mainly with the future tense, one can find it in the past with non specific subjects:

(18) Just a hundred years ago a man was lucky to live to be 45.

The real conditioning factors that determine these interpretations are not morphological or syntactic; they are themselves semantic: it is not tense per se that influences the interpretation of *lucky* or of factive adjective complements, it is a form of genericity. There is no way that, in the current state of affairs, we can detect genericity directly.

Beaver concludes his corpus study stating

> I doubt that there is any general principle that would enable one to predict from the written form of an arbitrary sentence involving a cognitive factive whether the factive complement is presupposed by the author. Certainly, there is a tendency for the complement to be presupposed. And certainly there are types of sentence involving cognitive factives, notably in the first and second person, for which the complement is rarely if ever presupposed. But the grey area, the range of cases for which no small set of formal features of the text would tell you whether the complement is presupposed or not, is just too big.

This conclusion, however, is not very satisfying from a computational point of view. Whereas the situation might be complex, there is a need for approaches that are more sophisticated than the one described in the beginning of this section but less defeatist than Beaver's[4]. We will continue to run into this unsatisfactory situation as long as we don't have systems that can directly couple NLU to real world experiences. At this point we have to work with morphological and syntactic proxies. Past tense, for instance, is a good proxy for episodical as distinct from generic interpretation, as discussed in (Mathew and Katz, 2009). But, being proxies, our features in this domain can only give us probable inferences. Whatever system that is built on them needs to provide for means to override them.

---

[4]Beaver goes on discussion some factors that might play a role in spoken language and points to an information structure based solution. The ingredients of that solution will not be computationally available for some time to come.

References

David Beaver. 2010. Have you noticed that you belly button lint colour is related to the colour of your clothing? In R. Bauerle, U. Reyle, and T. E. Zimmermann, editors, *Presuppositions and Discourse: Essays offered to Hans Kamp*, pages 65–99. Elsevier.

Lauri Karttunen. 1971. Implicative verbs. *Language*, 47:340–358.

Lauri Karttunen. 1974. Presupposition and linguistic context. *Theoretical Linguistics*, 1(1):181–194.

Lauri Karttunen. 2012. Simple and phrasal implicatives. In **SEM 2012*, pages 124–131, Montréal, Canada. Association for Computational Linguistics.

Lauri Karttunen. 2013. You will be lucky to break even. In Tracy Holloway King and Valeria dePaiva, editors, *From Quirky Case to Representing Space: Papers in Honor of Annie Zaenen*, pages 167–180. CSLI Publications, Stanford, CA.

Paul Kiparsky and Carol Kiparsky. 1970. Fact. In M. Bierwisch and K. E. Heidolph, editors, *Progress in Linguistics*, pages 143–173. Mouton, Hague.

Karin Kipper-Schuler. 2005. *Verbnet: a broad-coverage comprehensive verb lexicon*. Ph.D. thesis, University of Pennsylvania.

Angelika Kratzer. 2012. *Modals and Conditionals. New and Revised Perspectives*. Oxford University Press, Oxford, U.K.

Terence Langendoen and Harris Savin. 1971. The projection problem for presuppositions. In C.J. Fillmore and D.T. Langendoen, editors, *Studies in Linguistic Semantics*. Holt, Rinehart and Winston, New York.

Bill MacCartney and Christopher D. Manning. 2009. An extended model of natural logic. In *Proceedings IWCS-8 '09 Proceedings of the Eighth International Conference on Computational Semantics*, pages 140–156. University of Tilburg.

Thomas A. Mathew and E. Graham Katz. 2009. Supervised categorization for habitual versus episodic sentences. In *The Sixth Midwest Computational Linguistics Colloquium 2009 at Indiana University Bloomington*.

Rowan Nairn, Cleo Condoravdi, and Lauri Karttunen. 2006. Computing relative polarity for textual inference. In *ICoS-5*, pages 67–76.

Neal R. Norrick. 1978. *Factive Adjectives and the Theory of Factivity*. Niemeyer.

F. R. Palmer. 2001. *Mood and Modality*. Cambridge University Press, Cambridge, U.K.

Christopher Potts. 2005. *The Logic of Conventional Implicatures*. Cambridge University Press, Cambridge, United Kingdom.

Roser Saurí and James Pustejovsky. 2009. Factbank 1.0. Linguistic Data Consortium, September.

Roser Saurí and James Pustejovsky. 2012. Are you sure that this happened? assessing the factuality degree of events in text. *Computational Linguistics*, 38(2):261–299.

Roser Saurí. 2008. *A Factuality Profiler for Eventualities in Text*. Ph.D. thesis, Brandeis University.