

# Believing in Words

Herman Cappelen and Josh Dever

## §1 Lexical Theories of the Attitudes

The semantic puzzles posed by propositional attitude contexts have, since Frege, been understood primarily in terms of certain substitution puzzles. We will take as paradigmatic of such substitution puzzles cases in which two coreferential proper names cannot be intersubstituted *salva veritate* in the context of an attitude verb. Thus, for example, the following sentences differ in truth value:

(1) Lois Lane believes Superman can fly.

(2) Lois Lane believes Clark Kent can fly.

despite the fact that "Superman" and "Clark Kent" pick out the same individual.<sup>1</sup>

Equivalently, the following sentence may be true:

(3) Lois Lane believes that Superman can fly, but that Clark Kent cannot fly.

despite the coreferentiality of the names. (It will at times be convenient to appeal to this conjunctive attitude report in order to fix a single context of utterance.)

Substitution failures such as these create a puzzle when conjoined with the assumptions (a) that attitude reports report a binary relation between an individual and some object of that individual's attitude and (b) that that object of the attitude is determined by the content of the complement sentence in the attitude report. If all of the terms in two complement sentences (e.g., "Superman can fly" and "Clark Kent can fly") have the same semantic content, then, *prima facie*, they ought to generate the same object of believe and, *a fortiori*, materially equivalent attitude reports. Frege, famously, attempts to defuse the puzzle by positing a semantic value of sense in addition to that of reference, and thereby distinguishing the semantic contents of the two complement sentences.

Our focus is on an alternative class of theories that we label *lexical theories of the attitudes*. Such theories can be roughly characterised as theories according to which the lexical material used in the attitude report is semantically relevant. The theories we have in mind include those proposed by Burdick (1982), Harman (1972), Higginbotham (1986), (1991), Kaplan (1990), Larson and Ludlow (1993), Larson and Segal (1996), Pietroski (1994), (1996), Richard (1990), and Segal (1989).

---

<sup>1</sup> There are those, of course, who hold that, *contra* Frege, coreferential names can always be intersubstituted *salva veritate*. Although we ourselves have some sympathy for this position, we will henceforth always assume that propositional attitude contexts do not generally license such substitutions, since our target here is the viability of one particular genre of explanation of such substitution failures.

More precisely, a lexical theory is any theory which meets the following four conditions:

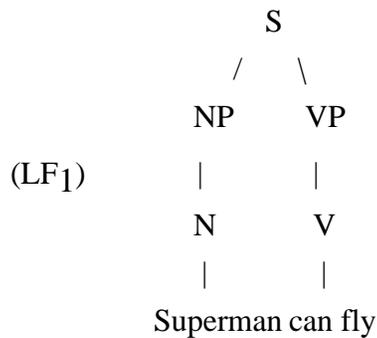
- (A) Attitude reports report a relation between an individual and some object of that individual's attitude.
- (B) The object of the attitude is determined by the content of the complement sentence in the attitude report.
- (C) Lexical material is in some contexts a *functional component* of the objects of the attitudes.
- (D) The *Cognitive Ken Constraint* holds.

We begin by explaining what we mean by a *functional component* and the *Cognitive Ken Constraint*.

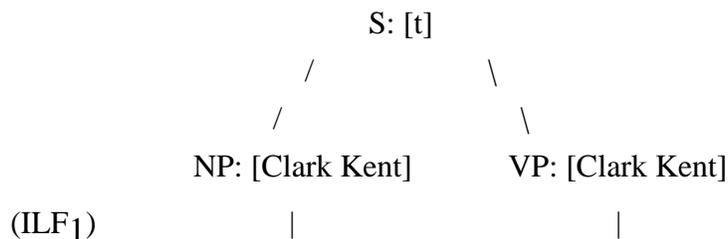
### Functional Components of the Attitudes

According to all the theories mentioned above, the objects of the attitudes include lexical material. We'll briefly describe the relevant features of so-called ILF theories and Richard's theory.

According Harman (1972), Higginbotham (1986), (1991), Larson and Ludlow (1993), Larson and Segal (1996), and Segal (1989)) objects of the attitudes are taken to be *interpreted logical forms (ILF)*. An interpreted logical form is a logical form in the usual sense -- a level of syntactic representation designed to serve as input to the process of semantic interpretation (see, e.g., Chomsky (1981), (1995)) -- augmented by coupling of each syntactic node with the semantic value of that node. For example, if the complement sentence of (1) above has a logical form such as:



it will have an ILF such as:



N: [Clark Kent]	V: [Clark Kent]
Superman: [Clark Kent]	can fly: [Clark Kent] <sup>2</sup>

An ILF incorporates at least three types of information: the semantic content of the complement sentence, the syntactic organization of that content, and the lexical means by which that content is introduced.

According to Richard (1990) the semantic value the complement clause of a belief report is what he calls *Russellean Annotated Matrices* or RAMs. RAMs consist of the expressions occurring in the complement clause and their referents. So the RAM for "that Superman can fly" is

<<"Superman", Superman>, <"can fly", <being able to fly>>

What these two strategies share with each other (and with other lexical theories) is the suggestion that the objects of the attitudes contain lexical material (we'll call objects of the attitudes containing lexical material *Lexical Objects of the Attitudes*, LOA's). The introduction of LOAs might seem to have the advantage of immediately showing why the report in (1) is not equivalent to the report in (2). (1) claims that Lois stands in the belief relation to a LOA containing, among other things, the word "Superman" (and not the words "Clark Kent"), while (2) says that she stand in that relation to an object containing, among other things, the words "Clark Kent".

However, the context sensitivity of attitude reports necessitates a qualification and the introduction of the notion of an *active context*. Lexical theories take the necessary relation between the believing agent and the believed LOA to be a highly context-sensitive one, similar to Davidson's "same-saying" relation.<sup>3</sup> Not every aspect of a LOA is in every context required to bear some important relation to the mental life of the reported agent. Thus, for example, in at least some contexts the attitude report:

(4) Galileo believed that the earth moves.

can be true even though the LOA of (4) contains lexical material which played no part in Galileo's mental life -- the relevant role in his mental life having been filled by the lexical material "eppur si muove". In these contexts the mere fact that the *semantic* content of the relevant LOA bears an important relation to Galileo's mental life suffices to guarantee the truth of the attitude report; in other contexts, the lexical differences between us and

---

<sup>2</sup> We follow Larson and Ludlow (1993) in the details of representing ILFs. Both the syntactic analysis used in (LF<sub>1</sub>) and the semantic analysis added in (ILF<sub>1</sub>) are clearly sketchy, but the ways in which these analyses might be made more precise will prove irrelevant to our concerns.

<sup>3</sup> See, e.g., Larson and Segal (1996), 422-424; Larson and Ludlow (1993), 339-342; and Richard (1990), 33-41.

Galileo might suffice to falsify (those utterances of) the report. Determination of what features of a given ILF are relevant, and the ways in which they are relevant, will typically be the result of what Larson and Ludlow calls a 'negotiation' between speaker and audience.

We will call a component  $\alpha$ -- semantic, syntactic, lexical, or otherwise -- of an LOA a *functional* component if there is some *possible context* in which two attitude reports  $A_1$  and  $A_2$ , differentiated only in that one contains  $\alpha$  while the other does not (replacing it with some appropriate component  $\beta$ ), differ in truth value. If there is any context in which it is true that:

(1) Lois Lane believes that Superman can fly.

but false that:

(2) Lois Lane believes that Clark Kent can fly.

then the lexical component "Superman" of the LOA of the complement sentence of (1) and the lexical component "Clark Kent" of the LOA of the complement sentence of (2) are functional. Given a pair  $\alpha, \beta$  of LOA components, call a given context  $C$  an *active* context for  $\alpha$  and  $\beta$  if that context is a witness to their functionality -- if, that is, there are two possible attitude reports, differentiated only in that the LOA's of their complement sentences contain  $\alpha$  and  $\beta$  respectively, which differ in truth value. As a convenient abbreviation, we will subscript names of utterances to indicate when they occur in active contexts. Thus " $u_{A:S,CK}$ " will name an utterance made in an active context for the lexical pair "Superman", "Clark Kent".

What lexical theories have in common, then, is not that the lexical material is relevant in *all* context, only that there are *some* contexts in which the lexical material used in the report is a functional component of the objects of the attitudes.

### **The Cognitive Ken Constraint**

For a lexical theory to be complete it must specify what *sort of relation* agents need stand in to the LOA. The lexical information included in LOA's, as discussed above, are intended to explain certain failures of material equivalence of attitude reports in contexts active for that lexical material. Therefore the semantics of the verbs of attitude reporting - - believes, doubts, wishes, etc. -- must appeal to a relationship between the attitude-holding agent and the object of his attitude which is *sensitive to* (i.e., the holding or not holding of which can be determined by) those pieces of lexical information.

Following the lead of lexical theorists, we will not attempt to give a full explanation of the factors which contribute to the holding or failing of the attitudinal

relation. We will instead focus on one minimal constraint on that relation, what we will call the *Cognitive Ken Constraint*:

(CK) Given a report of the attitudes of agent A made in a context C active for  $\alpha$  and  $\beta$ , A must stand in some psychological/cognitive relation to at least one of  $\alpha$  and  $\beta$ .

Thus, for example, if we make reports (1) and (2) above in a context in which they differ in truth value, then, since we have thus identified an active context for the lexical pair "Superman", "Clark Kent", Lois Lane must stand in some cognitive relation to at least one of these two names. One consequence of this constraint is that (1) and (2) cannot differ in truth value if Lois Lane has never been exposed to either of these two names for Clark Kent, no matter how much else she may know about the individual Kent. We take it that this constraint is intuitively unobjectionable. If the reports are genuinely reports of the attitudes *of the agent* how could aspects of the object of the attitude *make a difference* to whether that attitude holds if those aspects are wholly outside the cognitive ken of the reported agent?<sup>4</sup>

Some lexical theorists explicitly endorse CK. When describing a context that in our terminology is active for "Twain" and "Clemens", in which Mutt truly utters (5):

(5) Odile believes that Twain is dead

Richard says:

...the speaker and audience are focusing on a specific name, term or other way of representing something. ... This mutual focus produces a restriction on the use of "Twain" to represent something about the how of Odile's belief. In Mutt's case a restriction like -Use "Twain" to represent "Twain"-is operative (p.135).<sup>5</sup>

Larson and Ludlow, however, seem to reject CK. Our arguments against lexical theories rely heavily on CK, so we will spend the rest of this section justifying the claim that lexical theorists, indeed all theorists, should accept this constraint.

Larson and Ludlow say:

---

<sup>4</sup> While we are assuming that all ILF theorists (and, below, all lexical attitude theorists) accept direct reference, note that the plausibility of the CK constraint does not depend on direct reference. Were one Fregean about names, one might think that "Superman" carries a sense something like *superhero in tights and cape* while "Clark Kent" carries a sense something like *mild-mannered reporter*, and thus that Lois, even if she had never heard the names, might have (1) true of her and (2) false if she had seen Kent in his superhero guise flying through the sky and also seen Kent in his reporter guise carrying on his prima facie flightless life. But if one adopts this Fregean position, then the senses of "Superman" and "Clark Kent" also feature in the ILFs, and the ILFs of the complement sentences of (1) and (2) are differentiated semantically (via those senses) as well as lexically, so we fail to have here an active context for the lexical material alone. If one is a direct reference theorist, the above situation does yield an active context, but there should be no tendency to think that (1) and (2) differ in truth value absent Lois' exposure on those names.

<sup>5</sup> For further discussion of Richard's notion of words, see §4 below.

If our perspective differ from most other theories of attitude ascription it is in our emphasis on the relation between the ascription and the hearer, rather than on the relation between the ascription and the agent to whom the attitude is ascribed. In our view, it is the studying of the former relationship what will yield the biggest dividends in the understanding of belief ascription (p.342).

On this view, the LOA of the complement clause (in this case an ILF) is constructed not to describe the belief state of the believer, but to help the audience model the contextually relevant features of that belief state. This can be done by introducing active components in the LOA that the believer is unfamiliar with. Consider the following example. Lex has never heard the names "Superman" or "Clark Kent", but has seen a man in tights and cape flying through the sky, and a mild-mannered reporter sitting at his desk. Speaking to Lois, who has heard the names, I utter<sub>A:S,CK</sub>:

(6) Lex believes that Superman can fly but that Clark Kent cannot.

The lexical material is active here not because I'm reporting some cognitive relation Lex bears to the names, but rather because I know that the associations Lois has with those names will enable her accurately to model Lex's belief system based on what I said. She will, for example, know that Lex will be unsurprised to see a man in tights and cape leap into flight, but would be shocked to see the mild-mannered reporter become airborne.<sup>6</sup>

We say Larson and Ludlow *seem* to deny CK because they say about the relationship between the speaker and the audience that it is "...fundamentally a matter of *usage* and not content, and that the correct account of these phenomena falls outside the domain of semantics proper and into pragmatics." (p.339).

This claim can be taken in two ways. One the one hand it can mean that the speaker-audience relationship is entirely irrelevant to *semantics*; that it has no bearing on what is said by a belief report. If this is the correct interpretation, their remarks do not amount to denying CK. If so, Larson and Ludlow are directly subject to our subsequent arguments.

However, we suspect that this is not what they have in mind. The alternative interpretation is this. In the truth conditions of attitude reports there will be no explicit mention of speaker-audience relationships (in that sense this is not part of the semantics)<sup>7</sup>. It is, however, part of the meaning of the expression "x believes y" (where x

---

<sup>6</sup> Note that Lois's associations with the names "Superman" and "Clark Kent", which I exploit in enabling her to model Lex's belief state, are not part of the meanings of the names. If they were, then the context would not be active for the lexical material "Superman" and "Clark Kent", since the two LOAs would be differentiated semantically as well as lexically.

<sup>7</sup>Just as they don't think the truth conditions for "believes" should appeal to similarity, but nevertheless don't deny that the belief relation appeals to similarity.

is an agent and y an ILF) as it occurs in the statement of the truth conditions for belief reports, that x stands in the following relationship to y: *y helps the audience of the utterance model the contextually relevant features of x's beliefs*. On this view, the contextual features would not actually be mentioned in the truth conditions of the attitude report (in that sense they are not part of the semantics), but those who know what "x believes y" means, know that these contextual features (i.e., the relationship between the speaker and the audience) must be taken into account when evaluating a belief report.

On this interpretation, Larson and Ludlow are committed to the view that "x believes y" is *true* just in case y helps the audience model the contextually relevant features of x's beliefs.<sup>8</sup> This would amount to denying CK because an ILF could certainly help someone model x's beliefs even though x stands in no cognitive relationship to that ILF.

Larson and Ludlow give no arguments for this view; they simply state it at the end of their paper. As a result we are left in the dark about what has led them to this rather unorthodox position, and there are not positive arguments to respond to. We will, however, briefly outline three arguments against this view of the semantics for propositional attitude verbs.

First, Larson and Ludlow interpreted as above are committed to the view that (1) is about the report's intended audience and that audience's relationship to Lois. But why should we think that (1) is about anyone but Lois? Sometimes such counter-intuitive positions can be justified, but the burden of proof is certainly on those who advocates such views. In the absence of positive argument for the view (Larson and Ludlow give no such arguments), it is not a position that has a great deal going for it.

Second, if we can appeal to associations the audience has with given names, we can also appeal to mis-associations the audience has. But appeal to mis-associations will, if truth of a belief report amounts to successful exploitation of those mis-associations toward a goal of inducing accurate belief modeling, count as true some obviously false claims. Suppose, for example, I want to say (truly) that Lois thinks Nicolas Cage is a lousy actor. Suppose further that I know that my intended audience consistently gets Nicolas Cage and John Travolta confused. Finally, suppose Lois does not think that John Travolta is a lousy actor. According to Larson and Ludlow, an utterance of (7) could, in this context, be true because it helps the intended audience to model one of Lois beliefs.

(7) Lois thinks John Travolta is a lousy actor.

---

<sup>8</sup> Just as "I'm hungry" is true just in case the speaker is hungry.

(7), however, is not true under these circumstances. (7) may be pragmatically useful in getting the audience to think the right way about Lois, just as in general falsehoods can be useful in communicating truths. But that pragmatic utility cannot be read into the semantic content of (7), and hence exploitation of that utility provides no reason to doubt CK.

Third, on the interpretation of Larson and Ludlow that we are now considering, they claim that a belief report says that the ILF expressed by the complement clause of the report will 'cause a hearer H to form a certain theory about the belief structure of an agent A" (p.340). What is the content of such a 'theory of the belief structure of an agent"? The surprising thing is that it can't say anything about what Lois believes because if in that theory it says 'Lois believes that p' what that really means is: *p will help the audience of this utterance to model her beliefs*. It is not clear to us what it is to have a theory of her beliefs, when that theory says nothing about what she believes<sup>9</sup>. Hence it is not clear what it means to say that a belief report helps the audience model Lois' beliefs.

We realize that this is a subject that deserves further discussion, but think we have said enough to justify limiting our arguments to those lexical theorists who accept CK.

### **The Structure of Our Objection**

Our purpose is to challenge the assumption that LOAs are useful objects of the attitudes and hence that lexical theories of attitude reports are viable. Our central strategy will be to mimic a worry which arises for Fregean sense-based theories of the attitudes. Frege, in positing an additional semantic layer of sense to account for semantic features of attitude reports, opened up new ontological questions about the nature of senses, and it is the perceived inability adequately to answer these questions that have persuaded many people to move away from sense-based accounts of attitude reports. Our suggestion is that the incorporation of lexical material as functional components of LOAs opens similar ontological questions for lexical theorists -- we are now in pressing need of an account of what these lexical materials might be. Moreover, we think that there is no adequate answer to these new questions. No account of what lexical items are will allow them to play the role that lexical theorists need them to play.

In order to show this, we will consider here three broad approaches to the ontology of words, and show that each of the three, when placed within the LOA framework sketched above, meets insurmountable obstacles. First, we will consider the

---

<sup>9</sup> Larson and Ludlow could answer: Well, in some contexts (for instance the context of the theory of her beliefs) there is no appeal to the audience, there is simply a direct claim about Lois. If this is their answer, we suppose this means that they accept CK in those contexts. If this is their view, their position is susceptible to our arguments below.

"common-sense" view that words are sounds or patterns of letters, and more sophisticated variants on this common-sense view. Second, we will consider the possibility that contemporary linguistic theory provides us with a lexical ontology which will successfully undergird the LOA appeal to words. Finally, we will consider a rather more idiosyncratic view, recently endorsed by Kaplan (1990), Millikan (1984), and Richard (1990), that words are metaphysically complex causal chains of inter- and intra-personal stages. Having rejected all three of these approaches, we will find ourselves forced to reject the lexical project along with them.

## **§2 A Common-Sense Ontology of Words**

Utterances are public entities composed of various sounds, shapes, etc. We will consider first what we take to be the "common sense" view on lexical ontology: the view that words are the types of sounds, shapes, etc. tokens of which compose utterances.

### **Sign Systems and Their Relations**

In discussing this view it will prove useful to have access to two pieces of terminology for talking about signs and relations between signs: "sign systems" and "the same-sign relation".

- *Sign Systems:*

When we use a language, say English, we can do so in many different physical media. We can write English or we can speak it, we can write it using the roman alphabet (and if so, in indefinitely many fonts), in Braille, in Morse code or wave it in semaphore. These are different ways of producing the words of a language. We call such ways *sign systems*. There is no natural limit to kinds of sign systems natural languages can be expressed in. New sign systems are constantly developed. A good case study for the development of new sign systems is Stephen Hawking, who now communicates, in English, through a sign system based on eye movements.<sup>10</sup>

No language is essentially tied to one particular sign system or set of sign systems. In most natural languages, spoken language takes on more practical significance than other sign systems. However, if as the result of some global catastrophe, the human race went mute and forgot about spoken language, you could still read this paper, and you would be reading a paper written in English. You could write a reply to it, and that reply could be in English. No one particular sign system or set of sign systems is essential for the continued existence of English.

---

<sup>10</sup>Hawking suffers from a degenerative neurological disorder which leaves him with insufficient motor control for normal speech. He communicates using a computer setup which tracks his eye movements and, based on them, produces spoken English.

As a result, speakers who share no sign systems can share a language. Consider A, B and C. A speaks, but is illiterate (i.e., does not write or read). B is deaf/mute and only reads and writes. C is blind and deaf and only communicates in Braille. A, B and C can all be English speakers, even though they are incapable of communicating for lack of a common sign system.

- *Same-Sign Related*

Sign systems are correlated in such a way that signs within different sign systems count as the same sign.

*oyster*

and

OYSTER

and

oyster

all have the same letters in the same configuration and make up the same word. You can wave this configuration of letters in Semaphore, or produce it in Braille dots or in Morse code. Certain patterns of sounds also count as utterances of this sign, as do certain hand movements (in sign language).

When two signs in different sign systems are correlated in such a way that they are the same sign, we'll say that these signs are *same-sign related* (or SSR). The nature of the same-sign relation is complex and interesting, but a detailed study of it is beyond the scope of this paper. For our purposes, the most important feature of the same-sign relation is its conventionality. One important upshot of this conventionality is that there is never any absolute answer to the question: is token  $T_1$  same-sign related to token  $T_2$ ?

This question can only be answered relative to a set of conventions for how to correlate sign systems.

### **Lexical Information Only From the Sign System of the Belief Report**

We are considering the view that the ontology of the lexical component of LOAs is provided by our common-sense lexical ontology. Having fleshed out that ontology in terms of sign systems, we ask in what sign system the lexical material of a given LOA appears. The first suggestion will be that the sign system of that material is inherited from the sign system of the belief report. A spoken report appeals to a LOA containing sound patterns; a written report to a LOA containing shapes; a Braille report to a LOA containing dot configurations. Consider, then, a spoken utterance  $u_{A:S,CK}$  of (3):

(3) Lois Lane believes Superman can fly, but that Clark Kent cannot fly.

The lexical theory says that *u* is true only if Lois stands in the belief relation to a LOA containing phonological material and hence stands in a certain psychological/cognitive relationship to the phonological type instantiated by a spoken token of "Superman can fly, but Clark Kent cannot fly".

This suggestion is implausible for a rather obvious reason. It presupposes that the belief reporter must share a sign system with the reportee in order for *u* to be true. *u* can certainly be used to say something true about a blind person who only communicates in Braille. There can be no requirement on the truth of *u* that Lois has can use the same sign systems as the utterer of *u*.

### **Belief Reports Across Sign-Systems**

The lexical material that enters into LOAs must, then, be understood as encompassing information from a wide range of sign systems, not just the sign system the report is made in. But how wide a range of sign systems, and how is that range chosen?<sup>11</sup>

Since it is not a necessary condition on the truth of an attitude report that the reporter and the reportee share a sign systems, the answer cannot be that the LOA contains information from all the sign systems that the reporter correlates with the sign system the report is made in, nor can it contain lexical information from the sign systems the reportee correlates with the sign system used in the report.

What is needed is some way to connect the sign systems of the reporter and the reportee. But how do we do that when we can't rely on the way they correlate them? If facts about the reporter and reportee are inadequate to correlate these sign systems, perhaps we rely on the correlations in the broader linguistic community. The *view* would be that the lexical information in the LOA is from the set of sign systems that by community standards are correlated with the sign system used in the belief report. The lack of intersection between sign systems used by reporter and reportee thus ceases to be a problem, since there are community standards (sometimes unknown by both reporter and reportee) for correlating, say, spoken language and Braille. So assuming as before that the reporter only knows spoken sign systems and the reportee only Braille, a spoken utterance  $u_{A:S,CK}$  of (3) can be true because the reportee might stand in the appropriate psychological/cognitive relationship to the Braille information that by community

---

<sup>11</sup> Note that the answer cannot be "include all sign systems". Since the same sign system can be SSR to other sign systems in more than one way, if we want to accommodate the beliefs of speakers using different SSR standards, we would also have to include all possible correlations between sign systems. But then all possible lexical types would be included in every LOA, and the difference in lexical material could never serve to individuate LOAs.

standards is correlated with the spoken version of "Superman can fly, but Clark Kent cannot fly".

To see why such a theory would be mistaken, take a written utterance<sub>A:S,CK</sub> of (3) and assume that the report concerns a Lois that is an English speaker familiar only with Braille. Imagine that one night while Lois sleeps, the rest of the linguistic community gets together and decides to produce new correlations between written language and Braille. Under this new correlation the written version of "Superman can fly, but Clark Kent cannot fly" ends up being correlated with what is now the Braille correlate of "Gödel can fly, but Lesniewski cannot fly". Remember, "Superman can fly, but Clark Kent cannot fly", means just what it means now, the community has just decided that it wants the Braille correlate of that sentence to be different. Unbeknownst to Lois, the communal correlation between written language and Braille has changed.

Clearly, the community changing its correlations between Braille and written language is entirely irrelevant to the truth or falsity of *u*. However, under the suggestion now being discussed, such a change could affect the truth value of *u*. It might very well be the case that Lois stands in the belief relation to the LOA containing, among other things, the Braille version of "Superman can fly, but Clark Kent cannot fly", but not to such an ILF containing the Braille version of "Gödel can fly, but Lesniewski cannot fly". The Braille types "Gödel" and "Lesniewski" might play no role in her mental life, and that would suffice to make *u* false. This is an unacceptable result. Lois's beliefs are unaffected by how the community correlates written language and Braille. We can't change her beliefs simply by correlating sign systems in new ways.

We don't need such exotic cases to see the non-necessity of the present suggestion. All we need is to consider the possibility that Lois (who, remember, is familiar only with Braille) misspells the Braille version of "Superman can fly, but Clark Kent cannot fly". Suppose her Braille spelling of "Superman" is what, by conventional correlations, would be the Braille correlate of "Supperman". We can assume that the standard Braille version of "Superman" plays absolutely no role in Lois's mental life. This would, according to the theory under consideration, be sufficient to make (1) false. Again an unacceptable consequence of the lexical theory.

### **The Source of the Problem**

The source of the problem is this. What Lois believes is not a matter of communal convention and does not change simply by changing communal conventions. How sign systems are correlated is a matter of convention and making the truth value of belief reports contingent on these correlations between sign systems is therefore misguided. We

don't have a proof that lexical theories cannot get around this problem, but absent some ingenious suggestion, we suggest that all attempts to introduce phonological, graphological, Brailological etc. information will encounter problems very similar to those we have outlined above.

### **§3 Lexical Items in Contemporary Linguistic Theory**

Some might suspect that the difficulties, raised in the previous sections, with satisfactorily implementing the notion of a lexical item within lexical theories of the attitudes, results from an overly naive view of what words are. Both Larson and Ludlow (1993) and Larson and Segal (1996), for example, quite explicitly hold that the contents of the terminal nodes will not be words as understood in our folk theory -- i.e., the concatenations of sounds or letters with which we began the investigations of the last section -- but rather as entities whose nature will be explored and explained by the work of contemporary linguistics. In this section, we will argue that importing understandings of the lexicon from linguistics, far from allowing lexical theorists to escape the difficulties of the last section, actually compounds those difficulties.

It is *prima facie* plausible that there will be as much diversity among syntacticians concerning the nature of lexical items as there is concerning the nature of the syntactic theories built upon those lexical items. For much of their work, most linguists seem content to work with a rough-and-ready understanding of lexical ontology. Thus, for example, Larson and Ludlow in their discussion of ILF accounts of attitude reports give two canonical sources for contemporary linguistic views on the lexicon. Both of these canonical sources have only cursory comments on what *kinds of things* fill the lexicon. Haegeman (1991), for example, says:

We postulate that speakers of a language are equipped with an internal 'dictionary', which we shall refer to as the mental lexicon, or lexicon, which contains all the information they have internalized concerning the words in their language. As seen above, this mental lexicon will have to contain, among other things, information on syntactic categories. We assume that each word of the language known by a speaker will be listed in his mental lexicon with its categorical specification. For instance, a native speaker of English will presumably have a list containing the following information:

- |                      |                  |
|----------------------|------------------|
| 6a <i>meet</i> :     | verb             |
| 6b <i>employer</i> : | noun             |
| 6c <i>castle</i> :   | noun             |
| 6d <i>at</i> :       | preposition (37) |

Radford (1988), Larson and Ludlow's second canonical source, takes a similar, if somewhat more precise, line, appealing to the existence of:

*A Lexicon* (or dictionary) containing a list of all the words in the language, together with a specification of their idiosyncratic syntactic, semantic, phonological, and morphological properties. (337)

Note that both of these explanations of the mental lexicon give no indication what kinds of things the words *themselves* are. The lexicon is repeatedly described as a pairing of words with other information; barring some ontological significance to the italicization in the lexical list given, words are being taken simply as concatenated strings of letters (and the lexicon subsequently as a pairing of those strings with other (syntactic, etc.) information. This view of words, clearly, offers no advance over the "common-sense" view rejected in the last section.

That linguists are frequently sketchy in their views of lexical ontology is, of course, no great criticism of the work of linguists. When one's project is to capture the notion of grammaticality, it is natural to concentrate more on specifying a set of procedures which can take one from a collection of lexical items to all and only grammatical structures built on those lexical items, and less on the exact character of the initial building blocks. However, once lexical theorists take the work of contemporary linguists and embed it in a theory of the attitudes, thereby positing cognitive relations to the syntactic structures employed by linguists, those theorists acquire a pressing need, not felt by the linguists, for a detailed ontology of lexical items. Even when linguists are explicitly concerned with the nature of the lexicon, the kinds of concerns they have will typically be orthogonal to the concerns inherited by lexical theorists. Thus, for example, the detailed studies made by Di Sciullo and Williams (1987) of the internal morphology of words and by Pustejovsky (1995) of the lexical-semantic sources of polysemy, whatever their other virtues, will do nothing to differentiate coreferential proper names which lack internal morphology and share perfectly semantic properties.

### **Chomsky on the Lexicon**

Contemporary linguistic theory, then, offers no "magic bullet" to dispel the worries raised in the previous section. We will now set out one carefully developed view of the nature of lexical items in linguistic theory, turning to the views of Chomsky (in light, especially, of ILF theorists frequent invocations of Chomskian syntactic theory as a linguistic background for their semantic work). We will argue that this view only exacerbates the lexical theorist's difficulties. Chomsky's view on lexical items persists almost unchanged throughout the body of his works. Thus in Chomsky (1965) we are told:

In general, all properties of a formative that are essentially idiosyncratic will be specified in the lexicon. In particular, the lexical entry must specify: (a) aspects of phonetic structure that are not predictable by general rule (for example, in the case of *bee*, the phonological matrix of the lexical entry will specify that the first segment is a voiced labial stop and the second an acute vowel, but it will not specify the degree of aspiration of the stop or the fact that the vowel is voiced, tense, and unrounded); (b) properties relevant to the functioning of transformational rules (as the example of the preceding paragraph, and many others); (c) properties of the formative that are relevant for semantic interpretation (that is, components of the dictionary definition); (d) lexical features indicating the positions in which a lexical formative can be inserted (by the lexical rule) in a preterminal string. In short, it contains information that is required by the phonological and semantic components of the grammar and by the transformational part of the syntactic component of the grammar, as well as information that determines the proper placement of lexical entries in sentences. (87-88).

Similarly, Chomsky (1995) explains:

Items of the lexicon are of two general types: with or without substantive content. We restrict the term lexical to the former category; the latter are functional. Each item is a feature set. Lexical elements head NP, VP, AP, and PP and their subcategories (adverbial phrases, etc.). At D-Structure and LF, each such XP must play its appropriate semantic role, satisfying FI, as discussed earlier. The heads of these categories have (1) categorial features; (2) grammatical features such as  $\phi$ -features and others checked in the course of derivations, continuing to assume one of the interpretations of morphological structure discussed in section 1.1; (3) a phonological matrix, further articulated by the mapping to PF; (4) inherent semantic and syntactic features that determine s(ematic)-selection and c(ategorial)-selection, respectively. (54).

Chomsky's views on the nature of the lexicon expand on the introductory remarks of Haegeman and Radford in two ways. First, Chomsky replaces the notion of the lexicon as a list of correlation between words and packages of information with a less ontologically-loaded picture on which words are themselves merely packages of information. Second, Chomsky is considerably more detailed as to what information goes into a lexical entry.

On Chomsky's view, then, a word is a mentally represented structured informational/computational structure which can serve as an input to various other computational modules in the brain, primarily the syntactic, semantic, and phonological modules. Each of these computational modules will have various "default" procedures for handling certain kinds of inputs, so the informational content of the individual word need be only rich enough (a) to trigger the relevant default procedures and (b) to handle any

deviation from those default procedures. The informational complex which is the word "tree", for example, would have to have enough phonological information to allow the phonological module to produce the correct sound, but that complex will not need to specify that "tree" pluralizes by adding an "s" sound to the end of the word, and may not need to provide a complete pronunciation map to "tree", as certain features of the individual phonemes and their interaction may be deducible by the phonological module from general principles. A word with a deviant plural, such as "cactus", would require, as part of its informational complex, explicit instructions for deviating from the phonological module's default pluralization procedures.

Chomsky thus presents us with a detailed picture of the ontology of words. What remains to be seen is whether that picture is helpful in backing up the project of the lexical theorists. Consider again the classical cases of failure of intersubstitutability *salva veritate* of coreferential names. If "Superman" and "Clark Kent" are two coreferential proper names, then (given direct reference) their semantic and syntactic (i.e., categorical and other grammatical) features will be identical. If there are to be two qualitatively distinct informational structures (words), they must, on the Chomskian picture be differentiated via their idiosyncratic phonological properties.<sup>12</sup> The question thus becomes: can the difference in phonological information between two lexical entries always ground differences in attitudes toward claims containing those two lexical entries? We will give two reasons for answering this question negatively.

### **More Correlation Problems**

There are, of course, language users who do not use or understand phonological representations of language -- for example, deaf/mutes. Taking Chomsky's picture of lexical ontology at face value, it would seem that deaf/mutes are simply incapable of language use, since they can have no phonological information in their lexical entries. At best, they will be incapable of having in their lexicon numerically distinct coreferential proper names, since such names are differentiated only phonologically. Clearly this is not Chomsky's view, since deaf/mutes can have language, and are every bit as capable of possessing distinct words "Superman" and "Clark Kent" as the rest of us. We must

---

<sup>12</sup> Larson and Segal (1996) hold that "it is at least arguable that phrase markers" -- and hence, perhaps, ILFs -- "do not contain phonological information" (581). This view is driven by the desire to select the information content of ILFs by looking at the informational content of syntactically-positd LFs, coupled with a view which takes LF and PF (phonetic form) to lie at the termini of separate derivational branches of sentence formation. It is hard to see, however, how coreferential proper names could ever be functional on such a view, since there would be no information surviving in the ILF to differentiate them. Of course, if ILFs are enriched from the phrase structure marker in the ways suggested, although not explicitly endorsed, by Larson and Segal (453-454), that difficulty might be avoided.

instead take the Chomskian appeal to phonological information as short-hand for information feeding into whatever the primary utterance production module for a speaker is. We must allow for the replacement or supplementation of phonological information with other kinds of information by speakers who use other (non-phonological) methods of producing publicly-accessible utterances -- methods which may be as diverse as written language, sign language, semaphore, symbolic dance, or smoke signals.

Once the Chomskian theory is broadened to accommodate non-phonological language users, the problems of §2 re-emerge in their full force. Consider again the case of an agent truly making an utterance<sub>A:S,CK</sub> of:

(3) Lois believes that Superman can fly, but that Clark Kent cannot fly.  
and assume (a) that Lois Lane communicates only using Braille, and (b) that the reporting agent only speaks (and thus is unfamiliar with Braille). Then (3) will appeal to a psychological relation between Lois and a LOA containing lexical items as understood by Chomsky -- clusters of information, with the informational content of "Superman" and "Clark Kent" being differentiated (for this speaker) only phonologically. Since Lois stands in no cognitive relation to any phonological information, by CK phonological differences between "Clark Kent" and "Superman" cannot explain her differing relations to the claims that Superman can fly and that Clark Kent can fly. Thus the lexical theorist lacks an explanation of the truth of (3). The lexical theorist can, of course, try to tell some story on which the phonological information introduced by the reporter in (3) gets supplemented by Braille-style information, information which would then match the content of the (augmented-Chomsky-style) lexical entries "Superman" and "Clark Kent" in Lois's head. But any successful such story will have to address questions of how Braille structural information and phonological information are correlated, and the arguments of the previous section show that there is no successful way to address these questions.

### **A New Problem for Lexical Theorists**

Appeal to contemporary linguistic theory for views on lexical ontology, then, does nothing to extricate the lexical theorist from the difficulties introduced in the previous section. On the contrary, moving the lexical items from the public sphere into the head of the speaker only introduces the potential for new problems. To see this, consider now a case in which an agent makes a true utterance<sub>A:S,CK</sub> of:

(3) Lois believes that Superman can fly, but that Clark Kent cannot fly.  
in which both the reporting agent and Lois are competent users of spoken English. Inclusion of the public sound constructs "Superman" and "Clark Kent" in the ILF would

thus successfully differentiate Lois's attitudes toward the flying of Superman and Clark Kent. Appeal to internal phonological information, however, may not provide an explanation.

Thus consider the possibility that Lois has a *deviant phonological module*. This module, when receiving (e.g.) the informational content "pronounce long o", does not trigger (as it does in us) a series of muscle movements which terminate in a rounding of the mouth and a production of a long "o" sound. Instead, it triggers a series of muscle movements which terminate in a movement of the tongue to the base of the palate and the production of a "t" sound. If Lois's phonological module thus associates phonological instructions with muscle movements (and hence sound productions) in a non-standard way, but her lexicon is also appropriately non-standard, she will produce the same series of sounds for the same words as we do. But an ILF theory which appeals to her cognitive relation to internally represented phonological instruction sets will get her attitudes wrong, because, due to her deviancy, her lexicon looks nothing like ours, despite being in the same (phonological) sign system.

#### **§4 The Historical Chain Theory of Words**

David Kaplan and Mark Richard agree with the authors discussed above that the objects of beliefs contain linguistic material. They agree with us in that this obliges them to have an explicit ontology of words. They also agree with us in finding the traditional notion of a word inadequate for their purpose.

At the beginning of "Words", Kaplan asks:

Could it be that the elusive cognitive difference between believing that Hesperus is Hesperus and believing that Hesperus is Phosphorus rests on nothing more than syntax?

Kaplan wants to answer this question in the affirmative, but says that this "... speculation lead me to conclude that I had to go back to basics and rethink not just the semantics of names, but their very syntax, the metaphysics of words: How should words be individuated? What is the nature of a word?" (p.94)

In answering these questions, Kaplan develops a radically new notion of a word, a notion inspired by the causal theory of reference. He says:

I think that the token/type model is the wrong model for the occurrence/word distinction (i.e. the utterance/words distinction)...I want us to give up the token/type model in favor of a stage/continuant model (98).

According to Kaplan, "utterances and inscriptions are *stages* of words, which are *continuants* made up of these interpersonal stages along with more mysterious *intrapersonal* stages"(p.98). Someone introduces (which for Kaplan is to create) a word (in the simplest case by a dubbing). She then passes this on to others by producing sounds or inscriptions (these are the *interpersonal* stages of common currency words). This input is stored 'in' others speakers' memory (these are the *intrapersonal* stages of common currency word), who later pass the word on to others by producing utterances or inscriptions. And so it goes on. The common currency word is the entire big thing starting with the introduction/creation, including all the *inter-and-intrapersonal* stages<sup>13</sup>.

Mark Richard endorses a similar view of lexical ontology. He the objects of propositional attitudes to be what he calls 'ramified Russellian propositions' (RAM's). RAMs contain lexical material, but, Richard emphasizes that he does not ascribe to any traditional notion of a linguistic entity: "when I speak of two tokens being tokens of the same name, I certainly do *not* mean that they are *simply spelled or pronounced the same*." (p.183). His alternative is influenced by Kaplan (see note ...for a discussion of the relationship between Richard's view and Kaplan's view):

Part of what makes for sameness of name in the requisite sense is being part of the same "casual chain" of transmission. So the outside condition is one that is in part broadly casual in nature. To be parts of the same representation, proper name tokens have to be residues of the same causal or historical chain, in the sense in which this notion is used in the causal theory of names. (183-4)

### **Objections to LOA Containing Causal Chains**

Setting aside initial questions as to whether the historical chain theory of words is a good theory of words (but see Cappelen (1999) for a detailed criticism of Kaplan), we will focus on the question of whether the entities Kaplan and Richard describe (whether they be words or not) can serve as components of the objects of propositional attitudes in a successful lexical theory. We will focus specifically on Richard's theory.

Our objection to this theory does not require appeal to the substitution puzzles, but can be brought out by considering a simple belief report such as (1), made by Max.

(1) Lois believes that Superman can fly

As other lexical theorists, Richard says that belief reports are context sensitive. It is only given the right contextual features (where these include "the intentions of the user, his

---

<sup>13</sup> Millikan's view is similar. She says: "...what makes tokens of a word or sentence tokens of the *same* word or sentence (in the same language) is, in the first instance, *history*, not form or function" p.73 (see also p.75)

interest, and his beliefs about his audience's interests" p.136) that the lexical material, i.e. the causal chains, become relevant to the truth or falsity of a belief report. Let us imagine then, that Max utters (1) while he 'focuses' (this is the word Richard uses when on p.135 he uses a related example) on the word "Superman" and intends to communicate that Lois uses "Superman" to represent Superman. In our terminology, the utterance of (1) is made in a context that is active for "Superman". Given these contextual features, Max's report can be true, according to Richard, only if Lois's use of the word "Superman" is a part of the same causal history that is somehow picked out by the occurrence of "Superman" in the complement clause of Max's report. It follows from this that Max's utterance of (1) would be false even when (a)-(e) are true.

- (a) Lois uses "Superman" to refer to the same object as Max does in (1) (and we can even assume, with Kaplan and Richard, that Kripke's causal theory of reference is a correct picture of how the referent of "Superman" is fixed).
- (b) Lois pronounces, writes, etc., "Superman" in the same way as Max does.
- (c) Lois and Max associates exactly the same information with "Superman"
- (d) Louse would assent to the sentence "Superman can fly"
- (e) Lois believes of the referent of "Superman" that it is can fly.

Whatever the context is (i.e., whatever intention Max may have had when uttering (1), whatever he may have been focusing on) it is counterintuitive that Max's utterance of (1) could be false when (a)-(e) are true. That, however, is what Richard's is committed to. Max's utterance of (1) could be false even when (a)-(e) are true if the referent of "Superman", was dubbed "Superman" twice. Imagine little Superman first being dubbed "Superman" by his mother and then dubbed "Superman" by his adoptive parents in an unrelated dubbing. This could generate two independent causal chains and hence two different common currency words. It is possible then that Max's use of "Superman" is a part of one of these chains, and Lois's use of "Superman" is a part of the other.

What is in effect the same counterintuitive result can be brought out by thinking about belief reports in modal contexts. Consider Max's claim in (2), made with the same intentions and focus as in the previous case:

- (8) Lois could have believed that Superman can fly

Richard's view, when extended to this case, seems to imply that (8) could be true only if there is a possible world in which Louse is part of the same causal chain as Max is actually a part of when using "Superman". But that does not seem plausible. (8) could be made true by a world in which the name "Superman" was introduced in a way different from how it was actually introduced (say by a different person at a different place) and where the subsequent "causal chain", i.e. the resulting common currency word, developed

in a different way from how it has actually developed. Our intuition, contrary to Richard's theory, is that such a world could make any utterance of (8) true.

### **Conclusion**

The lexical theorist, we have suggested, is as much in need of a theory of words to ground his appeal to LOAs as the Fregean is in need of a theory of senses to ground his appeal to thoughts. Not just any theory of words, moreover, will do -- it must be a theory which will make sense of the lexical theorist's account that certain attitude reports (in certain contexts) are made true or false in part by the words (as understood on that theory) which helped constitute the LOA toward which the attitude is directed. Given the character of our use of natural language, which we can use in an indefinite and open-ended number of physical media, and in arbitrary, conventionally determined, and alterable ways in each of those media, it is difficult to see how any lexical ontology will isolate an entity fixed enough to play a substantive role in the mental life of the agent being reported, but ductile enough connect the mental life of that agent with the arbitrarily and unpredictably selected physical medium of the particular attitude report. While we obviously have not exhausted the full range of possible views on lexical ontology, we have hopefully said enough to show that it is not a trivial task to find a theory of words which can do what lexical theorists need words to do.

It is not unusual for philosophers of language and linguists to treat the notions of a lexical element, an expression, and a word somewhat loosely. For some purposes that might be reasonable, but in many cases it is not. We talk of words and sentences having a meaning, referring, being passed on in causal chains, being referred to by a quotation, etc. It is obviously important in all these cases to have a clear idea of what words and sentences are, i.e., it is important to have an explicit ontology of words. There is, however, a remarkable lack of attention paid to the nature of linguistic entities in the philosophy of language. The discussion of lexical theories revealed that insufficient attention to these issues covered up serious difficulties. We suspect the same may be true of other areas of philosophy of language.

Herman Cappelen  
Vassar College

Josh Dever  
The University of Texas at Austin

## References

- Burdick, H (1982). A Logical Form for the Propositional Attitudes. *Synthese* 52, 185-230.
- Cappelen, H (1999). Intentions in Words. *Nous* 33, 92-102.
- Chomsky, N (1965). *Aspects of the Theory of Syntax*. Cambridge: The MIT Press.
- Chomsky, N (1981). *Lectures on Government and Binding*. Dordrecht, Foris.
- Chomsky, N (1995). *The Minimalist Program*. Cambridge: The MIT Press.
- Di Sciullo, A. and E. Williams (1987). *On the Definition of Word*. Cambridge: The MIT Press.
- Haegeman, L (1991). *Introduction to Government and Binding Theory*. New York: Basil Blackwell.
- Harman, G (1972). Logical Form. *Foundations of Language* 9, 38-65.
- Higginbotham, J (1986). Linguistic Theory and Davidson's Program in Semantics. In LePore (ed.), *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, 29-48.
- Higginbotham, J (1991). Belief and Logical Form. *Mind and Language* 6,4, 344-369.
- Kaplan, D (1990). Words. *Aristotelian Society Supplement* 64, 93-119.
- Larson, R. and P. Ludlow (1993). Interpreted Logical Forms. *Synthese* 95,3, 305-355.
- Larson, R. and G. Segal (1996). *Knowledge of Meaning*. Cambridge, The MIT Press.
- Millikin, R (1984). *Language, Thought, and other Biological Categories*. Cambridge: The MIT Press.
- Pietroski, P (1994). Similarity and Innocent Semantics. McGill University Cognitive Center Manuscript #294.
- Pietroski, P (1996). Fregean Innocence, *Mind and Language* 11,4, 338-370.
- Pustejovsky, J (1995). *The Generative Lexicon*. Cambridge: The MIT Press.
- Radford, A (1988). *Transformational Grammar*. Cambridge: Cambridge University Press.
- Richard, M (1990). *Propositional Attitudes: An Essay on Thoughts and How We Ascribe Them*. Cambridge, Cambridge University Press.
- Segal, G (1989). A Preference for Sense and Reference. *Journal of Philosophy* 86, 73-89.