

## "Lexicography and Ethnographic Semantics"

Charles J. Fillmore  
University of California at Berkeley

### Introduction

For speakers of English, knowing how to use the preposition *in*, in its most basic location-specifying meaning, requires having access to a conceptual structure that we can refer to as *containment* in terms of which it is possible to locate one entity, linguistically, with reference to the interior of another entity. Knowing the preposition *on* and how to use it (the books on the shelf, the mirror on the wall, the fresco on the ceiling) requires a schema involving *surface contact* and *support*. Speakers of English also know implicitly that these same conceptual structures are also exploited for understanding other prepositions and preposition-complexes in English, such as *into* and *out of* alongside of *in*, and *onto* and *off of* alongside of *on*.

(Such primitive conceptual structures are undoubtedly available, cognitively, to speakers of any language, but they are among a small number of schemata which shape the ways in which speakers of English most naturally communicate about spatial relations. The spatial schemata employed in the semantic structure of the system of function words can vary strikingly from language to language, as is brilliantly demonstrated in papers by Melissa Bowerman of the Max Planck Institute for Psycholinguistics in Nijmegen on spatial location expressions in English, German, Dutch, and Korean.<sup>1</sup>)

Speakers of English who know the word *purgatory* are aware of a conceptual structure of considerable complexity, connecting many parts of one variety of Christian doctrine. Such a structure includes notions of sin and retribution, heaven and hell, grace and salvation, and a whole host of others, all of these intimately connected, conceptually, with the notion of *purgatory*.

Knowledge of the psychoanalytic concepts of *id*, *ego*, and *superego* presupposes an awareness of Sigmund Freud's theory of primitive psychic energies and the manner of their control and modification in the maturing individual. None of these terms can be understood without understanding the concepts linking them together. It would obviously make no sense to define each of these terms separately, or without at least indirect reference to the huge complex of ideas developed by Freud.

---

<sup>1</sup>Bowerman (1989, to appear)

The background conceptual structure behind the meaning of a word (in a given sense) or a group of semantically related words is something we can refer to as a *frame*. It should be clear that the full description of the meanings of lexical items – most lexical items, I believe – will have two aspects, a "frame-external" aspect, providing information about the frame, and a "frame-internal" aspect, which specifies the categorizing, identifying, or describing role which the word has within its frame.

The effort to discover and characterize the frame-external features of a word's meaning can be seen as a kind of ethnography. What needs to be discovered is the system of beliefs, experiences, or ready-made conceptualizations on the part of the speakers of the language, which are the necessary underpinnings of the ways they speak and the ways they "think for speaking" (Slobin 1991). A theory of word meaning that sees the need to include analyses of frame structures in an account of the organization of a lexicon can be spoken of as a variety of *frame semantics*.<sup>2</sup>

Since the frame-external information is often unbounded, the question of the desirability of bringing such information into lexical descriptions is sometimes confused with a putative problem of distinguishing *dictionary* and *encyclopedia*, or the equally confusing (but I think distinct) issue of distinguishing *language* from "the world".

Sometimes dictionaries give their readers no access to information about the relevant frames. A definition one finds in the Collins English Dictionary – with no subject field tag – reads as follows: "a regular stream of vortices shed from a body placed in a fluid stream". (The term being defined is **Karmen vortex street**.) An innocent reader confronting this definition will wonder how fluid streams differ from ordinary streams, what it means for a body to shed a vortex, etc., and this reader will not get quick help by looking up the words *fluid*, *stream*, *body*, *vortex*, or *shed*.

(Since nobody but my long-gone high school English teacher would encourage people to learn new words by finding them in a dictionary, this observation is not intended as a criticism of the writer of that entry! Presumably a person who is reading the kind of text that contains the phrase **Karmen vortex street** can be counted on to have the necessary background.)

Sometimes lexicographers direct attention to the frame by means of domain labels such as *Theology*, *Navigation*, or *Psychoanalysis*. The following is from Collins English Dictionary:

**ego 2. Psychoanal.** the conscious mind, based on perception of the environment from birth onwards: responsible for modifying the

---

<sup>2</sup> On frame semantics see Fillmore 1982, 1985, 1989, Fillmore & Atkins 1992, 1994.

antisocial instincts of the id and itself modified by the conscience (superego).

The label *Psychoanal.* is a reference to the larger conceptual framework, and the definition itself shows something of the interconnections with two other basic concepts within that framework. A dictionary reader who finds the definition inadequate knows what to do to acquire the missing background.

And sometimes lexicographers attempt to combine frame-internal and frame-external information in single defining statements. Again, from Collins.

**reincarnation** n. 1. the belief that on the death of the body the soul transmigrates to or is born again in another body

Here the definer wishes to communicate the idea that this account of what happens to the soul upon the death of the body is a part of a belief system. In a dictionary prepared for people whose religious beliefs accept reincarnation, the word would not be defined as a belief. The definition is odd, in any case, since while it makes sense to say that so-and-so believes in reincarnation, it does not make sense to say that so-and-so believes in a belief. Collins offers a separate definition just for believers:

**reincarnation** n. ...2. the incarnation or embodiment of a soul in a new body after it has left the old one at physical death

In this paper I wish to support a kind of lexicographic research and practice which clearly separates frame-external and frame-internal information, while including both within the lexicographer's assignment. My position is superficially similar to what Bo Svensén suggests in his recently translated book *Practical Lexicography*, emphasizing the need to distinguish the two kinds of information while recognizing the need, at times, to coordinate them. (p. 163)<sup>3</sup>

My view on the need to link language and the world is perhaps closest to that of Keith Allan, who writes:

If we are to say anything worthwhile about their meanings, the contents of the senses of certain words must draw on background information about the entities spoken or written of: this

---

<sup>3</sup> Svensén's discussion is built around the following quotation from H. E. Wiegand: Die Lexicographie sollte von einem Sprachbegriff ausgehen, der zwischen Sprache und Welt deutlich unterscheidet, Sprache und Welt aber nicht strikt trennt. Zwar ist Schreiben über Wörter vornehmlich sprachbezüglich: aber ohne ausgewählte sachbezügliche Information kann die Bedeutung vieler Wörter nicht angemessen erläutert werden.

information may be based on any or all of experience, convention, custom, myth, and language use. ... Influence is exerted from a host of imagistic, associative and formal as well as pragmatic factors that coalesce and mutually reinforce one another. ... If semantics is to go beyond translating symbols into yet more esoteric symbols, it must begin to reflect the richness of human experience that is intrinsic to language understanding: that is, semantics should start to represent what intelligent reflective layfolk understand by "meaning" in language. To accomplish such a goal, semantic representations need to be correlated with human experience as it is ordinarily expressed in natural language.

Allan (1992, pp. 371.372)

### The Data for Semantic Inquiry

The primary data for all synchronic linguistic inquiry in general consist of the noises people make and the interpretations people give to those noises.<sup>4</sup> To begin at the beginning, we can take it as uncontroversial that any complete theory of language must in principle be capable of explaining the relationship between the utterances of a language and the interpretations given to such utterances by the language's speakers.

The variability and elusiveness of the primary data, and the necessarily multi-layered character of the resulting explanations, make this task extremely difficult. The explanations for the mapping between utterance and interpretation are multi-layered because there are many co-existing influences on the production and interpretation of speech. Differences relating to the interpreters' experience and attentiveness, or the perceived relevance of contextual features, can cause interpretations to vary from individual to individual and from occasion to occasion. And the actual components of interpretations are often extremely difficult to pin down, since they include much that is imagistic, associational, emotional, nuance-rich, narrowly context-bound, etc.

A full account of the ability to interpret language utterances, therefore, involves many disciplines, as well as appeals to experiences and memories for which no "disciplines" exist. Part of the job of the *language specialist* is to abstract out of this (apparent) chaos just those parts that have to do with the conventions of language, as opposed to other things; and the particular job of the *lexical specialist* is to identify those aspects that have to do with the

---

<sup>4</sup>The reader should make necessary adjustments for non-acoustic input. Reference to the interpretation end of linguistic processing is for methodological reasons. There is also, of course, the relation between the speaker's intentions and the linguistic product of those intentions; but reliable observations at this end are not easy to come by.

knowledge that native speakers have of the lexical primes of the language – elements of linguistic form that have to be learned one at a time, as opposed to those that are understood as the result of the interaction of various elements. The job of identifying information that is specifically linguistic, and specifically lexical, is not always easy.

At some level all workers concerned with the semantic aspects of language can be thought of as having access to a *corpus of interpreted language samples*, the body of primary data with reference to which their research problems are defined and their descriptions are formulated.

I intend this notion of corpus quite broadly, covering a wide range of realities. For instance, the set of language samples could be a very large carefully selected digitally accessed collection of texts created for very special purposes, housed in Philadelphia, Birmingham, Oxford, Copenhagen, Gothenburg, Pisa, or Paris; it might be files of citation slips in the publishing house of some dictionary publishers; or it might simply be an implicit collection of linguistic forms that the researchers as native speakers would intuitively accept as belonging to their language.

The interpretations might simply consist of the analysts' native-speaker understanding of the texts in question, implicit and inarticulate for the most part but capable of being teased out during the inquiry phase. Or, in the case of a language currently being investigated, the interpretations could be something that investigators are in the process of discovering.<sup>5</sup> Finally they could conceivably stand as explicit meaning representation of some sort, possibly in the form of a translation into some secondary symbolic system, a formal or natural language.

A fundamental consideration for any semantic system is the idea of the multi-layered mapping, in such corpora, between the text – the linguistic form – and the interpretation. In the path between what people say and the interpretation we put upon what they say we will find:

- (i) references to experience with the world: its shapes and colors, actors and props, institutions and practices;
- (ii) strategies of common sense reasoning, including reasoning about world knowledge, the nature of communication, the ongoing discourse, etc.;
- (iii) knowledge of the semantic import of grammatical constructions;

---

<sup>5</sup>I don't intend this reference to the ongoing discovery process to be thought of as limited to "jungle linguistics" - the study of languages that have not been hitherto studied. Linguists find surprises daily in the lexical semantics of well-studied languages.

and, adding the part that we are especially interested in here,

(iv) knowledge about the lexical primes of the language: morphemes, conventionalized complex words, idioms, etc.

The main burden of my paper is to urge consideration of a proposal for coordinating information of types (i) and (iv) in this list.

### Three Relevant Professions

As a way of sharpening our understanding of these issues, let us consider the work and interests of three classes of people professionally concerned in one way or another with the lexical aspects of this mapping: (1) lexicographers, (2) lexical semanticists, and (3) computational linguists specializing in Natural Language Processing (NLP).

The lexicographers I have in mind are concerned with the utilitarian, rather than (say) the archival, functions of dictionaries. One of their goals is a product whose users feel that they know what kinds of information can be found in it, and who are rewarded, reasonably often, by finding in the dictionary information that helps them understand the meanings and uses of unfamiliar words that they encounter.

The lexical semanticists I include in this comparison group are those who see their job as that of discovering and organizing everything that speakers know about how the words in their language contribute to their "envisionment" of the texts containing them.

And the NLP researchers I have included in this assembly are interested in designing formal systems capable of drawing the same inferences from linguistic texts that human interpreters do, and are in the process of constructing lexicons which provide a major part of the information needed.

How do workers in these three areas differently conceive of the mapping from text to interpretation, and what role do they assign to themselves in the work of understanding that process? In particular, what is their view of the specific role that information about lexical primes has in the process, what is their view of the role that "world knowledge" plays in the process, and how are these two contributions coordinated?

Of these groups, the NLP researchers clearly cannot avoid reference to human experiences and interests. They may be able to circumscribe the kinds of knowledge they are required to import into their system by limiting the domain of the discourse they examine – to newspaper accounts of traffic

accidents, for example – but they cannot simply decide to ignore so-called "encyclopedic" knowledge. On ideological grounds they might make certain theoretical assumptions about a strict separation of linguistic from non-linguistic knowledge, but the NLP assignment itself does not require this.

Academic linguists run the full range from those who see no boundary whatever between lexical knowledge and world knowledge (Lakoff? Langacker?) to those for whom the main project of lexical inquiry is to isolate precisely that information which is associated with linguistic forms in total independence of facts about the things or experiences or phenomena that these linguistic forms have to do with (Cruse? Leech?). The purist's view is that linguistic knowledge is autonomous, belonging to an independent "language faculty", and that the linguist's job is to describe and characterize only those abilities or that knowledge that comes with being a speaker of a language, independent of whatever other abilities and knowledge one might have by being human, by being a member of a specific culture, or by being an observant participant in an ongoing interaction. The fact that the process of learning a language, or the situation of knowing a language, is synchronous with having many other independent abilities, or knowing many other independent facts, both motivates such a concern for purity and makes its achievement difficult.

In contrast to NLP workers and academic linguists, lexicographers, it seems, can go either way. There appear to be three main areas of decision about "encyclopedic" information in lexicographic work. One has to do with decisions about the admission of proper names: letting the dictionary provide minibiographies of famous figures in history and/or mythology, or combining the dictionary proper with a gazetteer. A second has to do with decisions about the extent to which a particular dictionary should contain terminology from science and technology. The third concerns the provision of scientific descriptions of natural kinds and natural phenomena (**zebra, gold, jade, water, etc., high tide, lunar eclipse, etc.**), or equivalences for the terminology of weights and measures (**foot, pound, degree Fahrenheit, etc.**). The issues I am raising in this paper include, but go beyond such questions.

### **A Simple Example**

A simple example will help us to focus on the kinds of problems that I see in connection with the separation of linguistic from non-linguistic knowledge and the lexical specialist's obligations. Suppose that our corpus contains the clause given in (1)

- (1) She came to a red light

and suppose we ask about the nature of the mapping between that sentence and its everyday interpretation. Somebody who does not know what is going

on at this point in a narrative clearly needs to know that in the signalling systems that cities set up for controlling automotive and pedestrian traffic-flow at important intersections or pedestrian crossings, green lights are used to indicate that those facing it are free to proceed, and red lights are used to get the traffic approaching it to stop, and then to wait until the green light goes on. The interpretation includes a social reality that the protagonist of sentence (1) at this point in the narrative is facing an obligation to stop her vehicle.

Our first question, then, is whether a dictionary (or "lexicon") created according to the needs of each of our three professions can be expected to contain information that could lead a user, human or machine, to the full interpretation of the clause, and whether such information should be introduced in association with the word **red**, or the phrase **red light**.

Since a computer does not have the kinds of experiences that the rest of us have had, the NLP team would definitely want to build into their systems an ability to derive such information, and it would of course be necessary to start from the linguistic form, in particular the phrase **red light**, perhaps reinforced by information about the larger phrase **come to a red light**. If the sentence continues in the manner of (2)

(2) She came to a red light, but she kept on driving

the system's inference engine should be expected to generate a number of tentative partially specified expectations and inferences for this portion of the narrative, perhaps assumptions about the driver's temporary inattentiveness, or the urgency of her errand.

Most linguistic purists, on the other hand, are likely to feel that the semiotic function of the colors red and green in systems of traffic lights worldwide are facts about the world, not facts about the *meanings* of the *English words red, green or light*. It just happens that communities throughout the world use red and green lights at intersections to serve certain traffic control purposes, and knowing and using facts like these is clearly distinct from knowing the lexicon of English. The linguist's responsibilities stop short of the full interpretation.

Lexicographers, given their more practical goals, might have reasons for deciding either way.

Before consulting any dictionaries about this question, the guess I made was that they would indeed have entries for **red light** and **green light**, but only in order to provide the needed link to the metaphorical meanings of these phrases. I formed this opinion by reasoning about the occurrence, in many dictionaries, of the definition "one that reads" or "a person who reads



or is reading" as the primary sense of the noun **reader**. The reason seemed to be that since the word has other senses, it would be wrong not to include its basic sense; or perhaps the reason was so that the extended senses of the word could be seen as specializations of the primary sense. Given the productivity of the agentive suffix and the familiarity of the simple verb **read**, a dictionary that in principle offered definitions of such words would be wasteful.

What I found in the dictionaries I examined was (1) that those that gave traffic-light definitions of the phrasal entries **green light** and **red light** did not show any explicit connection between these literal senses and the metaphorical senses based on them, and (2) that the dictionaries that did motivate the metaphorical sense explicitly referred to information about traffic lights, not information about the primary "meanings" of the phrases. (Excerpts from some of these the entries are given below.)

Let us turn to one of these metaphorical uses. Suppose we find in our corpus a sentence like (3)

(3) Our project was given the green light.

The interpretation, of course, is that our project was approved, was given permission to proceed.

The lexical specialist needs to decide whether the phrase with **green light** is a conventional way of expressing what it expresses – and therefore deserves a place in a description of the language – or whether with this sentence the writer is merely appealing to the reader to make use of a common experience to get at the intended figurative meaning. The linguistic purist might say (at first) that this is merely a normal instance of figurative language, a variety of metaphor. People who know the relevant facts – the argument would go – can come up with the right interpretation: they don't need the semanticist to do that work for them. Again, the work of the linguist as linguist ends early in the path from form to interpretation.

NLP researchers who need a short-term solution to this problem might simply identify the phrase as meaning 'to give approval'; those who want their systems to recognize, or to be able to work out, the metaphoric interpretation might want to do otherwise.

All of the dictionaries I consulted did in fact list a special sense for **green light**, in its figurative use. A number of good arguments can be given for taking the position that this expression belongs in a dictionary. One is the wide familiarity of the phrase with this interpretation, suggesting that it is a conventional way of saying what it says; another is the collocation with **give**, **have** or **get**, and the definite article, and somewhat more distantly, with the word **project**. (Both in the dictionaries and in my explorations with native

speakers, the kind of approval that is spoken of with metaphorical **green light** is approval for "a project".)

Linguists, I think, would in the end agree with this decision. But once we decide that **give someone the green light** in its 'approval' interpretation belongs in a dictionary, we face some more decisions.

First, should the phrase merely be defined as involving approval *tout court*, or should information about the *motivating context* be provided?

Second, if we make the latter move, should that information be seen as indicating a relation between *senses* of the phrase? (That is, is it a part of the *polysemy structure* of the phrase **green light**?) Would such information be a true part of lexical semantics, or is it to be thought of along the lines of interesting stories about word?<sup>6</sup>

### Traffic Lights and Dictionaries

All five of the dictionaries I examined had separate entries (or sub-entries) for the phrases **green light** and **red light**, if only for the metaphoric uses.

In describing the various practices with these phrases, let us speak of (i) the colored lights in such a signalling system and (ii) their signalling functions ('proceed' vs. 'stop') as *form* and *function*, respectively. We can then say that only the American Heritage Dictionary clearly identified each of these phrases with both the form *and* the function of the colored traffic lights. Notice the first senses in the following entries.

#### AHD3

**green light** *n* 1. The green-colored light that signals traffic to proceed. 2. *Informal*. Permission to proceed.

**red light** *n* 1. The red-colored light that signals traffic to stop. 2. *Informal*. A command to stop.

The categories are specified both in terms of the physical characteristics of their members (green light, red light) and the signalling functions they serve.

The Collins English Dictionary seems to take the function as primary, but adds information about the color of the lights after "esp." Webster's

---

<sup>6</sup>This is similar to Svensén's comments on the parenthetical comment in a definition of Leporello: "a long strip of paper folded concertina-wise (after the long catalogue of amours recited by Don Giovanni's servant Leporello in Mozart's opera)" (Svensén 1993, p. 165)

Ninth Collegiate Dictionary gives the same treatment to **red light**. Notice the first senses in each of the following:

Collins

**green light** *n* 1. a signal to go, esp. a green traffic light. 2. permission to proceed with a project.

**red light** *n* 1. a signal to stop, esp. a red traffic signal in a system of traffic lights. 2. a danger signal. 3. an instruction to stop or discontinue.

W9

**red light** *n* (1849) 1: a warning signal *esp*: a red traffic signal 2. a cautionary sign: DETERRENT

(I have never learned whether the "esp." in a dictionary entry is intended to express a statistical generalization or to point to a semantic prototype.)

For each of these phrases, the Concise Oxford Dictionary mentioned only the function. To judge from the first senses in the entries excerpted below, an arm-waving traffic officer could presumably be spoken of as giving green and red lights merely by pointing.

COD8

**green light** 1 a signal to proceed on a road, railway, etc. 2 *colloq.* permission to go ahead with a project.

**red light** 1 a signal to stop on a road, railway, etc. 2 a warning or refusal.

One can imagine that the compilers were counting on the reader to begin with a compositional meaning of the phrase.

For **green light** Webster's 9th Collegiate Dictionary and Webster's New World Dictionary give only the metaphorical use, but each of them *motivates* the expression with "from..." or "after..." followed by reference to both the form and the function of the green traffic light. WNW notices the collocation with **give and get**.

W9

**green light** *n* [fr. the green traffic light which signals permission to proceed] (1937): authority or permission to proceed esp. with a project

WNW 2nd coll ed

**green light** [after the green ("go") signal of a traffic light][Colloq.]  
permission or authorization to proceed with some undertaking;  
usually in **give (or get) the green light**.

One doesn't expect dictionary compilers to take a stand on the nature of metaphoric sense extensions, but one can't help noticing that none of the dictionaries explain metaphoric senses as extensions from non-metaphoric *senses* of the same term. In these two cases, the motivation is from the practice, not the words. In all other cases, the separate senses are simply given in a list.

For **red light** Webster's New World Dictionary gives the functional description, with "a red lamp, flare, etc." after "specif.", but then it gives "a red stoplight" as a second sense.

WNW

**red light** 1. any danger or warning signal; specif., a red lamp, flare, etc. 2. a red stoplight.

On the question of whether (**have/get/give someone**) **the green light** is a proper lexical item one can imagine opposing opinions. At the one extreme there might be those who believe that we are simply dealing with a metaphor that doesn't require explanation in a dictionary; and at the other extreme, it is quite conceivable that there are some speakers of English who use this phrase in its intended meaning without actually thinking about traffic lights. For such people it is a so-called "frozen metaphor", and therefore a part of their linguistic competence.

In giving an account of the metaphorical meaning of **green light**, the necessary frame information becomes quite specific. Since in the traffic light situation, red and green lights alternate, it is clear that somebody who is waiting for the green light is stopped. In the figurative use, then, the people interested in going ahead with the project have been standing still, prevented from going ahead, waiting for the event of "getting the green light" to occur, waiting for the light to turn green.

The situation is clearly different when we consider the phrase **run a red light** or **run through a red light**, a phrase that describes what "she" of sentence (2) is described as having done. Since these expressions have no straightforward "compositional" interpretation, they must be idioms, and hence they deserve a place in our dictionary.<sup>7</sup> In this case, an interpretation that did not include reference to the actual traffic light situation would clearly be mistaken, since such expressions are intended "literally" in the sense that

---

<sup>7</sup> Although I have just claimed that this idiom deserves a place in "our dictionary", I must admit that I haven't found it in anybody else's.

the image the user needs is of a vehicle violating a very specific rule of traffic. An explanation of the meaning here has to communicate an understanding of the nature and workings of electric traffic lights.

One other word that seems to belong to our story is **amber**. I suspect that there are speakers of English for whom the word **amber** is associated mainly with its use in talking about traffic lights, and I also suspect that if it weren't for this ready-made linguistic association, there would be no particular reason to use the word **amber** in this context rather than **yellow**. It should be noted that while the American dictionaries fail to identify **amber** as connected with traffic lights, Collins and COD8 both give form-and-function definitions, not for a phrase **amber light** (in parallel with **red light** and **green light**) but for **amber** as a noun.

#### Collins

**amber** *n.* 4. an amber traffic light used as a warning between red and green.

#### COD8

**amber** 2 a yellow traffic-light meaning caution, showing between red for 'stop' and green for 'go'.

Now since it is *necessary* to say something about "the world" in the entry for **run a red light** (and, according to Collins and COD8, in that for **amber**), and it is *useful* to give that same information in the explanation of **(have/get/give someone) the green light** – that is, since there are reasons in some parts of the lexicon to refer to the institutional and artefactual background that motivates the existence of these terms – a reasonable argument could be made for linking all relevant dictionary entries with information about this system. That list of items requiring this connection would include a number of semantically quite complex terms, such as **protected left turn**, for English, and **grüne Welle** for German.

Since it would be wasteful to include a full account of traffic signal systems in the definitions of each of the words in this set, we need a way of making such information available, without requiring it to be presented as part of the entry for each of the relevant words.

### Ethnographic Semantics

There is an important reason why we have had the luxury of worrying about whether information about traffic signal systems does or not have a place in the design of a dictionary of English, and that is that we already know every relevant fact about this essentially universal semiotic system. But a lexicographer from Mars building a dictionary of English would be considered irresponsible not to include the necessary cultural background. An English-

Martian dictionary would have to incorporate, or be attached to, an ethnography that described the lifeways of English-speaking people and identified the ways in which members of the culture linguistically dealt with aspects of those lifeways.

We often call on fictitious Martians to help us exoticize that which is close and familiar to us, but we find such efforts less compelling now that we know there is no likelihood of articulate beings living on that planet. So a genuine exotic context might be more useful in making my point. In a study of Trobriand Islanders' terms for body and mind, Gunter Senft quotes Malinowski's discussion of Trobriand notions of "mind" and "memory".

The mind, *nanola*, by which term intelligence, power of discrimination, capacity for learning magical formulae and all forms of non-manual skill are described, as well as moral qualities, resides somewhere in the larynx. The natives will always point to the organs of speech, where the *nanola* resides. ... The memory, however, the store of formulae and traditions learned by heart, resides deeper, in the belly. A man will be said to have a good *nanola* when he can acquire many formulae, but though they enter through the larynx, naturally, as he learns them, repeating word for word, he has to stow them away in a bigger and more commodious receptacle; they sink down right to the bottom of his abdomen.

(Malinowski 1922 408f, Senft 1993 pp. 1-2.)

It seems obvious that no ethnographic semanticist preparing a dictionary of the language of Trobriand Islanders would find it satisfying to give a minimal "definition" of *nanola* as, say, 'mind' arguing that the facts about its specializations (including the memorization of magical formulae) and its location (in the larynx) belong in an encyclopedia of Trobriand culture, an encyclopedia that is in no way connected with the dictionary. We can't really understand the word, I would claim, if we didn't understand the accompanying beliefs.

The piece of "ethnography" connected with a dictionary that gives clear understandings of the use of the language connected with traffic signals would have to describe the physical, institutional, and legal concepts that make up the form and function of this institution. One can imagine a combined ethnography and dictionary which provided this information for the Martian; one can imagine an electronic resource which linked dictionary entries with encyclopedia entries; and one can imagine an efficient print dictionary that included key words expecting readers to consult their own knowledge of the domain. It would be wasteful, of course, for all of the details of the frame to be included in each relevant entry, but at some level or other, the world knowledge about the system has to be understood as

conceptually a part of the information that ought to be available through a dictionary.

### Frame Discrepancies

In ordinary dictionaries, reference to facts about traffic signals can be kept to a minimum for defining the terms we have just looked at. In the same way, reference to the details of the commercial transaction scenario and the properties of a money economy need only the slightest allusion in the definition of words that index such frames – **buy, sell, pay, spend, charge**, etc. – and the workings of gravitational force, through which we understand verticality, does not need to be described in definitions of **up and down, ascend and descend, raise and lower, top and bottom, high and low**, etc. – because all of the dictionary users that we can imagine have mastered the details of such frames.

Certain traditions of dictionary-writing are problematic precisely because not all speakers of the language share the same interpretive frames, even in areas which are not thought of as terminological. For example, some of us do not have a religion, and those who do, do not all have the same religion. The possibility of frame conflict between compiler and user can be illustrated clearly with religious terminology.

If believing monotheists read a definition of **God** as "the chief object of worship in many religions", they would be right in complaining that that's not what the word means. On the other hand, if atheists read a definition of **God** as "the Supreme Being who created and maintains the universe", they could complain that the producers of the dictionary are using language that presupposes something that they find objectionable. A frame-external description cannot satisfy someone who takes the frame for granted; a frame-internal definition cannot satisfy someone who rejects the frame.

Sometimes dictionaries try to have it both ways. I mentioned earlier that the Collins English Dictionary gives two senses of **reincarnation** (repeated here), one containing the phrase "the belief that" and the other not – one definition is for the outsider, one for the insider.

**reincarnation** n. 1. the belief that on the death of the body the soul transmigrates to or is born again in another body 2. the incarnation or embodiment of a soul in a new body after it has left the old one at physical death

The second of these definitions is the one that defines the word for people whose belief-world includes the process described. But as I suggested earlier, there is something wrong with the first definition, namely, that no linguistic

separation is made between the content of the belief and the fact that it is a (i.e., somebody's) belief.

With religious terms, dictionaries sometimes provide indirect access to the needed background information with domain labels such as *Hinduism*, *Theology*, or *Christianity*. But such practices are not consistently maintained.

Through the domain label *Christianity*, Collins presents **venial sin** and **mortal sin** as concepts within Christian belief systems. Of course one might object that the frame-external information is not sufficiently detailed: not all Christian doctrines include these notions.

Collins

**venial sin** n *Christianity*: a sin involving only a partial loss of grace. Compare mortal sin.

**mortal sin** n *Christianity*: a sin regarded as involving total loss of grace. Compare venial sin.

The American Heritage Dictionary assigns **venial sin** to the Roman Catholic Church, and in its definition **mortal sin** gives useful examples of the category and is clear about consequences.

AHD3

**mortal sin** n. *Theology*. A sin, such as first-degree murder or perjury, that is so heinous it deprives the soul of sanctifying grace and causes damnation.

**venial sin** n. *Roman Catholic Church*. An offense that is judged to be minor or committed without deliberate intent and thus does not estrange the soul from the grace of God.

All of these definitions refer to grace, which is also defined as a Christianity-internal notion.

Collins

**grace** n 8. *Christianity*: a. the free and unmerited favour of God shown towards man b. the divine assistance and power given to man in spiritual rebirth and sanctification

**Original sin**, where all of these problems got started, on the other hand, is introduced in Collins without definition-external reference to a particular belief system, but is ascribed to Christianity in AHD..

Collins



**original sin** n a state of sin held to be innate in mankind as descendants of Adam

To judge from the language of the Collins definition, this is just the way things are. The hedge "held to be..." in this definition invites the inference that the definers have no doubts about the existence of this universal state of sin, but they do allow as matters of controversy its innateness and its origin in a decision made by our ultimate ancestors. (This hedge, of course, gives a wrong understanding of the concept itself.)

#### AHD3

**original sin** n. According to Christian theology, the condition of sin that marks all human beings as a result of Adam's first act of disobedience.

The American Heritage Dictionary here chose to give the frame-localizing information in the defining phrase rather than as a subject tag.

The connections between **sin** and **grace** on the one hand and **hell** and **heaven** on the other hand are not made explicit by the Collins lexicographers.

#### Collins

**hell** n 1. *Christianity*: (sometimes *cap.*) a. the place or state of eternal punishment of the wicked after death, with Satan as its ruler. b. forces of evil regarded as residing there.

**heaven** n 1. (sometimes *cap.*) *Christianity*: a. the abode of God and the angels. b. the place or state of communion with God after death

The American Heritage Dictionary sees **hell** as belong to "many religions" but has no frame-external marking on **heaven**.

#### AHD3

**heaven** n. Often Heaven. a. The abode of God, the angels, and the souls of those who are granted salvation. b. An eternal state of communion with God; everlasting bliss.

**hell** n. 1. a Often Hell. The abode of condemned souls and devils in some religions; the place of eternal punishment for the wicked after death, presided over by Satan.

With the words **heaven** and **hell** we become aware of some of the lexicographer's difficulties with this family of terms. At some level we would like a dictionary informed by a theory of frame semantics to show how the concepts and categories it introduces are related to each other, so that, for

example, notions like grace and salvation and heaven, sin and damnation and hell, would all be connected; but since **heaven** and **hell** are concepts found in many religions, this would require an assumption that the words have separate meanings in each of those religions.

Alternatives to heaven and hell are purgatory and limbo. Both Collins and AHD attribute **purgatory**, to Roman Catholic beliefs, while **limbo** is taken to be a more general notion.

#### Collins

**purgatory** n. 1. *Chiefly R. C. Church.* a state or place in which the souls of those who have died in a state of grace are believed to undergo a limited amount of suffering to expiate their venial sins and become purified of the remaining effects of mortal sins.

**limbo** n. 1. *(often cap.) Christianity.* the supposed abode of infants dying without baptism and the just who died before Christ.

#### AHD3

**purgatory** n. 1. *Roman Catholic Church.* A state in which the souls of those who have died in grace must expiate their sins.

**limbo** n. 1. Often Limbo. *Theology.* The abode of just or innocent souls excluded from the beatific vision but not condemned to further punishment.

The Collins writers felt it necessary to include the hedges "are believed to" and "supposed" even though the belief-context was provided with the subject labels.

The word **God** is assigned to *Theology*, in Collins; it has no frame tag in the American Heritage Dictionary, but the belief context is shown with the phrase "conceived as." (There must have been some interesting in-house discussions at Collins leading to the use of upper-case initials – Supreme, Being, Creator – in their definition.)

#### Collins

**God** n l. *Theol.* the sole Supreme Being, eternal, spiritual and transcendent, who is the Creator and ruler of all and is infinite in all attributes; the object of worship in monotheistic religions.

#### AHD3

**god** n. 1. **God.** a. A being conceived as the perfect, omnipotent, omniscient originator and ruler of the universe, the principal object of faith and worship in monotheistic religions. b. The force, effect, or a manifestation of this being.

The Collins lexicographers present **Satan** without any qualifications: no hedging inside the defining statement and no domain-label covering the whole thing. The definition just tells you who he is. The American Heritage Dictionary assigns the concept to theology.

#### Collins

**Satan** n 1. the devil, adversary of God, and tempter of mankind; sometimes identified with Lucifer (Luke 4:5-8).

#### AHD3

**Satan** n. *Theology*. The profoundly evil adversary of God and humanity, often identified with the leader of the fallen angels, the Devil.

We have seen cases where the external-frame information is indicated with a domain label (e.g., *Christianity*), and we have seen cases where it is alluded to by a hedge inside the defining phrase (e.g., "held to be"...). There are also cases of definitions which need such external reference but which lack them, making them essentially uninterpretable, similar to what we saw with **Karmen vortex street** earlier. The Chambers Dictionary definition of **reincarnate** is an example:

#### Chambers

**reincarnate** *v.t.* to cause to be born again in another body or form: to embody again in flesh.

I am sure that if I did not have some independent notion of reincarnation, for which I could make reference to beliefs about a "soul" (a term requiring its own external framework) that originally inhabited one body leaving that body at death, I could not have imagined the conditions under which something can get "embodied again in flesh", and I don't think that looking up the words **embody** or **flesh** in that same dictionary would have been able to help me.

Before leaving the spiritual domain, I permit myself to observe that even in an area in which real-world facts are hard to come by, Collins has chosen to add, in the **angel** entry, some useful encyclopedic information about angelic social stratification.

#### Collins

**angel** n. 1. *Theol.* one of a class of spiritual beings attendant upon God. In medieval angelology they are divided by rank into nine orders: seraphim, cherubim, thrones, dominations (or dominions), virtues, powers, principalities (or principedoms), archangels, and angels.

I have no interest in criticizing dictionaries for their treatment of religious terms. My fascination with this terminology has been mainly because this is an area in which it is important to keep track of the difference between what a word means and the fact that the word is a part of a large and complex package of beliefs. This task of maintaining this separation is difficult because there is no convenient mechanism for doing this. If labels like *Theology* and *Hinduism* were consistently used, and definition-internal hedges about beliefs were avoided, the problem could be partly solved; but such labels generally represent categories that are too broad for the meaning to be properly anchored in its own proper belief system.

### A Frame-Informed Dictionary

I believe that a dictionary should make it easy for the reader to know what background frames motivate the category a given word represents. In the case of scientific and technical vocabulary, this may not seem like a problem, since the people who use terminological dictionaries presumably are already trained in the basics of the relevant discipline. In the case of the most general vocabulary, this is not seen as a problem, since everybody who uses the dictionary already has access to the relevant frames.

However, we are not only interested in practical dictionaries and everyday users. If we return to the interests of the NLP researcher, we can remind ourselves once again that a computer needs to be provided with the frames that the rest of us already possess, and so lexical information that anchored a text in a conceptual structure that allowed precise inferences would have to be regarded as useful in systems seeking to achieve some level of automatic language understanding. The concept of "frame" has long played an important role in NLP research.

But more than that, the frames that underlie word meanings should become the basis for the recognition of semantic relations among words, and among word senses. The concept of antonym, for example, covers a very broad range of relations, and their nature can be clarified if the semantic frames the words are situated in are made clear. The frames that underlie word meanings should become the basis for recognizing semantic differences across languages, especially since these sometimes permit generalizations across frame characteristics.

We can imagine an electronic lexical resource which links word definitions with information about frames (in some ways this can be thought of as linking entries in a dictionary with entries in a very special kind of encyclopedia), and we can imagine lexicography projects that are devoted to establishing these links by discovering the nature of the frames. Such projects, to the extent that they try to uncover the semantic frames underlying

the general vocabulary, are not frivolously engaged in designing a data-base for some eventual English-Martian or English-Flatlandic dictionary, but are laying the groundwork for understanding the ways in which the words in our language are connected with each other, the ways in which semantic near-equivalences can differ from each other across languages, and the ways in which the vocabulary of a language is an index of the culture of its speakers.

(In my oral presentation I hope to be able to present a brief description of one aspect of the DELIS project, devoted to the description and exploitation of the semantic frames underlying the vocabulary of sensation and perception.)

### References

- Bowerman, Melissa (1989), "Learning a semantic system: what role do cognitive predispositions play?", in H. K. Rice & R. L. Shieffelbusch (eds.), *The Teachability of Language*. Baltimore: Paul H. Brooks.
- Bowerman, Melissa (to appear), "The origins of children's spatial semantic categories: cognitive versus linguistic determinants", in J. J. Gumperz & S. C. Levinson (eds.), *Rethinking Linguistic Relativity*. Cambridge: Cambridge University Press.
- Fillmore, Charles J. (1982), "Frame semantics," in The Linguistic Society of Korea, *Linguistics in the Morning Calm*, Seoul: Hanshin, pp. 111-1137
- Fillmore, Charles J. (1985), "Frames and the semantics of understanding," in *Quaderni di Semantica* 6.2.222-254.
- Fillmore, Charles J. (1992), "'Corpus linguistics' vs. 'Computer-aided armchair linguistics'", *Directions in Corpus Linguistics* (Proceedings from a 1991 Nobel Symposium on Corpus Linguistics, Stockholm), Berlin: Mouton de Gruyter, pp. 35-60
- Fillmore, Charles J. and Beryl T. Atkins (1994), "Starting where the dictionaries stop: the challenge for computational lexicography" in B. T. Atkins & A. Zampolli, eds., *Computational Approach to the Lexicon*, Oxford University Press
- Fillmore, Charles J. and Beryl T. Atkins (1992), "Towards a frame-based lexicon: the case of RISK", in Adrienne Lehrer and Eva Kittay (eds.), *Frames and Fields*, Erlbaum Publishers, pp. 75-102
- Malinowski, Bronislaw (1922), *Argonauts of the Western Pacific*, London: Routledge and Kegan Paul

Senft, Gunter (1993), *Body and Mind in the Trobriand Islands*, Working Paper No. 17, Cognitive Anthropology Research group at the Max Planck Institute for Psycholinguistics.

Slobin, Dan I. (1991), "Learning to think for speaking: native language, cognition, and rhetorical style". *Pragmatics* 1.7-25.

Svensén, Bo (1993), *Practical Lexicography: Principles and Methods of Dictionary Making*. Oxford: Oxford University Press.