The Responsibilities of the Compleat Lexicographer
(Abstract of Invited Talk)
Charles Fillmore
University of California - Berkeley

One can imagine three sorts of expectations of a maximally efficient lexical resource, i.e., the dictionary of the future.

- Situation 1: I find an unfamiliar word in a text and wish to learn how to interpret the part of the text containing that word. The dictionary I use should tell me what meanings the word has, and should help me decide which of them are compatible with the given context and how each of those meanings can integrate with the meaning of the rest of the context expression. If the word has certain conventional associations that the author is counting on readers to pick up on, I'd expect to be told what they are.

- Situation 2: I come upon a word and wish to make it my own. For this I need to learn what senses it has and what grammatical constructions it participates in for each of these senses, and I would like to see a reasonable variety of example sentences illustrating the word's uses. I would need to be informed about whatever pragmatic, regional and sociolinguistic associations exist for this word. But in addition to all of that, I would need to be shown what alternatives there are to this word P near-synonyms or words representing superordinate or subordinate categories P so that I could judge what effects might get created by my choosing this word rather than any of its alternatives.

- Situation 3: I wish to express some idea in the language but I don't know how to choose the word or phrase that best fits my intentions. I use some description of my intentions to find the appropriate words and their descriptions.

A system that meets expectation (1) serves the "decoding" function of a dictionary, a responsibility for lexicography that has never been questioned. A system that meets expectation (2) achieves part of the "encoding" function, a function that is in general not accomplished in ordinary dictionaries, and is rather poorly served by familiar thesauri and synonymica. Expectation (3) seems to me to be impossible, if only because I cannot imagine presenting to dictionary users an explicit and learnable notation for conceptual structures or communicative intentions, which could encompass the entire language and through which the user could search for the needed locutions. (For such purposes it is more reasonable that a language learner enter a dictionary set up for the efficient satisfaction of expectation (2), and use it to explore the ways in which the language in question has shaped its lexicon for the expression of ideas.)

Nobody needs to be told that existing print dictionaries, and their machine-readable versions, fail us in tasks (1) and (2), but I will say a few words about that anyway P but not as a criticism of dictionaries. I will then suggest ways in which an on-line lexical resource could in principle approach the requirements of tasks (1) and (2). A version for human users would allow text input, would parse the text, and would allow the user to connect from any word to the relevant dictionary entry. Each dictionary entry would pro-
vide access to information about the multiple semantic frames or knowledge structures against which the word's category (in each sense) is understood, and would make available the section of the grammar in which explanations are provided of the forms and functions of the grammatical constructions in which the word can participate. Each entry, for each sense will connect to related words or word-groups by way, perhaps, of a vastly enriched version of WordNet. For human users, access to concordance lines from natural corpora illustrating uses and constructions would be invaluable. (As long as we're dreaming, we may as well link proper nouns, and words that name artifacts, natural kinds and social institutions, etc., to appropriate encyclopedia entries.)

My emphasis will not be on the workings of such a system, but on the kinds of information it will need, much of it information that would be needed in a lexicon devised for NLP purposes as well. For the grammar, we need a framework whose structures are close to the surface; for the semantic structure we need much more than feature lists and simple notions of "argument structure"; and the system of lexical relationships will be informed by both "franal" and subtle grammatical information.