HARD-SCIENCE LINGUISTICS AS A FORMALISM TO COMPUTERIZE MODELS OF COMMUNICATIVE BEHAVIOR

BERNARD PAUL SYPNIEWSKI

Rowan University - Camden, Broadway and Cooper Street, Camden, NJ 08102 USA
E-mail: syp@eticomm.net

Hard Science Linguistics (HSL) is a new linguistic theory, first worked out in Yngve [1], developed from the insights into human language gained during the history of machine translation and similar efforts. Unlike most linguistic theories, HSL concerns itself with the details of how people communicate rather than on how sentences are parsed. From the start, HSL developed with an eye toward making its results scientifically valid, not in the sense of other linguistic theories but in the sense of physics or chemistry. Both the historical basis in machine translation for HSL and the attention paid to a scientifically valid formalism make HSL an attractive candidate for the development of large-scale computer models of communicative behavior. In this paper, I will use some "mixed domain" terminology in order to more quickly explain HSL in the space available.

1 Introduction

The Cold War need to translate large volumes of scientific and military publications into English spurred the earliest attempts at developing machine translations systems. Initial attempts concentrated on the grammar of sentences and began with word for word translations. Problems with this and similar approaches appeared in short order Yngve [2]. Most approaches to machine translation have assumed that translation is <u>exclusively</u> a problem of language to be addressed grammatically Barr, Cohen and Feigenbaum [3]. While researchers acknowledged that context was a difficult problem that needed to be solved, context was mostly seen as grammatical context. Though some researchers and philosophers, such as Austin [4] recognized the behavioral elements in the problem of context, they often overlooked the consequences of these elements.

Most linguistic theories tacitly assume that language can be studied by itself, without reference to the societal matrix in which it exists. Linguistics generally treats language understanding as equivalent to the understanding of grammar. Artificial intelligence has adopted this outlook. While there is much to be said for natural language processing systems that understand the construction of sentences, we should not confuse these systems with systems that try to understand how language is used by human beings in their everyday lives. Our systems must understand more than grammar. Most linguistic theories do not provide us with an understanding of anything other than grammar.

Despite the success of Generative Transformational Grammar (GTG), linguistics does not have a sound scientific basis. Linguistic discourse is philosophical discourse with roots in the ancient Aristotelian and Stoic grammatical and logical traditions. Most linguists do not produce scientifically testable results because general linguistics does not provide a scientifically sound formalism; indeed, while many linguists pay lip service to the need to make linguistics scientific, there is no general expectation that linguists will produce scientifically acceptable results.

HSL consciously developed a formalism that could produce scientifically sound linguistic results. One of the more controversial themes of HSL is the de-emphasis of the importance of grammar, what HSL refers to as the "linguistics of language". HSL models "communicative behavior", i.e., language in a social context. Language, for HSL, is purposeful rather than merely grammatical. HSL provides a method for unifying traditional linguistic and extra-linguistic issues in a scientifically acceptable way. In its brief history, HSL has been used to model complex social phenomena such as business negotiations Brezar [5], ethnic stereotypes, Cisło [6], the analysis of textbooks, Coleman [7], Czajka [8], and criminal litigation, Sypniewski [9] as well as more traditional linguistic concerns such historical linguistic change, Mills [10], Malak [11] fillers, and Rieger [12].

2 Some Methods and Tools Provided by Hard-Science Linguistics

HSL provides the researcher with a number of tools to describe the interaction between individuals and their environment. Briefly, some of those tools are:

1. An individual may interact with other individuals by playing a role part in a linkage. A linkage is a theoretical framework in which communicative behavior takes place. A role part is a description of the linguistically relevant behavior that an individual performs in a particular linkage. An individual may play a role part in several different linkages, which may or may not overlap in space or time. For example, an individual may be both a father and a little league coach at the same time. The role parts exist in different linkages that interact while a little league game is in progress.
2. Every linkage has a setting. A setting is a description of the linguistically relevant environment in which a linkage exists. Settings may have props, linguistically relevant objects. For example, the amount of feedback in an auditorium's sound system may affect the communicative behavior of speakers on stage.
3. Linkages have tasks, which, in turn, may have subtasks. Tasks and subtasks are descriptions of linguistically relevant behavior, somewhat analogous to functions in computer programming.
4. Individuals have properties that may be affected by communicative behavior or that may have an effect on the communicative behavior of others. The loudness

of a speaker's voice or the speaker's command of the language of the audience may be reflected in properties of the speaker or listener's role part.

5. HSL uses its own notation (procedures, properties, and other elements) to construct plex structures. Plex structures describe the building blocks of an HSL model.

The researcher models people or groups communicating among themselves along with their relevant setting(s) by creating a linkage, enumerating its participants and describing their role parts, describing the sequence of tasks and subtasks that must take place, describing the setting in which the linkage exists, and the relevant properties of the role parts and setting. HSL insists that all models be based on observable communicative behavior stated so that the results of the model accurately predict behavior in the real world in a reproducible way.

3 Implications for Computer Science

HSL allows the modeler to develop a model of arbitrary complexity. Furthermore, the modeler is not restricted to describing language. HSL is based on the scientific principle that the world has structure. An HSL model of communicative behavior is more complex than any model based on any other linguistic theory. The payback from this complexity is substantial. Communicative behavior becomes more manageable, the findings more justifiable, and the model more reflective of the real world. Since HSL sees the world in terms of properties, structures, and function-like tasks, a thoroughly developed model may be easily ported to an appropriate computer language. A structured model of communicative behavior resembles familiar paradigms in computer science. Linkages may be modeled by interacting classes, with each class representing a task or subtask. It may even be possible to use the Unified Modeling Language to move a model from HSL to the computer. The event-driven programming paradigm may be able to express some of the dynamism inherent in HSL. This is still controversial among HSL workers because of the type of model HSL creates. Professor Yngve believes that it will be difficult to adequately model the parallelism of complex HSL models on a serial computer. Because HSL is in its infancy, this remains an experimental question.

4 Discussion of the Current Attempts to Build SIMPLEX

In the mid-1980s, Victor Yngve, then at the University of Chicago, began to develop a simulator called SIMPLEX for his linguistic theories, later to become HSL. Because of the size and capabilities of contemporary machines, Professor Yngve decided to write SIMPLEX in FORTH. SIMPLEX remained incomplete, partially because the underlying linguistic theory needed further development.

Presently, Professor Yngve, I, and others have resurrected SIMPLEX and intend to develop it beyond its mid-1980s incarnation. We intend to continue using FORTH for three reasons. First, we will be able to use the code already written since the basic FORTH language has retained stable. Second, the American National Standards Institute (ANSI) standardized FORTH after SIMPLEX was written. ANSI FORTH is now file-based, rather than block-based, as were the original FORTHs. Cross-platform development will thus be simpler using ANSI FORTH. Third, and most important, HSL is now a fully developed theory. We now have a goal for SIMPLEX.

SIMPLEX will be both a program and a programming language that will process plex structures. Because HSL models the real world and not just the grammar of sentences, HSL provides a methodology for representing parallel tasks. One of the reasons that FORTH proves useful is that FORTH originated as a computer language to handle multiple synchronous and asynchronous tasks. Accurately representing parallel tasks is one of the biggest challenges for SIMPLEX and one of its biggest potentials.

At the time this paper was written, SIMPLEX is still in its infancy. However, the development of SIMPLEX is substantially advanced even though the computer code is, roughly, where it was in the mid-80s. We now have a complete formalism and methodology to model; this was not the case at the time that the original SIMPLEX was written. We are currently porting code from the old block structure to a file structure. In the process, we are testing various FORTHs on different platforms to identify compatibility problems. Because FORTH is a very efficient language, it is likely that SIMPLEX will run on machines that are significantly less powerful than today's desktop standard. We are testing different FORTHs on different platforms in order to determine the minimal configuration needed for SIMPLEX. There is significant interest in HSL in countries where state of the art computing equipment might not always be available. Some preliminary tests with the original version of SIMPLEX show that SIMPLEX may prove useful on Palm Pilots and similar devices.

Our goal is to create a cross-platform computer program that will accept files of plex structures, analyze them, and simulate them on a desktop computer. A researcher will then be able to see the model in action and modify it whenever necessary. Once we finish porting SIMPLEX from block-oriented to file-oriented FORTH, we will begin developing sections of the simulator that will process specific HSL structures. SIMPLEX, when fully developed, will become a major tool for HSL researchers who wish to verify the plex structures and findings that they have developed.

5 Acknowledgements

I wish to thank Victor Yngve for his critical review of the manuscript.

References

1. Yngve, Victor H., From Grammar to Science. (John Benjamins, Philadelphia, PA, 1996).
2. Yngve, Victor H., Early MT Research at M.I.T. - The search of adequate theory. In *Early Years in Machine Translation, Amsterdam Studies in the Theory and History of Linguistic Science,* vol. 97 ed. by W. John Hutchins, (John Benjamins, Amsterdam/Philadelphia, 2000) pp. 38-72.
3. Barr, Avron, Cohen, Paul R. and Feigenbaum, Edward A., The Handbook of Artificial Intelligence, vol. 4 (Addison-Wesley, Reading, MA 1989) pp. 223–237.
4. Austin, J. L., How to Do Things with Words, 2nd ed. (Harvard U. P., Cambridge, MA, 1975).
5. Brezar, Mojca Schlamberger, A Business Negotiation Analysis in the Scope of Hard-Science Linguistics, In Yngve and Wąsik [13].
6. Cisło, Anna, The Victorian Stereotype of an Irishman in the Light of Human Linguistics, In Yngve and Wąsik [13].
7. Coleman, Douglas W. Data and science in Introductory Linguistics Textbooks, Paper presented at the LACUS Forum XXVII, Houston 2000.
8. Czajka, Piotr, Human Needs as Expressed in Educational Discourse on the Basis of Textbooks in Linguistics, In Yngve and Wąsik [13].
9. Sypniewski, Bernard Paul, A Hard Science Linguistic Look at Some Aspects of Criminal Litigation in Contemporary New Jersey. Rowan University-Camden Campus, ms.
10. Mills, Carl, Linguistic Change as Changes in Linkages: Fifteenth-Century English Pronouns, Paper presented at the LACUS Forum XXVII, Houston 2000.
11. Malak, Janusz, Mayday or M'aider. A Call for Help in Understanding Linguistic Change, In Yngve and Wąsik [13].
12. Rieger, Caroline L., Exploring Hard Science Linguistics: Fillers in English and German Conversations, Paper presented at the LACUS Forum XXVII, Houston 2000.
13. Yngve, Victor H. and Wąsik, Zdzisław (eds.), Exploring the Domain of Human-Centered Linguistics from a Hard-Science Perspective (Poznan, Poland: School of English, Adam Mickiewicz University, 2000).