In J.M. Abe and J. I. da Silva Filho (Eds.), Logic, Artificial Intelligence and Robotics. (2001) Proceedings of the Second Congress of Logic Applied to Technology--LAPTEC 2001, held at Sao Paulo, Brazil, November 12-14, 2001. Amsterdam: IOS Press, 238-254.

397

Logic, Artificial Intelligence and Robotics
J.M. Abe & J.I. da Silva Filho (Eds.)
IOS Press, 2001

Semantic Computation by Humans, Computers and Robots

Patrick Suppes
Stanford University

Abstract.

The era of voice recognition by a great variety of technological devices is just beginning. No doubt speech-recognition rates will continue to improve and there will be increasing acceptance of this important channel of communication between human users and devices, from personal computers, PDA's, and private automobiles to kitchen ovens and refrigerators. Much of this communication will be successful. In Section 1, a detailed example from machine learning of robotic natural language will try to make clear why this is so. But the general reason is clear. Well-defined sublanguages of a natural language can be given a restricted and unambiguous semantic interpretation. Their semantic structure is close to that of an elementary formal language. In contrast, open-ended conversations, with the semantic ambiguities characteristic of such language use, present different and much more difficult problems. But progress on these problems will be critical for much wider use of speech in interacting with devices. In Section 2, some conjectures about how the brain handles such language are discussed, together with some detailed data on brain processing of language. Section 3, the final one, is devoted to considering whether device-implementation of conversational language is likely or not to use similar semantic computations.

1 Machine Learning of Regimented Natural Sublanguages

I, together with younger colleagues, have been developing a theory of machine language learning in the context of robotic instruction of elementary assembly actions [1, 2, 3]. More specifically, our situations of language learning concern actions like moving to objects, picking up objects and moving objects in environments. Typical objects are screws, nuts, washers, and sleeves of various colors, sizes, and shapes, related to each other by common spatial relations. Typical commands to be comprehended are *Get a screw.*, *Go to the black washer behind the screw.* and *Put the screw in the left hole.* The system developed so far deals successfully with English, Chinese, and German, as well as several other languages.

We have taken what we believe is a new tack in the approach to machine learning by using in a very explicit way principles of association and generalization derived from classical psychological principles. The principles we used were, however, much more specific and technically developed.

The fundamental role of association as a basis for conditioning is thoroughly recognized in modern neuroscience and is essential to the experimental study of the neuronal activity of a variety of animals. For similar reasons its role is just as central to the learning theory of neural networks, now rapidly developing in many different directions. We have not, however, made explicit use of neural networks, but have worked out our theory of language learning at a higher level of abstraction. In our judgment the difficulties we face need to be solved before a still more detailed theory is developed.

The classical psychological principles of learning used here have been thought by linguists to be wholly inadequate as the basis for a theory of language learning. Nothing could be further from the truth. Skinner's naive formulation of the problems of language learning [4] was rightly attacked by Chomsky [5], but no serious alternative learning theory has been offered by linguists even today.

First, we briefly describe our approach to machine learning of natural language. Second, we focus on the problem of denotation that is important in our use of probabilistic association of words and their meaning. Third, we outline the background cognitive and perceptual assumptions of our machine learning work. Fourth, we formulate explicitly our two general axioms of association and denotation, but do not state the additional axioms describing the full learning process. These may be found in the publications already cited, with some changes being made over time.

1.1 Our Approach to Machine Learning

Without going into all the details, we want to convey a rather clear intuitive sense of the process of learning of natural language in terms of the various events that happen when an utterance is given to a robot. (In this and succeeding sections we shall refer to robots, but it should be understood that the basic program of machine learning would apply without serious modification to other applications. Following standard learning usage, we shall often speak of trials where, of course, we mean that the trial begins with a command in the form of an utterance to be executed by the robot.

The most important way to describe conceptually the learning process our program embodies is in the description of the state of memory of a robot at the beginning of each trial. There are four aspects of this memory that are changed due to learning. The first is the association relation between words of a given language and internal symbols that have as denotations actions, objects, properties and relations in the robot's world. A central problem is to learn in each language what word is properly associated with a given internal symbol. A second aspect of the memory that changes is the denotational value of a given word, which will affect its probability of being associated. The third part that changes is the short-term memory that holds a given verbal command for the period of the trial on which it is effective. This memory content decays and is not available for access after the trial on which a particular command is given. This means that at the beginning of the trial, before a command is given, this short-term buffer is empty. What we have said thus far, could, with some stretching, fit into classical theories of association, but for language learning it is quite evident that the association relation and some simple features of short-term memory are certainly not enough.

The fourth aspect is the important one of learning grammatical forms. Consider the verbal command *Get the nut.* This would be an instance of the grammatical form *A the O,* where *A* is the category of actions and *O* is the category of objects. This form actually represents a mild oversimplification, because we do not have just a single category of actions. There are several subcategories, depending upon the number of arguments required, and certain other natural semantical requirements as well. The example will illustrate how things work, however. The grammatical forms are derived by generalization only from actual instances of verbal commands given to the robot. No prior knowledge of any sort of the grammar of the natural language to be learned is available to the robot. Also important is the fact that associated with each grammatical form as it arises from generalization are the associations of the words which have been the basis for the generalization, along with their internal representations. For example, if *Get the nut* were the occurrence in which the grammatical form just

stated was generated, then also stored with that grammatical form would be the associations *get* ~ $g and *nut* ~ $n, where $g and $n are internal symbols whose denotations are known to the robot. When incorrect associations are deleted by further learning, the grammatical forms based on such associations are also deleted.

1.2 Problem of Denotation

In the probabilistic theory of machine learning of natural language which we have been developing, we have encountered in a new form a standard problem in the analysis of the semantics of natural language, namely, how to handle words that are nondenoting. We do not mean nondenoting in some absolute sense, but relative to a fixed set of semantic categories. These categories in the robotic case are, roughly speaking, the categories of actions, objects, properties and relations. It may well be that in some elaborate set-theoretical semantics of natural language, nondenoting words like the definite article *the* denote a complicated set-theoretical function, but the relevance of such an elaborate semantics to language learning is doubtful. In the robotic context, we have something simpler and closer to the common man's view of what denotations are. We take as denoting words color and object words, common nouns, familiar concrete action words, *etc.*. We take ordinary prepositions in English and sometimes other devices in other languages to denote relations in most cases.

When a child learning a first language or an older person learning a second language first encounters utterances in that new language, ther is no uniform way in which nondenoting words are marked. There is some evidence that various prosodic features are used in English and other languages to help the child. For example, in many utterances addressed to very young children, the definite or indefinite article is not stressed but rather the common noun it modifies, as in the expression *Hand me the cup.* But such devices do not seem uniform and in any case are not naturally available to us in our machine-learning research, where we use written input of words without additional prosodic notation.

As has already been made clear, a central feature of our approach to machine learning is the probabilistic association between words of the natural language being learned and denoting symbols of the internal language. It is appropriate that at the beginning all words are treated equally, and so the associations are formed from sampling based on a uniform distribution. On the other hand, after many words have been learned and a good deal of language has been acquired by the robot, it is very unnatural, and also inefficient, if the robot is now given, for example, the esoteric command *Get the voltmeter,* to have the internal symbol $*vol* be associated with equal probability with the definite article *the* and *voltmeter* – we assume here that the association of *get* is already correctly fixed. After much experience, what we want is that there is very little chance of associating the definite article *the* with any denoting symbol.

1.3 Background Cognitive and Perceptual Assumptions

Before explicitly formulating the learning principles of association and denotation we use, we first state informally assumptions we make about the cognitive and perceptual capacities of the class of robots, albeit as yet quite limited, we work with.

Internal language. The robot has a fully developed internal language, which it does not learn. It is technically important, but not conceptually fundamental, that in our case this language is LISP. When we speak here of the internal language we refer only to the language of the internal representation, which is itself a language at a higher level of abstraction, relative

to the concrete movements and perceptions of the robot. It is the language of the internal representations held in memory that provides the direct interface to the natural-language learning. In fact, most of the machine learning of a given natural language can take place through simulation of the robot's behavior by using just the language of the internal representation. The first associations learned are between the internal representation in memory of a coerced action and a contiguous verbal utterance in the natural language being learned.

The fundamental importance of this internal representation of a coerced action can be recognized by considering a parallel case of animal learning. When a dog is trained to *Get the paper* or *Get the ball* by being led through the desired action or by some related technique, the residue in memory of what we term the coerced action is surely drastically abstracted from the perceptually rich context of the demonstrated action desired, and it is that abstracted internal representation in memory that must be associated to the verbal stimulus in order for the dog later to perform the desired action upon hearing the verbal command. We are a long way from knowing even the general structure of the dog's internal representation in memory of the action. In this limited sense, life with a robot is much easier, for we ourselves create the form of its internal representation.

Objects, relations and properties. We furthermore assume the robot begins its natural-language learning with all the basic cognitive and perceptual concepts it will have. In other words, our first-language learning experiments are pure language learning. Any learning of new concepts is delayed to another phase. For example, we have assumed that the spatial relations frequently referred to in all, or at least all the languages we consider in detail, are already known to the robot. This is quite contrary to human language learning. For example, probably in no widely used natural language at least, do children at the age of thirty months use or fully understand the relations of left and right. To avoid misunderstanding, we emphasize that we consider it an important future task to have the robot also learn the familiar spatial and temporal relations.

Actions. What was just said about objects and relations applies also to actions, represented in English by such verbs as *pick up, get, place, etc.*. The English, of course, must be learned, but not the underlying actions.

Associations and grammatical forms. Before stating any formal principles of learning, we feel it is desirable to describe as informally and intuitively as possible the learning setup we use. Consider the English command *Pick up the screw,* no part of which has as yet been learned by the robot. The learning steps may be roughly schematized as follows:

(i) By coercion, or simulation of coercion, the robot creates in memory an internal representation of the coerced action of picking up the screw;

For statement of learning principles we show this internal representation, not as a LISP expression, but just as a schematic function $I(\ldots)$ of the denoting terms in the LISP expression. Here, by *denoting terms* we mean the names in the internal language of the actions, objects, properties and relations mentioned. The internal representation of *Pick up the screw* is then *I($p,$u,$s)*, where $\$p$ = the action of picking, $\$u$ = the direction up and $\$s$ = screw.

(ii) By contiguity the robot associates the verbal utterance and the internal representation

$$\textit{Pick up the screw} \sim I(\$p, \$u, \$s),$$

where $\sim$ is the symbol we use for association;

(ii) By probabilistic association, the robot associates the internal denotations with the English words, with one possibility the following incorrect result:

$$pick \sim \$s, up \sim \$p, screw \sim \$u.$$

We need to observe the following:

a. We assume from the beginning the robot knows word boundaries, as delineated by the typed input. This is an example of an assumption that is natural for robots, but clearly false for very young children;

b. For our simple example, there are 24 possible ways of associating the three internal symbols to the four denoting words in the English utterance. We initially assign to each of these 24 possibilities equal probability, but as trials continue, modify the probability by dynamic changes in denotational values, as is explained later in detail.

(iv) After the associations are made, by the principle of generalization, which we call the category generalization, each word is assigned the category of its associated internal symbol. In the present case $pick \in O$ – the category of objects, $up \in A$ – the category of actions and $screw \in R$ – the category of relations. A grammatical form is then also generalized from the verbal command:

$$O\ A\ the\ R$$

which, like the assigned categories, is wrong for English, but remember that this is just the starting point of learning. With this grammatical form is associated its internal representation $I(A, R, O)$ which characterizes its meaning.

(v) A new command is presented as the next step, say *Pick up the nut.* By coercion the internal representation $I(\$p, \$u, \$n)$ is created (see (i) above). The robot then first searches its memory to see if any of the words uttered are associated to one of the internal denotations. Here the result is $up \sim \$p$, and also the classification of *the* as a nondenoting word is found. There are then six possibilities of probabilistic association for *pick, the* and *nut.* Note that the earlier incorrect association of *pick* with $\$s$ does not appear here, which means that at this stage of learning it will be changed. So, let us suppose the new associations are

$$pick \sim \$u, nut \sim \$n.$$

We also have as a new grammatical form

$$R\ A\ the\ O$$

which though incorrect, now has only the confusion of the associations of *pick* and *up* as its source. To correct these associations we must separate the constant pairing of *pick* and *up*, which is what we do. In any case, we form at once the association to the internal representation:

$$R\ A\ the\ O \sim I(A,\ R,\ O).$$

(vi) Learning stops whenever the following steps of interpretation can be successfully completed upon giving the robot a verbal command:

a. An association to an internal denotation or a nondenoting classification is found in memory for each word;

b. The category of each word is found in memory;

c. The grammatical form resulting from (b) is found with an associated internal representation in memory;

d. The command is correctly executed on the basis of the internal representation.

1.4 The General Axioms of Association and Denotation

We state the axioms in a general form, but we assume already that each word a of the target natural language has a denotational value $d_n(a)$ on each trial. This value changes from trial to trial according to the two different models presented in the next section.

Probabilistic association. *On any trial n, let a natural language sentence s be associated to σ, its internal representation, let $\{a_i\}$ be the set of words of s not associated to any internal denoting symbol of σ, let $d_n(a_i)$ be the current denotational value of each such a_i and let $\{\alpha_j\}$ be the set of internal denoting symbols not currently associated with any word of s. Then:*

(i) an element α_j is uniformly sampled without replacement from $\{\alpha_j\}$;

(ii) at the same time an element a_i is sampled without replacement from $\{\alpha_j\}$ with the sampling probability

$$p_n(a_i) = \frac{d_n(a_i)}{\sum_{\{a_i\}} d_n(a_i)};$$

(iii) the sampled pairs are associated, i.e., $a_i \sim \alpha_j$;

(iv) sampling continues until either the set $\{a_i\}$ or the set $\{\alpha_j\}$ is empty.

Denotational value computation. *If at the end of a trial a word a in the presented sentence is associated with some internal symbol α, then $d(a)$, the denotational value of a, increases and if a is not so associated $d(a)$ decreases. Moreover, if a word a does not occur on a trial, then $d(a)$ stays the same unless the association of a to an internal symbol α is broken on the trial, in which case $d(a)$ decreases.*

A more detailed denotational axiom is the following:

If, at the end of trial n, a word a_i in the presented verbal stimulus is associated with some denoting internal symbol α_j of the internal representation σ of s at the end of the trial, then

$$d_{n+1}(a_i) = (1 - \theta)d_n(a_i) + \theta, \quad 0 < \theta \leq 1$$

and if a_i is not so associated,

$$d_{n+1}(a_i) = (1 - \theta)d_n(a_i).$$

Moreover, if a word a_i does not occur on trial n, then

$$d_{n+1}(a_i) = d_n(a_i).$$

unless the association of a_i to an internal symbol α_j is broken on trial n, in which case

$$d_{n+1}(a_i) = (1 - \theta)d_n(a_i).$$

Table 1: Comprehension Grammars

Chinese, English, German	
1. $O \rightarrow [DA]\ S$	
2. $O \rightarrow [IA]\ S$	
3. $S \rightarrow P\ S$	
4. $S \rightarrow OBJ$	
5. $G \rightarrow R\ [PO]\ O$	
6. $D \rightarrow R$	
7. $P \rightarrow P\ [\&]\ P'$	
8. $P \rightarrow [\neg]\ P$	
9. $P \rightarrow P\ [\vee]\ P'$	
Chinese, English	English, German
10. $A \rightarrow [ADV]\ A_4\ D\ O$	12. $A \rightarrow [ADV]\ A_1\ O\ [COP]$
11. $A \rightarrow A_4\ O$	13. $A \rightarrow [ADV]\ A_2\ G\ [COP]$
	14. $A \rightarrow G\ A_3\ O$
	15. $A \rightarrow A_3\ O\ [COP]\ G$
	16. $A \rightarrow [ADV]\ A_4\ [COP]\ O\ D$
	17. $S \rightarrow S\ [RP]\ P$
	18. $S \rightarrow S\ [RP]\ G$
Chinese	German
19. $A \rightarrow [ba]\ O\ A_1$	26. $A \rightarrow A_4\ [einen]\ S\ D\ [der]\ P\ ist$
20. $A \rightarrow [xianzai]\ G\ [na4]\ A_2$	27. $A \rightarrow A_4\ [nun]\ S\ D\ die\ G\ ist$
21. $A \rightarrow zai\ G\ A_3\ O$	28. $A \rightarrow A_4\ die\ S\ R\ von\ dem\ S\ D\ [der]\ G\ ist$
22. $A \rightarrow [ba]\ O\ A_3\ [dao1]\ G$	29. $A \rightarrow A_4\ die\ S\ R\ der\ S\ D\ die\ P\ ist$
23. $A \rightarrow [ba]\ O\ A_4\ D$	
24. $S \rightarrow G\ [de]\ S$	
25. $G \rightarrow O\ [de]\ R$	

1.5 Comprehension Grammars Generated

Comprehension grammars have been generated so far for English, Chinese, German, French, Dutch and Korean. Mainly because the first languages of the authors are English, German and Chinese respectively, we will concentrate on these three languages with remarks about the others. The corpus used has consisted of 456 commands, the last 60 of more complicated ones. In the first group commands range from the simple *Get the screw.* to *Place the screw on the plate.* In the last 60 a typical more complicated command is the following: *Put a small nut on the screw behind the washer left of the plate.*; *Ba yige xiaode luomu fang zai nage ban zuobian de dianquan houmian de nage luosiding shang.*; *Tu eine kleine Mutter auf die Schraube hinter dem Dichtungsring links von der Platte.*

We now turn to the three grammars generated for English, Chinese and German. The grammatical rules are written in context-free notation, in Table 1, where A is the category of actions, D is that of directions, G of regions, O of objects, P of properties, R of relations and S of sentences. Brackets are used for congruence classes of words which semantically function the same way in comprehension, possibly from different languages. For example, $[DA]$ is the congruence class of definite articles in Chinese, English and German used in our corpus. The full list is given in Table 2. The symbol G denotes the option to omit a word of the congruence class.

Using the axiom on congruence to collapse rules that differ only in the occurrence of semantically equivalent nondenoting words, the number of rules for English is 18, for Chinese 18 and for German 20. What is of perhaps greater interest is the analysis of the structure

Table 2: Congruence Classes

$[DA]_1$	=	(*the; nage, zai, zhege; das, dem, den, der, die*),
$[IA]_2$	=	(*a; yige; eine, einem, einen, einer*),
$[PO]_5$	=	(*of,* ϵ; ϵ; *von,* ϵ),
$[\&]_7$	=	(*and; he; und*),
$[\neg]_8$	=	(*not; busi; nicht*),
$[\vee]_9$	=	(*or; huozhe; oder*)
$[ADV]_{10}$	=	(*now, so,* ϵ; ϵ)
$[ADV]_{12}$	=	(*now, so,* ϵ; *nun,* ϵ)
$[COP]_{12,13,15,16}$	=	(ϵ; *ist,* ϵ)
$[ADV]_{13}$	=	(*now, so,* ϵ; *dann, jetzt, nun,* ϵ)
$[ADV]_{16}$	=	(*now, so,* ϵ; ϵ)
$[ADV']_{16}$	=	(ϵ; *jetzt, nun,* ϵ)
$[RP]_{17}$	=	(*that is, which is; der, die, ist die*)
$[RP]_{18}$	=	(*that is, which is,* ϵ; *der, die, die sich,* ϵ)
$[ba]_{19}$	=	(*ba, xianzai ba,* ϵ)
$[xianzai]_{20}$	=	(*chao, name, xianzai*)
$[na4]_{20}$	=	(*na4, na4li*)
$[ba]_{22}$	=	(*ba,* ϵ)
$[dao1]_{22}$	=	(*dao1, zai,* ϵ)
$[ba]_{23}$	=	(*ba, xiazai ba,* ϵ)
$[de]_{24}$	=	(*de, de nage*)
$[de]_{25}$	=	(*de,* ϵ)
$[der]_{26}$	=	(*der, die*)
$[einen]_{26}$	=	(*einen, jetzt eine*)
$[ist]_{26}$	=	(*ist,* ϵ)
$[nun]_{27}$	=	(*die, nun die*)

of common rules in comparison with special rules for the particular languages. The basic numerical data are these. There are 9 rules that are common to English, Chinese and German where 'commonality' means that the category such as definite article now includes words from each of the three languages, in this particular case for example, *the, nage, zai, zhege, das, dem, den, der*, and *die*. In addition, there are 2 rules common to English and Chinese that are not in the German grammar, there are 7 rules common to English and German that are not used in Chinese, and no rules common only to Chinese and German.

What is remarkable about rules 1–9 in Table 1 is that none of them are high level rules for the generation of complete commands. The first 4 deal with the generation of object phrases, Rule 5 with the generation of a description of a region, Rule 6 with a direction, and the last 3 with properties. Note that the rules common to English and Chinese are high level rules for generating commands, and that 5 of the 7 rules common to English and German are high-level rules for generating commands. The remaining 2 are for generating object phrases. Five of the 7 rules in Table 1 special for Chinese are high-level rules for generating commands.

The big surprise is that no special grammatical rules are required for English.

In Table 2 the congruence classes for the three languages are shown. The subscript on each class shows the grammatical rule of Table 1 on which it depends. For example, $[ADV]_{10}$ depends on Rule 10, which is relevant only for English and Chinese. The congruence class for Rule 1 is intuitively incorrect. The reason is simple to explain. The Chinese particles *xianzai, dao1, zai,* and *ba* do not generally co-occur with object phrases, as the rules suggest, but with other categories of words. For example, *xianzai* has a meaning that is close to the English adverb *now*, the particle *dao1* occurs with verbs, as does *ba*. What is important is

that at the level of the commands we are considering, the nonintuitive Rule 1 is satisfactory for the purposes of comprehension. It would of course lead to very bad Chinese if applied to the production of utterances. We emphasize once again that in considering these Chinese examples, we are generating comprehension by uniform procedures that are the same across all languages. We are not claiming this can be done for production or even that for the ultimate reaches of comprehension it will be satisfactory.

2 Brain Processing of Words and Sentences

In our earlier analyses of brain-wave representations of words and sentences, based on electroencephalographic (EEG) data, we averaged over trials, as well as subjects, and made a discrete fast Fourier transform (FFT) to the frequency domain [6, 7, 8]. We then searched for a filter that optimized correct recognition of the words or sentences being processed. Using filters to eliminate noise from signals is widespread in many kinds of signal processing. When speech or music constitute the signals, such filters not only work well, but are practically necessary, because of the large number of component waves.

The optimal filters we found in our earlier studies usually fell well within the range 2- to 15-Hz. So, using the standard software for discrete fast Fourier transforms with a sampling rate between 600 and 1,000 Hz, we, in fact, were ordinarily using a filter that contained less than 60 discrete frequencies. This relatively small number of frequencies immediately suggests an alternative to filtering for our EEG-recorded brain waves. This is to look at the frequencies with comparatively large amplitudes, select a small set of these, and use their superposition instead of a filter. (The work summarized here comes from Suppes and Han [15].) So, for each observation i the superposition wave S_i is just

$$S_i = \sum_{j=1}^{n} A_j \sin(\omega_j t_i + \varphi_j),$$

where A_j, ω_j and φ_j are the amplitude (in microvolts), frequency (in radians/s) and phase (in radians) of the jth sine wave. Superposition of continuous light waves is familiar from classical optics in the study of diffraction and interference. Here we use discrete wave representations to match our discrete fast Fourier transforms. For simplicity of comparison, we report only relative amplitudes, not microvolt calibrations, for all three experiments, performed with three different EEG systems. We note that the least-squares criterion of fit used in all our analyses is invariant under a change of the units in which amplitude is measured.

For the reason already stated, such superpositions are uncommon in standard signal processing, but in our special low-frequency environment they can work very well. Moreover, to represent a word by a small set of superposed pure sine functions gives a definite sense of the minimum number of parameters needed for the invariant brain waves that seem to characterize rather well the words we have studied. To be explicit, each sine wave j in the superposition is characterized completely by A_j, ω_j and φ_j. For superpositions made up of five frequencies this yields a 15-parameter representation, which is certainly not enough to represent the spectral analysis of any spoken word, and is thereby testimony to the apparently simplifying transformations imposed by the auditory system on a sound-pressure wave as it reaches the cortex.

When we referred to "invariant" brain waves above we had in mind the extensive averaging over trials and subjects used to obtain an invariant result. Such averaging can eliminate more than noise, for the unaveraged signals may well contain much additional information,

such as individual associations, not needed to identify the word itself. At present, our success in correct recognition of what word or sentence is represented depends on using such averaging.

Our focus entirely on waves, with no reference to populations of neurons and their intermittent spiking, as the real communication setup physically, may make some readers skeptical of our results. We certainly believe the waves we find in the observed EEG data arise from the spiking activity of many neurons, as their source. Whether or not there is a more fundamental way, at least in principle, to observe more directly, for cognitive purposes, the collective activity of the neurons handling speech, although it is an important question.

2.1 *General Methods of Analysis*

After applying an FFT to the averaged EEG data, we selected high-energy frequencies for the superposition wave representing a given word. Our criteria for selecting these frequencies were the following:

(i) We excluded any frequencies below 1.5 Hz as being too low, or, in the case of the large direct-current amplitude at 0 Hz, as being irrelevant for classification or prediction;

(ii) Based on our earlier studies, we excluded any frequencies equal to or greater than 20 Hz;

(iii) A frequency selected must be a local maximum in its amplitude;

(iv) When two local maxima were separated by only one frequency in the discrete FFT, we used only the one with higher amplitude, and if the amplitudes were the same, we selected the higher frequency;

(v) For superposition of n frequencies, a frequency selected must be one of the n highest local maxima, subject to the exclusion of (iv).

Subjects were numbered consecutively, and sometimes used in more than one of the earlier studies. The same numbering is used in this article. The sensors referred to follow the nomenclature of the standard EEG 10-20 system.

2.2 *Results*

In our first experimental study [6], we presented auditorily seven words to subjects in 100 randomized trials for each word. We recorded with the standard 10-20 EEG system subjects' brain waves beginning shortly before the onset of each verbal stimulus. Using the EEG data of the best sensor C4 for recognizing the seven words *first, second, third, yes, no, right* and *left,* we applied the method of superposition described above. First, we averaged together the unfiltered EEG data for subjects S3, S4 and S5, which yielded 300 trials for each stimulus word presented auditorily. These data were the test samples. We next applied an FFT to these averaged data for each word, and we selected, for each word, using the criteria stated earlier, the seven frequencies, i.e., sine waves, with the highest energy from the less than 60 frequencies, computed by the discrete FFT, with 0.662 Hz the difference between successive frequencies. As an example, the frequency-domain graph of the relative amplitude for the auditory word *first* is shown in Fig. 1.

Using now as a prototype for each word the superposition of the seven selected sine waves, we classified the test samples consisting of the unfiltered but averaged data described

above. The criterion of fit, as in our previous work, was minimum least squares of all observations over a selected temporal interval. For a number of intervals we correctly classified all seven words.

To test how few superposed sine waves were required, we next systematically reduced the number of frequencies used in the superposition, by deleting first, the highest frequency from the seven for each word. With six superposed sine waves we also correctly classified the test samples for all seven words. Continuing, by deleting always the highest remaining frequency, we continued to classify correctly all seven words, for five, four or three frequencies. Finally, using only the single remaining lowest two frequencies for each word, we correctly classified six of the seven words.

Table 3: Three frequencies selected for each of seven words

Freq in Hz	Relative Amplitude	Phase in degrees	Freq in Hz	Relative Amplitude	Phase in degrees
	First			*Yes*	
2.649	10.86	-10.8	3.974	14.51	-69.79
4.636	13.74	26.8	5.298	10.90	151.69
5.961	14.47	-100.7	7.285	7.50	130.8
	Second			*No*	
3.312	11.01	105.0	2.649	14.02	52.8
5.961	20.16	-110.6	5.961	11.68	-93.9
9.935	4.89	137.1	9.935	4.94	-99.8
	Third			*Right*	
4.636	17.03	12.6	1.987	10.57	-166.1
6.623	17.23	21.1	3.312	13.14	154.8
8.610	12.72	-5.3	4.636	13.00	46.2
				Left	
			3.312	15.70	-170.9
			4.636	9.54	36.9
			7.285	9.57	139.3

In Table 3, we show the superposed three lowest frequencies, their amplitudes and their phases for each of the seven words. The selected frequencies all fall between 1.9 and 10.0 Hz, with more variation in phase than amplitude. The three selected for *first*, as shown in Table 1, are easily identified also in Fig. 1. In Fig. 2, the superposition of the lowest two,

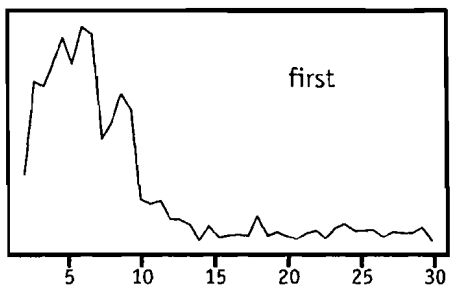

Figure 1: Graph in the frequency domain of the FFT of the averaged data for the word *first*, with discrete frequencies shown on the x axis in hertz and the amplitudes of the frequencies on the y axis in relative amplitude.

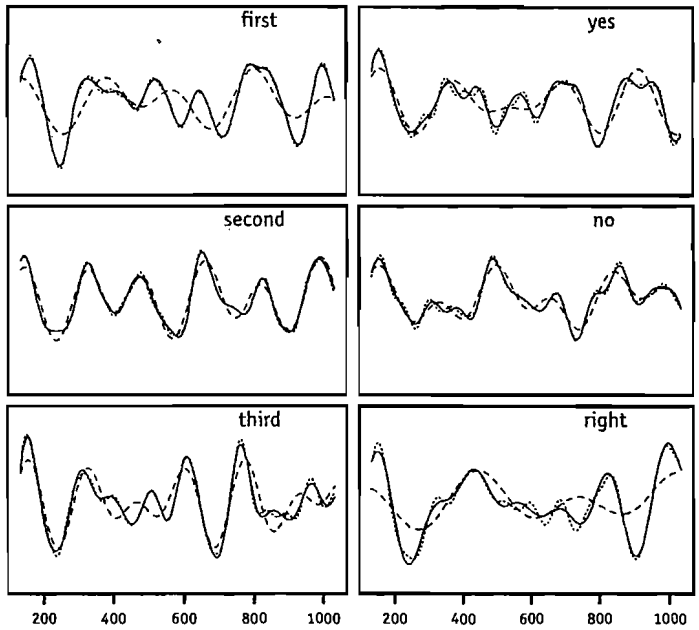

Figure 2: Comparison of superposition of the two lowest frequencies (dashed line), superposition of five frequencies (solid line), and superposition of seven frequencies for six of the seven words (dotted line). The x axis is measured in milliseconds after the onset of the stimulus and the y axis in relative amplitude.

the lowest five and the seven frequencies are shown in panels for six of the seven words. We omit the seventh one, for *left*, only to save space. The graphs for Fig. 2 are for one of the temporal intervals optimal for all superpositions, except the superposition of two frequencies. The temporal interval used was for 897 ms, beginning 132 ms after onset of stimulus. The interval for two frequencies was 750 ms, beginning at 132 ms. x

First, a completely different approach to superposition would be to estimate, purely statistically, for each prototype, a fixed number of best frequencies with optimally fitted amplitudes and phases. But such a purely statistical approach, with no physical constraints, produces frequencies with wildly varying amplitudes that seem obviously implausible, even though, through interference of one frequency with another, very good fits are produced. In contrast, by using frequencies found in the Fourier analysis of the data, the case for their physical reality is clear.

On the other hand, using pure sine waves to represent the fundamental frequencies may seem unrealistic. Because there is a lot of evidence that the images of the stimuli we used last only for a second or so at most, a more obvious mathematical choice would be some kind of wavelet representation with a short temporal domain. There are now many types of wavelets, with accompanying software, available for such purposes [9, 10]. In the end this may turn out to be the right approach, but our earlier efforts [7] to use wavelets did not improve on the Fourier results. For the present, we think the use of pure sine waves is an acceptable approximation for the short temporal length we consider, and, in fact, may be hard to improve on. A different problem that must be dealt with in a more complete model of

the brain's processing is the damping out of the activated image. But this extension presents no fundamental difficulty.

2.3 *Natural Computations*

The contrast between the style, if not the subject matter, of the first two sections of this paper are stark. The discourse and marshalling of concepts in the first section are very much in the spirit of formal work on languages, as can be found in logic and computer science, even though the target of the machine-learning effort has been on the learning of natural language, but of course a very regimented sublanguage. Such problems do have an empirical aspect, namely, it is impossible to know in advance how well the learning schemes of association, denotation and grammatical generalization will work. Only by running a program can one draw serious conclusions about the learning, at least at the present stage of development of the theory.

The contrast between Section 1 and Section 2 brings out clearly enough how complicated it is to understand thoroughly the processing of natural language by the brain. The work described in Section 2 has been an intense computational effort to get through the complicated data recorded by standard EEG methods to identify well-defined brain-wave representations of words and sentences. The example given here is a simple one, just for reasons of keeping the exposition limited. In many ways the most successful work has been not with words, but with entire sentences, as reported in references here to these 3 articles [7, 8, 11]. In the case of sentences, the contrast with Section 1 is really stark. Subjects were asked to judge the truth or falsity of simple questions about the geography of the world. Typical sentences used in the experiments were *Berlin is not north of Rome, Warsaw in not the capital of Austria,* and *the largest city of Poland is Moscow.* In many ways, the sentences used were as regimented as those in the robotic experiments of Section 1. But here the difference was that we were attempting to identify the processing xof the sentences in EEG-recorded brain waves. We were, for the very best subjects, surprisingly successful. For example, we were able for the best subject, to correctly classify 93 of 100 different sentences [11].

On the other hand, the limitation of this result is also very clear. We are as yet only able to speculate on how we naturally compute the truth value of such sentences. Certainly, the sentences themselves are not sitting in memory. Most of them subjects probably have never heard before, or, if they have, they certainly have forgotten them, and have not stored them; they are long since forgotten. Simple semantic computations must be made by subjects, upon hearing or reading any of these simple sentences about geography. How does a subject compute their truth or falsity, that is, how do subjects make the appropriate semantic computation? It is to be stressed as strongly as possible that formal theories of truth derived from the work of Tarski and others in logic many years ago do not provide any real answer. It was, of course, no part of Tarski's intention to provide a theory of natural semantic computation, and only the most general aspects of that work would seem to have relevance to the actual computation carried out in the brain processing of natural-language sentences.

Something a little more than speculation can be said about these natural computations, and it surely will become a subject of intense focus in this century. The associative processes known to occur naturally and widely in the brain, speculation about which were already an important part of Hobbes' and Hume's theory of the mind, developed especially with considerable thoroughness in Hume's *A Treatise on Human Nature,* [12]. Now a very large psychological literature, a growing linguistic literature and a growing neuroscience literature exist on the processes of association of the brain. The concept of an associative network is the best current framework for these natural semantic computations. It is important to stress

that the scientific literature in psychology over the past 50 years on natural processes of association is rich and varied. It does not mainly address the problem of computation that is of focus here, but it provides an excellent basis on which to develop theoretical ideas that can be checked against some observable brain processes. How far those observations will take us with present technology is still an open question. From a direct computational standpoint, in many ways the current work in linguistics is formally more helpful, but it is not possible to try to survey that relevant literature here.

I turn now to my last topic about the features of natural language important for conversation, but which are as yet neither a part of robotic language development nor of empirical brain studies to any serious extent.

3 Semantics of Conversation

It is scarcely news that no robots or computers of any form are capable of conducting, at any time, extensive conversations of a natural sort on any subject. The nihilistic view, and anti-scientific one, is that such conversations will never be possible. This is certainly not the view that I hold, but all the same, what I want to stress here is the depth and difficulty of solving the scientific problems that stand between the present state of affairs and being able to endow robots and other computers with appropriate speaking and listening capabilities.

I mentioned at the beginning that the era of realistic speech recognition has begun. But this does not mean the same thing as having natural conversations with our favorite devices. The kind of work exemplified in the first section on machine learning will take us a long way toward making practical use of speech for information and control.

In this concluding section, I organize my remarks about what I think are the main problems on which we must make serious headway before we can have natural conversations with devices. My remarks fall under two main headings: ambiguity and prosody.

Ambiguity. Perhaps the most striking thing about the semantic quality of natural conversation is the presence of ambiguity in every direction. To begin with, many of the words we use naturally and frequently have a variety of distinct meanings, what is sometimes technically called in linguistics *polysemy*. Famous examples are to be found among some of the most stinging criticisms of artificial intelligence in the past decades. What are the methods by which we determine whether *bank* means riverbank or an institution of finance? And that is only when it is used as a noun, not as a verb. Moreover, once we have fixed in some way the particular meaning of each word in a sentence we find, when we apply standard grammatical parsers, that there are often for sentences of any length, an astronomical number of correct parses, and in a very high percentage of cases, at least two.

The main reason we have great difficulty in thinking about how to solve these problems for robots and other devices so that we can provide them programs, or learning methods for acquiring programs, is our ignorance of how we solve them in human speech.

There is a definite tendency on the part of some logicians, and undoubtedly some computer scientists as well, to think that the right approach to these problems is to live with our ignorance and to simply regard it as a defect of natural language that it is so saturated with ambiguity. But this is a naive strategy for more than one reason. The first is that it is hopeless to have a program of changing the way natural language itself is used by people speaking to each other. A modified program to have people speak in a very regimented sublanguage of their natural language in dealing with their devices will also not be a successful strategy. Competitors will continually be striving to meet the natural demand that the devices talk and speak as well as our normal human companions. I don't mean with this remark to suggest that there won't be good intermediate solutions of the kind I have already referred to. It is

just that the ultimate goal here, both scientific and practical, will be to achieve the kind of easy natural discourse so common between us.

The second and deeper reason for claiming that the strategy of changing natural language is mistaken is this. It is not anything like a defect of natural language that it is saturated with ambiguity. The ambiguity provides a kind of instability, as in the case of other unstable systems or chaotic systems, to be taken advantage of in efficient control and communication. In this case it is a matter of communication. What takes place in general terms is a quick reduction of the ambiguity by the context of the talk and the intentions of the speakers and listeners. But the general context, social and physical, as well as the intentions of speakers and listeners, are difficult things to introduce into our formal analysis of language in the tradition of formal semantics. A successful scientific theory must be prepared to deal with a much wider range of subject matter than is traditional in semantics, as conceived by logicians and even most computer scientists. Now it is a familiar move in philosophy to say that the considerations just mentioned belong to pragmatics, not semantics. But that I dismiss as a mere terminological move. However we label it, considerations of context and intentions are central to our own intuitive understanding of speech, and a problem to master at a scientific level.

It is also important to mention that ambiguity is not restricted to the lexicon and grammar. Our production and perception of spoken language, in terms of the sounds meant to represent the phonemes and syllables of speech are often produced or received badly. The problems of ambiguity at this level are at least as plentiful and, at a fundamental level, still difficult to understand in terms of the correction procedures naturally applied in the brain processing of the listener, and often in the correction of mistakes by the speaker. There is much to be said about these phenomena. I will not say more now.

Prosody. Prosody refers to the organization of pitch and loudness in producing variations in the melody and sonority of spoken utterances and, sometimes, longer passages of speech. A description of prosody by phonologists is an important and complicated topic for linguistics in the study of any natural language. My focus here is not on the technical description of how a prosodic contour is produced by a speaker, but what those prosodic contours of pitch and loudness are used to produce in the minds of listeners. The cries of babies and the joyous cheers of older children, as well as the expressions by all of us of sadness, anger or fear, mainly rely upon the prosodic features of our speech. How the prosodic organization of speech, which at one level can be given a near physical description, is processed in a meaningful way in the brains or minds of listeners, is still little understood. As you might expect, there are brain recordings showing immediate response of the brain to unusual prosodic features, for example, sharp changes of pitch or of loudness. This is rather like similar EEG recordings, for many years now, of the effects on the brain's processing of semantic anomalies. The gross evidence of such anomalies, just as in the case of prosody, is still a very great distance from how anything of any subtlety is actually processed in these matters in the cortex, the most likely part of the brain involved.

Irony. The remarks just made on various uses of prosody are superficial, but at least they touch upon familiar subjects. Often, the case that I want to discuss last, is less frequently mentioned, but of great importance in sophisticated conversations of almost every kind. I am thinking of the use of prosody to give a statement, but especially a response to a statement, an ironic turn that often reverses the literal semantic meaning of the words being used. Socratic irony, i.e., the pretense to ignorance, represents a well-known philosophical tradition. Among recent philosophers, one of the best with serious comments on irony is Paul Grice [13], (pp. 53–54). Subtle as the remarks are, they do not go far enough or cover sufficiently the many complexities of the use of irony. Let me give just two examples in conclusion.

The first is the strong contrast between attempts to use irony in written language, as opposed to spoken speech. A novelist famous for the conversations between his characters is Henry Green [14]. But what can he do when he wants to produce some way of noting that a character's remarks are meant ironically? Here is a typical awkward device that is needed.

> "Does anyone else know of this?"
>
> "Auntie does."
>
> "Of course," Edge took her up with a heavy irony that was wasted, because the girl did not notice.
>
> Green 1951, p. 157

It is not appropriate here to survey the many ways that novelists mark the use of irony, but to stress the contrast with spoken speech is essential. In this great age of cinema, television and video we all can recognize how easy it is for an actor on the stage or in front of the camera, especially in front of the camera, to portray irony by an easy prosodic variation or by a lifting of eyebrows or a subtle movement of other facial muscles.

My second example is something that is not carried far in Grice's analysis. That is the complexity that arises from the social setting of more than two persons. Two knowing people talking to an innocent third can continually make subtley ironic remarks without the innocent third person being aware of what is going on between the other two. So here, the point that Grice makes about irony, namely the one mentioned above, that in irony, the true intended meaning is usually the opposite of the conventional meaning of the words spoken is often not accurate. The two knowing conversationalists, while in their responses affirming to the third conversationalist, the innocent one, the conventional meaning are denying it between themselves. This play of affirmation and denial, sincerity and irony, ricochets through conversations in all walks of life and about all subject matters.

This subtle use of irony is wonderfully done in modern films, even better than on the stage. A mere knowing lift of an eyebrow while saying something apparently sincere, or even a slight change in prosodic contour, is enough to alert the audience to the sense of irony that has been brought into play.

Whether on the stage, before the camera, or in the chair across the room, a muttered "Indeed" or "You don't say" or "Whatever" can be marked with irony, often to the delight of all and the malice of none. Getting such performances from our digital devices will in the end be a work of art, as well as of serious science.

References

[1] Suppes, P., Böttner, M. and Liang, L. (1995) Comprehension grammars generated from machine learning of natural language. *Machine Learning*, **19**, 133-152.

[2] Suppes, P., Böttner, M. and Liang, L. (1996) Machine learning comprehension grammars for ten languages. *Computational Linguistics*, **22**, 329-350.

[3] Suppes, P. and Liang, L. (1996) Probabilistic association and denotation in machine learning of natural language. In A. Gammerman (ed.), *Computational Learning and Probabilistic Reasoning*. Sussex, England: John Wiley & Sons, Ltd., 87-100.

[4] Skinner, B. F. (1959) *Verbal Behavior*. New York: Appleton.

[5] Chomsky, N. (1959) Review of B. F. Skinner, *Verbal Behavior*. *Language*, **35**, 26-58.

[6] Suppes, P., Lu, Z.-L. and Han, B. (1997) Brain wave recognition of words. *Proc. Natl. Acad. Sci.* **94**, 14965-14969.

[7] Suppes, P., Han, B. and Lu, Z.-L. (1998) Brain-wave recognition of sentences. *Proc. Natl. Acad. Sci.* **95**, 15861-15866.

[8] Suppes, P., Han, B., Epelboim, J. and Lu, Z.-L. (1999) Invariance between subjects of brain wave representations of language. *Proc. Natl. Acad. Sci.* **96**, 12953-12958.

[9] Daubechies, I. (1992) *Ten Lectures on Wavelets*. Philadelphia: Soc. Indust. Appl. Math.

[10] Bruce, A. and Gao, H.-Y. (1996) *Applied Wavelet Analysis with S-Plus*. New York: Springer.

[11] Suppes, P., Wong, D.K., Perreau Guimaraes, M., Uy, E.T. and Yang, W. (to appear) High statistical recognition rates for some persons' brain-wave representations of sentences. *Proc. Natl. Acad. Sci.*

[12] Hume, D. (1739) *A Treatise on Human Nature*. London: John Noon.

[13] Grice, P. (1989) *Studies in the Way of Words* Cambridge, MA: Harvard University Press, pp. 53-54.

[14] Green, H. (1951) *Concluding*. New York: The Viking Press.

[15] Suppes, P. and Han, B. (2000). Brain-wave representation of words by superposition of a few sine waves. *Proc. Natl. Acad. Sci.* **97**, 8738-8743.