Neuropsychological predictions of a model of unsupervised language learning

David M. W. Powers
Artificial Intelligence Laboratory
The Flinders University of South Australia
GPO Box 2100, Adelaide SA 5001 — Australia

Language is about patterns. As babies, we see and hear patterns in the world around us and in the speech we hear. As humans, our language and tool making abilities distinguish us from the rest of the animal kingdom. In fact, the patterns around us can be roughly divided into two kinds: those that have been invented by human brains and those that are the result of natural laws.

It seems that the fact that our brain is so good at detecting and predicting patterns is precisely what has allowed us to be so creative. This project pursues a hypothesis that language is an outgrowth of the way that our brain looks at the world, and that the mechanisms which we use to understand the world are the mechanisms we use in learning - and creating - language. This is the hypothesis that has become the cornerstone of the field of Cognitive Linguistics, which rejects Generative Linguistics' notion of an autonomous language organ or language acquisition device, but treats language as an outgrowth of cognition.

Recent work in natural language learning has shown that a variety of unsupervised learning techniques can discover linguistic patterns at the level of sounds (phonemes/characters), minimal meaning and prosodic units (morphemes/syllables), larger grammatical and prosodic units (phrases and clauses), and semantic classes. Interestingly, these computer learning models make a number of predictions which should be experimentally verifiable through psycholinguistic and neuropsychological testing of subjects. We have recently begun to investigate the implications of these models, to propose tests able to confirm or deny some the underlying experiments, and to perform EEG and ERP experiments in which we seek confirmation of our model's predictions about subjects' brain responses to linguistic data, by measuring their electroencephalographic responses ('brainwaves').

The models we are investigating make the assumption that language and general cognition are built on the same fundamental pattern matching mechanisms. We specifically deny the claims of modularity and innateness: that language is a separate module distinct from the rest of cognition and having the basic principles of universal grammar built in, and we proceed on the basis that language and ontology are acquired by unsupervised learning using very general cognitive mechanisms which are innately biased towards understanding the world rather than specifically for language. We hypthesize that pattern matching is the primary learning mechanism, and that constancy and spatio-temporal proximity provide the primary basis for similarity. That is, in the physical words, objects or systems that are indivisible entites tend not to vary suddenly or exhibit discontinuity in either space or time. We therefore seek to model nearness in spatio-temporal space as nearness in cognito-perceptual space, and presume that this is achieved in some kind of a connectionist cognitive framework. However, connectionist models are essentially a special case of a broader class of statistical techniques, and our study is therefore broader than the connectionist models. Given that any perceptual unit exists in a spatio-temporal context which is captured in some form of cognito-perceptual space (or sensory-motor space), we assume that any putative cognito-perceptual unit has a context which includes neighbours in time, space, frequency and other feature spaces which have been provided by earlier stages of processing.

Let us assume now that we are considering a particular such unit. We have a putative unit (actually a coherent class identified by some pattern of features at a lower level) which may be a member of a class at the present level, or may be part of a cluster which is the true unit and is member of a class identified by a particular pattern in the context at the current level). A class is identified as the set of units which occur in essentially the same context, but since contexts will very seldom be identical, we use a measure and a threshold of similarity to decide what should be regarded as similar and thus what should cluster together as a class.

To illustrate the kind of specific predictions of our model we would expect:

1. Optimal learning of particular kinds of patterns occur in language and ontology at the same age;
2. Synactic structures and features are recognized before semantic relations;
3. Closed class words are recognized before open class words, since the model exhibits this behaviour;
4. Prosody is recognized before content, and closed class words produce prosody-like effects;
5. Grammar rules are more actively employed by the speaker than the hearer;
6. Grammar piggybacks on structure, and relationships are not specified more precisely than necessary.

Initial 'brainwave' experiments using subliminal and supraliminal stimuli have demonstrated that we can distinguish conscious (Left and Right Hemispheres) and un-/pre-conscious (Right only) processing . Previous word-level research has identified semantic distinctions (e.g. whether a person is saying yes or no) and syntactic and subcategorization distinctions (e.g. noun/verb, open/closed, active/stative). The basic idea behind our current experiments is that closed class units are very special. They are the first words classified in word level simulations and analogous closed class units are discovered first at various levels from the grapheme/phoneme to the morpheme/word/phrase in grapheme and phoneme level simulations. Note that this conflicts directly with the usual assumptions that content is more important and is learned first. Our model predicts that the child actually learns to recognize closed class units and structures before it makes semantic associations. We should therefore see evidence of these units before the child speaks, in areas proximal to the auditory rather than just the speech cortex.