What if Chomsky were right?

Roland Hausser

Universität Erlangen-Nürnberg
Abteilung Computerlinguistik (CLUE)
rrh@linguistik.uni-erlangen.de

The outcome of scientific research depends on how a phenomenon is viewed and how the questions are phrased. This applies also to the nativist view of language acquisition. As a complement to MacWhinney's discussion of nativism from the viewpoint of cognitive psychology, I would like to devote this commentary to the question of the title from the viewpoint of computational linguistics.

Formally, the nativist approach has been based on a distinction between finite and infinite sets. Chomsky defines a language as an infinite set of strings (sequence of word forms) and a grammar as a filter which picks the grammatically correct strings from the free monoid[1] over the finite lexicon of the language. Language acquisition is described in terms of a language acquisition device (LAD) which has the task of selecting from the infinite set of possible grammars the one which is correct for the language in question.

The 'logical problem of language acquisition' is how the LAD can select a grammar which is correct for an infinite language, even though the data presented to the LAD (observed sentences) are necessarily finite. This problem is only made worse by Chomsky's alleged degeneracy of input and poverty of negative evidence, focussed on by MacWhinney.

Given that humans can obviously learn language anyway, something in addition to a finite set of data is required. According to Chomsky, it is some innate universal grammar, common to all languages. Differences between languages are attributed to different parameter settings of the universal grammar.

As empirical proof for the existence of a universal grammar we are offered language structures claimed to be learned *error-free*. They are explained as belonging to that part of the universal grammar which is independent from language-dependent parameter setting. Structures claimed to involve error-free learning include

1. structural dependency
2. C-command
3. subjacency

[1] The free monoid over a set of words, e.g. {a, b}, is the infinite set of all possible sequences consisting of these words, e.g. aa, ab, ba, bb, aaa, aab, aba, abb, baa, bab, bbb, etc. The free monoid over a finite set is infinite because there is no restriction on the length of the sequences.

4. negative polarity items
5. that-trace deletion
6. nominal compound formation
7. control
8. auxiliary phrase ordering
9. empty category principle

In the first half of his paper, MacWhinney carefully examines each of these, and shows that there is either not enough evidence to support the claim of error-freeness, or that the evidence shows that the claim is false, or that there other, better explanations.

Alternatively, let us assume for a moment that Chomsky is right in the sense that the nativist approach to language acquisition is a scientifically fruitful approach. What would be the outcome of a successful completion of his research program?

There would be an explicitly defined language acquisition device containing an explicitly defined universal grammar. Presented with a finite amount of language data, the LAD would automatically select or construct the correct language-specific grammar. This grammar would be capable of formally deciding whether or not any string of words of the language in question is a grammatical sentence. Furthermore, in the LAD's process of developing the correct grammar in concord with the input data, this grammar would make or allow for the same errors, for example overgeneralization, as observed in children.

Such a system, if it could be built, would be *predictive*. Just as astronomy can precisely predict the future positions of a planet, the LAD could predict the well-formedness of a string not previously encountered, relative to different stages of language acquisition.

There is an important difference between astronomy and language acquisition, however: the prediction of astronomy is relative to constellations observed in the sky, while the prediction of the LAD is relative to the intuitive grammaticality judgements of native speakers. Furthermore, the movement of the stars has no social purpose, while the production and interpretation of language is for *communication* in the sense of transferring information from the speaker to the hearer.

Therefore, predicting grammaticality relative to the development of language acquisition is not enough. The real goal of linguistics is a model of how natural language communication works.[2] This model must be objectively verified by building machines (robots) which can communicate freely in natural language.

Could a successful completion of the nativist program at least contribute to the enterprise of building talking robots by delimiting the set of human languages? For this, the nativist analysis of language form (grammar) would have to follow language function (communication), in line with the most general law of evolution, which it doesn't.

Conversely, could the systematic construction of artificial agents contribute to the explanation of language acquisition in small children? As a case in point, con-

[2] For a detailed description of such a theory see Hausser 1999.

sider the structural dependency constraint. Leaving aside the unacceptable nativist assumption that speakers 'move' things in a sentence (which is in conflict with the time-linearity of natural language interpretation and production) we may ask why the yes-no interrogative corresponding to

a) The man who is running is coming.

is

b) Is the man who is running [] coming?

and not

c) * is the man who [] running is coming

The ungrammaticality of c can be explained without postulating some universal grammar. From the viewpoint of communication, c simply doesn't make sense semantically: what is being questioned is foreground information; therefore it can't be stuck into a relative clause.

This analysis is different from MacWhinney's explanations presented in the second half of his paper, namely limiting the class of grammars, revised end state criterion, conservatism, competition, cue construction, monitoring, and indirect negative evidence. While these explanations and methods are welcome for the project of modeling natural language communication, they are not sufficient by themselves.

The crucial step of moving from a finite set of data to the grammar of an infinite language can be fully explained neither by postulating some universal grammar (Chomsky) nor by a combination of auxiliary principles (MacWhinney). Instead, an explanation of language acquisition requires an explicit modeling of the child's more and more capable attempts to interpret and produce language expressions meaningfully.

This model must include the utterance situation as seen by the child, defined in terms of agents, objects, relations, and clear communicative intentions. The purpose of an utterance, either during interpretation (mother to child) or production (child to mother), is a much stronger influence on the grammatical structure of the expression used than whether or not this expression has been encountered before, be it as positive or as negative evidence.

Bibliography

Hausser, R. (1999) *Foundations of Computational Linguistics, Human-Computer Communication in Natural Language*. 2nd Edition 2001, pp. 578, Berlin, New York: Springer-Verlag.