

What research questions does the article address?

In this article, Behrens discusses a detailed three-year case study following a young boy, Leo, as he learns (what will ultimately be) his native language, German. The study takes place in Leo’s home in Leipzig, Germany, beginning when Leo is about 1 year, 11 months and continuing onward until just short of Leo’s fifth birthday.

One of the main goals of this study is to try to collect data that might shed more light on accepted notions about language acquisition, that children “acquire highly complex language very fast and seemingly without effort.” Behrens questions: “just how fast is fast,” and he suggests that the question of language mastery can only be answered with a “fuller understanding of what the end-state of acquisition is like.” So it seems that one of Behrens’ underlying goals in the study is to obtain data that might contribute to the constructionist viewpoint, by determining (paraphrasing his words) “[if (and when) children use the distributional information in the input (from adult speakers) in their language acquisition, or if (and in what domains) children’s language production ‘deviates from what they hear.’”

Toward that end, Leo’s speech (output) and that of the adults (his parents and a research assistant) is recorded for a period of three years, amassing a corpus of nearly 1 million words (403,034 words from Leo; 458,564 from the adults). Behrens analyzed the data gathered (after tagging and coding in various ways, as described in the paper) across three linguistic domains—parts of speech, NP constituency, and “constituency of verb phrase” for purposes of “comparing the input-output relationship between child and adult language production” to determine “the course of language acquisition” as “measured against the explicit comparison to the target-state, ie., the adult-like distribution and use of linguistic structures.”

In other words, what do the speech patterns of the adults around Leo look like (distribution patterns), how does Leo’s speech compare to theirs, and when does Leo actually start talking like the adults around him?

Why do you find the research question interesting?

As I think you mentioned in class, language acquisition holds a great deal of interest to most of us in the linguistics field (or perhaps you were meaning “us” as human beings)? Either way, seeing infants, toddlers, and young children learn to navigate their way through the world using language is simply one of the most fascinating aspects of human development and something I personally really enjoy.

Also, as a newborn starts developing, there are just a couple of major “benchmarks” that parents seem to want to keep track of in the first couple of years or so—sitting up, crawling, first steps, and first words. (And long before these, perhaps “sleeping through the night ;-) Parents tend to worry if their child doesn’t seem to do any of these things “soon enough,” and language especially seems to be perceived as a measure of ‘brightness’ (at least, among the parents I’ve known.) So such an in-depth study of a single child, particularly one who was identified as a “late talker,” is an interesting example, since ultimately, the distributional frequency of Leo’s language by the end of the study

matches that of the adults. In other words, being a “late talker” didn’t seem to hinder Leo’s progress, so this study provides data that might be used to offset preconceived notions about children’s inability to talk at an early age as a sign of some defect or potential limitation (although of course, that’s not the purpose of this study).

What conclusion about the question does the article draw?

According to Behrens, his study confirms much of what is already known about German-language development (that children “start with nouns and other non-inflected elements”) before incorporating complex structures (such as case marking, for example). Although the input from the adults is consistent and stable over the course of the study Leo’s usage does not start patterning like that of the adults until after about 6 months into the study and beyond, for example:

- | | |
|---|-------------------|
| | LEO’S AGE: |
| • From simple nouns and verbs to more adult-like speech | 2 years, 7 months |
| • Using verbs in a way that approximates adult speech | 2 years, 5 months |
| • Using different types of noun phrases (det+n, etc) | 3 years |

Nonetheless, Behrens concludes that “the data presented here indicate that the distributional properties of adult language indeed exercise a strong power in shaping the child’s language use.” He then backpedals a bit on that conclusion, by suggesting that since Leo’s output “does start differently from the adults, frequency cannot be the only factor to account for language development. From an emergentist perspective, one looks for perceptual, functional, and social factors to explain children’s early language because children have access to more information than just type/token-information (social cues, semantic, and pragmatic cues).”

But none of these factors were looked at in this study, and so it seems odd to bring them up in the conclusion, particularly since this seems like the point he was trying to make in the study—so is he taking issue with his own conclusion, or is he discounting the basic premise of the input-output model? The concluding section of this paper seems to bring in a lot of additional insight not introduced at the start of the paper, so the motivation for the study seems to have gotten muddled (or perhaps it’s just in my own mind) by the end of the paper.

Finally, Behrens suggests that this study provides indirect evidence regarding speed of acquisition, and that “acquisition is slow, when measured against the rich input,” estimating that Leo “has heard more than 550,000 verb tokens” before he actually starts using verbs “just before his second birthday.” In Behrens’ view, it seems this is woefully slow given the amount of (stable) input from the adults.

Do you find the authors’ approach satisfactory? If not, how else would you do it?

As you pointed out during class, the discrepancy between the adult data (input) and Leo’s output during the early stages of the study dispute the notion that language acquisition is data-driven, and so his study doesn’t provide solid evidence in support of the constructionists’ viewpoint. In fact, as you said, the contrast of Leo’s early output compared to that of the adults would provide evidence of just the opposite.

Despite its flaws, what I did like about this paper is that Behrens describes all the setup and preparation for the study in a fair amount of detail—the specifics of the hardware used, setup and testing equipment, a one-month trial run, and so forth, so I found the paper interesting in terms of

how the study was conducted. I also thought the somewhat detailed discussion of the process he and his team used to tag and code the corpus was informative and provides some ideas as to how this type of research might be done, as well as references to various other linguistic tools (the communicative development inventories (CDI), SONIC-Chat, CLAN-programme *mor* tagger, and so on). Mention of these tools in the paper prompted a quick web search, which turned up plenty of resources available for the CHILDES database related tools (SONIC-Chat, *mor*), so I make a note for myself for future reference. (On the other hand, I would have liked maybe a bit more information in the paper about the CDI¹ because that's not freely available. So the role of CDI, what it looks like, how it was used in terms of this study is a bit unclear to me.)

A final comment: just last week I viewed a TED talk given by Patricia Kuhl² about her research into the way infants using statistics in the process of learning their language. This theory was completely new to me (in chatting with Yasamen Shahabi, I've learned that this is her area of research for her thesis, so I realize you're well aware of this research, but the TED talk is an interesting 10-minute diversion). At any rate, I mention this only because this research area (of language acquisition) seems really fascinating, so much more so than the research approach taken in Behrens' paper, so I've now lost some of my initial enthusiasm (with respect to "why do you find this research interesting") for this paper.

¹ "The CDI: Words and Gestures (Infant form) is designed for use with 8- to 16- month old children. The CDI: Words and Sentences (Toddler form) is designed for use with 16- to 30-month old children. Either form may be used with older, developmentally-delayed children. Further information about these forms including normative data is provided in the MacArthur Communicative Development Inventories User's Guide and Technical Manual. The manual as well as the forms can be ordered from Brookes:..." http://www.sci.sdsu.edu/cdi/full_e.htm

² http://www.ted.com/talks/patricia_kuhl_the_linguistic_genius_of_babies.html