

# Environmental and Genetic Influences on Language Development

Connecting environmental and genetic explanations for nonlinear language development in  
infancy and toddlerhood

Federica Bulgarelli<sup>1</sup>

<sup>1</sup>University at Buffalo

[fbulgare@buffalo.edu](mailto:fbulgare@buffalo.edu); 346 Park Hall, Buffalo NY 14260 716-645-0226

Word count: 1493

Acknowledgements: This work was mentored by Alys Doyle.

## Environmental and Genetic Influences on Language Development

Language development is not linear, regardless of the lens through which it is examined. Across both comprehension (i.e. receptive vocabulary) and production (i.e. expressive vocabulary, grammar), development is characterized by slow and gradual progress followed by sudden “spurts”. For example, comprehension is thought to improve drastically around 14 months of age; while infants exhibit knowledge of a few words prior to this time, their word recognition quickly improves after this point [1]. Expressive vocabulary knowledge is also slow at first; typically developing infants go from knowing just a few words to knowing at least 50, but typically over 100, between 18-24 months [2]. Grammatical development follows a similar trajectory. Infants begin producing words in isolation, and as they start learning verbs and prepositions, begin combining them into simple sentences. Exactly what leads to these seemingly sudden, and measurable, changes in language development is of interest to researchers across disciplines.

In the current issue in *Biological Psychiatry*, Verhoef and colleagues [3] investigated the genetic basis of receptive and productive vocabulary at two time points spanning these periods of non-linear growth: early-phase (infancy, 15-18 months) and later-phase (toddlerhood, 24-38 months). They found that while all vocabulary measures spanning 15-38 months were moderately heritable, genetic correlations between early phase expressive and later-phase receptive vocabulary were negligible. Such data suggest the existence of distinct genetic influences operating at these different time periods. Further, while expressive vocabulary both in early-phase and later-phase was genetically linked to literacy related measures, only later-phase expressive and receptive vocabulary were genetically correlated with intelligence and educational attainment across the lifespan. Thus, while both early and later language development milestones are important for lifelong cognitive and educational functioning, Verhoef et al., (2024) suggest that they may stem from distinct genetic bases.

If different sets of common genetic variants influence infant and toddler vocabulary skills, we would predict spurts in language knowledge around the toddler age might be related to mechanisms that may not be available to, or used by, learners until then. This possibility is consistent with the literature on constraints used for language learning. Prior to 18 months of age, children are thought to acquire words through a brute-force approach, learning word-object pairs through slow associative learning mechanisms. Around 18 months, infants are thought to begin making use of learning constraints, such as mutual exclusivity, to launch their word learning. Mutual exclusivity, a learner’s assumption that objects should have one label and labels should refer to one object, can help infants quickly learn new labels for objects for which they lack a label, without needing to have them named directly. Whether mutual exclusivity does not develop until 18 months or is just not used until then has been an open question. Regardless, the fact that learners begin to utilize this mechanism to boost word learning after 18 months could be consistent with the idea of different genetic basis from that which governs associative learning mechanisms supporting earlier word learning [see 4, 5].

Another knowledge gap from the language development literature that is potentially relevant to the idea of distinct genetic influences on language development in infancy vs. toddlerhood relates to Late Talkers. Late Talkers are children 18-30 months of age who exhibit typical receptive language skills but delayed expressive milestones (e.g. do not produce at least 50 words by 24 months); yet, they go on to “catch up” and not receive later language disorder diagnoses. While some demographic factors such as preterm birth and socio-economic status increase the chances of late language emergence at the population level, research to date has not been able to determine exactly why some children exhibit these short-lived expressive language challenges [6]. Given the results of Verhoef and colleagues, Late Talkers could show delays in expressive language skills due to individual differences in when and how genes underlying these skills come online. Incorporating examination of genetic variation into studies of Late Talkers could provide further evidence for this possibility.

## Environmental and Genetic Influences on Language Development

Given evidence that genetic contributions do not explain the entirety of language development, it is also useful to consider the role of environmental experiences. Studies of environmental influences have taken different approaches to quantifying language input and relating language exposure to word learning. A recent meta-analysis found that both language quantity (e.g. number of words) and quality (e.g. vocabulary diversity and syntactic complexity) predicted children's language outcomes [7], suggesting that 20-30% of the variance in language outcomes can be explained by properties of linguistic input. Looking at more specific factors, in a recent analysis predicting the age at which infants will say specific words, we found that how often individual words were heard accounted for 7% of the variance, and how variable those words sounded acoustically accounted for an additional 5% of the variance [8]. Thus, environmental influences have a sizable effect on language outcomes, both when measured at the vocabulary level and at the individual word level.

A potential challenge to the effects of the environment on language development is that it is often impossible to separate genetic from environmental influences. That is, biological parents are typically contributing both genes and language input to their children. Since genes relate to linguistic behaviors (as shown by Verhoef and colleagues), then parents could be more talkative because of their own genes, which they have passed on to their children who will therefore also be more talkative. Coffey et al., (2021) [9] has tried to tease apart the effects of genes and the environment by studying the effects of parental language input for a group of adoptee children, who therefore do not share parental genes. They found that parental language input is still correlated with vocabulary growth, even in the absence of genetic confounds.

Since the environment matters, marked changes in language skills around 18 months could be due to children's environments changing to support new language learning strategies. However, longitudinal analyses of infants' linguistic input suggest remarkable consistency in environmental input over time. Bergelson (2020) [1] showed that from 6-17 months, noun input to infants remained highly stable (though there was high individual variance), regardless of whether the analysis looks at the total number of nouns, how easy it is to identify what object is being referred to, and the syntactic frames that noun occurred in. And yet, infant's in-lab comprehension improved drastically around 14 months. Thus, it is likely something about the learner, rather than the environment, that changes. In line with Verhoef and colleagues (2024), one thing that might change is the underlying biology that supports the ability to make better use of the information at their disposal.

One limitation of the Verhoef et al., (2024) study is that they only included children of European descent. This limitation is common in genomewide studies and the field is actively seeking to address this issue. Nonetheless, the sample composition limits the generalizability of conclusions. The issue of representation has been an increasing focus in the behavioral language development literature as well, as the majority of the field's findings to date have also relied on Western, Educated, Industrialized, Rich and Democratic (WEIRD) samples. Recent efforts to combat this issue have involved extending particularly naturalistic recording methods to less represented communities. As a concrete example, Casillas et al (2021) [10] measured linguistic input to children in a Tseltal village in Southern Mexico, and found that children's linguistic input varied considerably from that of North American infants often studied, in that children rarely received speech directed to them (termed child-directed speech). Despite these differences, many aspects of language development remain consistent, for example children in this sample meet milestones such as production of their first word at approximately the same time as North American infants. Thus, while child-directed speech is thought to be an important factor for language learning in North American samples, it is certainly not required for typical language development. Many different experiences thus support language learning, and these may vary cross-culturally and across racial and ethnic groups. Given that experience and

## Environmental and Genetic Influences on Language Development

genetics interact, we may expect differences in the genes that underly language development at different ages to also vary as a function of ancestry.

In summary, Verhoef et al., (2024) provide novel insights into how different genes may shape language development at earlier and later phases of development. These findings are consistent with the literature on language development showing non linear improvements around toddlerhood across a variety of measurements. Broadly surveying the literature on the effects of the environment on language development highlights the importance of considering both genetic and environmental influences in understanding individual language learning trajectories. While it may be a challenging feat, future research should consider investigating the genetic and environmental bases of language development together, allowing, for example, to elucidate whether variable effects of language input on language outcomes are mediated or boosted by genetics. Combined with more representative samples, these results set the stage for examining an additional factor that may play a role in rapid changes in language development in early childhood.

## Environmental and Genetic Influences on Language Development

Financial Disclosure: Dr. Bulgarelli reported no biomedical financial interests or potential conflicts of interest.

### References:

1. Bergelson, E. (2020). The comprehension boost in early word learning: Older infants are better learners. *Child development perspectives*, 14(3), 142-149.
2. Bloom, P. (2001). Roots of word learning. *Language acquisition and conceptual development*, 159-181.
3. Verhoef, E., Allegrini, A. G., Jansen, P. R., Lange, K., Wang, C. A., Morgan, A. T., et al. (in press): Genome-wide analyses of vocabulary size in infancy and toddlerhood: associations with ADHD, literacy and cognition-related traits. *Biol Psychiatry* XX:XXX-XXX.
4. Markman, E. M. (2014). Constraints on word learning: Speculations about their nature, origins, and domain specificity. In *Modularity and constraints in language and cognition* (pp. 59-102). Psychology Press.
5. Bohn, M., Tessler, M. H., Merrick, M., & Frank, M. C. (2021). How young children integrate information sources to infer the meaning of words. *Nature human behaviour*, 5(8), 1046-1054.
6. Rescorla, L. (2011). Late talkers: Do good predictors of outcome exist?. *Developmental disabilities research reviews*, 17(2), 141-150.
7. Anderson, N. J., Graham, S. A., Prime, H., Jenkins, J. M., & Madigan, S. (2021). Linking quality and quantity of parental linguistic input to child language skills: A meta-analysis. *Child Development*, 92(2), 484-501.
8. Bulgarelli, F., & Bergelson, E. (2023, June 5). Linking acoustic variability in the infants' input to their early word production. <https://doi.org/10.31234/osf.io/su768>
9. Coffey, J. R., Shafto, C. L., Geren, J. C., & Snedeker, J. (2022). The effects of maternal input on language in the absence of genetic confounds: Vocabulary development in internationally adopted children. *Child development*, 93(1), 237-253.
10. Casillas, M., Brown, P., & Levinson, S. C. (2020). Early language experience in a Tzeltal Mayan village. *Child Development*, 91(5), 1819-1835.