

Speech, gesture and the origins of language

WILLEM J. M. LEVELT

Max-Planck-Institut für Psycholinguistics, Wundtlaan 1, 6525 Nijmegen, The Netherlands. E-mail: pim@mpi.nl

During the second half of the 19th century, the psychology of language was invented as a discipline for the sole purpose of explaining the evolution of spoken language. These efforts culminated in Wilhelm Wundt's monumental *Die Sprache* of 1900, which outlined the psychological mechanisms involved in producing utterances and considered how these mechanisms could have evolved. Wundt assumes that articulatory movements were originally rather arbitrary concomitants of larger, meaningful expressive bodily gestures. The sounds such articulations happened to produce slowly acquired the meaning of the gesture as a whole, ultimately making the gesture superfluous. Over a century later, gestural theories of language origins still abound. I argue that such theories are unlikely and wasteful, given the biological, neurological and genetic evidence.

Psychology and the genesis of language

The evolution or genesis of spoken language has always been a core concern for students of the psychology of language, or psycholinguists as they have been calling themselves since the 1960s. In fact, the psychology of language was invented by linguists to account for the evolution of language. Levelt¹ described this state of affairs as follows:

When, after the period of Idealism, logic lost its ground as a basis for the understanding of grammar, linguists began to view language as a spontaneous product of nature. Language is a natural product of consciousness, according to leading nineteenth century linguists such as Herder, Steinthal, and Lazarus. The great challenge for such linguistics was to explain the origin, structure and variety of languages as arising from an 'Ursprache', spontaneously generated by human awakening consciousness. That romantic, evolutionary perspective shifted the weight of explanation in linguistics to psychology. 'Glückliche Fortschritte in der Sprachwissenschaft setzen eine entwickelte Psychologie voraus' (Fortunate

advances in linguistics presuppose a developed psychology), wrote Heymann Steinthal in 1855.²

Knoblauch,³ in his thorough account of this early history, shows that, regrettably, such a ‘developed psychology’ simply didn’t exist. Steinthal had little else than Herbart’s atomistic individual psychology to refer to, and the occasional self-proclaimed psychologist of his time was not interested in this kind of challenge. In other words, the ‘psychology of language’ was invented by linguists *pour le besoin de la cause* – it was little else than linguistic folk psychology.

But matters changed when Wilhelm Wundt took up the challenge, beginning in the 1880s and eventually resulting in his two-volume *Die Sprache* of 1900.⁴ For Wundt too, the principal aim was to explain spoken language and languages as a spontaneous product of consciousness, as an expression of mind. The ultimate linguistic explanandum was the phylogenesis of language and Wundt set out to create a theory of microgenesis, the generation of words and sentences in the individual mind, to provide the theoretical foundation. Modern psychology of language began with a theory of ‘speaking’. The deep motivation for that theory was to provide the psychological foundations for an explanation of the evolutionary origins of language.

Historically speaking, therefore, the psycholinguist’s primary ‘window on language genesis’ was the microgenesis of utterances in the speaker’s mind. This window has not lost any of its relevance since Wundt’s time, but it certainly deserves more attention next to the other psycholinguistic perspectives, the genesis of language in the child’s mind and the online genesis of interpretation (intention recognition) in the listener’s mind. The generation of utterances, i.e. the act of speaking, psychologically involves several levels of processing, which operate in parallel and with relative autonomy.^{5,6} At the primary, conceptual level the speaker retrieves and unfolds information whose expression can effectively transmit some communicative intention to an interlocutor. For Wundt, this gradual mental unfolding of predications is a centrepiece of his theory. It is his major effort to explain the online generation of syntax from a semantic/conceptual base. Here he was the first to use phrase structure diagrams as we still do today, a notion he himself introduced in linguistics.⁷ From the evolutionary point of view, this successive attentional focusing on the whole idea (*Gesamtvorstellung*) and gradually on its constituents and sub-constituents in their mutual relations (which Wundt called ‘apperception’) must have been an ability essential to the genesis of language. When brain damage limits this attentional apperceptive span, the generation of complex utterances is indeed severely limited.⁸ The mental unfolding of simple predications (though hardly in syntactic patterns) is most likely within a chimpanzee’s abilities.⁹ A core issue for any theory of language

origins is how early humans developed a spoken medium for communicating their intentions.

This spoken medium corresponds to another level of processing in the speaker. Speakers generate overtly articulated utterances, consisting of words and phrases that systematically express the intended information. An ultimate step here is to prepare a ‘syllabification’ over successive words. This syllabification is indeed a rather autonomous process, which, in many languages, easily ignores the integrity of successive words (as for instance in *he’ll-bur-nit* for ‘he will burn it’). The speaker’s incremental encoding of the utterance’s articulatory shape is a process of high complexity and impressive speed.¹⁰ Among primates this is unique to *Homo sapiens* and hence a particular challenge to any theory of language genesis. It has always amazed me that, since Wundt and even earlier, the dominant response to this challenge has been to ignore it and to argue that language evolved in a gestural mode.

Gestural origins of language – a persistent misconception

Let us consider, as a typical early example of such reasoning, Wilhelm Wundt’s grand tour-de-force. It was to explain the phylogeny of language from a universal, original sign language. The temptation was, admittedly, irresistible. Spoken languages are all different, the apparent outcome of an extensive cultural evolution. But sign language was still one and undivided. Wundt’s informants had convinced him (counterfactually as we now know) that all deaf communities essentially use the same, primitive gesturing system for communication. A universal fallback system, not a full language, but one that can serve in basic, pragmatic communication. This universality of sign language cannot be an accident, Wundt must have thought; it must have been at the origin of language.

This was certainly not an original idea. In his 1644 book, Bulwer¹¹ called gesturing the ‘onlely speech that is naturall to Man’. He had many followers, among them Jean-Jacques Rousseau.¹² In fact, the idea of a gestural origin of language has been kept alive to the present day. Prominent recent examples are Corballis¹³ and Arbib.¹⁴ I keep being surprised by this need to use gesture as a *deus ex machina* for the explanation of spoken language genesis. Occam’s razor can be very helpful in clearing the air.

First, assume you have explained the evolution of a gestural language. The next step to take is the explanation of spoken language genesis. This will, among other things involve the complete knocking out of a specialized manual articulatory system and the creation of a vocal articulatory system with its phonological base. It is like building a car by first building a ship and then removing the sail, putting it on wheels and adding a combustion engine. This is the recurring unsolved problem for the gesturalists. Wilhelm Wundt tells us about articulatory gestures

as expressive movements, which are part of the larger, meaningful bodily gesture. They are not there to produce any specific sound; rather the sound is just a consequence of this articulatory component of the gesture as a whole. As a consequence, the speech sound has only an indirect relation to the meaning expressed in the total gesture, etc, etc. But the real work to be done in such a theoretical framework is to explain causally the construction of a spoken, self-sufficient and quite arbitrary lexicon from the primitive, pre-existing meaningful gestural lexicon. Since that work, i.e. explaining the genesis of a spoken vocabulary, will have to be done anyhow, why do it twice, first building a flexible gestural lexicon, then building a totally different vocal lexicon? Modern gesturalists, just like Wundt, neither explain why the once universally available, communicatively entirely useful gestural languages fell into equally universal disuse, nor do they take up McNeillage's¹⁵ challenge of presenting a detailed account of the ultimate evolution of speech, as if that final step comes for free. Rather they keep arguing that signing is our more primitive, original form of communication. McNeill,¹⁶ in discussing Goldin-Meadow's¹⁷ discovery of spontaneous development of a gestural communication system among deaf children, concludes as follows: 'It is consistent with the theory that vocal speech overlies an ability of manual communication. When vocal speech is blocked the original system remains'. This, I have argued,¹⁸ is a non-sequitur. It is equally consistent with other theories, notably the one that claims that if the vocal-auditory channel is blocked, the most flexible alternative will be used for linguistic communication.

Second, the selective advantage of a gestural language is unconvincing. Such a language is functionally dead during, on average, 12 hours a day. It is of little use for gatherers in bushy areas, as our ancestors were. It heavily interferes with tool use, harvesting, cooking and other important daily routines that involve the use of hands. Not a practical system at all.

Third, and most important, *Homo sapiens* had a biological endowment that provided enormous advantage over all other primates: it had cortical control over vocalization. This important argument is usually ignored in the current literature. The classical work by Ploog^{19,20} and Jürgens^{21,22} has shown that there are two systems in our brains that control vocal behaviour. The old system, which we share with all other primates and land-living vertebrates, comprises limbic structures, all of which funnel into the peri-aqueductal grey of the midbrain. It allows for emotional vocalization, a repertoire of vocal signals relating to specific motivational states. But it lacks cortical control. The second, new system, is the neocortical voice pathway as part of the pyramidal tract. It began to emerge in primates but became qualitatively different in *Homo sapiens*. It provided man with full voluntary control of the voice and of articulation. This speech system kept working in tandem with the old 'emotional' system, which created the wonderful

ability to express voluntarily a wide spectrum of emotional states by using the voice. This is not yet language, of course. The argument, rather, is that this second system must have provided us, as social animals, with substantial selective advantage.

The recent discovery of a FOXP2 gene with a specifically human mutation²³ adds to this argument in the following way. The gene is a transcription factor which, among many others, affects genes responsible for the development of pathways involved in the control of fine orofacial movements, including fine articulation. The authors argue that the mutation became widespread some 200,000 years ago or shortly after, i.e. when modern humans emerged; the genetic evidence suggests that this occurred under selective pressure. In other words, our ancestors had high stakes in the voluntary control of precise articulation. Apparently, speech, rather than gesture, became the chosen medium for linguistic communication in early *Homo sapiens*.

These and other arguments should focus our attention on where the biological action really is, namely in linking a conceptual/semantic system to the vocal-auditory communicative domain. In ontogeny these two systems initially develop quite independently, involving different regions in the maturing brain. During the babbling phase, from the seventh month onwards, there is sudden maturation of basic syllabic gestures, such as gi-gi and da-da. These babbles are entirely meaningless. This articulatory maturation is achieved through continuous audio-motor feedback, but without reward for correct imitation (this is in strong contrast with Wundt's⁴ claim that first articulations are pure imitations). Only by the end of the first year, does this syllabic repertoire get tuned to the target language.¹ This is during the period that the infant starts recognizing often recurring words in the speech stream.²² For the child these words are initially as meaningless as his own babbles. But the child now has the capacity to generate a set of possible words or proto-words. It is still a long way (a few years for the child) to develop this phonological system in full detail and to link such possible words to a system of meanings; it is an even longer route to enrich them with syntactic properties, but the balls to play with are all there.¹ Similarly, these balls were there for our ancestors. These are the biological endowments we should consider head-on when we meditate about the early evolution of language.

References

1. W. J. M. Levelt (1968) The genetic perspective in psycholinguistics. Or: where do spoken words come from? *Journal of Psycholinguistic Research*, **27**, 167–180.
2. H. Steinthal (1855) *Grammatik, Logik und Psychologie, ihre Prinzipien und ihr Verhältnis zueinander* (Berlin. Photographic reproduction (1968): Hildesheim: Georg Olms Verlag).

3. C. Knoblauch (1988) *Geschichte der psychologischen Sprachforschung in Deutschland von 1850 bis 1920* (Tübingen: Niemeyer).
4. W. Wundt (1900) *Die Sprache*. 2 vols (Leipzig: Engelmann).
5. W. J. M. Levelt (1989) *Speaking: From Intention to Articulation* (Cambridge, MA: MIT Press).
6. W. J. M. Levelt (1995) The ability to speak: from intentions to spoken words. *European Review*, **3**, 13–23.
7. W. Wundt (1880) *Logik. Eine Untersuchung der Prinzipien der Erkenntnis und der Methoden wissenschaftlicher Forschung* (Stuttgart).
8. H. H. J. Kolk (2001) Does agrammatic speech constitute a regression to child language? A three-way comparison between agrammatic, child, and normal ellipsis. *Brain and Language*, **77**, 340–350.
9. D. Premack (1977) *Intelligence in Ape and Man* (New York: Halsted).
10. W. J. M. Levelt, A. Roelofs and A. S. Meyer (1999) A theory of lexical access in speech production. *Behavioral and Brain Sciences*, **22**, 1–75.
11. J. Bulwer (1644) *Chirologia: on the Natural Language of the Hand* (London).
12. J. J. Rousseau (1782) *Essay sur l'origine des langues* (Geneva).
13. M. C. Corballis (2002) *From Hand to Mouth* (Princeton: Princeton University Press).
14. M. A. Arbib (2003) From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences*, **26**, 199–260.
15. P. F. MacNeilage (1998) The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences*, **21**, 499–546.
16. D. McNeill (1979) Language origins, in: M. von Cranach, K. Foppa, W. Lepenies, & D. Ploog (eds), *Humanethology* (Cambridge: Cambridge University Press).
17. S. Goldin-Meadow and H. Feldman (1975) The creation of a communication system: a study of deaf children of hearing parents. *Sign Language Studies*, **8**, 226–236.
18. W. J. M. Levelt (1979) The origins of language and language awareness, in: M. von Cranach, K. Foppa, W. Lepenies & D. Ploog (eds), *Humanethology* (Cambridge: Cambridge University Press).
19. D. Ploog (2003) Is the neural basis of vocalization different in non-human primates and Homo sapiens? *Proceedings of the British Academy*, **106**, 121–135.
20. D. Ploog (1986) Biological foundations of the vocal expression of emotions, in: R. Plutchik and H. Kellerman (eds) *Emotion: Theory, Research and Experience. Vol 3, Biological Foundations of Emotion* (New York: Academic Press).
21. U. Jürgens and D. Ploog (1985) On the neural control of mammalian vocalization, in: E. V. Evarts, S. P. Wise and D. Bousfield (eds) *The Motor System in Neurobiology* (Amsterdam: Elsevier).
22. U. Jürgens (2003) From mouth and hand to hand: on language evolution, *Behavioral and Brain Sciences*, **26**, 229–230.
23. W. Enard, M. Przeworski, S. E. Fisher, C. S. L. Lai, V. Wiebe, T.

- Kitano, A. P. Monaco and S. Pääbo (2002) Molecular evolution of FOXP2, a gene involved in speech and language. *Nature*, **418**, 869–872.
24. D. M. Houston, P. W. Jusczyk, C. Kuijpers, R. Coolen and A. Cutler (2000) Cross-language word segmentation by 9-month-olds. *Psychonomic Bulletin & Review*, **7**, 504–509.

About the Author

Willem Levelt is founding director of the Max Planck Institute for Psycholinguistics in Nijmegen, The Netherlands, and honorary Professor of psycholinguistics at Nijmegen University. His books include *On Binocular Rivalry* (1965), *Formal Grammars in Linguistics and Psycholinguistics* (3 Vols, 1974) and *Speaking: from Intention to Articulation* (1989). He is President of the Royal Netherlands Academy of Arts and Sciences and a member of various academies, among them the Academia Europaea and the US National Academy of Sciences.

