Multiword Sequences as Building Blocks for Language: Insights into First and Second Language Learning

Moderator: **Morten H. Christiansen** (christiansen@cornell.edu)
Department of Psychology, Cornell University, USA

Moderator: **Inbal Arnon** (inbal.arnon@gmail.com)
Department of Psychology, University of Haifa, Israel

Elena Lieven (elena.lieven@manchester.ac.uk)
School of Psychological Sciences, University of Manchester, UK

Alison Wray (wraya@cardiff.ac.uk)
School of English, Communication & Philosophy, Cardiff University, UK

Keywords: formulaic language, language acquisition, language processing, second language learning, usage-based learning

Introduction

Many grammatical frameworks view words and rules as the basic building blocks of language, with multiword sequences being treated as peripheral exceptions in the form of idioms, etc. (e.g., Pinker, 1999). The new millennium, however, has seen a shift toward construing multiword sequences not as linguistic rarities but as important building blocks for language acquisition and processing. Based on a growing bulk of evidence of sensitivity to multiword sequences in language learning and use (see Ellis, 2012, for a review), multiword sequences have come to figure prominently in many current approaches to language, including item-based learning (Lieven, 2010), formulaic language (Wray, 2008), usage-based language processing (Arnon & Snider, 2010), and chunk-based learning (McCauley & Christiansen, in preparation). This symposium brings together experts from these different approaches to language to explore the idea that first (L1) and second (L2) language learners differ with respect to their ability to use multiword building blocks to learn and process language, and that this difference affects learning strategies and outcomes.

Unlike young children, adult learners rarely reach native proficiency in pronunciation, morphological and syntactic processing, or the use of formulaic language and idioms (see Ellis, 2012, for a review). Yet adults do not have problems with all aspects of novel language learning: they seem to learn certain aspects of language (e.g., words) better than others (e.g., grammatical relations, formulaic expressions). Existing accounts of the differences between L1 and L2 language learning have tended to focus on biological, cognitive, and neural differences between children and adults. These accounts predict the general difference in proficiency between the two populations, but struggle to explain the specific patterns of language learning observed in children and adults.

Understanding the different paths and outcomes of L1 and L2 learning has wide-reaching implications for cognitive science in terms of what it means to know a language, how much of such knowledge is 'built-in', and how learning changes as a function of prior knowledge and experience. Crucially, while L1 acquisition, adult psycholinguistics, and L2 learning are often studied separately, we bring together insights from developmental psychology (Lieven), psycholinguistics (Arnon), computational investigations of language structure (Christiansen), and applied psycholinguistics (Wray) to present a diverse and rich perspective on multiword building blocks in language learning and use.

The symposium participants have all worked extensively on language acquisition and use. Lieven has been at the forefront of developing the usage-based approach to language learning and has conducted numerous studies on the nature of children's early language use and representation. Arnon has been studying both the processing of multiword sequences by adult native speakers and the way chunk-based learning can impact adult performance in artificial languages. Christiansen has conducted extensive psycholinguistic and computational work exploring the units of language learning and the way such units affect learning. Wray has worked broadly on formulaic expressions in both native and non-native speakers as well as more recently in the language of Alzheimer's patients. Together, the participants have published more than 70 papers relating to the role of multiword sequences in language.

Lieven: Multiword Sequences in L1 Acquisition

Theoretical and empirical reasons suggest that children build their language not only out of individual words but also out of multiword strings. These are the basis for the development of schemas containing slots. The slots are putative categories which build in abstraction while the schemas eventually connect to other schemas in terms of both meaning and form. Evidence comes from the nature of

the input (Cameron-Faulkner et al., 2003); the ways in which children construct novel utterances (Lieven et al., 2009); and the computational modeling of children's grammars (Bannard et al., 2009). However, nearly all this research is on English which is unusual in its rigid word order and impoverished inflectional morphology. There has also been much less research on the development of the 'meaning pole' in the form-meaning mappings of schemas. I will address both these issues using our recent studies in English, German, Polish and Chintang.

Arnon: Multiword Sequences in Adult Language Learning and Use

Prior studies have shown that native speakers are sensitive to the distributional properties of multiword sequences when processing language (see Ellis, 2012, for a review). Results are presented suggesting that this sensitivity also extends to language production and is not modulated by syntactic constituency: higher frequency phrases are phonetically reduced within and across syntactic boundaries indicating the prominence of sequence based information. A second study investigated whether such sensitivity to multi-word sequences might be harnessed to improve L2 learning. Adult learners showed better learning of an artificial language incorporating a grammatical gender system when first exposed to larger chunks (sentences) and only then individual words (noun-labels). This result suggests that L2 learning of grammatical gender languages may be improved by initially exposing learners to multiword sequences instead of isolated words, thus mirroring the sensitivity to multiword sequences in L1 acquisition and use.

Christiansen: Computational Investigations of Multiword Chunks in Language Learning

Computational modeling provides further means to investigate the use of multiword chunks by different types of language learners. The Chunk-Based Learner (CBL; McCauley & Christiansen, in preparation) gradually builds an inventory of chunks—consisting of one or more words—used for both language comprehension and production. The model learns incrementally from corpora of child-directed speech using simple distributional information and accommodating a range of developmental findings. Results are presented indicating that multiword chunks provide a useful basis for capturing children's productions across a number of different languages independent of their word order. When applied to L2 learner corpora, CBL reveals that the productions of such speakers rely less on multiword chunks compared to speech of both L1 learners and adult native speakers. Thus, these modeling results corroborate our hypothesis about the differential use of multiword building blocks by L1 and L2 learners.

Wray: Formulaic Expressions: Further Issues

Why have we not progressed further than we have, in understanding the role of formulaic sequences in L2 learning? This presentation will consider how certain assumptions underpinning the existing body of knowledge could constrain the research questions we ask. For instance, how safe is the assertion that non-native speakers rarely achieve nativelikeness (typically attributed to not mastering formulaic sequences)? How appropriate is it to gauge the formulaic language knowledge of adult L2 learners by comparing it to what, for native speakers, is anchored in the social and cognitive experiences of childhood? To what extent can we assert that (all) L1 speakers know the same things about how words fit together? How do recent proposals by Hanks (2013), Port (2007) and Sinclair (Cheng et al., 2009) that the word is not a reliable unit of form or meaning impact on the growing evidence that multiword strings might be?

Symposium Format

The symposium starts with a 5-minute introduction, followed by four 20-minute presentations (including time for questions), and concludes with a 15-minute general discussion.

References

Arnon, I. & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, *62*, 67-82.

Bannard, C., Lieven, E. & Tomasello, M. (2009) Modeling children's early grammatical knowledge, *Proceedings of the National Academy of Sciences, 106*, 17284-17289.

Cameron-Faulkner, T., Lieven, E. & Tomasello, M. (2003). A construction based analysis of child directed speech. *Cognitive Science*, *27*, 843-873.

Cheng, W., Greaves, C., Sinclair, J. M. & Warren, M. (2009). Uncovering the extent of the phraseological tendency: Towards a systematic analysis of concgrams. *Applied Linguistics, 30*, 236-252.

Ellis, N.C. (2012). Formulaic language and second language acquisition: Zipf and the phrasal teddy bear. *Annual Review of Applied Linguistics*, *32*, 17-44.

Hanks, P. (2013). *Lexical analysis: Norms and exploitations.* Oxford: Oxford University Press.

Lieven, E. (2010). Input and first language acquisition: Evaluating the role of frequency. *Lingua*, *120*, 2546-2556.

Lieven, E., Salomo, D. & Tomasello, M. (2009). Two-year-old children's production of multiword utterances: A usage-based analysis. *Cognitive Linguistics, 20*, 481-508.

McCauley, S.M. & Christiansen, M.H. (in preparation). *Language learning as language use: A computational model of children's language comprehension and production.*

Pinker, S. (1999). *Words and rules.* New York: Basic Books.

Port, R. F. (2007). How are words stored in memory? Beyond phones and phonemes. *New Ideas in Psychology,* 25, 143-170.

Wray, A. (2008). *Formulaic language: Pushing the boundaries*. Oxford: Oxford University Press.