**Session: One size fits all?**

The notions of “chunk” and “chunking” have become well-established in language processing and acquisition research. Chunks typically refer to multi-word units or partially abstract phraseological patterns. Examples discussed in the literature range from n-grams and their variations, like lexical bundles (Biber & Conrad 1999), fixed expressions (Moon 1998), concgrams (Greaves & Warren 2010), or PoS-grams (Stubbs 2007), to more diverse patterns like collocational frameworks (Renouf & Sinclair 1991), units of meaning (Sinclair 2004), formulaic sequences (Wray 2005), collostructions (Stefanowitsch & Gries 2003), and constructions (Goldberg 2006).

What all these takes on chunks have in common is their focus on repeated occurrences of form-meaning pairings in language use. In other words, they are concerned with analysing the output of past production processes as attested, for example, in corpora. Moreover, they assume very similar chunk sizes: larger than a word, but shorter than a clause.

By contrast, the focus of the present theme session departs from this tradition in two key aspects. First, all three papers focus on how perceivers segment the continuously incoming stream of auditory or written language input into chunks in real time – that is, they adopt a process perspective focusing on online comprehension. Second, the papers cover processing units both smaller than 4-5 words (e.g., chunks at the sublexical level) and larger than that (e.g., chunks at the clause level).

The session will be concluded with a general discussion focusing on theoretical and empirical issues arising from the papers: Is the incoming language stream segmented into chunks at multiple levels (from morphological via lexical and syntactic up to discourse) in parallel? If so, how do chunks of different grain size interact? Which variables can influence the segmentation behaviour of language comprehenders, and by which methods can online chunking be measured? Overall, this theme session hopes to arrive at a more nuanced and comprehensive understanding of the size of natural language processing units.

**References:**


