Dagmar Divjak (University of Birmingham)

Title
Behavioural Profiles: looking back, looking forward

Abstract
Theoretically grounded in usage-based models of language, Behavioural Profiling (Divjak & Gries 2006, Divjak 2006, Gries 2006) is a technique that helps analysts extract patterns supporting intricate meaning differences between linguistic units of any size from large amounts of data.

Behavioural Profiles are in essence manually created sparse vector representations of the context in which a linguistic structure of any size is used. Context may be described through various morphological, syntactic, semantic and lexical features, coded for each token of interest, to be used for comparing the distributional properties of the targeted linguistic item. Crucially, BPs are dynamic: they rely on co-occurrence counts between the linguistic features of forms encountered in a specific context such as tense, aspect, voice, number, case, etc. Furthermore, quantifying context based on features rather than mere token co-occurrence means that BPs use corpus data on multiple levels at once, focussing on the correlation between distribution patterns and functional characteristics of linguistic tokens. This data-heavy approach sets BPs apart from the more traditional strands of corpus linguistics, but also from descriptive linguistics which has historically relied on qualitative data and introspective methods. However, similarly to these approaches, BPs incorporate linguistic features. BPs have therefore found a very broad take-up in theoretical and applied linguistics, becoming a key method in the field (McEnery & Hardie 2012; Bębeniec 2024). They have been used in diachronic and synchronic work with a semantic focus, within and across languages, including translation studies, but they have also revealed valuable insights into morphology and syntax (for an overview, see Janda 2016), where they are known as "grammatical profiles".

Introduced now nearly 2 decades ago (Divjak 2004), it has gained in popularity and has yielded many interesting applications to linguistic phenomena within and across languages, synchronically and diachronically. In this talk I will focus on the kind of variations these applications have introduced, and discuss the extent to which they fit with the theoretical underpinnings of Behavioural Profiling. I will also look ahead to Large Language Models which are powered by so-called embeddings. Embeddings can be seen as a continuous variant of the categorical BPs. While BPs are discrete (encoding presence versus absence) and sparse (manual annotation limits the number of features that can be included), embeddings are real-valued and dense, and typically one or two orders of magnitude larger than standard BPs. Besides, it is assumed that embeddings package world knowledge (Grand et al. 2022), not encoded by BPs. Nevertheless, BPs have been shown to surpass LLMs in detecting shifts in meaning (Guilianelli et al. 2022), indicating their superior ability to capture fine-grained morphological and syntactic signals.