Language use varies according to a number of factors, from pragmatic and cognitive to social. For on-line production it has been shown that specific forms of variation directly serve rational communicative goals by offering ways to modulate information density (Jaeger/Levy 2007); and for comprehension, there is ample evidence that particular linguistic choices are associated with specific levels of surprisal (Levy 2008; Schulz et al. 2016; Delogu et al. 2017; Sikos et al. 2017). What is much less clear is what the communicative effects of interaction are, if any, i.e. of particular linguistic choices recurring across interactants, interaction instances and contexts. Spontaneously occurring linguistic co-adaptation or convergence among interactants in on-line situations is a widely studied phenomenon (see e.g. Coles-Harris (2017) for the phonetic level) but communicative effects have so far only rarely been discussed (see e.g. Hume/Mailhot (2013) for phonological effects). Similarly, conventionalization, i.e. the longer-term linguistic effects of repeated interaction, has hitherto hardly been considered from a communicative perspective. Widely acknowledged as a relevant process in language change (Schmid 2015), conventionalization provides a prerequisite for innovation (de Smet 2016) and leads to persistent change in the language system overall (Bybee/Hopper 2001) and, when pertaining to particular socio-cultural contexts, to the formation of varieties (registers, dialects (Ure 1982; Trudgill 2008).

We suggest here that conventionalization is a cornerstone in communication since it comes with significant surprisal and entropy-reducing effects (Harris 1991, 2002). To show this, we pursue an exploratory, corpus-based approach, focusing on scientific writing (Degaetano-Ortlieb/Teich 2018, Degaetano-Ortlieb/Teich forthcoming), a well-studied and fairly controlled domain, and its evolution across 300 years from the mid-17th century onwards. The data we use are the Proceedings/Transactions of the Royal Society of London. To capture lexical and syntactic aspects of linguistic change leading to conventionalization, we employ probabilistic language models (word and part-of-speech based n-gram models, topic models, word embeddings); and to evaluate the observed effects, we apply various measures of information content (surprisal, entropy, relative entropy). We find for instance that diachronically, within the scientific domain, relative entropy on n-gram models overall decreases, and topic-document entropy and entropy over word embedding clusters also go down, thus indicating conventionalization at both lexical and syntactic levels. For qualitative interpretation, we inspect the linguistic features that significantly contribute to these trends, resulting in stable average surprisal and low entropy on those linguistic choices that become characteristic of scientific language over time.