Exploring formulaic language in 17th-century Dutch-language newspaper articles

Leonardo Zilio and Paul Arblaster

Journalism is one sphere in which formulaic language is common, since it creates communicative shortcuts that enable news to be conveyed quickly and to be easily fitted into existing mental categories. Europe's first weekly newspapers began to be published in the early seventeenth century: in Germany from 1605, the Dutch Republic from 1618, and the Spanish Netherlands and England from 1620. These newspapers often reported the same news, sometimes on the basis of the same newsletters or simply by copying it from one another. The focus was on great public events: movements of fleets and armies, military engagements, arrivals and departures of ambassadors, the public life of royal families, and the publication of decrees or proclamations.

While the weekly production of thousands of words of text should in theory create a massive multilingual corpus, in practice survivals are patchy and unpredictable, and do not always overlap. One of the best-preserved news series of the early seventeenth century is that of the *Nieuwe Tijdinghen*, published in Antwerp in the years 1620-1629 (see Arblaster 2024). Its unusually good survival is due to some owners having issues bound into annual volumes to keep a year's overview of the news (Pettegree 2015). Between collections in Amsterdam, Antwerp, Ghent, The Hague, London, and above all the Royal Library of Belgium in Brussels, most issues survived and have been collectively catalogued both digitally [USTC N4-1 to N4-1460] and in print (Der Weduwen 2017). Many issues, particularly those held in Antwerp, Brussels and Ghent, have now been made available online, but only as images rather than text.

This study uses a relatively small sample, mostly transcribed in the late 1990s by Paul Arblaster while working on his doctoral thesis, which after revision was published as *From Ghent to Aix: How They Brought the News in the Habsburg Netherlands, 1550-1700* (Arblaster 2014). With the aid of computational tools, we investigate linguistic features of this 51k-word corpus, focusing specifically on recurrent expressions, terminology and formulaic language. We use corpus linguistics tools, such as AntConc (Anthony 2005) and Sketch Engine (Kilgarriff et al. 2008) to produce an initial overview of n-grams, collocations, keywords and multi-word terms, to then further organise and expand these lists using a semi-automatic approach.

The historical writing that is present in these documents means that the original lists extracted with automatic tools need to be manually validated to some extent, as many words that are not present in modern corpora will result in false positives for the analysis, for instance, of keywords and terms. After a manual revision, Python scripts are employed to search for variant spellings by using edit-distance algorithms. We also briefly test whether tools based on generative artificial intelligence can help in detecting textual regularities and terminology in these historical documents. The resulting lists of expressions and terms are then further contextualised with the help of concordances and example-sentences to form a repository of the formulaic language that is present in these historical newspaper articles.

References

Anthony, L. (2005, July). AntConc: design and development of a freeware corpus analysis toolkit for the technical writing classroom. In *IPCC 2005. Proceedings. International Professional Communication Conference*, 2005. (pp. 729-737). IEEE.

Arblaster, P. (2014). *From Ghent to Aix: How They Brought the News in the Habsburg Netherlands, 1550-1700*. Leiden: Brill.

Arblaster, P. (2024). Las noticias publicadas por Abraham Verhoeven en Amberes en 1621, in Manuel Borrego and Carmen Espejo-Cala (eds.) *El mundo en 1621: Avisos, relaciones de sucesos, conexiones culturales*. Besançon: Presses universitaires de Franche-Comté, 139-161.

Der Weduwen, A. (2017). *Dutch and Flemish Newspapers of the Seventeenth Century, 1618–1700*, vol. Leiden: Brill, 334-417.

Kilgarriff, A., Rychly, P., Smrz, P., & Tugwell, D. (2008). The sketch engine. *Practical Lexicography: a reader*, 297-306.

Pettegree, A. (2015). Tabloid Values: On the Trail of Europe's First News Hound, in Richard Kirwan and Sophie Mullins (eds.) *Specialist Markets in the Early Modern Book World*. Leiden: Brill, 15-34.

Short academic bios

Leonardo Zilio is an assistant professor in translation and interpreting technologies at the Université catholique de Louvain. He obtained his PhD degree in linguistics at the Federal University of Rio Grande do Sul, Brazil, in 2015 and later developed international post-doctoral research in Belgium, England and Germany. His main research interests lie in the interface between computational linguistics, digital humanities, translation and terminology.

Paul Arblaster, DPhil (Oxon), is a historian and translator who teaches in the Marie Haps Faculty of Translation and Interpreting, UCLouvain Saint-Louis Brussels, and the Louvain School of Translation and Interpreting. His research focuses on transnational communication, translation, persecution and religious mobility in early-modern Europe. He has also written a general *History of the Low Countries* (third edition 2018).