

Verbalization tasks in sensometrics: a review of the scientific bibliography through bibliometrics

Mónica Bécue-Bertaut ^{1,*}

1. Universitat Politècnica de Catalunya, Barcelona (Spain)

* Contact author: monica.becue@upc.edu

Keywords: Verbal-based techniques; Bibliometrics; Textual data; Correspondence analysis; Counts data.

In response to a demand for less time consuming methods, verbal based techniques are increasingly used in sensorial evaluation either as a complement to similarity-based methods such as napping or free sorting task or as a full-method such as flash profile, check-all-that apply (CATA) or open-ended questions.

The novelty of the data —counts and not ranking or intensity measures—gives rise to discussions about pros and cons, number and type of assessors (consumers or trained panellists), application domains, analysis methods, reliability and validity, interpretation and so on. In this work, we aim at showing how bibliometrics can bring a support to such discussions.

Bibliometric studies have been increasing in the recent years because of the availability of electronic data bases collecting the scientific literature and the diffusion of efficient methods for statistical analysis of texts (Bansard et al. 2007; Ohsumi, 2009).

Among these methods, correspondence analysis (CA; Benzécri, 1981; Lebart et al, 1998) plays a relevant role due to its potentiality to organize and synthesize a large amount of texts. In this case, CA starts from the table crossing abstracts and words. Through CA, the words are organized into a lexical space such as two words are all the more closer that they are used in similar contexts. Symmetrically, the abstracts are mapped into a space such as two abstracts are all the more close (distant) that their content is similar (different). Similar contents can be expressed through different words because CA is able to take advantage of synonymy relationships. These relationships do not need to be introduced as an external information through ontologies but are automatically assessed by the own CA. Passing from the abstracts space to the words space through the classical CA transition relationships allows for identifying the vocabulary responsible of similarities or differences among the abstracts.

When the corpus is chronological, multiple factor analysis for contingency table (MFACT; Bécue & Pagès, 2008) allows for introducing the chronology into the analysis and underlining the changes observed in the scientific domain.

In this work, we analyze the abstracts published in the food science journals available at the Web of Science referring the use of verbalization tasks in the evaluation of sensory characteristics of products. Each abstract is completed with its title, keywords, authors, journal and publication year. This analysis offers a starting point to establish the state of the question as well as on the different collection modes, the new statistical methods that are conceived or introduced from another field and the type of results that they provided.

References

Bécue-Bertaut, M. & Pagès, J., 2008. Multiple Factor Analysis and Clustering of a Mixture of Quantitative, Categorical and Frequency Data. *Computational Statistics and Data Analysis*, 52, 3255-3268.

Bansard, J-Y., Rebholz-Schuhman, D., Cameron, G., Clark, D., van Mulligen, E., Beltrame, F., del Hoyo Barbolla, E., Martín-Sánchez F., Milanese, L., Tollis, I., Van der Lei, J. & Coatrieux, J-L. (2007). Medical informatics and bioinformatics: a bibliometric study. *IEEE Trans Inf Technol. Biomed.* 11, 237-243.

Benzécri, J.P., 1981. *Pratique de l'analyse des données*, Vol. 3, *Linguistique & Lexicologie*. Dunod, Paris.

Lebart, L., Salem, A. & Berry, L.(1998). *Exploring textual data*. Kluwer, Dordrecht.

Ohsumi, N. (2009). Mining of textual data, recent tend and its direction. http://wordminer.comquest.co.jp/wtmtips/pdf/20090910_1.pdf