Corpus-based terminography: popularizing discourse and the role of communicative contexts in specialised corpora building

The lack of, or reduced, understanding of the consumer, as far as food items with health claims that are available on the market are concerned has led to the proposal of a terminological database targeted to non-experts, with the aim of making scientifically valid and accurate information on the so called functional food systematically available. The resource was conceived, built and populated within the framework of a corpus-based approach for terminography, i.e., the applied field of Terminology, which besides the textual, also comprises the conceptual and communicative dimensions.

The singular nature of this project lies in the fact that texts on functional food targeted to the consumer are not only written by researchers, professors or science communicators, but also by actors from the food industry and journalists, who produce texts according to given communicative intentions: if the target public is heterogeneous, so are the text producers. Consequently and upon familiarization with the special subject field, the following question was raised: how to select and organise texts on functional foods produced by different actors, in the corpus, in order to be able to identify and extract term candidates and contexts rich in conceptual information, as a basis for definition writing, to populate the resource with information that suits the needs of the consumer?

The answer to that question was based on the hypothesis that the identification of communicative contexts, i.e., the circumstances of discourse production, considering text producers, communicative intention and target public, in which popularising discourse on functional food is produced could be intimately related to text selection criteria and the design of the specialised corpus.

Therefore, texts from different genres were selected, with the aim of obtaining a representative sample of popularizing discourse on functional foods, and organized within the corpus according to the communicative context in which they were produced. Those texts were then separately analysed, using WordSmith Tools, so that terminological information identified and extracted could then be compared: quantitatively and according to relevance. This comparison aimed to evaluate the adequacy of inclusion of texts produced in three different communicative contexts in the corpus, of corpus design and, as a consequence, to assess the need to redesign the corpus, in order to obtain more and better results. A reference corpus of texts targeted at subject field specialists, and thus representative of scientific discourse on the field, was also built, again for comparison purposes and to verify if information not present in the corpus of study and relevant for the resource was identifiable.

Results of this study are going to be disclosed and concluding remarks on the process of designing a specialised corpus, representative of popularizing discourse according to communicative contexts for terminographical purposes, will be discussed.


