

Lotte Tavecchio
Vrije Universiteit Amsterdam

A corpus-based contrastive analysis of English and Dutch sentencing patterns

With English being the international language of communication, many Dutch texts that are aimed at an international readership are now produced in English. The quality of these English texts and the extent to which the English used is influenced by the writer's native language is a much discussed topic by both linguists and language professionals. This raises questions about whether there are indeed essential structural differences between such closely related languages as English and Dutch, and, if so, what the main differences are on a grammatical, rhetorical and discourse level.

In my research project, I address these questions by investigating and comparing the ways in which English and Dutch text sentences are structured in four different genres, namely academic prose, newspaper articles, short stories and public information leaflets. For this purpose I have developed a discourse annotation system which has been implemented in a self-compiled corpus of English and Dutch texts. In this corpus sentences are not only segmented into discourse units, where punctuation marks are considered to play an important role in marking unit boundaries, but the paratactic or hypotactic relationships that exist between these units and the way in which these units are realized syntactically are also annotated. This makes it possible to uncover recurrent sentencing patterns in both languages, and subsequently to relate such patterns to rhetorical design and to investigate the extent to which different patterns may be related to syntactic differences in the two language systems involved.

In this paper, I will outline the design of the discourse annotation system and its implementation in a contrastive corpus of English and Dutch texts. The focus will be on the annotation process and the main difficulties that arose during this process. In addition to this, I will also concentrate on the uses of this discourse annotation system: what types of questions can be answered by applying this system and what types of structural differences between the English and Dutch written language can this system reveal.