Italian article and word order in Polish. A contrastive study

Word order is one of the frequently mentioned means which in Polish language can correspond to the article use. The aim of the study was to examine the relation between the position of NP in Polish sentence and the use of article in parallel Italian NP using corpus data. It was assumed that it should be at least similar to the relation described by Szwedek (2010) for Polish and English. He argues that Polish NPs in sentence-final position will have non-coreferential interpretation which correlates with the use of indefinite article in English and Polish NPs in sentence-initial position will have coreferential interpretation which corresponds to the use of English definite article.

For the study's purpose, a small parallel Polish-Italian corpus of 2M tokens was created. It gathers initial parts of 40 narrative texts with their translations. The corpus is balanced as for translation direction and single text size. The study is based on a sample of the most frequent animate, concrete and abstract nouns which were randomly extracted from the corpus together with the context (500 excerpts for each noun class). Each excerpt was annotated with different features which may influence the NP position in Polish clause, such as e.g. source language, number, Italian article, NP syntactic function, clause type, noun class, topic persistence, lexical definiteness markers in Polish). Multiple correspondence analysis was used as an auxiliary exploratory method to visualize potential correlations between the features.

The expected relation between the position of Polish NP and the article used in parallel Italian NP is not present in the examined sample. However, it appears for smaller subsets of excerpts. The test χ^2 indicates that such relation is statistically significant for subjects, animated nouns, NPs related to recurring topic or when there is a choice between definite and indefinite article in Italian clause. In all those cases the relationship seems rather weak, it's the strongest for NPs related to recurring narration topic with ϕ coefficient 0,296. All of it suggests that the word order in written Polish can correspond to the Italian article only in a limited number of cases, most likely when the NP referent plays an important role in narration as agens.

References:

Chesterman, Andrew. 2005. *On Definiteness: A Study with Special Reference to English and Finnish*. Cambridge: Cambridge University Press.

Desagulier, Guillaume. 2017. *Corpus Linguistics and Statistics with R. Introduction to Quantitative Methods in Linguistics*. New York: Springer.

Glynn, Dylan. 2014. Correspondence analysis. Exploring data and identifying patterns. W: Glynn, Dylan & Justyna A. Robinson (red.), *Corpus Methods for Semantics: Quantitative studies in polysemy and synonymy*, 443-485. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Gebert, Lucyna. 1996. Riflessioni sull'articolo mai nato nelle lingue slave. W: Benacchio, Rosanna, Francesca Fici & Lucyna Gebert (red.), *Determinatezza e indeterminatezza nelle lingue slave*, 11-25. Padova: Unipress.

Levshina, Natalia. 2015. *How to do Linguistics with R. Data exploration and statistical analysis*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

McEnery, Tony & Richard Xiao. 2007. Parallel and Comparable Corpora: What is Happening? W: Rogers, Margaret & Gunilla Anderman (red.), *Incorporating Corpora: The Linguist and the Translator*, 18-31. Clevedon: Multilingual Matters.

Szwedek, Aleksander. 2010. *The Thematic Structure of the Sentence in English and Polish*. Łódź-Warszawa: SWSPiZ.

Topolińska, Zuzanna. 1976. Wyznaczność (tj. charakterystyka referencyjna) grupy imiennej w tekście polskim. Polonica 2. 33-72.