Exploring Patterns of Textual Cohesion in Multilingual Corpora and their Application for Teaching Language and Translation

Svitlana BABYCH
Centre for Translation Studies
University of Leeds
Leeds, United Kingdom
s.babych@leeds.ac.uk

Use of cohesive ties (Halliday & Hasan, 1976) is essential for developing competence in text production and reception. However, language learners tend to transfer explicit markers of cohesion from their mother tongue, which are often not matched by typical usage patterns of the target language. For this reason it is important to put cohesive ties into multilingual context, which benefits both language learners and trainee translators.

Still, traditional teaching materials do not systematically cover or present to students the complex issue of cohesion:

1. Typically they concentrate on developing writing skills and are limited to providing unstructured lists of the linking words and phrases (connectors), which are illustrated by examples that work on the sentence level. This does not support students in using these cohesive devices on the text level. As recent studies have shown (Leńko-Szymańska, 2008) this leads to overuse of connectors in students' writing. Also, the lists of connectors do not provide contrastive analysis of their usage across languages.

2. Usually there is no discussion of the methodology of how the lists were created and often the compilation method is unsystematic and introspective, which for some languages results in incidental choices of connectors, and does not ensure the completeness of coverage.

There is a clear demand for developing a systematic methodology of discovering and analysing cohesive devices and their patterns in corpora. These devices then need to be adequately presented in teaching materials in cross-lingual context on textual level. Recent research in Corpus Linguistics addresses the issue of *lexical textual cohesion* (Flowerdew & Mahlberg, 2009)

My study focuses on *conjunction* as a cohesive device and presents a corpus-linguistic framework for systematically deriving, analysing and evaluating *lexico-grammatical* patterns of conjunction as means of textual cohesion in multilingual context for the purposes of teaching languages and translation:

1. I propose a method of systematically discovering textual connectors from large corpora. The working hypothesis is that connectors which function above the sentence level usually appear in paragraph- or sentence-initial position (c.f. Biber et al., 1999). Frequency lists of sentence-initial words and multiword expressions (up to five words long) were automatically collected from large Ukrainian, Russian and English corpora of current electronic media, including news, blogs and comments (about 200 million words for each language). Textual connectors were manually identified among the

most frequent items in these lists, which ensured coverage of most typical single- and multiword connectors.

2. I then evaluated corpus-derived lists against traditional grammar-based lists of textual connectors, extending it by approximately 80% with the corpus-based data, which shows inherent limitation of the grammar-based resources.

3. I develop a search tool, based on Corpus Workbench (Christ, 1994), for exploring the patterns of connectors in real texts. Cohesive profiles of texts are represented as patterns of language-independent semantic categories of connectors, so it is possible to visualise rhetorical structure of the text across different languages.

My paper shows potential of corpus-based techniques for systematically deriving, analysing and presenting linguistic resources for pedagogical purposes.

References

Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education Limited.

Christ, O. (1994). 'A modular and flexible architecture for an integrated corpus query system.' Presentation at COMPLEX'94, Budapest, Hungary.

Flowerdew, J., & Mahlberg M. (eds). (2009). *Lexical Cohesion and Corpus Linguistics.* Philadelphia: John Benjamins.

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.

Leńko-Szymańska, A. (2008). 'Non-native or non-expert? The use of connectors in native and foreign language learners' texts.' *Savoirs et savoir-faire dans l'apprentissage et l'enseignement d'une langue étrangère* 27: 91-108.