Using annotation and specialised electronic corpora to facilitate the reading of journal articles in the field of Law

Pascual Pérez-Paredes1 (pascualf@um.es)
María Sánchez-Tornel1 (mstornel@um.es)
José M. Alcaraz Calero² (jose.alcaraz-calero@hp.com)

1: Universidad de Murcia
2: Cloud and Security Lab, Hewlett-Packard Research Laboratories

Abstract

In an effort to balance their training in specialised vocabulary acquisitiveness and reading, as well as their understanding of the compositional and rhetorical structure of the rhetorical moves of RAs, a corpus of RAs was annotated so as to foster the development of learner linguistic awareness. The corpus consisted of 31 RA published in 2009 in the Oxford Journal of Legal Studies and the total amount of words was 393,052. This paper explores the design and implementation of this innovative experience during the 2010-2011 academic year.

1. Legal English for research purposes

A comprehensive analysis of any professional genre must consider and integrate text-internal as well as text-external aspects of language use (Bhatia, 2004). Beyond the well-known difficulties of legal language (Bhatia, 2002) and its main features (Alcaraz Varó, 2007), the rhetorical structure of a RA is a reflection of the professional practices of scholars in the field of law. Linguists consider that a genre is an instance of a sociolinguistic activity through which members of certain discourse community achieve their communicative purposes (Swales, 1990). A genre is defined then by its shared communicative purposes and manifested by its particular structural and linguistic features. The Research Article (RA) is the main channel of scientific or scholarly communication, which turns it into the largest genre for PhD candidates in law.

2. Corpus compilation

Corpus compilation: research articles (RA)
Domain: general legal studies
Year: 2009
Number of RAs: 31
Number of words: 393,052
Scope of the journal: The Oxford Journal of Legal Studies is published by the Faculty of Law of the University of Oxford. It publishes an annual issue of refereed articles and reviews within the field of law, in a wide range of legal disciplines. Its legal interest is exclusive but not exclusive. In addition to traditional questions of legal interest, it covers all fields of law, from the relationship of law to other disciplines. The corpus was a selection of 31 articles from the 72 published in 2009.

3. Corpus annotation

The corpus was marked and annotated using BackBone Annotator (3.8). BACKBONE Annotator is a tool that assists annotators in the process of manual annotation of corpora. It allows for the management of multiple corpus files and corpora. Annotators can create import or export different corpora and perform the same or different annotation schemes on them. Our annotation relies on three key aspects:

1. Annotation is unit-bound, not text-bound. Our RAs are divided into paragraphs and then rhetoric moves are identified. This approach allows for the identification of different rhetorical moves. The corpus was annotated using the annotation software BACKBONE Annotator. The annotation process was guided by the genre-based approach of Bhatia (1982) and its main features (Alcaraz Varó, 2007). The corpus was annotated using the annotation software BACKBONE Annotator. The annotation process was guided by the genre-based approach of Bhatia (1982) and its main features (Alcaraz Varó, 2007).

2. Annotation is based on the Text Encoding Initiative (TEI)-compliant XML. This gives our XML files the flexibility to be read and understood by different tools and platforms. The corpus was annotated using the annotation software BACKBONE Annotator. The annotation process was guided by the genre-based approach of Bhatia (1982) and its main features (Alcaraz Varó, 2007).

3. Our annotation approach has a multipurpose nature, in this case RAs rhetorical moves. The annotation informs PhD students on the main discoursal characteristics of a given section. The annotation includes information on the rhetorical structure of RAs, which turns it into the largest genre for PhD candidates in law.

4. Corpus availability

Once annotated, the resulting corpus was published on a server and learners (n=15) were provided with scaffolded guided activities that sought to promote a problem-solving, data-driven learning approach. Learners were presented with activities that required using the annotated corpus and navigating the rhetorical moves therein.

5. Learners’ evaluation & conclusion

Students (n=15) completed a questionnaire once they submitted their answers. Learners found the resource useful (Q1) and said that reading concordances (Q2) and having access to annotated resources facilitates their understanding of rhetorical moves in RAs (Q3). The evaluation of the learners seems to be in line with previous findings on using corpora for the teaching of specialised languages (Fan & Xunfeng, 2002; Yoon & Hinve, 2004).

Q1. I found this resource very useful as an aid for reading and comprehension
Q2. Concordance lines helped me to better understand the different moves
Q3. Concordance lines helped me to better understand the different moves in research articles
Q4. This type of resource is very useful in research articles
Q5. If you had the chance of using this resource for free, would you use it again?

References