



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

Pascual Pérez-Paredes<sup>1</sup> (pascualf@um.es)

Maria Sánchez-Tornel<sup>1</sup> (mstornel@um.es)

José M. Alcaraz Calero<sup>2</sup> (jose.alcaraz-calero@hp.com)

<sup>1</sup>:Universidad de Murcia

<sup>2</sup>:Cloud and Security Lab. Hewlett-Packard Research Laboratories

### Abstract

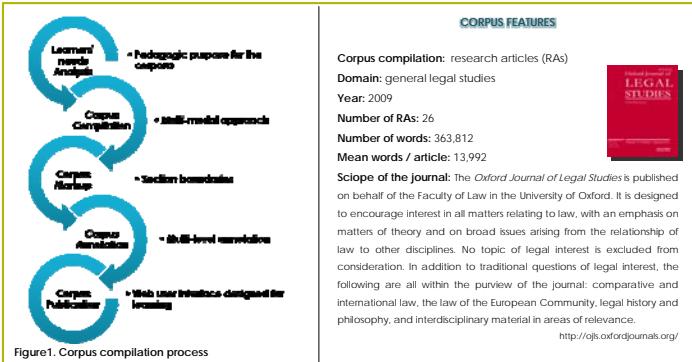
Mastering specialised languages is increasingly becoming a pressing need for professionals in the academic and business practices. The experience we report involves students seeking a PhD in Law at UM. Within their training, these learners need to go beyond mainstream English learning for general communication, their primary concern being reading and understanding research for their academic, scholarly purposes. In this context, nothing to gains with the rhetorical structure of research articles (RAs) is of great relevance for the development of the learners' linguistic research skills.

In an effort to balance their training in specialised vocabulary acquisition and reading, as well as their understanding of the complexities of the rhetorical structure of RAs, a corpus of law RAs was annotated so as to foster the development of learner rhetoric awareness. The corpus consisted of 26 RAs published in 2009 in the Oxford Journal of Legal Studies and the total amount of words was 343,812. This poster 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, 1982) and its main features (Alcaraz Varo, 2002), 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 target genre for PhD candidates in law.

### 2. Corpus compilation



### 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 (Pérez-Paredes & Alcaraz, 2009). BACKBONE Annotator allows for the management of multiple corpus files and corpora. Annotators can create import or select different corpora and perform the same or different annotation schemes on each of 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 and annotated.
- (2) Annotation is based on the Text Encoding Initiative (TEI)-compliant XML. This gives our XML files the processing and structuring power of XML plus the uniformity and standardisation of TEI.
- (3) Our annotation approach has a multipurpose nature, in this case RAs rhetorical moves. The annotation informs PhD students on the main discursive characteristics of a given section.

**ANNOTATION CATEGORIES**

- Information move
  - Background about theory
  - Research aims
  - Niche
- Finding
- Referring to previous research or authority
- Claim
- Recommendation
- Language, argument and discourse organisation
- Article organisation
- Aim of article
- Argumentation building

Figure 2. Backbone Annotator

### 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.

Figure 3. User interface of the corpus of legal research articles

### 5. Learners' evaluation & conclusion

Students (n=7) 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 & Hirvela, 2004).

Q1. I found this resource very useful as an aid for reading and comprehension

Q2. Concordance lines helped me to better understand the use of lexis

Q3. Concordance lines helped me to better understand the different moves in research articles

Q4. This type of resource is very useful for the study of Legal English

Q5. If you had the chance of using this resource for free, would you use it again?

### References

- Alcaraz Varo, E. (2007). *El inglés jurídico*. Barcelona: Ariel
- Bhatia, V. K., (1982). *An investigation into the formal and functional characteristics of qualifications in Legislative writing and its application to English for academic legal purposes*. (Doctoral dissertation). University of Aston, Birmingham, U.K.
- Bhatia, V. K. (2004). *Worlds of written discourse: a genre-based view*. London: Continuum International.
- Fan, M. & Xunfeng, X. (2002). An evaluation of an online bilingual corpus for the self-learning of legal English. *System* (30)1, 47-63
- Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.
- Yoon, H., & Hirvela, A. (2004). ESL student attitudes towards corpus use in L2 writing. *Journal of Second Language Writing*(13) 4, 257-283.