

From ‘our methods apply equally well’ to ‘the model does a very poor job’: A corpus-based study of academic value-marking

Davide Simone Giannoni
University of Bergamo

Abstract

Successful communication depends on the participants’ ability to establish a common ground, embracing shared perceptions of what is good or bad, suitable or unsuitable, affective or ineffective in a given context. While the range of speech acts expressing *evaluation* in academia has been extensively researched, little is known of the underlying *values* they encode. This paper illustrates the theoretical and methodological approach employed in a recent study of academic writing (Giannoni, 2010), based on a 1 million word corpus of English research articles from ten disciplinary areas. Using candidate items with 100+ wordlist occurrences, the most prominent value-marking categories were investigated through a combination of automated and manual procedures. Four of these (relevance, size, novelty and goodness) were singled out for further investigation, based on lexical sets and word groups including synonyms, antonyms and different parts of speech. The paper also highlights some of the challenges inherent in the methodology and suggests directions for further research. The significance of such data should not be underestimated, as evaluative categories reflect value systems instrumental to the epistemology/ideology that permeates every academic discipline.

1. Introduction

In recent years applied linguists have paid special attention to the evaluative plane of language, which is rooted in “the lexis of judgement and subjectivity” (Thompson and Hunston, 2000: 16). Evaluative speech acts are primarily communicated through wordings chosen in response to the communicative context, according to the interactants’ needs and sensitivities. Their study can shed light on the co-existence of meanings which are at the same time personal and communal, autonomous and negotiated. Academic discourse has proved a particularly fertile area for this kind of investigation (cf. Hunston and Thompson, 2000; Del Lungo Camiciotti and Tognini Bonelli, 2004; Dossena and Jucker, 2007) because of the perlocutionary role of value judgements in the construction of scientific arguments. The lexis of evaluation contributes primarily to the expression of what constitutes ‘worthiness’ (Thetela, 1997) amongst scholars affiliated to a given discipline or subdiscipline. Many of these aspects are “construals of experiences in context on binary scales between positive and negative” (Downes, 2000: 104). This means they can be classified respectively as facilitating or hindering a scholar’s search for knowledge. The textual investigation of evaluative meanings is no easy task, however, because a large number of realisations are polysemous, vague or indirect. When corpus-analytical tools are employed, the challenge is even greater and even Biber (2006: 99) recently complained that “nearly any word could be analyzed as reflecting an evaluation, making it hard to identify a closed set of words used to convey specific attitudes and evaluations”.

1.1. From evaluation to value

The view adopted here is that every evaluative act implies an axiological variable or ‘value’ along which the evaluation is performed. The range of such variables is limited and amenable to textual investigation. If we recognise this link, the search for academic values, as communicated through discourse, is not only possible but also desirable, for it offers insights into the axiological basis of disciplinary cultures. The need to turn from the plane of evaluation to its underlying values is implicit in Thompson and Hunston’s argument that “every act of evaluation expresses a communal value-system, and every act of evaluation goes towards building up that value-system. This value-system in turn is a component of the ideology which lies behind every text” (2000: 6). Affiliation to a certain discipline implies compliance with (largely unwritten) norms that regulate access to the community. Scholars interact, debate and share their findings following well-trodden routines and textual avenues, within a cultural-semiotic framework provided by their research culture. As Becher aptly puts it, “disciplines are cultural phenomena: they are embodied in collections of like-minded people, each with their own codes of conduct, sets of values and distinctive intellectual tasks” (1981: 109). Values are therefore an essential component of academic proficiency, whose understanding can facilitate or hinder access for new members. In this sense they act as gatekeepers to the disciplines and “graduate students need to learn to work within the value-systems of their target communities” (Swales, 1990: 218).

1.2. Analytical hurdles

When attempting a corpus-based investigation of academic values, there are several difficult choices that the analyst should consider. The first concerns the terminology used in the literature to define such variables, which seems to lack any consistency: ‘parameters of value’ (Thetela, 1997) to ‘parameters of evaluation’ (Thompson and Hunston, 2000), ‘axiological classes’ (Felices Lago and Hewitt, 2004), ‘semantic categories’ (Pérez-Llantada Auría, 2008), ‘evaluative dimensions’ (Bednarek, 2009), ‘semantic groups’ (Nishina, 2011), and finally ‘values’ (Breeze, 2011). Resorting to different terms to the same referent is rather common in the humanities and social sciences; unsurprisingly, it also affects the literature on evaluation, where “a number of terminological problems still remain, several with more than terminological implications, that is, they point to lack of clarity in the conceptualisation of the object of research” (Mauranen, 2004: 204). The present study will broadly adopt the term *value* to indicate any aspect of research to which the parent disciplinary community assigns an interpretation whose polarity reflects its practices, standards and beliefs.

The other challenges are of a more methodological nature and concern the way textual data is obtained and processed, i.e. a study’s choices and ‘flavours’ (Hunston 2011). Some attempts to identify values in discourse (cf. Felices Lago and Hewitt, 2004; Bednarek 2009) have adopted a top-down procedure based on predetermined sets of conceptual categories through which the data is deductively interpreted; conversely, my approach here is bottom-up, as it seeks to extract analytical categories inductively from the data. Another difficult choice is the suitable degree of specificity or taxonomic discreteness to be used when classifying values: this is not only a matter of personal judgement but also subject to the number of occurrences in hand. I have tried to strike a balance between generality (which leads to interpretive vagueness) and overspecificity (which makes it difficult to compare results across studies).

When ‘corpus-crunching’ tools are applied to the investigation of evaluative lexis, the researcher is confronted with further hurdles, most notably which parts of speech deserve attention. Studies of this kind (cf. Pérez-Llantada Auría, 2008; Breeze, 2011; Nishina, 2011) often limit the lexical range to qualifiers, and occasionally adverbs. In so doing, the role of verbs and nouns encoding the same semantic-evaluative categories is overlooked: e.g.

interesting (adj), *interestingly* (adv), but also *interest* (v/n). This paper adopts an inclusive approach extending to all the main parts of speech.

To examine the occurrences of a given lexical item contrastively, it is of course essential to normalise data (usually as n / 10,000 words) but this is not sufficient unless, where possible, the data itself is validated. While for very large corpora the manual inspection of concordance entries is usually impractical, the 1 million word corpus assembled for this study allows an in-depth investigation of concordance lines to minimise errors due to the inclusion of homographs and other spurious items.

The last point worth mentioning has no straightforward solution. It concerns the interpretation of results, which in our case are largely due to cross-disciplinary variation. All things being equal, the argument goes, any major discrepancy between sections of the corpus is likely to reflect epistemological distinctions, that is the way scholars in different academic fields conceptualise research and its goals. This method follows a well-established tradition (cf. Hyland, 2000; Fløttum et al., 2006) but should be handled with care, because textual differences observed across domains may subsume other factors, i.e. the range of topics, subdisciplinary areas and subgeneric variants.

2. Corpus construction

Representative multidisciplinary corpora based on academic journals are relatively easy to assemble, thanks to the wide availability (online and offline) of such publications. What poses most problems is the selection of disciplines warranting inclusion, in other words the coverage of the corpus. The very notion of disciplinary field is a slippery one, as scientific knowledge is regularly reorganised/reassessed by researchers and institutions. Hyland (2009: 27-28) aptly sums up the situation in the following terms:

The idea of discipline is a troubled one. The concept is challenged by post modernism [...] and by institutional changes such as the emergence of modular degrees and practice-based degrees like nursing and social work. Like most communities, disciplines experience constant change and frequent conflict. New disciplines like biotechnology and gerontology spring up at the intersections of existing ones and achieve international recognition, while others like philology and astrology decline and disappear. Institutional and commercial forces as much as intellectual ones seem to drive these changes and shape what gets recognised as a discipline in a particular context rather than remaining just an ‘approach’. Forensic linguistics seems to be making great strides towards disciplinary status, for example, while corpus linguistics is not. In other words, disciplines are neither stable nor agreed.

Some of the earliest and most influential taxonomies of knowledge are those adopted by librarians for recording library holdings by discipline. The Dewey Decimal Classification (DDC, 2003) follows a hierarchical structure based on ten classes of disciplines, which in turn are divided into divisions and sections. Figure 1 overleaf illustrates how Linguistics (one of the divisions of Language) is organised into eight sections, ranging from ‘Writing systems’ to ‘Verbal language not spoken or written’. The DDC remains one of the most authoritative models for librarians, despite its idiosyncratic structure and lack of consistency.

It is important to recognise that the disciplinary coverage of a corpus reflects the kind of assumptions made before its construction. The framework followed here is relatively traditional, based on four categories used in Becher and Trowler (2001) and/or Hyland (2000), plus a fifth one for the mathematical sciences. It covers a total of ten disciplines representing five ‘sciences’: Biology (BIO), Physics (PHY) [natural sciences]; Engineering (ENG), Medicine (MED) [applied sciences]; Anthropology (ANTH), History (HIST) [human sciences];

Economics (ECO), Sociology (SOC) [social sciences]; Mathematics (MATH), Computer Science (CS) [mathematical sciences].

To ensure that the texts assembled were sufficiently representative of each domain, the most prestigious journal in each discipline was chosen, based on its Impact Factor score as listed in the Journal Citation Reports (JCR 2005) database. While this was generally a straightforward procedure, in four instances the data was indexed only by subdiscipline: for instance, there was no single entry for Engineering, but a choice between Biomedical E., Chemical E., Civil E. and many others. The decisions eventually made were as follows: Medicine→General and Internal; Computer Science→Cybernetics; Physics→Particles and Fields; Engineering→Civil.

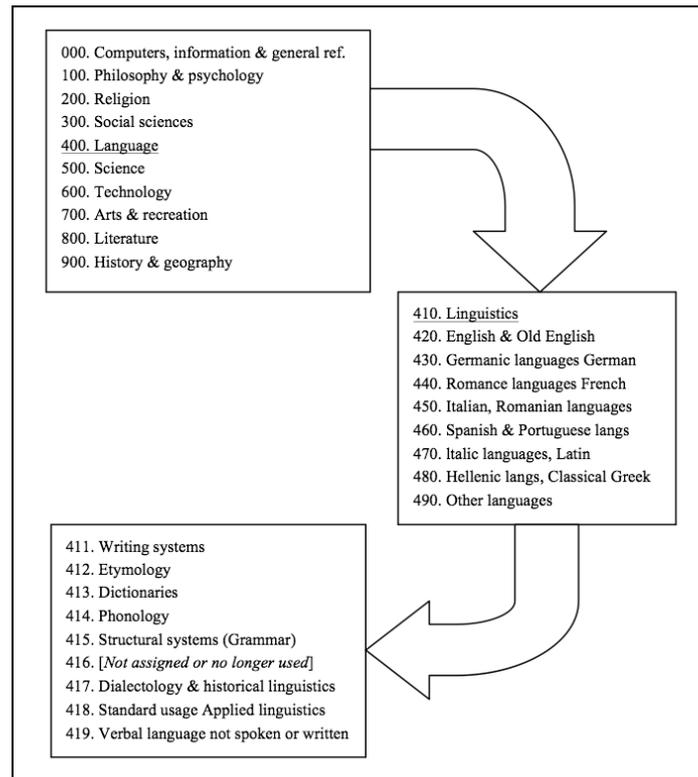


Figure 1: 'Language' and 'Linguistics' in DDC (2003)

Ten contributions identifiable as research articles (RAs) were downloaded from each journal and stored for computer processing, making a total of 100 RAs. Before uploading the texts to a well-known concordancer (Scott, 2007), they had to be sanitised to ensure that accessory content was not included in the corpus. The procedure included:

- Removal of authors' names and contact details, references, foot/endnotes, charts/diagrams and appendices, page headers and footers, including page numbers.
- Removal of the abstract or opening summary (where present).
- Removal of the Supporting Information section appended to biology articles.
- Substitution of mathematical functions, equations and formulae with the string *expression*; such occurrences were a major feature of mathematics RAs, as shown in the example overleaf (Figure 2).

The resulting corpus measured almost 1 million words (986,185 tokens), with the longest RAs in MATH (17,187 tokens per text), CS (13,804) and SOC (12,477) and the shortest

in MED (3,997), PHY (6,303) and BIO (7,989). This means the corpus was sufficiently large to yield data for an in-depth investigation of value-marking items in a single genre. At the same time, it was not too large to make direct investigation of concordance lines unmanageable, so that the analyst to steer a middle course between the search for quantitatively significant and qualitatively valid evidence.

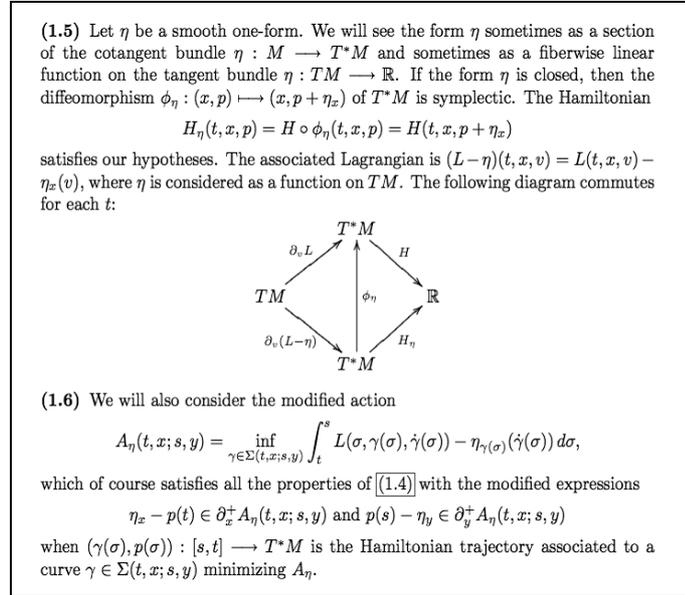


Figure 2: Excerpt from a MATH text

3. Identification of value markers

The first decision made at this point was to focus on *explicitly evaluative* lexis, leaving out realisations whose semantic opacity or pragmatic complexity defy the scrutiny of corpus-analysis techniques (cf. Römer, 2008). Using WordSmith Tools, a global wordlist was generated and all items with 100+ occurrences inspected manually to identify potential value-marking words (candidates). This produced a total of 83 candidates, accounting together for some 20,000 tokens, i.e. approximately one word in 50. The candidates were then grouped into semantic categories labelled as follows: relevance, size, goodness, novelty, value, timing, impact, complexity, generality, completeness, appeal.

For a closer examination of such evidence, the top four categories were selected for further analysis. A concordance was generated for each item, checked line by line to exclude irrelevant occurrences (e.g. *well* used as a noun or *new* in New York) and the list of validated value-marking instances adjusted. Finally the items included in each category were integrated by adding formally/semantically-related lexemes that did not occur in the original list of candidates (e.g. *good* \rightarrow *bad*; *significant* \rightarrow *significance*). Whenever possible, lexical groups were formed considering all possible parts of speech, though at times potentially available items (e.g. *unproblematically*) were then found to be absent from the corpus.

3.1. Relevance

This value indicates “how closely related something is to the topic or field being discussed” (Swales and Burke, 2003: 5) and was found to be the most prominent in the corpus. Its markers were distributed as follows (items from the original candidate list are underlined)

here and below): *importance, important, importantly, unimportant* [32% of total]; *major, main, principal, mainly, principally, minor* [31%]; *significant, significance, significantly, insignificant* [21%]; *key, crucial, vital, crucially* [9%]; *relevant, relevance, irrelevant* [7%]. Examples of omitted (i.e. evaluatively neutral) uses include: *religion should be thought of as the **vital** expression of groups* (SOC2); *the emphasis on patience and self-discipline is the **key** to the effectiveness of Catholic schools in the United States* (ECO8).

3.2. Size

Markers in this category were the second most frequent. They encode judgements that evaluate the evidence in hand (or other aspects of the knowledge-production process) as ‘large’ or ‘small’, rather than specifying its dimensions. The range of items identified in the corpus comprises: *large, larger, largest* [52%]; *small, smaller, smallest* [42%]; *big, bigger, biggest* [6%]. Examples of clearly irrelevant and thus omitted uses: *for the population at **large**, the nation-states also provides incentives* (SOC); *to foster collective action (**Small** 2004; Warren 2004)* (HIST); *make its **Big** Mule leadership less interested in further growth* (SOC).

3.3. Goodness

These markers point to the core meaning of positive evaluation and formed the third most common category. They point to facets viewed as generally desirable (or, conversely, as undesirable) without specification of their exact nature. The relevant items were: *good, better, best, well, improve*, bad, poor, worse, worst, badly, poorly, worsen** [64%]; *problem/s, problematic, problematically, unproblematic* [21%]; *positive, positively, negative, negatively* [9%]; *error/s, erroneous, erroneously* [4%]; *right, rightly, wrong, wrongly* [2%]. Because of the wide repertoire considered, some uses had to be omitted: *nationalists were simply **well** ahead of the Austrian state* (HIST); *let p_0 be the **positive** number of the initial assumption* (MATH); *including these six highly corrupt mayors in the sample creates a **negative** relationship between reelection rates and corruption* (ECO); *diaphyseal breadths are biased to the **right*** (ANTH); *there is a **good** deal of research* (ECO); *we consider u the unique forward maximal solution to the Cauchy **problem*** (MATH); *means of species with small samples were strongly affected by sampling **error*** (ANTH); *resources available to residents of **poor** urban neighborhoods* (SOC).

3.4. Novelty

The last category considered identifies a referent as more or less valuable, based on some unprecedented quality viewed as a (potential) improvement on previous referents within the same class. Its repertoire consisted of three groups of items: *new, newer, newly, novel, novelty/ies, fresh, old, obsolete, antiquated* [73%]; *recent, recently, latest* [18%]; *innovate*, innovation/s, innovative, renew** [9%]. Examples of omitted uses are: *the Puebla plant is the sole world producer of the **New** Beetle* (ECO); *direct radiocarbon dates, however, proved them to be only 5000 years **old*** (ANTH); *in view of the **fresh** condition of most artifacts* (ANTH); *in their **latest** review, Olson and Olson (2003) summarize this issue* (CS); *those features or values that are more **recent** are not better than older ones* (HIST).

4. Results and discussion

The distribution of these four values across the corpus (Figure 3) highlights two levels of differentiation between domains. First of all, the total amount of explicit marking considered appropriate by RA writers in the field, which reaches its lowest in MATH (6.9 occurrences / 10,000 words) and BIO (13.6), and peaks in CS (24.3) and PHY (23.7); secondly the kind of

value prevalent in each discipline. If we combine the two, the following picture emerges: ANTH is size-focused and moderately explicit; BIO is novelty-focused and scarcely explicit; CS is relevance/goodness-focused and highly explicit; ECO is relevance/goodness-focused and highly explicit; HIST is balanced and highly explicit; ENG is balanced and highly explicit; MATH is size-focused and scarcely explicit; MED is relevance-focused and moderately explicit; PHY is size-focused and highly explicit; SOC is relevance-focused and scarcely explicit.

If grouped by scientific area, the results indicate that relevance is most valued in the applied sciences (ENG+MED), size in the natural sciences (BIO+PHY), goodness in the social sciences (ECO+SOC) and novelty in the human sciences (ANTH+HIST). While the readers is referred to Giannoni (2011) for a detailed interpretation of these figures, their main implication is that evaluative lexis pervades research writing in ways that vary considerably from one disciplinary culture to another.

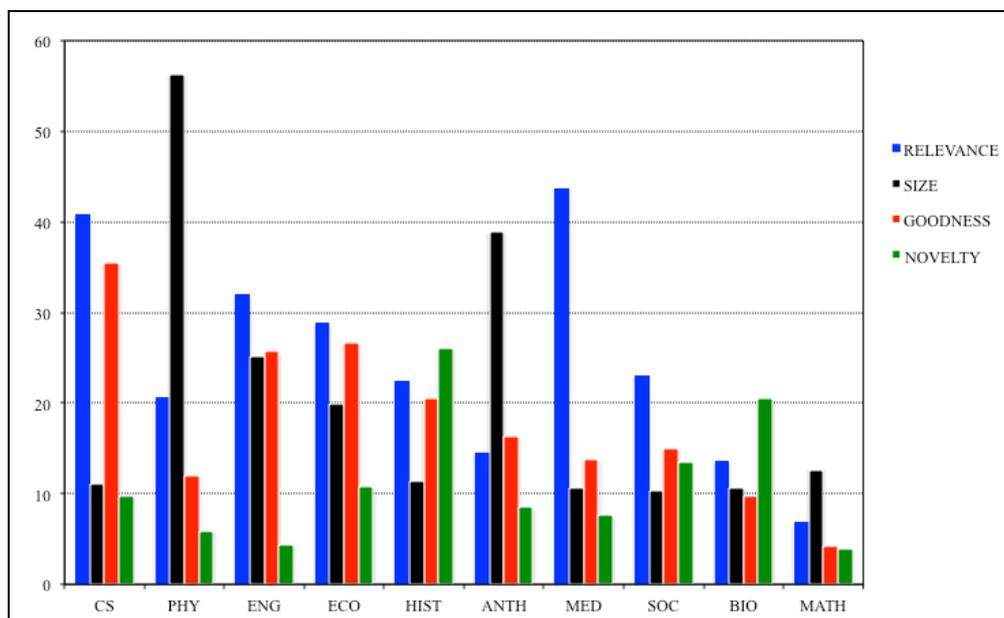


Figure 3: Explicit value-marking across disciplines

While not particularly noticeable *per se*, when assessed in a corpus these items appear to be particularly salient, thus disproving the claim that evaluative adjectives such as *good*, *poor*, *great*, *considerable*, etc. are “too weak and too imprecise for use in scientific writing” (Wilkinson, 1991; cited in Pérez-Llantada Auría, 2008: 130). Admittedly, the range of lexemes and phrases capable of encoding meanings related to the value system is potentially unlimited, but concentrating on the most frequent categories of explicit markers makes it possible to examine a manageable number of realisations close to the semantic core of the target values.

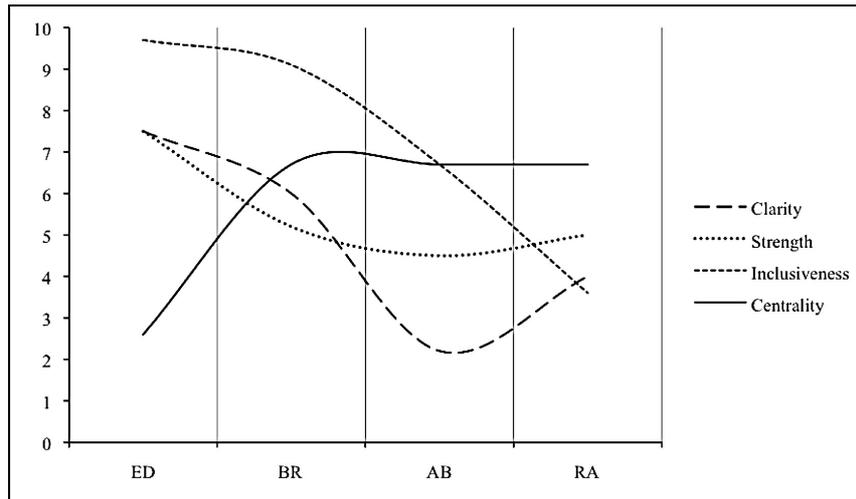


Figure 4: Evidence of intergeneric variation (from Giannoni 2009: 177)

It would be wrong, however, to interpret all of the data as discipline-specific, without considering the role of generic constraints due to the type of text considered. RAs are not objective accounts of disciplinary values/beliefs but represent the ‘public face’ of science, the way scholars adhere “to accepted ways of thinking and writing [...] through agreed procedures such as experimental demonstration, replicability and falsification of results” (Hyland 1997: 21). There is indeed some evidence that genres regulate the type of value foregrounded by their creators: a recent study (Figure 4) of metaphoric value-marking lexis points to considerable differences in this respect between journal editorials (ED), book reviews (BR), abstracts (AB) and research articles (RA).

5. Conclusion

The identification of value-related items in a written corpus is not an impossible feat, but merely a matter of discreteness. Accounting for every possible realisation is clearly out of the question, but by restricting the target to overt realisations, it is indeed possible to explore how scholars signal (un)worthiness in their research. Most of the lexical items employed are so mundane as to go easily unnoticed, but perhaps their argumentative power stems exactly from such an apparent insignificance. The fact that (like so many other features of language) the four values described in this paper are not unique to academia, may partly contribute to this impression.

For Miceli and Castelfranchi (1989: 191), values are “precious for the maintenance of social cohesion” because they are accepted without need for justification by their recipients. Nobody is asked to explain why ‘honesty’ or ‘success’ are desirable. Shared belief in a value is enough for its maintenance – an aspect that explains why the occurrences observed in the corpus generally omit to specify why an evaluated entity is judged to be *important*, *novel* or *well* performed. This means that scholars reproduce normative research-focused meanings whose truth cannot be challenged because it lacks detail.

The applications of corpus-based value-mapping techniques are potentially unlimited, not only because they can uncover aspects of the ‘ideology’ (Thompson and Hunston, 2000) underlying a text but also because of the need for axiological proficiency in EAP courses. Novice writers need to replicate the degree of explicitness appropriate to their field by deploying the right language to highlight values that are crucial to their peers. Comparing published texts with learner corpora (as in Breeze, 2011) can provide insights in this direction.

Finally, it would be interesting to explore differences in value marking across different linguistic/rhetorical backgrounds, as disciplinary networks become increasingly international and ‘deterritorialised’ (cf. Canagarajah 2002) through the medium of English.

References

- Becher, T. (1981). “Towards a definition of disciplinary cultures”. *Studies in Higher Education*, 6, 109–122.
- Becher, T. and P.R. Trowler (2001). *Academic Tribes and Territories*. 2nd ed. Buckingham: SRHE & Open University Press.
- Bednarek, M. (2009). “Dimensions of evaluation: cognitive and linguistic perspectives”. *Pragmatics & Cognition*, 17, 146–175.
- Breeze, R. (2011). “Disciplinary values in legal discourse: a corpus study”. *Ibérica*, 21, 93–116. Available at http://www.aelfe.org/documents/05_21_Breeze.pdf.
- Canagarajah, A.S. (2002). “Multilingual writers and the academic community: towards a critical relationship”. *Journal of English for Academic Purposes*, 1, 29–44.
- DDC (2003). *Dewey Decimal Classification and Relative Index*. Edition 22. Dublin, OH: OCLC Forest Press.
- Del Lungo Camiciotti, G. and E. Tognini Bonelli (eds) 2004. *Academic Discourse – New Insights into Evaluation*. Bern: Peter Lang.
- Dossena, M. and A. Jucker (eds) (2007). *(R)evolutions in Evaluation*. Special issue of *Textus*, 20. Genoa: Tilgher.
- Downes, W. (2000). “The language of felt experience: emotional, evaluative and intuitive”. *Language and Literature*, 9, 99–121.
- Felices Lago, Á. and E. Hewitt (2004). “Personal qualities of applicants in job advertisements: axiological and lexical analysis of samples in English from the Spanish press”. *LSP & Professional Communication*, 4, 8–26. Available at <http://rauli.cbs.dk/index.php/LSP/article/download/2037/2037>.
- Fløttum, K., T. Dahl and K. Torodd (2006). *Academic Voices. Across Languages and Disciplines*. Amsterdam: John Benjamins.
- Giannoni, D.S. (2009). “Disciplinary values in English academic metaphors”. *Linguistica e Filologia*, 28, 173–191. Available at <http://hdl.handle.net/10446/379>.
- Giannoni, D.S. (2010). *Mapping Academic Values in the Disciplines: A Corpus-Based Approach*. Bern: Peter Lang.

- Hunston, S. (2011). “Doing analysis in discourse and corpus: the case of evaluative language”. Plenary talk given at CL 2011 – Discourse and Corpus Linguistics Conference. Birmingham, 20-22 July 2011.
- Hunston, S. and G. Thompson (eds) (2000). *Evaluation in Text. Authorial Stance and the Construction of Discourse*. Oxford: Oxford University Press.
- Hyland, K. (1997). “Scientific claims and community values: articulating an academic culture”. *Language and Communication*, 17, 19–31.
- Hyland, K. (2000). *Disciplinary Discourses. Social Interactions in Academic Writing*. Harlow: Pearson Education.
- Hyland, K. (2009). “Constraint vs creativity: identity and disciplinarity in academic writing”. In M. Gotti (ed.) *Commonality and Individuality in Academic Discourse*. Bern: Peter Lang, 25–52.
- JCR (2005). *Journal Citation Reports*. New York: Thomson Reuters. Online resource available at <http://www.isiwebofknowledge.com>.
- Mauranen, A. 2004. “Where next? A summary of the round table discussion”. In G. Del Lungo Camiciotti and E. Tognini Bonelli (eds) *Academic Discourse – New Insights into Evaluation*. Bern: Peter Lang, 203–215.
- Miceli, M. and C. Castelfranchi (1989). “A cognitive approach to values”. *Journal for the Theory of Social Behaviour*, 19, 169–193.
- Nishina, Y. (2011) *Evaluative Meanings and Disciplinary Values: A Corpus-Based Study of Adjective Patterns in Research Articles in Applied Linguistics and Business Studies*. PhD thesis. University of Birmingham. Available at <http://etheses.bham.ac.uk/1540>.
- Pérez-Llantada Auría, C. (2008). “Stance and academic promotionalism: a cross-disciplinary comparison in the soft sciences”. *Atlantis*, 30, 129–145.
- Römer, U. (2008). “Identification impossible? A corpus approach to realisations of evaluative meaning in academic writing”. *Functions of Language*, 15, 115–130.
- Scott, M. (2007). *Oxford WordSmith Tools 5.0*. Oxford: Oxford University Press.
- Swales, J.M. (1990). *Genre Analysis. English in Academic and Research Settings*. Cambridge: Cambridge University Press.
- Swales, J.M. and A. Burke (2003). “‘It’s really fascinating work’: differences in evaluative adjectives across academic registers”. In P. Leistyna and C.F. Meyer (eds) *Corpus Analysis. Language Structure and Language Use*. Amsterdam: Rodopi, 1–18.
- Thetela, P. (1997). “Evaluated entities and parameters of value in academic research articles”. *English for Specific Purposes*, 16, 101–118.

Thompson, G. and S. Hunston (2000). "Evaluation: an introduction". In S. Hunston and G. Thompson (eds) *Evaluation in Text. Authorial Stance and the Construction of Discourse*. Oxford: Oxford University Press, 1–27.

Wilkinson, A. (1991). *The Scientist's Handbook for Writing Papers and Dissertations*. Englewood Cliffs, NJ: Prentice Hall.