

This is a pre-print version of the following research article:

Moreno, Ana I. and Swales, John M. (2018). Strengthening move analysis methodology towards bridging the function-form gap. *English for Specific Purposes*, 50, 40-63. [<https://doi.org/10.1016/j.esp.2017.11.006>].

[Pre-print version: <http://hdl.handle.net/10612/10667>]

Abstract

Move analysis is a text analytical approach first developed by John Swales (1981) to investigate the underlying generic structure of research articles (RAs) in terms of moves-and-steps for pedagogical purposes. A widely shared aspiration of move analysts has been to identify the linguistic features characterizing the various RA moves not only in English, but also across languages. One shortcoming blocking this advancement is the lack of multilingual corpora fully annotated for their specific communicative functions in a coordinated and reliable manner. In this paper, we describe and discuss a methodology for analysing the various RA sections for their generic structure up from the step level in two languages and across a wide range of disciplines, using the discussion section as a test case for illustrating that methodology. Among the topics treated are establishing criteria for choosing a suitable sample of comparable RA discussions across the two languages, designing a model for annotating the section's moves and steps, creating an accessible computer-assisted coding scheme, achieving good levels of inter-rater reliability, and obtaining validation from expert informants and writers. In essence, this is a methodology paper offered as a working model for other EAP researchers undertaking similar analyses in future.

Keywords

Genre analysis, move analysis, communicative functions, generic structure, academic writing, applied linguistics

1. Introduction

Move analysis is a text analytical method developed by Swales in 1981 as an essential component of his genre analytical framework (1990). In his approach, moves are “discoursal or rhetorical units performing coherent communicative functions in texts”,

whose linguistic realizations may be very variable in length and in other ways (Swales, 2004: 228-229). Steps, on the other hand, are the multiple text fragments that “together, or in some combination, realize the move” in such a way that “the steps of a move primarily function to achieve the purpose of the move to which it belongs” (Biber et al., 2007: 24). Moves and steps mainly differ in that interpretation of a given text fragment at the step level is usually articulated in more specific terms (e.g. ‘indicating a gap’) than at the move level (e.g. ‘establishing a niche’).

Swales’ original motivation for developing this text analytical scheme was to help advanced students for whom English is not their first language to improve their reading and writing of RAs in English. Many researchers have applied versions of this method of analysis in order to uncover the underlying generic structure of not only RA sections but also many other academic, professional and general genres (see a review in Biber et al., 2007). A major aim of these move analysts has been the identification of the linguistic features characterizing the various RA rhetorical moves (e.g. Cortes, 2013; Cotos et al., 2017; Kanoksilapatham 2005; Le and Harrington, 2015; Swales, 1981), often for pedagogic purposes. We refer to this research gap as ‘the function-form gap’. From our applied genre perspective, filling this gap involves establishing the most salient types of text items, or patterns, occurring in a specific rhetorical context in an RA, or any other genre, that may lead a competent reader to interpret a given communicative function in a highly predictable manner. This goal is applicable to all text fragments realising a relevant communicative function except, of course, when the function is not signalled by any specialised text item, or pattern, as is the case of implicit, or inferred, causal logical functions (see Moreno, 2003a: 119, 138).

The data shown in recent studies of RA generic structure (e.g. Annuai and Wannaruk, 2013; Cotos et al., 2017; Yang and Allison, 2003: 381-383) clearly suggest that the step is a more appropriate level for investigating the function-form gap than is the move. However, the field has still some way to go in this respect, especially in languages other than English, due to the paucity of large-scale corpora of RAs reliably annotated at the step level (cf. Cotos et al., 2017; Del Saz Rubio, 2011), and this despite all the technological advances now available (e.g. Anthony, 2003). Furthermore, it still remains unclear which is the minimal formal unit for annotating moves-and-steps (cf. a proposition, in Connor and Mauranen, 1999; the sentence for moves and the phrase (or clause) for steps, in Cotos et al., 2015, 2016; or a sentence or paragraph, in Crookes,

1986), and whether functional interpretation best proceeds top-down or bottom-up. In this context, a group of experienced EAP researchers drawn from a number of Spanish universities was set up in 2010 as the ENEIDA¹ Team. One of their goals was to annotate a large sample of RAs reliably, giving priority to the identification of steps as functional coding units. So far, the team have proposed working move-and-step schemes for all the empirical research articles (ERA) sections in a wide range of disciplines and two languages.

The major aim of this paper is to reflect on the challenges faced by the ENEIDA annotators, or coders (see acknowledgments), in developing such move-and-step schemes for annotating ERAs at the step level as well as on the solutions adopted to improve reliability and validity. In the next section, we comment on the evolution of the move-and-step concepts, briefly introduce the aims of the ENEIDA Project followed by relevant results obtained so far, and explain why we choose to use the Discussion section to illustrate the kind of challenges faced in the process of annotation.

2. Move analysis and the ENEIDA Project

2.1. Move analysis

By the time of *Genre Analysis* (Swales, 1990), several things were becoming clearer about move analysis, if they were not yet explicitly stated. First, a move was a rhetorical construct, the linguistic realization of which could be as short as a clause and as long as a paragraph (and/or sometimes repeated in cycles). Second, the function of a move was realised by the presence of one or more specific functions, or steps (Swales, 1990: 141). Third, the identification of move boundaries (i.e. the text items signalling the beginning of a move, or the transition from one move to the next; see also Paltridge, 1994: 296) could be uncertain, but was aided by a combination of bottom-up search for lexical or syntactic signals and a top-down close reading of the text for topic breaks or shifts in content. Fourth, there was a place for specialist disciplinary experts to verify the analysts' interpretations, given their deeper knowledge of the text subject matter and their stronger intuitions regarding the typical rhetorical structure and language used in good papers in

¹ Equipo Nacional de Estudios Interculturales sobre el Discurso Académico (Spanish Team for Intercultural Studies on Academic Discourse).

their fields (e.g. Tarone et al., 1981). More recently, triangulation has typically involved interviews (sometimes text-based) with various participants, very often authors, but also including editors, reviewers and expert disciplinary writers (e.g. Hyland, 2012). Fifth, following Crookes (1986), there might be a place for additional analysts (or raters) who could confirm the findings of a primary investigator, their required training being open to question.

A challenge to these emerging procedures was provided by Paltridge (1994). He concludes that “Hasan, Bhatia, Swales, and Crookes, thus, all draw essentially on categories based on *content* to determine textual boundaries, rather than on the way the content is expressed *linguistically*” (original emphases) (Paltridge, 1994: 295). However, he does not discuss instances where linguistic features can indeed be seen by the reader as signalling a rhetorical shift, as with adversatives plus negative or quasi-negative language to signal a research gap (Swales, 2004: 229). More recently, Pho (2008) also questioned the standard combined procedures (e.g. Swales, 1990; Kanoksilapatham, 2005), arguing that identification of moves based on bottom-up linguistic signals and top-down content analysis leads to a certain circularity of reasoning. However, relating both kinds of evidence is a key element in hermeneutic methods, which Geertz characterizes as “a dialectical tacking between parts which comprise the whole and the whole which motivates the parts, in such a way as to bring parts and the whole simultaneously into view” (Geertz, 1980: 103). A different perspective on a combined procedure can also be inferred from Flowerdew’s (2002) reflection on the issue:

Although I refer to identification of schematic structure as the first stage in genre analysis, this is an idealization for the purpose of exposition. In actual fact, various interrelated levels of analysis go on at the same time: identification of communicative purpose(s), schematic structure, grammatical features, lexical features, etc. (p. 95)

In general, it would seem that Flowerdew’s approach is the one adopted, either overtly or covertly, in most move analyses (see also Bhatia, 2001; Nwogu, 1990). However, little information is typically provided about the identification processes followed by researchers (e.g. Hopkins & Dudley-Evans, 1988; Peacock, 2002) so that they can be accurately replicated. In any case, as it turned out, the ‘dance’ metaphor of moves and steps has emerged as a fortunate choice as it coincides well with later conceptions of

genre, especially following Frow (2006), wherein instantiations of genre are seen as active performances rather than more passive replications of classes of text.

One shared assumption about move analysis is that it needs to be done manually (Biber et al., 2007: 33), since interpretation of communicative functions is a cognitive task difficult to access and operationalize. In effect, this assumption partly explains why most of the analyses done to date typically deal with small corpora (exceptions are mainly recent, e.g. Cotos et al., 2015, 2017; Peacock, 2002), as manual analysis is time-consuming, resulting in what some consider to be a shortcoming in move analysis, the lack of elaborate quantitative studies (Biber et al., 2007: 39). A further obstacle is that none of the corpora of RAs in English fully annotated at the step level (e.g. Cotos et al., 2015, 2017; Del Saz Rubio, 2011) are available to external researchers. Not surprisingly in the last decade, a number of ‘corpus-based’ studies (Cortes, 2013; Kanoksilapatham, 2007; Le & Harrington, 2015) have attempted to bridge the function-form gap by using indirect methods (e.g. examining the functions of frequently occurring lexical bundles, or word clusters, or performing multidimensional analysis, respectively).

Although these indirect approaches indicate potential routes to bridge the function-form gap in RAs, they also raise other important concerns. The first one is that they are not able to distinguish between the most salient linguistic items, or patterns, helping readers to interpret a given step and those that simply co-occur in that function. For example, Cortes’ (2013) remarkable work on RA Introductions selects for study in functional terms those propositions characterised by the presence of lexical bundles of four or more words. However, it cannot be assumed that all the four or more word lexical bundles identified are the actual signals triggering the interpretation of those functions. For instance, despite their proven correlation, it is not easy to perceive the inferential link between the lexical bundle “by the presence of the” (p. 40) and the function *Reviewing items of previous literature* in an RA Introduction. The same could be stated about “at the same time” (p. 40) in relation to *Summarizing methods*.

The second concern is that the criteria often used for selecting keywords as a basis for investigating the lexical bundles, or clusters, make them unsuitable for an exhaustive investigation of the function-form gap at the step level. For instance, as Cortes’ study departs from taxonomies of linguistic items developed by previous studies using similar techniques, it leaves open the question of whether her corpus might contain other ways

of signalling steps that may have gone unnoticed in the previous studies (see Moreno, 1998; Parkinson, 2011 for evidence to substantiate this type of concern). In their turn, Le and Harrington (2015) study certain clusters in Discussion sections and their association to the steps in the *Commenting on results* Move. As the authors explain, for their selection of relevant keywords, they excluded all items that were not nouns or verbs. However, as is well known, certain logical functions such as *Explaining* may also be signalled by other types of text items (e.g. subordinators, see Moreno, 2003b: 276; Parkinson, 2011). In fact, as Lim (2010) shows, one of the commentary steps identified in the Results sections, *Explaining the findings*, which is also present in Discussion sections (Yang & Allison, 2003), is characterised by the use of “because” and “since”.

A third concern has to do with the methods for obtaining the lexical combinations based on selected keywords in order to study their association with given steps. In particular, focussing on lexical bundles of four or more words (as in Cortes, 2013) would overlook smaller word combinations that might also be salient in the identification of certain steps. For instance, one way of signalling the function *Pointing out negative features or limitations of the current study* in an RA Discussion section is by combining the exclusive pronoun “we” with the single item “lack*” as in “We also lacked information on potentially useful clinical variables” (our own data). Thus, restricting a study to longer word combinations would have not identified this formal pattern simply because it involves just two lexical items.

In its turn, Kanoksilapatham’s (2007) multidimensional study of move type-linguistic feature correlations in biochemistry RAs also raises issues about the specificity of conducting a multidimensional analysis of such overarching moves as *Consolidating results* (her Move 13), which include a great variety of steps, and, not unexpectedly, many kinds of linguistic realisation. Taken together, the reviewed studies suggest the need to keep annotating corpora manually (or hopefully in future in a computer-assisted manner) at least when the researchers’ goal is to identify the rhetorical steps in a genre and the most salient signals leading to their interpretation. As Cortes (2013) recognises, “such a corpus [i.e. tagged for moves or steps] would facilitate the analysis of any type of linguistic or organisational feature” (p. 37).

Finally, move analysis has been employed as a framework for investigating cross-disciplinary (Basturkmen, 2012; Cotos et al. 2015, 2017; Holmes, 1997) and cross-

cultural variation (Amnuai & Wannaruk, 2013; ElMalik & Nesi, 2008; Loi & Evans, 2010; Yakhontova, 2006) in the generic structure of various RA sections. As it happens, in her comparison of RA discussions in Applied Linguistics and Dentistry, Basturkmen (2012) obtains interesting cross-disciplinary differences “at step and substep levels (rather than move level)” (p. 143). Amnuai and Wannauruk’s (2013) study of RA applied linguistics discussions published in English-medium international and Thai journals also finds no difference in the occurrence of such a complex move as *Commenting on results*, whereas their data about steps within that move indicate some degree of cross-cultural variation. For their part, ElMalik and Nesi (2008) conclude that, although medical research articles written in English as L1 by British writers and in English as L2 by Sudanese writers show the same moves, the distribution and realisation of steps within moves varies across these two groups. Once again, taken together, all these findings suggest that move analyses might obtain more revealing results if RAs were also annotated at the step level. Thus, information about the segmentation and labelling procedures used should at some point be provided in greater detail than is usually offered so that studies may be replicated and their reliability proven. To see how this might be done, we now turn to the ENEIDA Project.

2.2. *The ENEIDA Project*

The ENEIDA project (Ref.: FFI2009-08336/FILO) (2010-2014) has collected data from multiple interrelated sources to investigate Spanish researchers’ needs in English for research publication purposes (ERPP) as well as their motivations, attitudes, writing difficulties and possible strategies that may help them to get published in English-medium international journals (Moreno et al., 2011). In essence, its ultimate aim is to design useful ERPP pedagogic resources from cross-cultural and intercultural perspectives for Spanish researchers, drawing on results obtained through a combination of qualitative corpus-driven, quantitative corpus-based and ethnographic methods utilised in three different types of studies: Needs analyses (Phase 1), cross-cultural studies (Phase 2), and intercultural studies (Phase 3).

Analyses of the responses from 1717 Spanish post-doctoral researchers from four universities and one research-only institution in Spain surveyed through the ENEIDA questionnaire (Moreno et al., 2013) yielded, among others, the following results in Phase 1. First, it was found that the ERA is the most relevant publication genre in all knowledge

areas surveyed (Moreno et. al, 2011). Second, although nowadays Spanish researchers are highly motivated to publish their ERAs in English (López-Navarro et al., 2015), it is evident that they are having difficulty in getting them accepted by international Anglophone journals, especially due to writing issues (e.g. Martín-Martín et al., 2014). Not surprisingly, the informants also express their willingness to receive training in writing for publication in English (e.g. Burgess et al., 2014). Third, the aspect of writing they are more concerned about is the linguistic expression of specific communicative functions, such as “ways to clearly express my interpretation of the results of my study,” or “strategies to express the relevance of my contribution to the field more clearly” (e.g. Gea-Valor et al., 2014: 56). As can be seen, those concerns are articulated at the step level, and also refer to the phraseology employed to express the steps. Finally, for the ENEIDA survey respondents, an important goal of future training sessions should be to help the participants achieve a better understanding of the differences between versions of ERAs in English-medium and Spanish-medium journals (e.g. Gea-Valor et al., 2014), investigated in Phase 2. This may be partly because, as Nwogu (1997) observed twenty years ago, “most research article writers are familiar with the IMRD format, but not all are conscious of the fact that there exists an internal ordering of the information presented in the various sections of the research article” (p.119).

To able to compare the generic structure of ERAs across these two languages before analysing the particular phraseology signalling each step, the methodological framework of this phase has involved the following sub-goals. First, creating two suitable corpora of ERAs published in international journals in English and in Castilian Spanish-medium journals. Second, designing a clear protocol for segmenting all the sections into coherence units from the ‘step perspective.’ Third, developing a codebook for annotating the various ERA sections. Given the methodological challenges faced in achieving such sub-goals in texts from a wide range of disciplines and two languages, the present paper will only focus on the challenges and the corresponding solutions to annotate the corpora for their steps (and moves) so that other EAP researchers pursuing similar goals may take them into account. In other words, despite the pedagogic orientation of the ENEIDA project, we will not attempt to offer specific quantitative results regarding the identified steps or their signals, nor will we show how such results can be translated into pedagogic resources. These issues will be for other papers. We now explain why we have chosen the Discussion section to illustrate the ENEIDA coders’ procedures.

2.3. *The ERA Discussion section*

The first reason for our focus on this section is that the Discussion section of ERAs has proved particularly problematic for EAP scholars, as reviews by Amirian et al. (2008), Basturkmen (2012), Cotos et al. (2016), Peacock (2002), and Swales (2004) demonstrate. In our view, one problem lies in the great variability of move conceptualisations proposed by existing empirically-based models about this part-genre in English (e.g. Basturkmen, 2009, 2012; Berkenkotter & Huckin, 1995; Cotos et al., 2016; Holmes, 1997; Hopkins & Dudley-Evans, 1988; Kanoksilapatham, 2005, 2007; Nwogu, 1997; Peacock, 2002; Yang & Allison, 2003). First, it is striking that the number of moves varies from three in Berkenkotter and Huckin (1995) through seven in Yang and Allison (2003) to eleven in Hopkins and Dudley-Evans (1988). Second, there seems to be less consensus on how to conceptualise the moves than the steps. For instance, half of the reviewed studies (Basturkmen, 2009, 2012; Holmes, 1997; Hopkins & Dudley-Evans, 1988; Yang & Allison, 2003) label the first move as *Background information* or *Information* (in Peacock, 2002). In contrast, the other half choose to conceptualise this move very differently, as can be deduced from the label chosen (e.g. *Occupying the niche*, in Berkenkotter & Huckin, 1995; *Re-establishing the territory*, in Cotos et al., 2016; *Contextualising the study*, in Kanoksilapatham, 2005, 2007 or *Highlighting overall research outcome*, in Nwogu, 1997).

It could be argued that these different conceptualisations may have been related to the nature of the disciplines under study, as all researchers in the first group except for Hopkins and Dudley-Evans (1988) analyse social science Discussion sections, while most researchers in the second group deal with natural science Discussion sections. However, they seem to be more related to the researchers' decisions to apply either Hopkins and Dudley-Evans' (1988) original model or to mirror Swales' (1990) ecological metaphor. In fact, each of the studies analysing a wider range of disciplines (Cotos et al, 2016; Peacock, 2002) follow one or the other tradition. Interestingly, when studies go down to the step level their way of conceptualising the function of fragments is more similar, as can be deduced from the labels they use. For instance, one of the Move 1 steps is labelled as *Reference to previous research* (Holmes, 1997; Hopkins & Dudley-Evans, 1988; Peacock, 2002), *Referring to previous literature* (Kanoksilapatham, 2005, 2007); *Background about theory* (Peacock, 2002) or *Drawing on general background* (Cotos et al., 2016). Another step is labelled as *Summarising the study* (Yang and Allison, 2003);

Re-stating research aims/methodology (Peacock, 2002); *Presenting generalisations, claims, deductions or research gaps* (Kanoksilapatham, 2005, 2007); *Drawing on study specific background* (Cotos et al., 2016). This generally suggests that steps may be better indicators of shared psychological realities than moves and that annotating Discussion sections for their steps before conceptualising the moves might help us to arrive at a clearer picture of what is happening in this part-genre.

The picture becomes even fuzzier after contributions from cross-cultural studies, which tend to apply existing models, are taken into account. One group of studies has compared Discussions written in English by either “proficient” or “native speakers” to those written in another language as L1 and in English by “non-native speakers” (e.g. Persian in Amirian et al., 2008; or Spanish in Williams, 2011). For instance, Williams’ (2011) study of the factors affecting the structure and discourse style of the Discussion section in biomedical empirical articles reveals that the most influential factor is the type of study undertaken, producing two distinctive discourse styles. These circumstances suggest the need to take into account the study type factor in new corpora compilations.

Another group of studies have compared the generic structure of ERA Discussions written in English by “native” speakers to those written in English as L2 by speakers of another language (e.g. Thai in Amnuai & Wannaruk, 2013; Sudanese in ElMalik & Nesi, 2008; or unspecified languages in Peacock, 2002). These studies as a whole identify rhetorical variation at the step level and suggest the influence of cultural factors. For example, ElMakik and Nesi (2008) found that the Sudanese writers did not always make use of such steps as *Stating specific outcome*, *Interpreting outcome*, *Indicating significance of outcome* and *Contrasting recent and previous outcomes* in Move 10 (*Explaining specific outcome*), while they appeared to be obligatory for the British writers. It is difficult to compare their results with those of others as this study, which replicates Nwogu’s (1997) model, seems to understand some of the steps in less standard ways. Furthermore, the authors’ interpretation of some steps does not seem to be the same as that intended by the proponent of the model, raising reliability issues. For instance, the examples given in ElMalik and Nesi (2008: 90) and Nwogu (1997: 132) for the same step *Stating specific outcome* seem to be different in nature.

Another reason for focussing on the Discussion section is our concern about the usefulness of the explanations given by certain studies to the cross-cultural variation

identified. For instance, ElMalik and Nesi (2008) explain that their Move 10 steps “require writers to take an independent view, only supported by the strength of their own argument, an approach well understood in individualist cultures which promote self-expression and personal choice but arguably at odds with accepted behaviour in a collectivist society” (p. 93). Although this kind of speculation seems plausible for some of the steps, in our view, it might not apply to all of them as a block and it would therefore be pedagogically more useful to understand the reasons for cross-cultural variation at each individual step.

The third reason for choosing the ERA Discussion section is that there is mounting evidence that this section is often the most difficult for Spanish authors to write (Moreno et al., 2012; Perez-Llantada et al., 2012). This has been explained by the fact that it involves “mastery of rhetorical strategies for conveying interestingness, motivation and promotion-related elements and, above all, foregrounding the significance of the research while being aware of the provisional nature of the scientific claims” (Perez-Llantada et al., 2012: 124). Glasman-Deal (2009) also perceptively comments that, whereas the typical Introduction is designed to draw the reader into the research topic and thereby appreciate its significance, in the Discussion, the purpose is to lead the reader beyond the topic by showing its significance for the research world and often the “real” world outside it. It is not surprising then that editors regularly note (C. Feak, B. Paltridge p.c.) that adequate pieces of local research often fail to get final acceptance in international journals because the authors cannot successively project the potential value and interest of their studies for the wider readership.

Of course, having an appropriate structure and content for a Discussion section is not a sufficient condition for success. Many other factors can lead to a manuscript’s rejection, such as flawed methodology, egregious claims, inadequate treatment of previous research, and inappropriate phraseology, among others (see Moreno et al., 2011). In other words, the kind of results obtained from analysing the generic structure of ERAs offer a road map, but this road map does not in itself ensure successful navigation of the route. Yet, after analysing the moves and move cycles used in 252 RA Discussions written by “NS” *versus* “NNS” writers, Peacock (2002: 483) suggests that “the most probable reason for the differences may be found in the suggestions of Vassileva (1997) and Yakhontova (1997) that NNS research writers have difficulty with genre conventions that differ from their L1” [i.e. differences in academic writing and genre conventions, respectively]. This

supports our future goal of determining in what sense and to what extent the generic conventions of Discussion sections vary across English-medium international journals and Spanish-medium national journals.

As a whole, the present review suggests the need to create a comprehensive move-and-step scheme for Discussion sections that allows researchers to annotate cases more reliably, compare generic structure across contexts more meaningfully and explain cross-cultural variation in more useful terms. This goal could be best achieved if we first described what writers actually do with their words at the step level when they are writing their RA Discussion sections in as transparent a manner as possible, seeking to explain why writers have made certain decisions. To this purpose, a comprehensive scheme with well-defined categories was first developed to code the Discussion section as part of the ENEIDA project. Then, this scheme was taken as a model for the rest of the sections, to maintain the internal consistency of the coding system. This is the fourth reason why we use examples from the Discussion section to illustrate the procedures followed in all sections.

We now review and reflect upon the unique methodological challenges faced by the researchers involved in Phase 2 of the ENEIDA Project. First, there is the issue of compiling a corpus, either on paper or electronically, in such a way that it can produce reliable, sustainable and comparable results that can be transferred to pedagogic resources (section 3). A second issue is how best to carry out a structural analysis of the genre exemplars, partly with regard to segmenting the text into meaningful units and deciding upon and labelling the communicative functions therein, and partly with regard to deciding the level of detail and elaboration in the model (section 4). The third major issue in such an enterprise is what roles can and should be played by independent analysts (section 5) and by specialist (or content) informants (section 6), and what can be done by way of selection and/or training of analysts and informants.

3. Compiling relevant corpora of research articles

The choice of corpora for ERPP has been a topic of some debate for some years, but consensus now seems to have largely coalesced around a preference for specialized corpora (e.g. Flowerdew, 2004; Nesi & Gardner, 2012). In effect, this has come to mean

corpora that are constrained to particular genres or part-genres, often with sub-corpora devoted to different disciplinary areas (e.g. Hyland, 2004). Difficulties and uncertainties do increase when cross-cultural and cross-linguistic comparisons are made of, say, articles in English with those in another language (Connor & Moreno, 2005), especially given recent trends towards the Anglicization of the Research world (Cargill & Burgess, 2008; Perez-Llantada, 2012). Indeed, apart from a few pockets in the humanities, there are probably today no *international* peer-reviewed non-Anglophone journals left (Lillis & Curry, 2010: 14–16). For example, the venerable and highly-respected German language journals in the natural and medical sciences all switched over to using the English language some time ago (Ammon, 2001). In Spain, the situation in the natural and medical sciences is similar (Martín-Martín et al., 2014; Moreno et al., 2011), and has also recently affected the social sciences. For instance, although 71% of the social scientists responding to the ENEIDA survey still see the relevance of publishing their results in Spanish-medium journals, they often do not do so due to the lack of high impact factor journals in Spanish (Gea-Valor et al., 2014).

While it is fashionable to compare high-impact and prestigious English-language journals with whatever leading journals there may be in another language, there is likely to remain some imbalance in readership, citational uptake, rejection rate, reviewer and editorial scrutiny, etc. between the two sub-corpora. Although the alternative of choosing small, regional English-language journals would go some way toward rectifying this imbalance, such Anglophone minor publications would not probably provide useful models for those EAL researchers hoping to ‘break into the big time’. Admittedly, publishing in a small journal may well be what graduate students *qua* graduate students can reasonably aspire to, but in today’s competitive research environments, such ‘waystage’ products are unlikely to be what the consumers of projects like ENEIDA are looking for.

Being aware of the difficulties with compiling ERAs in Castilian Spanish, especially in the experimental disciplines, the team set a reasonable goal of 30 sufficiently comparable pairs of ERAs in English and Castilian Spanish. The decision about the corpus size was partly based on Cortes’ (2013: 42) suggestion that the more focused or restricted the corpora are, the more frequent the lexical bundles are likely to be. Given the proven correlations between lexical bundles and steps (Cortes, 2013), the frequency of the targeted set of steps was also expected to be relatively higher in well-focussed corpora. For instance, the frequency of fragments pointing out a limitation of the current study is

much more likely to be higher in a corpus of Discussion sections than of other ERA sections. For this reason, focussing on the Discussion section would require a smaller number of texts to identify a ‘good’ sample of the targeted step than if we focussed our attention on any other section. The same rationale would apply to the other sections.

3.1. Choosing comparable corpora for pedagogic purposes

The pedagogic orientation of the ENEIDA project also led the team to establish that the ERAs in the corpora for the cross-cultural studies should represent good matched samples of writing both in English-medium international journals and in Castilian Spanish-medium journals. Who better than expert disciplinary informants to recommend good samples of comparable ERA pairs? Consequently, our selection procedure consisted in emailing targeted researchers from the five institutions participating in the study. This email (2013) explained to them our research and pedagogic goals and asked them to recommend well-written comparable ERA pairs in their own fields to the team (see also Cotos et al., 2017). The result was that 54 researchers from the targeted knowledge areas, i.e. Life sciences (LFS), the Healthcare Sciences (HCS), the Social Sciences (SCS) and Other Natural Sciences (ONS), eventually sent their proposals by email and filled in an online questionnaire about various aspects of their recommendations, including their perception of their quality on a five-point Likert scale. To orient the informants in their search, they were requested to recommend, among other article types, ERAs that:

- Reported empirical studies, one originally in an English-medium international journal and one in a Castilian Spanish-medium journal, published between 2000 and 2010 by researchers other than themselves in journals in which they, or their colleagues/graduate students were considering publishing in the future or had considered suitable for publishing in the past;
- Were perceived as good samples of writing irrespective of the impact factor of the journal; and
- Were comparable in terms of overall topic and study type (e.g. experimental, descriptive, survey, comparative, longitudinal, multiple case study...), with neither being a translation of the other article in the pair.

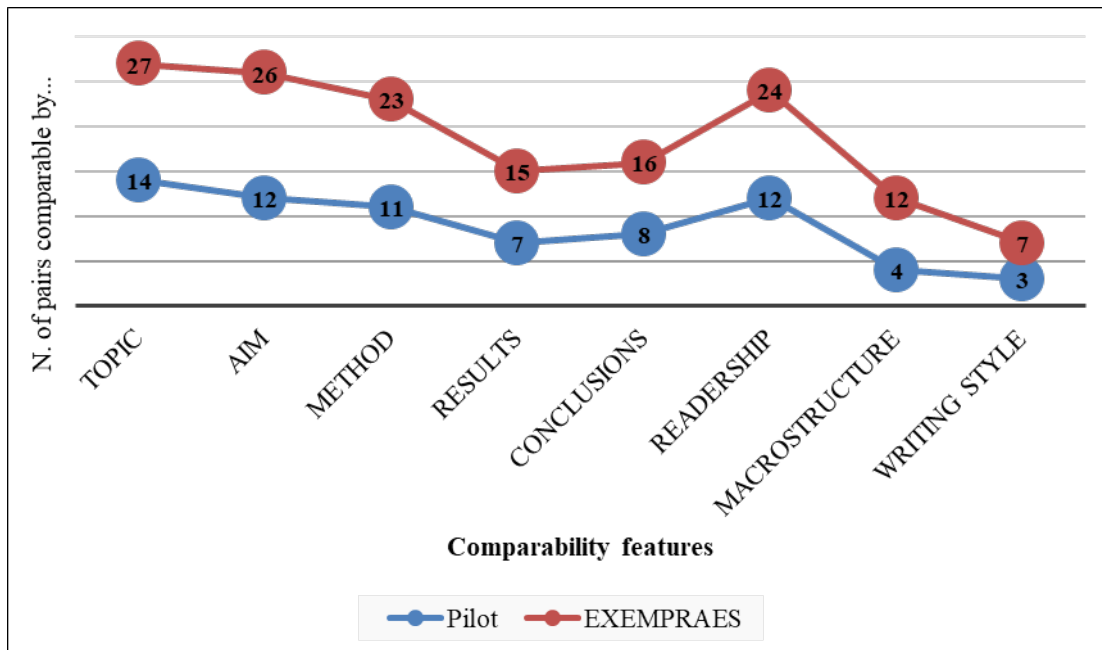
These requirements constitute the features characterizing the ENEIDA target population of ERAs for the cross-cultural studies. However, after a screening process, some proposed pairs were excluded for the following reasons:

- for not being empirical (e.g. review articles);
- for one being a translation of the other;
- for the English-medium international journal article being authored exclusively by a Spanish author, as this was the type of article compiled for Phase 3 of the project;
- for the collaborator being a co-author; or
- for being of poor digital quality, making it difficult to convert the text into .txt format automatically.

The potential informants had been sent guidelines as to what was considered ‘good’ scientific writing. These guidelines basically referred to widely accepted principles of scientific and general communication such as containing the expected types of sections, orderly content, accurate citations, logical argumentation, relevant concepts expressed with clarity, etc. The idea behind these rather basic principles was that they could serve to remind the informants what the ENEIDA Team was looking for.

3.2. The EXEMPRAES Corpus

The resulting corpus was named Exemplary Empirical Research Articles in English and Spanish (EXEMPRAES). It comprises 32 pairs of ERAs from 2000-2010: eight from the LFS, seven from the HCS, eleven from the SCS and six from ONS. The red line in Graph 1 below indicates that the articles in the vast majority of pairs were perceived comparable in terms of overall topic and aim (26). The collaborators also considered the recommended articles comparable in terms of either overall topic (27) or aim (26), plus at least one of the following general aspects of an RA the questionnaire asked them about: the type of target readership, the type of research method, the type of conclusions, and the type of results. We used their answers to figure out which general aspects could best reflect the informants’ own conception of the ‘study type’ notion (Williams, 2011). As Graph 1 shows, this notion seems to correlate best with the overall topic (27), type of aim (26), type of target readership (24) and type of research method (23). Interestingly, the recommended pairs were perceived much less comparable in terms of general macrostructure (see section 4.1) and writing style.



Graph 1. Comparability of the pairs in the EXEMPRAES and the pilot corpora

3.3. The pilot sample

The coordinator then selected a stratified pilot sample of 15 pairs to be used at various stages of the piloting. These stages are specified in the right column of Table 1. The word *design* indicates that the pairs in the corresponding rows were used in the design of the model; *validation*, that they were used in the validation of the coding scheme, either for pedagogic purposes or by members of the relevant discourse community; *training*, that they were used to train an independent analyst; and *Test 1* and *Test 2*, that they were used to perform the tests aiming to prove inter-rater coding reliability.

As can also be seen in Graph 1 (blue line), the pilot pairs show relatively similar levels of comparability to the whole sample. However, in drawing the pilot sample, the coordinator just aimed to strike the best possible balance in terms of knowledge areas (LFS=4; HCS=3; SCS=5; ONS=3) and section types (see Graph 2 below). Another challenging issue was the size of the pilot sample, which was resolved by performing the first analyses on the first 10 ERA pairs (see Table 1) and adding more pairs until it was found that the sample had been saturated. This saturation was judged to have occurred when no new rhetorical steps were identified after analysis of three new consecutive ERA pairs. In the case of the Discussion section, saturation occurred after analysing 15 pairs. These pairs thus constitute the pilot sample on which the initial qualitative cross-cultural move-and-step analyses were carried out.

Table 1. Use of the sample pairs from the EXEMPRAES Corpus in the study

N	Pair code	Discipline	Eng. ERA n	Sp. ERA n	Eng. D/C N	Sp. D/C N	Stages in the study
1	SCS4	Sociology	8635	10626	942	1200	Design/Validation
2	SCS3	Psychology	8137	5110	1395	1279	Design
3	SCS8	Applied Economics	12099	15135	1313	1674	Design
4	LFS8	Botanics (Relations)	7980	5058	2196	138	Design
5	LFS6	Zoology (Organisms)	4047	2323	1226	500	Design
6	HCS2	Endocrinology	4935	3706	1226	746	Design
7	ONS2	Chemistry	5180	4245	289	74	Design
8	ONS5	Agricultural Sciences	4652	4003	735	150	Design
9	LFS2	Molecular Biology (Functions)	4224	1697	1297	159	Design/Validation/Training
10	HCS4	Physiotherapy	4109	2687	935	508	Design/Training
11	SCS2	Pedagogy	5961	3924	521	1029	Test 1
12	HCS7	Ophthalmology	3131	1819	1238	650	Test 1
13	SCS11	Applied Economics	7200	8358	917	853	Test 2
14	LFS3	Botanics (Organisms)	13233	9270	4571	1925	Test 2
15	ONS4	Chemistry	6127	2812	583	619	Test 2
Total number of words			96881	80773	19412	11504	Pilot

To achieve greater levels of comparability of the moves and steps (Connor & Moreno, 2005), the team decided to establish the cross-cultural comparisons across functionally equivalent ERA sections. Each ENEIDA coder focused on the analysis and comparison of a different section (Moreno, 2014a). We now move onto the next problematic issue.

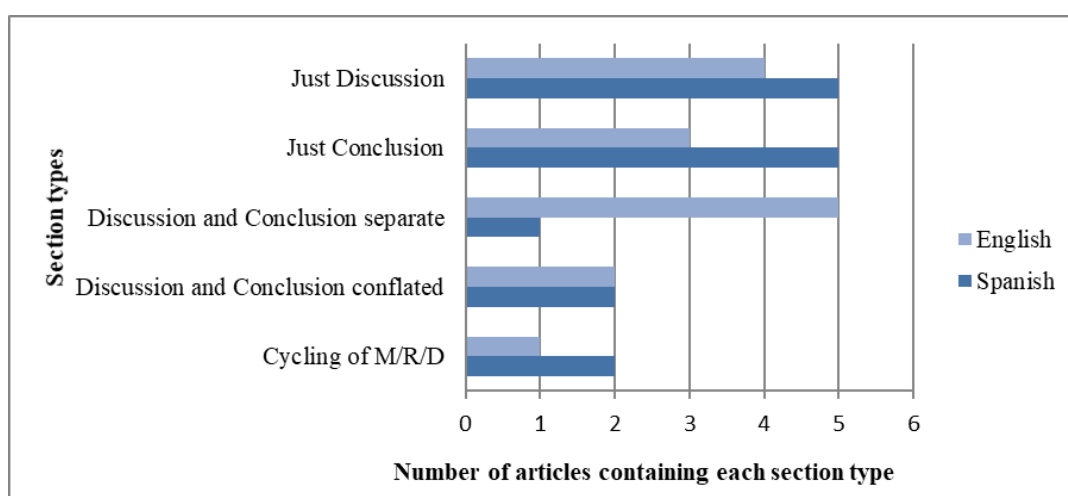
4. Carrying out a structural analysis of the genre exemplars

Carrying out a structural analysis of the genre exemplars to develop move-and-step schemes for all ERA sections involved three phases: subdividing the ERAs into functionally meaningful sections (4.1); segmenting the sections into meaningful fragments from the step perspective according to a segmentation protocol (4.2), and

categorising, or labelling, the segments obtained in each section according to specific but coordinated organising systems, or “coding schemes” (Tesch, 1990: 113) (4.3).

4.1. Subdividing the ERAs into functionally meaningful sections.

Usually the purpose of each ERA section was inferred unequivocally from the section heading: abstract, introduction, methods, results, discussion, and conclusion. However, on other occasions it was necessary to browse their content before this could be accurately specified. The major challenge during this phase concerned separating the Discussion section from other sections, given the great variability in the way ERAs tend to close after reporting results, a feature already observed by numerous previous studies (e.g. Swales, 1990; Yang & Allison, 2003). Indeed, in the EXEMPRAES Corpus, discussions do not always follow results in a separate section (cf. Lim, 2010) and precede other closing sections such as conclusions, but they may also conflate with results or with conclusions, or may occasionally be found in cycles together with methods and results in a final section. Thus, to achieve the maximum degree of comparability possible, it was decided that the Discussion and other closing (D/C) sections would be taken as a broad functional block (see Graph 2, excluding D/C sections conflated with results), on the assumption that they are not written to report results the first time around, but to discuss, interpret and conclude from the previously-reported results (Yang & Allison, 2003). This solution also turned out to be in accordance with Cotos et al.’s (2016) study.



Graph 2. Closing sections in the pilot sample from the EXEMPRAES Corpus

We now describe and comment on the stages followed in order to arrive at a coding scheme for the D/C section. In the design stage, the main analyst used the first eight article pairs to segment this section into meaningful fragments (see section 4.2). To develop a preliminary coding scheme, she interpreted their communicative function at the step level (4.3.1). Then the level of elaboration of the scheme, i.e. the number of mutually exclusive categories, was established for potential pedagogic purposes (4.3.2). After reducing the number of categories with the help of software (4.3.3), the main analyst performed an intra-rater reliability test based on pairs 7 and 8 (4.3.4), and settled on a working coding scheme, grouping the steps into moves (4.3.5). At the end of the design stage, she drafted the segmentation protocol (4.2.1) and codebook (4.3.6). Then, she used pairs 9 and 10 to train an independent analyst (5.1) and five more pairs grouped into two sets (11-12 and 13-14-15) to perform inter-rater coding reliability tests. In this process, various revisions and specifications of the codebook were performed (4.2.2 and 4.3.7) until good results were obtained (5.2). Finally, she used the English ERA of pair 1 again and the Spanish ERA of a new pair from EXEMPRAES to validate the coding scheme with members of the corresponding discourse community (6). We now move on to expounding the procedures used for identifying steps and establishing textual boundaries between segments.

4.2. Segmenting the sections into meaningful fragments

It has become standard practice to segment texts into moves, these being considered discourse units containing at least one proposition (e.g. Connor & Mauranen, 1999). However, some studies have equated this notion with grammatical units such as a sentence or paragraph (e.g. Crookes, 1986; Holmes, 1997; Hopkins & Dudley-Evans, 1988; Peacock, 2002) or have taken anomalous decisions by annotating “at sentence-level for moves and at phrasal-level for steps” (Cotos et al., 2015: 55). In the ENEIDA Project, following Swales (2004: 228-229), a special emphasis was made to treat moves as truly ‘functional’ rather than ‘formal’ units, and to annotate each and the same text fragment both at the step and move level. Thus, the main difference to Swales’ original account is that interpretation was situated primarily at the step level before going up to the move level.

4.2.1. The segmentation protocol

The purpose of the segmentation protocol was to make the rules for segmenting all the sections explicit. To be as systematic as possible, the main analyst broke each section into relevant text fragments, as in (1), where the text in parentheses after each segment indicates the analyst's interpretation (4.3.1). This interpretation was arrived at after reading the previous sections of the ERA (see justification in Basturkmen, 2012: 137). The analyst then continued with the process until she reached the end of the section.

(1) [Normal subjects were used]₁ (This is restating an aspect of the method: the participants used)/[to obtain a homogenous normal group for intra- and inter-therapist comparison,]₂ (This is justifying the previously-stated feature of the method)/[but whether the kinematics of manipulation are the same in normal subjects and patients is still unclear]₃ (This is evaluating the current state of knowledge in broad terms)/[and should be addressed in future studies.]₄ (This is making a recommendation for future research)/

As fragment (1) shows, segmentation was not necessarily sentence-based. This made the identification of certain boundaries more difficult. The main problems, illustrated in 4.2.2, arose when the fragments did not contain a verb or they were prepositional phrases, retrospective metadiscourse items, part of a citation or embedded in other meaningful segments. Additional challenges were faced when the propositions were perceived as relevant in relation to some neighbouring proposition rather than to the overall purpose of the section, or when they were not actually moving the text forward.

4.2.2. Dealing with fuzzy boundaries: extra segmentation rules

In view of these problematic fragments, the standard notion of a step had to be redefined in the following way to be able to apply the segmentation protocol consistently. A step is a text fragment containing 'new propositional meaning' from which a specific communicative function can be inferred 'at a low level of generalization by a competent reader of the genre' and is perceived as 'essential to advance the text' in the direction expected 'to achieve the purpose(s) of the (part-)genre in which it appears.' 'Propositional meaning' is understood as the text material referring to the world outside the text, which often occurs together with 'metadiscourse meaning', more concerned with the text and its reception, in the same sentence (Hyland, 2005: 40-41). A step can be realised by a proposition, a proposition complex or an even larger fragment of text. Thus, for a

fragment to be considered a step and be segmented separately, its new propositional meaning should be perceived as “essential to the movement [of the text]” (Hasan, 1984: 88), and contain at least one verb, whether finite, non-finite or elliptical, or a nominalization easily convertible into a verb phrase.

As can be seen in example (2) below, fragments (2.1), (2.2), (2.3), (2.4) are all propositions that contain a verb in a finite or non-finite form, whereas (2.5) contains a verb in an elliptical form that can be easily recovered: “all {*of which **are**} appropriate steps”. In all cases, the propositions are adding essential new propositional meaning.

(2) [The data from...**allowed** us to revise...]₁ (This is stating a positive feature of the study)/[by **adding**...]₂ (This is stating a contribution of the study)/ [and **rewording** items...]₃ (This is stating another contribution of the study)/ [as well as **develop**...]₄ (This is stating a further contribution of the study),/ [all * appropriate steps in the further development of the instrument]₅ (This is stating a positive feature of the study)

It might be argued that propositions (2.2), (2.3) and (2.4) could have been coded as just one fragment as they have the same functional value. However, each one includes new propositional meaning from which a communicative function can be inferred and, most importantly, our cross-cultural approach was interested in finding out how many statements of contribution were typically present in D/C sections across the two writing cultures. For this reason, each statement of contribution had to be coded separately. To overcome coding uncertainties such as those described above, the segmentation protocol created several rules deriving from such a redefinition of the step, as follows. Text fragments were not coded as separate if they were of the following types:

- A prepositional phrase

Where the prepositional phrase “in accordance with...” in (3) below does not constitute a proposition in itself.

(3) [**In accordance with prior research**, we found a strong positive relationship between...and...] (This is comparing a result with previous research)

In fact, the interpretation of this phrase in combination with the upcoming proposition leads the reader to infer the same step as that inferred from the proposition in (4), where the corroborative tie-in with other research receives more emphasis.

(4) [**This finding is in accordance with** other studies of similar patients [15...26]] (This is comparing a result with previous research)

- A retrospective metadiscourse item

(5) [**The results quoted** are all based on the data from...] (This is restating a key feature of the study)

Although the noun phrase “the results quoted” in (5) contains an elliptical verb in a reduced relative clause “the results {*which **have been**} quoted”, it is a retrospective metadiscourse item (Hyland, 2005) which does not add ‘new’ propositional meaning nor allows readers to infer any new relevant communicative function until it is processed in combination with the proposition it introduces.

- A part of a quotation or of a reported proposition

(6) [**Como sugería Barron (1963):** «si definimos la originalidad como..., y si definimos la inteligencia simplemente como..., entonces... Esto es, los problemas muy difíciles de resolver requieren una solución que sea original» (p. 219).] [Translation: (henceforth, Trans.): “**As suggested by Barron (1963):** «if we define originality as...and if we define intelligence as..., then... That is, those problems that are difficult to solve require an original solution» (p. 219).”] (This is making a comparison with previous research)

Where the whole fragment, including the evidential, i.e. (Trans.): “As suggested by Barron (1963):...” was coded as one only step since the whole block was perceived as just making one essential point.

Exceptionally, a few fragments were coded twice for technical reasons.

- Embedded segments

(7) [We suggest that there may be two reasons for this]₁ (This is announcing two possible reasons)./[The first could be that, [**although**...],₂ (This is providing background information without citations) it also facilitates knowledge leakage—]₃ (This is

speculating about one of the reasons)/[and this could pose a problem...]₄ (This is reacting to that reason)/

Where the embedded proposition (**although...**) in (7.2) was coded as part of the embracing unit (7.3), beginning at “The first” and ending at “leakage”, in order not to break the latter into two segments, but then it was coded as a separate unit to record its own function (see also Cotos et al., 2017). Luckily, this was rather infrequent and it did not affect the number of units coded but the number of words in the embracing unit.

However, some fragments could not even be ascribed using our redefinition of a step. Two major problematic types arose: announcing and elaborating functions.

- Announcing function (AF)

The function of these fragments was to announce rather than expound new propositional meaning. They typically included a prospecting item (Sinclair, 1993) whose full meaning could only be determined if the analyst went on reading across segment boundaries. They were used to announce various types of fragments:

- immediately upcoming (sub)sections. For example, (8) [**Discussion**], where the noun can be considered a nominalization easily convertible into a verb phrase: {We now **discuss** our results};
- external sources or discontinuous fragments. For instance, (9) [**For an analysis of ... see: Pešić et al. (2011)**];
- immediately upcoming moves/steps. For example, (10) [There are **several limitations to the present study**], announcing propositional meaning; (11) [Por último, resulta oportuno hacer **algunos comentarios** sobre este estudio] [(Trans.:) Lastly, it is relevant to make **a few comments** on this study], announcing a discourse function; and (7.1) [We suggest that there may be **two reasons for this**], announcing a logical relational function (Moreno, 2003a).

- Elaborating Function (ELF)

Other fragments were found ‘problematic’ because, although they were perceived as relevant elaborations of meaning, it was in relation to a neighbouring move/step without contributing to moving the text forward. In that sense, they were perceived as secondary to, or at the service of, some move/step rather than as steps themselves. Three types of elaborations were identified: justifications, as in (1.2), illustrations, as in (12.2), or

clarifications, as in (14) (supplementary Table 2, henceforth, Table 2S), beginning at “es decir,” (Trans.: “that is,”).

(12) [...in regions with low levels of social capital, it is necessary for firms to invest more in accumulating their own firm-specific social capital.]₁ (This is making a recommendation for future practice)/ [For instance, they can promote meetings, partnerships, and communication with other firms and organizations—both inside and outside the local region.]₂ (This is illustrating how the recommendation can be implemented)/

For this reason, unlike Hopkins and Dudley-Evans (1988), we did not consider announcing and elaborating functions as moves/steps. Nor did we code them as part of them (cf. Cotos et al., 2015), as this would have interfered with the team’s goals of bridging the function-form gap for future pedagogic application. This decision created a further problem: how to treat these fragments, which did not fit in well in a move-and-step scheme? The solution consisted of developing a more comprehensive model of communicative functions (4.3.4) in a way that allowed meaningful comparisons across the two writing cultures involved. We now move on to expounding the procedures followed for categorizing the meaningful fragments obtained, illustrated with examples from the D/C section.

4.3. Categorizing the fragments obtained

To categorise, or label, all the segments obtained reliably, each ENEIDA coder developed a codebook for each section in a coordinated manner. Each codebook included various elements: a) the hierarchical list of communicative functions, or coding scheme, with their labels, and codes (or abbreviations); b) detailed definitions for each communicative function; and c) examples of each communicative function in the two languages, with highlighted signals aiding interpretation, wherever possible. The kinds of information offered within each element of the codebook for the D/C section is illustrated in Appendix B.

4.3.1. Interpreting the communicative function of the fragments

In order to produce a preliminary coding scheme, a corpus-driven approach to text interpretation was taken, so as not to impose taxonomies of steps (or moves) from the previous literature (see section 2.3). This involved considering each meaningful text

fragment in context and interpreting its communicative function as specifically as possible to situate it at the step, or occasionally, (sub-)step level. This kind of pragmatic interpretation (Paltridge, 1995) was articulated using –ing verb forms after a second reading, written at the top of each segment (see fragment (1) above). This procedure yielded a very high number of categories: 60. As it was obvious that such a scheme would not be suitable for pedagogic purposes, the main analyst set out to group the 60 categories obtained into more overarching steps and these, in turn, into various moves according to their superordinate function.

4.3.2. Establishing the level of specification of the labels for pedagogic purposes

The next challenge, then, was to identify the most appropriate level of elaboration of the scheme that potential learners could handle, thus validating it for pedagogical purposes. To do so, the main analyst presented the basics of her segmentation protocol and preliminary coding scheme, together with the English ERAs in pairs 1 and 9 in the pilot sample, to the co-author of this paper. After analysing the texts independently, the two co-authors discussed their segmentation and labelling. Drawing on the second author's experience teaching and writing materials to teach this part-genre and on recommendations given by the qualitative research literature, they decided that the number of categories in the final coding scheme should not exceed 25. This implied not going down to the sub-step level, despite the relevance of doing so in some cases. For instance, two sub-steps (*Expressing agreement, similarity, consistency with previous research* and *Expressing disagreement or difference with previous research*) of the step *Comparing with previous research* displayed distinct semantic prosodies. For example, if we compare fragment 12 in supplementary Table 1 (henceforth, Table 1E), where the signal is "In accordance with prior research, we found..." to fragment 11 in Table 2S, where the signal is (Trans. :) "these results do not coincide with those obtained in the meta-analysis carried out by Kim (2005)," it can be noted that the polarity of one proposition is positive while that of the other is negative. Thus, in pedagogical contexts it may be relevant to contrast them. For a similar reason, it may be relevant to contrast the three sub-steps of the step *Explaining results and discussing effects: Explaining results, Discussing effects* and *Speculating about the reasons for results*.

4.3.3. Reducing the number of labels into a workable coding scheme

To make the task of reducing the number of labels easier, the main analyst organized the initial list into a workable coding scheme using the qualitative analysis software

NVivo10. To do so, the article pdfs were first converted into Word or .txt format and then cleaned. Cleaning the documents was an arduous task as it involved eliminating all features except for the core text, such as tables, graphs, photographs, page numbers, headers, reference lists and appendices. This was done by a technician, hired with project funding. The main analyst then imported the texts into the software and coded the same eight pairs of D/C sections in NVivo10 according to her preliminary coding scheme. Next, while annotating another two pairs, 9 and 10, she proceeded to re-read, regroup, resize, merge, relabel and define the communicative functions thus obtained, aiming to obtain mutually exclusive and self-explanatory categories. Finally, she explored the data from cross-cultural and cross-disciplinary perspectives in NVivo and merged those categories that did not yield apparent differences into more comprehensive ones.

4.3.4. Performing an intra-rater reliability test

To control for human error and check the internal consistency of her own annotations, one month later, the main analyst annotated the same pairs of D/C sections (9-10) once more within NVivo10. She then examined differences between her two sets of annotations, and revised certain aspects of the model. This was the end of the design stage. Later on, there were two inter-rater reliability tests (section 5.2) (the first being reported in Moreno, 2014b, and the second, in Moreno 2015) and a number of interviews with expert informants (section 6), leading to a few more revisions of the codebook. However, because of space limitations, we now describe the most recent.

4.3.5. The revised coding scheme of communication functions in the D/C section

The hierarchical list of functions identified in D/C sections can be seen in Table 2 below. The first column contains the abbreviations, or codes, adopted for each function, or label. The 25 categories of specific communicative functions (in lower case), which make up the revised coding scheme, are grouped into seven superordinate, or general, functions (in capital letters): announcing functions (at the top of the table); five move types, or ‘nuclear functions,’ in the middle of the table, and elaborating functions (at the bottom). The labels adopted for each specific function attempt to describe what writers are actually doing with their words at a low level of generalisation in D/C sections.

Table 2. Coding scheme of communicative functions in ERA Discussion (and/or other closing) (D/C) sections

Revision of the coding scheme in Moreno (2014b and 2015)

CODE	COMMUNICATIVE FUNCTION
AF	ANNOUNCING (FUNCTION)
SEC	Announcing (sub)sections
EXT	Announcing or referring the reader to external sources
MSP	Announcing moves, steps or propositional meaning
BGI	BACKGROUND INFORMATION
KFS	Re-stating key features of the current study
RWC	Reporting background information with citations
POC	Providing background information without citations
SUM	SUMMARIZING OR RESTATING KEY RESULTS
RES	Presenting results neutrally
CRES	Contrasting with other results in the study
HRES	Highlighting results
COMM	COMMENTING ON KEY RESULTS OR OTHER FEATURES
MEAN	Establishing the meaning of results
COMP	Comparing with previous research
EXPL	Explaining results or discussing effects
PRED	Making predictions
REACT	Reacting to results or other features
EV	EVALUATING THE CURRENT STUDY OR OTHER RESEARCH
LIM	Pointing out negative features or limitations of the current study
STATE	Evaluating the state of knowledge or practice in broad terms
CONTR	Stating the contribution of the current study
POS	Pointing out positive features of the current study
GAD	Noting specific gaps in knowledge or deficiencies in other research or practice
IMP	DRAWING IMPLICATIONS
REC	Making recommendations for future research or practice
APP	Suggesting the applicability of results or usability of outcomes
HYP	Hypothesizing for future research
ELF	ELABORATING (FUNCTION)
JUST	Justifying what is stated in a neighbouring proposition
EXEM	Exemplifying what has been stated in a previous proposition
CLAR	Clarifying what has been stated in a previous proposition

In Table 1E and Table 2S (see supplementary material), we illustrate how two entire D/C sections, one in English and one in Spanish, were segmented and labelled for pedagogic purposes using this version of the scheme. The accompanying labels indicate the general and specific communicative function performed by each segment, most of which would correspond to moves and steps, respectively. The signals leading to the interpretation of the various functions have been highlighted in bold face throughout each text. In the case of implicit functions, the recovered signals that would have led to each interpretation

more clearly have been introduced in brackets preceded by an asterisk, without implying that they should have been used in the actual texts.

Importantly, the present hierarchy of functions does not attempt to make statements about the order in which the various functions appear in D/C sections. The moves are presented in what seems to be a logical order of exposition (see ‘The Discussion in Essence’ Poem in Moreno, 2015) and the steps within the moves are in their likely order of frequency in English, based on preliminary quantitative results. However, their position in the D/C sections may be variable, as will be shown in a future study. Although the resulting list of functions may look similar in many ways to those offered by previous researchers as a whole, our scheme can be considered more comprehensive, as there is just one step, *Outlining parallel or subsequent developments* (Holmes, 1997), that we have not identified probably because the EXEMPRAES corpora do not include the discipline of History.

Moves conceptualizations in the present model are different to those made in previous models in that they are not formulated before identifying the steps (cf. Biber et al., 2007: 33). Instead, moves are constructs created at a later stage by the researchers to group similar steps according to their more overarching text function. Inspired by Hopkins and Dudley-Evans (1988), the moves in the model could be said to perform the following broad functions mainly: transactional (*Background information* and *Summarizing or restating key results*), interactional and logical (*Commenting on key results or other features*), interactional (*Evaluating the current study or other research*), and logical (*Drawing Implications*).

When compared with the models offered by those researchers who also develop the step level (e.g. Basturkmen, 2012, Cotos et al., 2016; and Yang & Allison, 2003), our model is different in a number of respects. First, it includes a series of relevant steps (see below) that are not distinguished or identified by Yang and Allison (2003) (the ones in italics) or that have not been identified by Cotos et al. (2016) either (the ones in bold). The functions include ***Reporting background information with citations***, ***Providing background information without citations***, ***Contrasting with results in the same study***, ***Highlighting results***, ***Making predictions***, ***Reacting to results or other features***, ***Evaluating the state of knowledge or practice in broad terms***, ***Stating the contribution of the current study*** and ***Noting specific gaps in knowledge or deficiencies in other research or practice***. We

believe that exploring most of these functions from a cross-cultural perspective will allow us to reveal insights with high pedagogical value.

Another substantial difference is that the present model has been expanded to embrace other types of communicative functions, such as announcing and elaborating functions, which also play a significant role in text construction. The reason why announcing functions have been placed at the top of the scheme is that they always come before the fragment announced. The only study that identifies an announcing function, *Previewing the discussion 'road map,'* is Cotos et al. (2016) (see also Cortes, 2013). However, Cotos et al. consider this function as a step (Step 4), a solution that would not have been consistent with our definition of move/step, since announcing units do not really add new propositional meaning and thus do not contribute to moving the text forward.

Elaborating functions are placed at the bottom of the scheme as they are considered at the service of moves/steps. It is true that previous models identify similar functions like *Justifications of soundness of results despite limitations* (see Cotos et al., 2016: 44, under their list of “content realisations”). However, we have preferred to dissociate these functions from any particular step/move, as they may be at the service of a great variety of moves/steps. For instance, in the EXEMPRAES corpus we have not only identified justifications “of the need for the future work recommended,” as Hopkins and Dudley-Evans (1988: 118) do, but also justifications of deductions, observations, limitations, aspects of the method, comparisons, restatements of key features, and of many more types of steps. Other researchers have also considered *Exemplification* and *Justification* (see Hopkins & Dudley-Evans, 1988: 118) or *Exemplifying* (Kanoksilapatham, 2005) as separate functions. However, they treat them as moves, which would have again been inconsistent with our definition of a move. Lastly, none of the reviewed studies has identified the elaborative function of *Clarifying*.

The present scheme also differs from others that keep elaborative functions, such as *Examples* and *Justifications*, within the scope of the unit they relate to (e.g. Nwogu, 1997; Yang & Allison, 2003). For instance, the example provided by Nwogu (1997: 133) to illustrate the step *Indicate the significance of the outcome*, i.e. “These results are particularly important because they come in the wake of several observations suggesting...” would have been segmented into two fragments, the second beginning with the subordinator “because”, in our study. The reason would have been that, in our view,

such a fragment performs two different communicative functions, each signalled by distinctive text items. While in the first proposition the writers are *Highlighting (these) results* as more remarkable than others by means of the items “particularly important”, the second one is *Justifying* such a step. Yang and Allison (2003) recognize that a more finely-tuned analysis than their own might be motivated by other purposes; and indeed our distinction is pedagogically relevant in writing contexts where authors are expected to justify their claims. Thus, keeping track of (un-)justified claims will allow the team to investigate them further. This kind of pedagogical rationale also explains our distinction between *Reporting background information with citations* and *Providing background Information without citations*, a distinction that is not made in previous studies.

Our detailed methodology has also led us to identify instances of steps that some previous models would have not. For instance, one full sentence classified by Nwogu (1997: 132) as *State a specific outcome*, i.e. “The current study **which was free of all temporal uncertainty and selection biases** indicated that...” (our emphasis), would have been segmented into two fragments, each with its own relevant communicative function, in our study. In particular, the whole sentence would have been interpreted as *Establishing the meaning of results*, and the relative clause (in bold face) embedded in it would have been interpreted as *Pointing out positive features of the current study* (see section 4.2.2).

Furthermore, our approach differs from others in that it gives priority to the function performed by each fragment irrespective of its position in the text (cf. Biber et al., 2007: 33). For instance, Basturkmen (2012) observes that some examples within Move 1 have a “strong promotional function.” Precisely for that reason, a segment like “The present study presents ‘real world’ data from general practice...” (Basturkmen, 2012: 138) would have been labelled in our study as *Pointing out positive features of the current study*, and therefore as Move 4 *Evaluating the current study or other research*, rather than as Move 1 (see also segments (2), (4) and (6) in Table 1E). For a similar reason, the function *Summarizing results*, identified by Yang and Allison (2003), has been merged into *Summarising or restating key results* in the current scheme. In particular, if we disregarded the position of the “*Summarising*” segments in the text, it was difficult to distinguish them from either those *Re-stating results* or those *Establishing the meaning of results* (cf. (26) to (8a) in Table 2S). In other words, our methodological approach has not only uncovered hidden cases, but it has also served to make the scheme more consistent and comprehensive, while simplifying it in some respects.

The rationale behind the choice of labels we have used in our model has been to facilitate the following goals in our research program relating to the reasons for likely cross-cultural and cross-linguistic variation at the step level. One, to compare the generic structure of D/C sections across English and Spanish meaningfully. Two, to discuss communicative functions with our informants efficiently. Three, to bridge the function-form gap better in the two languages. For instance, in relation to the first goal, the answers elicited from specific questions like “Why are you comparing your study with a previous study here?” will, in our view, be more revealing cross-culturally than the ones elicited by a general question like or “Why are you making this comment on your results here?” In relation to goal two, communication with informants will be more efficient (i.e. fewer explanations will be needed) if we use labels like *Background Information* instead of *Occupying the niche* to refer to the overall function of a step within such a move. In relation to goal three, it will be easier to demonstrate the inferential link between the lexico-grammatical realisation of most coding units and specific steps (e.g. *Re-stating key features of the current study*) than between them and moves (e.g. *Background Information*.)

4.3.6. *The codebook for the D/C section*

The codebook is a document consisting of seven pages (Moreno, forthcoming) (for an extract, see APPENDIX B). Most steps were easy to label according to the codebook. However, a number of cases were more difficult to analyse and annotate. In what follows, we show which decisions were eventually adopted to code difficult cases.

4.3.7. *Complex decisions in assigning steps: extra coding rules*

Some fragments caused ‘labelling problems’ for a variety of reasons: because of the rhetorical context in which they appeared, because they were giving rise to borderline interpretations or simultaneous functions, because they were containing misleading signals or not containing signals at all.

- Playing different functions in different rhetorical contexts

This problem was solved by considering the rhetorical context in which the step occurred. For instance, although the hypothesis in (13.3) looks like speculation, it was interpreted as a hypothesis for future research (IMP_HYP), as indicated by the retrospective item “another avenue for future research” in the upcoming segment.

(13) [In this paper, we focus on...]₁ (BGL_KFS)/, [an important element in...]₂ (EV_POS)/[but regional social capital **might influence** the effectiveness of...]₃ (IMP_HYP)/.[**Another avenue for future research** would be to...]₄ (IMP_REC)/

- Giving rise to borderline interpretations

This occurred when it was difficult to tell two functions apart. For instance, predictions were sometimes difficult to distinguish from hypotheses, as in (14).

(14) [In other words, out of 1000 lambs born alive, about 22 more lambs **are expected to** die if they are...than if they are...] (COMM_PRED)/

This problem was largely solved by improving the definitions (see Appendix B) or merging similar functions into one overarching category (see section 6).

- Giving rise to various communicative functions simultaneously

When this happened, these fragments were coded only once (cf. Cotos et al., 2015), according to their most prominent communicative function, following previous scholars who make similar decisions (Crookes, 1986; Holmes, 1997: 325; Yang & Allison, 2003). One problematic subset, where various interpretations were possible, is represented by fragment (15.2).

(15) [Aun así, el debate continua]₁ (EV_STATE) [y **aún falta realizar más investigación**]₂ (IMP_REC) [para...]₃ (ELF_JUST)

(Trans.:) [Still, the debate continues]₁ (EV_STATE) [and **there is still the need to carry out more research**]₂ (IMP_REC) [to...]₃ (ELF_JUST)

While the analyst interpreted (15.2) as *Recommending future research*, one co-author of the corresponding ERA said it was written to reflect on the state of knowledge, although he acknowledged that their intention was to promote future research. In any case, to solve this kind of problem, the analysts' interpretation was given priority over the writers' possible intentions (cf. Cotos et al., 2015).

A closer look at a small set of cases revealed that the problem was not with labelling multifunctional fragments but with segmenting them accurately (cf. Basturkmen, 2012: 138), as in (16), where (16.2) is an elliptical proposition equivalent to “in {*that they} show” that may receive step status.

(16) [With regard to ..., **our findings corroborate previous research**]₁
(COMM_COMP)/ [**in showing** a growing gap in ...,]₂ (SUM_RES)/

Thus, although the fragment looks similar to (3), it was interpreted as a rhetorical effort to both compare with previous research and to re-state the authors' own findings.

- Containing misleading signals or not containing signals of functions

Cases of misleading or absent signals were also solved by giving priority to the analysts' interpretation, as in (17), where, despite the use of the signal "concluded", the writers were not interpreted as *Establishing the meaning of results* but as *Presenting results neutrally*.

(17) [**We concluded** that mean EQ-5D score in Iranian patients with T2DM was 0.70]
(SUM_RES)/

In contrast, while the signal "En síntesis" (Trans.: "In sum") in segment (26) in Table 2S leads the reader to interpret this fragment as *Summarising results*, it is really *Establishing the meaning of results*. Similarly, although segment (42) in Table 1E does not contain any explicit signal and may look like a result, after consultation with the external informant, the fragment was interpreted as *Explaining results...*. This interpretation was easier to perceive by recovering a relevant implicit signal { *This may be because... }

5. Testing the model

The next stage was to check whether other co-analysts could apply the coding scheme reliably, qualifying them to annotate further exemplars of the same part-genre.

5.1. Selecting and training an independent analyst

As it was difficult to find experts in each of the fields willing to annotate texts in this fashion, the main analyst of the D/C section recruited an independent coder, who was a competent academic reader both in English and Spanish. A Spanish lecturer in English at a Spanish University (see acknowledgements), holding two BA degrees, one in Psychology and one in English, the independent analyst was not specialized in most of the disciplinary fields represented in the study. However, she at least had taken various courses in research methods and statistics during her first degree. She first took a crash course in NVivo10 to learn how to annotate texts and was then trained in annotating pairs

9-10 of D/C sections according to the codebook and segmentation protocol thus far designed. Difficulties encountered during this process led the main analyst to write specifications for solving difficult cases (4.2.2 and 4.3.7).

5.2. *Performing inter-rater coding reliability tests*

To achieve a good level of inter-rater coding reliability (Krippendorff, 2004), two tests were adopted. In both cases, analysts were coding the corresponding texts according to the working coding scheme for the D/C sections (Moreno 2014b; 2015, respectively), though they were also free to use an ‘other’ category. These two tests were performed by running a coding comparison query in NVivo10, yielding two kinds of coefficients for each category, or code, and section, or source text: the percentage agreement and Cohen’s Kappa coefficient. As Biber et al. (2007) explain, the percentage agreement is a simple statistic merely reflecting “the number of agreements per total number of coding decisions, but it does not account for chance agreement among raters” (2007: 35). The Cohen’s Kappa coefficient is considered more useful than the agreement percentage because it is a chance-corrected measure, which takes into account the likelihood that the agreement between coders has occurred by chance (see also Kanoksilapathan, 2005; NVivo10 help, 2014; Orwin, 1994). To calculate average figures for each code across the sources, weighing them by their size, Excel was used following instructions from NVivo10 help (2014).

A very high level of agreement in average percentage terms was obtained in Test 1 (above 97%). However, the average Kappa coefficient was ‘poor’ (below 0.40) on six of the categories (Moreno, 2014b). Thus, differences in coding on those categories were discussed and minor aspects of the model revised. For instance, a new category was added: IMP_HYP (cf. Moreno, 2014b; 2015). Table 3 displays the average Kappa and agreement coefficients obtained for steps in Test 2. The announcing and elaborating functions led to 100% agreement.

Table 3. Agreement level after coding the steps in the D/C sections of the ERAs in the pilot EXEMPRAES Corpus by two independent coders (Test 2)

Communicative function₁	Average Kappa	Agreement (%)	A and B (%)	Not A and Not B (%)	Disagreement (%)	A and Not B (%)	B and Not A (%)
BGI	0.80	98.35	3.58	94.77	1.65	0.51	1.13
KFS	0.70	99.74	0.30	99.44	0.26	0.04	0.22

RWC	0.88	99.08	3.63	95.45	0.92	0.37	0.55
POC	0.50	99.54	0.23	99.32	0.46	0.14	0.32
SUM	0.74	98.25	2.60	95.65	1.75	1.12	0.62
RES	0.70	98.64	1.67	96.97	1.36	0.99	0.37
CRES	0.74	99.58	0.61	98.97	0.43	0.20	0.23
HRES	0.78	99.58	0.61	98.97	0.13	0.02	0.11
COMM	0.79	97.54	4.91	92.63	2.46	1.38	1.08
MEAN	0.75	98.69	2.03	96.67	1.31	0.76	0.55
COMP	0.84	99.67	0.66	99.02	0.32	0.31	0.02
EXPL	0.65	98.99	0.94	98.06	1.01	0.17	0.83
PRED	0.63	99.49	0.43	99.07	0.51	0.37	0.14
REACT	0.48	99.93	0.03	99.90	0.07	0.03	0.04
EV	0.84	99.13	2.32	96.81	0.87	0.38	0.49
LIM	0.82	99.87	0.28	99.59	0.13	0.07	0.06
STATE	0.83	99.51	1.19	98.32	0.49	0.23	0.26
CONTR	0.74	99.82	0.25	99.57	0.18	0.05	0.13
POS	0.75	99.87	0.19	99.68	0.13	0.02	0.11
GAD	*	*	*	*	*	*	*
IMP	0.82	99.37	1.49	97.88	0.63	0.23	0.40
REC	0.79	99.53	0.89	98.64	0.48	0.20	0.28
APP	0.38	99.84	0.05	99.79	0.16	0.12	0.04
HYP	0.67	99.67	0.33	99.34	0.33	0.09	0.24
Average agreement levels₂	0.71	99.51	0.79	98.72	0.47	0.24	0.24

1. See Table 2 for the meaning of the abbreviations
2. Average agreement level on steps (excluding *), weighed by source size

As can be seen, the average level of agreement is now over 99.5% (2nd column in table), which improves the value obtained in Test 1 and in previous studies (e.g. Peacock, 2002). Furthermore, the average Kappa coefficient (1st column) is now ‘good’ (0.71) overall, improving on the one obtained by Crookes (1986) for moves. Only IMP_APP shows a ‘poor’ Kappa coefficient (0.38), despite having obtained an average agreement percentage of 99.84%. This unexpected result could be explained by the scarcity of cases in this category. No new steps were identified after Test 2. However, a type of fragment classified as a substep under the category EV_STATE in the previous version of the scheme was upgraded to step status, *Noting specific gaps in knowledge or deficiencies in other research or practice* (EV_GAD), illustrated as follows:

(18) [Nectar volume and concentration has long been considered to be a useful predictor of...,]₁ (EV_STATE)/ [although **there are numerous caveats to the argument that** there is a simple relationship between...and...]₂ (EV_GAD):/

Here fragment (18.2) was not interpreted as evaluating the state of knowledge in broad terms, but as identifying a specific deficiency in a previous argument. Finally, it was necessary to validate the labels used with expert informants (Biber et al., 2007; Cotos et al., 2015).

6. Validating the coding scheme

To validate the scheme following Basturkmen (2012: 138), the main analyst interviewed the authors of two ERAs from the EXEMPRAES corpus. To do so in relation to English-medium ERAs, she emailed the segmented and labelled version of the D/C section in the English ERA of pair 1 (Table 1E) to its corresponding author, obtaining a response from the second author, as well as to an external expert reader in the field. To validate it in relation to Spanish-medium ERAs, she emailed the segmented and labelled version of the Conclusion section in the Spanish ERA of a pair outside the pilot study (Table 2S) to its first author, obtaining a response from the first and fourth authors. All had been requested to do two tasks. First, to state whether they agreed that the labels attached to each segment were expressing what they thought was their communicative function, in general and specific terms. Second, to assess whether the labels were self-explanatory without having access to the definitions in the codebook.

The expert reader explicitly stated that the labels were generally speaking self-explanatory, although he commented on a number of multi-functional fragments, corroborating one of the uncertainties in 4.3.7. He also mentioned the relevance of including the word ‘associations,’ or ‘correlations,’ as we had done in one of the definitions, if we used the phrase ‘causes and effects’ in a label, since relations between variables investigated in the social sciences are often not causal. As including more words in the corresponding label would have made it too long, we decided to choose a more comprehensive term to refer to all types of results (see below).

The authors of the research articles also made a few interesting comments. For instance, one of the authors of the Spanish article (Table 2S) perceived segments (10), where the signal is (Trans. :) “this result seems to go along the line of what threshold theory states...” and (11), where the signal is (Trans. :) “these results do not agree with those obtained by the meta-analysis carried out by Kim (2005),” as performing the same function. He

mentioned that both (Trans. :) “involved the interpretation that they agree or disagree with previous research.” This suggested that the functions *Expressing support to an existing theory or method* and *Comparing with previous studies...* in the previous version of the scheme might well be merged into one.

For her part, author 2 of the English article reacted to the analysis of segment (13), where the signal is “explain”, by saying that it was not speculating (cf. (34) “a possible explanation for...might be that...”). This suggested eliminating the word ‘speculating’ from the corresponding label. She also mentioned that she did not think the result in fragment (18) was “unexpected,” while the one in fragment 41 was “unexpected”, despite what the labels used said. This suggested eliminating such words from the labels (cf. Hopkins & Dudley-Evans, 1988). Neither did she understand well what the label *Elaborative function* meant. She finally commented that she did not “see a real difference in [sic.] interpreting findings and discussing causes and effects.” A similar observation is made by Basturkmen (2009), though Le and Harrington (2015) insist on making the distinction between these two functions as they seem to be realised by different lexico-grammatical means. Having observed not only the same lexico-grammatical variation but also a conceptual difference between the two kinds of coherence relations involved, i.e. internal *versus* external relations (see Hyland, 2005: 45) (cf. fragments 42 and 43 in Table 1E), we have maintained the distinction, as it is pedagogically useful, though we have revised the wording of the labels (see below).

In sum, the validation process indicated that the informants were actually interpreting the D/C sections in functional terms and that the labels were generally self-explanatory. Even so, the process helped us to revise various aspects of the codebook, whose clarity was finally double-checked with the external informant. For instance, we merged the two functions that had been perceived as very similar by the Spanish author (i.e. SUPP and COMP) into one (COMP), in accordance with Yang and Allison (2003). In addition, as most informants seemed to understand the label “*Elaborative function*”, we simply replaced it by *Elaborating* to express this overall function in an –ing form. Likewise, we substituted the label “*Discussing causes and effects; speculating*” by *Explaining results or discussing effects*, as the word *explaining* encompasses *speculating*, which involves using explanatory hypotheses or mitigated explanations, and the word *results* encompasses descriptive results, causal relations and associations.

We also replaced the label *Interpreting findings*, which we had used to replace Moreno's (2015) former label "*Deducing from findings*" following Yang and Alison (2003), with the label *Establishing the meaning of results*. In all likelihood, using such a vague term as 'Interpreting,' used also by Nwogu (1997), was the source of the conflation. In any case, a clear conceptual distinction can be made between fragments where authors draw logical inferences from key results (*Establishing the meaning of results*) and fragments where writers explain or attempt to explain the reasons for the results obtained or discuss the consequences of phenomena in the world outside the text (*Explaining results or discussing effects*). Finally, realising that we had used the same lemma "discuss" to name Move 3 (*Discussing findings and other features*) and the whole section (*Discussion and/or other closing sections*), we found it convenient to keep that lemma for the section label and to use "Commenting on results..." following Yang and Alison (2003), as part of the label for Move 3. This change then led us to rephrase Moreno's (2015) step "*Commenting on a finding or side-effect*" in Move 3 as *Reacting to results or other features* to avoid using the same verb, "comment," for a move and a step within that move.

7. Conclusions

In the present paper, we use the ERA Discussion (and/or other closing) sections to illustrate the difficulties encountered and the solutions offered by the ENEIDA project to develop accessible schemes for coding English and Spanish ERAs for their communication functions. Challenges were faced at different stages of the project requiring solutions consistent with the project aims. For instance, the problem of collecting comparable corpora of ERAs in Castilian Spanish and English was sorted out by resorting to interdisciplinary collaboration. While recent studies have used informants to name leading journals in their field (e.g. Peacock, 2002) or to confirm the quality of the journals included in the study (e.g. Del Saz Rubio, 2011), the ENEIDA invited content specialists to participate in the actual corpora compilation (see also Cotos et al., 2017). One advantage of this procedure is that the ERAs compiled can be considered good samples of writing from the perspective of the actual consumers of scientific information, rather than from the researchers' perspective. Another advantage is that the ERA pairs

are more likely to be comparable in such an influential factor in determining the structure and discourse style of the Discussion section as the study type (Williams, 2011).

The second major challenge emerged when carrying out a structural analysis of each section due to the ENEIDA's emphasis on situating interpretation at the step level in all the sections. This procedure may, in fact, strike the reader as somewhat anomalous as the great majority of ESP-related genre analysis studies promote or are built on top-down approaches with the moves being identified first (e.g. Peacock, 2002), but is in line with other corpus and computational linguistic analyses of scientific texts (Anthony, 2003; Teufel et al., 1999). Furthermore, as we have argued, the step is a more appropriate level for investigating the function-form gap. In effect, analysts were naturally driven to start with a bottom-up procedure, involving a closer reading of the wording in the fragments, to identify relevant specific topics. We thus concur with Flowerdew (2002) that perception of schematic structure cannot be dissociated from the bottom-up interpretation, at least at the step level.

Segmenting the texts into smaller fragments led to various problems of consistency. These were partly resolved by revising and extending the standard definition of a step and making the segmentation rules derived from such a definition more explicit, as well as providing extra specifications as to how to deal with fuzzy boundaries. Such a definition created the additional problem of leaving two sets of fragments out of the model: announcing and elaborating functions. Thus, a more comprehensive model of communicative functions was proposed. This decision was inspired by Hopkins and Dudley-Evans' (1988) heuristic approach in their search for a pedagogically useful model that could integrate different kinds of communicative functions (transactional, interactional and logical). Its advantage is that it will enable the ENEIDA Team to bridge the function-form gap in all communicative functions relevant to participating in this part genre more clearly.

Our initial emphasis on function and content rather than form led to a non-sentence-based segmentation protocol, where the minimal unit of analysis was the proposition. This is in overall agreement with other clause-based functional models (Connor & Mauranen, 1999; Mizuta et al., 2005). In that respect, our methodological approach also contributes to the continuing debate about communicative function boundaries (Paltridge, 1994). The ENEIDA procedures confirm that the criteria used for segmenting and labelling discourse

units from a functional perspective are primarily cognitive. Indeed, they allow analysts to make sense of the text as a whole, model it, and even make predictions about the structure of similar texts. The analysis involved is guided by mental processes utilised to identify topic breaks or shifts in the text and to interpret the pragmatic meaning of the resulting segments in the context in which they occur to perceive their function. This supports Paltridge's (1994: 1) conclusion that "there are non-linguistic... reasons for generic staging in texts." However, the pragmatic interpretation of each relevant fragment cannot take place until its wording has been processed and in that sense the analysis is linguistically-driven, as in Cortes (2013).

As segmenting and labelling in interpretative qualitative research is often considered unreliable and arbitrary (Orwin, 1994; Paltridge, 1994; Kanoksilapatham, 2005), various measures were taken to mitigate these weaknesses. First, a few extra rules were established as part of the codebook, improving the consistency of coding decisions. Second, even though the main analyst was satisfied with the model that she had initially designed (i.e. the revised model after her intra-rater reliability test), to guarantee that other analysts working independently segmented and labelled the same texts in similar ways, two inter-rater reliability tests providing Cohen's Kappa values were necessary. This suggests that intra-rater reliability measures or inter-rater reliability measures offering only percentage results in similar studies may not provide a good standard of reliability and should be supplemented by inter-rater reliability tests offering Cohen's Kappa values (e.g. Crookes, 1986; Kanoksilapatham, 2005). Our methodological approach also sought to guarantee that the labels were valid for pedagogical purposes and well understood by expert informants. The entire pilot corpora will need to be coded again according to the revised codebook before the annotations can be used for other purposes (see below).

Although there are various possible ways of doing a move analysis (Biber et al., 2007), we would like to offer the ENEIDA methodological approach to other EAP researchers pursuing similar goals as ours, as it has produced good results in two languages, English and Spanish. This means that the codebook can be applied in a reliable way by reasonably well-trained competent academic readers who are not expert in the fields involved. This achievement offers hope to the field by confirming that this kind of rigorous analysis can be done without the direct involvement of content informants in the actual text analyses. We recognize that the procedures laid out in this paper are, in their entirety, probably too complex and time-consuming to be adopted by an individual EAP teacher. However, we

feel that some awareness of the methodological options presented should help to improve the quality of future move-and-step models of genres for specific purposes offered by other ESP researchers for potential application in English and other language teaching contexts.

One possible limitation of the coding scheme developed so far for the D/C sections of ERAs is the size of the pilot sample on which it is based. However, after analysing the last three pairs, no new steps were identified, so the coding scheme can be considered relatively stable. The ENEIDA Team has so far performed preliminary cross-cultural comparisons of the various article sections using such a type of scheme, revealing insightful quantitative differences at rhetorical levels of analysis (e.g. Moreno, 2015). Having obtained comparable text fragments across similar English and Spanish ERA sections will not only allow the team to compare aspects of their content and organization (e.g. Cotos et al., 2015; Yang & Allison, 2003), but also phraseological (e.g. Biber et al., 2007; Cortes, 2013; Cotos et al., 2016, 2017; Le & Harrington, 2015; Nwogu, 1997), grammatical and stylistic patterns in a meaningful way. Identifying such kinds of patterns will certainly help the team to bridge the function-form gap not only in English ERAs independently, but also in Spanish and across the two languages. Furthermore, it will be possible to compare the linking resources typically used to signal move boundaries across comparable sequences of moves-and-steps.

We believe that the cross-cultural results obtained from applying the methodology proposed here to the analysis of the EXEMPRAES corpus will be useful to create pedagogic resources from a cross-cultural perspective to help Spanish scholars compose their ERA D/C sections more efficiently. A good example is the poem composed by the main author, inspired by preliminary results (Moreno, 2015). Designing further tasks for Spanish scholars and proving their efficiency in learning contexts should be the topic of further research. Admittedly, although this study is ultimately motivated by an interest in finding practical solutions for Spanish scholars, those solutions can only be partial, as other difficulties faced by them in the publication process (Moreno et al., 2011) may be in addition to any ability to conform to the rhetorical conventions expected of good D/C sections in English-medium journals.

Finally, the revised ENEIDA annotations could well be used to retrain AntMover (Anthony, 2003) to code the whole EXEMPRAES compilation semi-automatically.

Although this option would still require a qualified researcher to relabel the misclassified steps, it would allow the ENEIDA Team to carry out the quantitative studies of generic structure that many scholars have long hoped for (Biber et al., 2007) much faster than if the texts were annotated manually from scratch in a corpus-based fashion. For the moment, the revised annotations will feature in the *EXEMPRAES Corpus*, an open access online search interface developed by Anthony and Moreno (forthcoming), allowing users to search language items in refined ways, i.e. by section, move, step, language, knowledge area and disciplinary field, for teaching, learning and for cross-disciplinary and/or cross-linguistic research.

Acknowledgements

This study would not have been possible without the collaboration of the following participants: Ramón Rodríguez (CSIC), a specialist informant, Sally Burgess (ULL), Pedro Martín-Martín (ULL), M^a Lluisa Gea-Valor (UJI), Rosa Lorés (UZ), Pilar Mur (UZ) and Enrique Lafuente (UZ), who compiled sections of the EXEMPRAES Corpus; 54 Spanish researchers who recommended ERAs for the EXEMPRAES Corpus; Hilo Digital (Madrid; al.regidor@gmail.com) who converted and cleaned the pdfs; Albano Fernández (ULE) who helped to organise the bibliography; Leticia Barrionuevo (ULE) who calculated the pilot ERAs impact factors; Marie-Hélène Paré (Barcelona), in her role as an NVivo instructor and methodological consultant; Sally Burgess, Pedro Martín-Martín, M^a Lluisa Gea-Valor in their role as ENEIDA coders, and Silvia Ramos (ULE), the independent analyst in the D/C section study, for their feedback to Ana I. Moreno during the coordination process; Javier Vidal García (ULE), a social scientist, in preparing the Excel sheet for calculating average reliability coefficients; Natascha Notte, Juan E. Jiménez and Eduardo Garcia, co-authors of the segmented ERAs presented in the supplementary materials, and Itesh Sachdev, an expert reader in the social sciences, during the validation process; as well as the two anonymous reviewers and Nigel Harwood for their most helpful and encouraging comments on previous versions of this paper.

Funding: This work was supported by the Spanish Ministry of Science and Innovation (Ref.: FFI2009-08336/FILO). Ana I. Moreno was the PI.

8. References

- Amirian, Z., Kassaian, Z., & Tavakoli, M. (2008). Genre analysis: an investigation of the discussion sections of applied linguistics research articles. *The Asian ESP Journal*, 4, 39-63.
- Ammon, U. (Ed.) (2001). *The Dominance of English as a Language of Science: Effects on Other Languages and Language Communities*. New York: Walter de Gruyter.
- Annuai, W., & Wannaruk, A. (2013). Investigating move structure of English applied linguistics research article discussions published in international and Thai journals. *English Language Teaching*, 6, 1-13.
- Anthony, L. (2003). AntMover (Version 1.0.0) [Computer Software]. Tokyo, Japan: Waseda University. Available from <http://www.laurenceanthony.net/>
- Anthony, L., & Moreno, A. I. (forthcoming). The EXEMPRAES Corpus, in A. I. Moreno (2013, new edition forthcoming). *The ENEIDA Website*. León: Universidad de León. Available from <http://eneida.unileon.es/>
- Basturkmen, H. (2009). Commenting on results in published research articles and masters dissertations in Language Teaching. *Journal of English for Academic Purposes*, 8, 241-251.
- Basturkmen, H. (2012). A genre-based investigation of discussion sections of research articles in Dentistry and disciplinary variation. *Journal of English for Academic Purposes*, 11, 134-144.
- Berkenkotter, C., & Huckin, T. N. (1995). *Genre Knowledge in Disciplinary Communication: Cognition/Culture/ Power*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Bhatia, V. K. (2001). Analysing genre: Some conceptual issues. In M. Hewings (Ed.), *Academic writing in context: Implications and applications* (pp. 79-92). Birmingham: University of Birmingham Press.
- Biber, D. Connor, U., & Upton T. A. (2007). *Discourse on the Move: Using Corpus Analysis to Describe Discourse Structure*. Amsterdam/Philadelphia: John Benjamins.
- Burgess, S., Gea-Valor, M^aLl., Moreno, A.I., & Rey-Rocha, J. (2014). Affordances and constraints on research publication: A comparative study of the language choices of Spanish historians and psychologists. *Journal of English for Academic Purposes*, 14, 72–83. doi:10.1016/j.jeap.2014.01.001
- Cargill, M., & Burgess, S. (2008). Introduction to the special issue: English for research publication purposes. *Journal of English for Academic Purposes*, 7, 75-76.

- Connor, U., & Mauranen, A. (1999). Linguistic analysis of grant proposals: European Union research grants. *English for Specific Purposes*, 18, 47-62.
- Connor, U., & Moreno, A. I. (2005). Tertium Comparationis: A vital component in contrastive research methodology. In P. Bruthiaux, D. Atkinson, W. G. Eggington, W. Grabe, & V. Ramanathan (Eds.), *Directions in Applied Linguistics: Essays in Honor of Robert B. Kaplan* (pp. 153-164). Clevedon, England: Multilingual Matters. [<http://hdl.handle.net/10261/13416>]
- Cortes, V. (2013). The purpose of this study is to: Connecting lexical bundles and moves in research article introductions. *Journal of English for Academic Purposes*, 12, 33-43.
- Cotos, E., Huffman, S., & Link, S. (2015). Furthering and applying move/steps constructs: technology driven marshalling of Swalesian genre theory for EAP pedagogy. *Journal of English for Academic Purposes*, 19, 52-72.
- Cotos, E, Link, S., & Huffman, S. (2016). Studying disciplinary corpora to teach the craft of Discussion. *Writing and Pedagogy*, 8.1, 33–64.
- Cotos, E. Haufman, S., & Link, S. (2017). A move/step model for methods sections: Demonstrating Rigour and Credibility. *English for Specific Purposes*, 46, 90–106.
- Crookes, G. (1986). Towards a validated analysis of scientific text structure. *Applied Linguistics*, 7, 58-70.
- Del Saz Rubio, M^a M. (2011). A pragmatic approach to the macro-structure and metadiscoursal features of research article introductions in the field of Agricultural Sciences. *English for Specific Purposes*, 30, 258-271.
- ElMalik, A., & Nesi, L. (2008). Publishing research in a second language: the case of Sudanese contributors to international medical journals. *Journal of English for Academic Purposes*, 7, 87-96.
- Flowerdew, J. (2002). Genre in the classroom: A linguistic perspective. In A. M. Johns (Ed.), *Genre in the Classroom: Multiple perspectives* (pp. 91-104). Mahwah, NJ: Lawrence Erlbaum.
- Flowerdew, L. (2004). The argument for using English specialized corpora to understand academic and professional language. In U. Connor, & T. A. Upton (Eds.), *Discourse in the Professions: Perspectives from Corpus Linguistics* (pp. 11-36). Amsterdam: John Benjamins.
- Frow, J. (2006). *Genre*. London: Routledge.
- Gea-Valor, M^a Ll., Rey-Rocha, J., & Moreno, A. I. (2014). Publishing research in the international context: an analysis of Spanish scholars' academic writing needs in the

social sciences. *English for Specific Purposes*, 36, 47–59.
doi:10.1016/j.esp.2014.05.001

Geertz, C. (1980). *Negara: The Theatre State in Nineteenth-century Bali*. Princeton, NJ: Princeton University Press.

Glasman-Deal, H. (2009). *Scientific research writing for non-native speakers of English*. London: Imperial College Press.

Hasan, R. (1984). The nursery tale as a genre. *Nottingham Linguistic Circular*, 13, 71-102.

Holmes, R. (1997). Genre analysis, and the social sciences: An investigation of the structure of research article discussion sections in three disciplines. *English for Specific Purposes*, 16, 321-337.

Hopkins, A., & Dudley-Evans, T. (1988). A genre-based investigation of the discussion sections in articles and dissertations. *English for Specific Purposes*, 7, 113-121.

Hyland, K. (2004). *Disciplinary discourses*. Ann Arbor, MI: The University of Michigan Press.

Hyland, K. (2005). *Metadiscourse*. London: Continuum.

Hyland, K. (2012). *Disciplinary identities: Individuality and Community in Academic Discourse*. Cambridge: Cambridge University Press.

Kanoksilapatham, B. (2005). Rhetorical structure of biochemistry research articles. *English for Specific Purposes*, 24, 269-292.

Kanoksilapatham, B. (2007). Rhetorical moves in biochemistry articles. In D. Biber, U. Connor, & T.A. Upton (Eds.), *Discourse on the Move: Using Corpus Analysis to Describe Discourse Structure* (pp. 73-119). Amsterdam: John Benjamins.

Krippendorff, K. (2004). *Content Analysis: An Introduction to its Methodology* (2nd ed.). Thousand Oaks: Sage.

Le, T. N. P., & Harrington, M. (2015). Phraseology used to comment on results in the Discussion section of applied linguistics quantitative research articles. *English for Specific Purposes*, 39, 45–61.

Lillis, T., & Curry, M. J. (2010). *Academic writing in a global context: the politics and practices of publishing in English*. London/New York: Routledge.

Lim, J. (2010). Commenting on research results in applied linguistics and education: A comparative genre-based investigation. *Journal of English for Academic Purposes*, 9, 280-294.

- Loi, C. K., & Evans, M. S. (2010). Cultural differences in the organization of research article introductions from the field of educational psychology: English and Chinese. *Journal of Pragmatics*, 42, 2814-2825.
- López-Navarro, I., Moreno, A. I., Quintanilla, M. A., & Rey-Rocha, J. (2015). "Why do I publish research articles in English instead of my own language? Differences in Spanish researchers' motivations across scientific domains." *Scientometrics*, 103, 939–976. DOI 10.1007/s11192-015-1570-1
- Martín-Martín, P., Rey-Rocha, J., Burgess, S., & Moreno, A. I. (2014). Publishing research in English-language journals: Attitudes, strategies and difficulties of multilingual scholars of Medicine. *Journal of English for Academic Purposes*, 16, 57-67. . doi:10.1016/j.jeap.2014.08.001
- Mizuta, Y., Korhonen, A. Mullen, T., & Collier, N. (2005). Zone analysis in biology articles as a basis for information extraction. *Medical Informatics*, 75, 468-487.
- Moreno, A. I. (1998). The explicit signalling of premise-conclusion sequences in research articles: a contrastive framework. *Text*, 18, 545-585. [<http://hdl.handle.net/10261/13671>]
- Moreno, A. I. (2003a). The role of cohesive devices as textual constraints on relevance: A discourse-as-process view. *International Journal of English Studies* (Monograph: Discourse Analysis Today), 3, 111-165. [<http://hdl.handle.net/10261/12399>]
- Moreno, A. I. (2003b). Matching theoretical descriptions of discourse and practical applications to teaching: the case of causal metatext. *English for Specific Purposes*, 22, 265-295. [<http://hdl.handle.net/10261/13383>]
- Moreno, A. I. (2014a). Rhetorical strategies to get published in international *versus* national journals: The devil is in the details. 2014 Languages for Specific Purposes Conference Academic Communication in a Digital Age. Centre for Academic English, Stockholm: Sweden. 12-14 June, 2014. [<http://su.diva-portal.org/smash/get/diva2:790570/FULLTEXT01.pdf>]
- Moreno, A. I. (2014b). The rhetorical structure of research article Discussion (and/or Conclusion) sections in international *versus* national journals. 2014 Languages for Specific Purposes Conference Academic Communication in a Digital Age. Centre for Academic English, Stockholm: Sweden. 12-14 June, 2014. [<http://su.diva-portal.org/smash/get/diva2:790570/FULLTEXT01.pdf>]
- Moreno, A. I. (2015). 'The Discussion Section in Essence' Poem. The effect of national/international audience on the rhetorical structure of research article Discussion sections. PRISEAL Conference, University of Coimbra, Coimbra, Portugal, 30 Oct-01 Nov 2015. [<http://buleria.unileon.es/handle/10612/4868>]

- Moreno, A. I. (forthcoming). The ENEIDA Codebook for annotating ERA Discussion (and/or Closing) sections, in A. I. Moreno (2013, new edition forthcoming). *The ENEIDA Website*. León: Universidad de León. Available from <http://eneida.unileon.es/>
- Moreno, A. I., Rey-Rocha, J., Burgess, S., Martín-Martín, P., Gea-Valor, M^a L., López-Navarro, I., Garzón, B., & Sachdev, I. (2011). Spanish Researchers Publishing In Scientific Journals: Motivations, Views, Strategies, Experiences and Training Needs. URL: [<https://buleria.unileon.es/handle/10612/1757>] [06/06/12]
- Moreno, A. I., Rey-Rocha, J., Burgess, S., López-Navarro, I., & Sachdev, I. (2012). Spanish researchers' perceived difficulty writing research articles for English-medium journals: the impact of proficiency in English *versus* publication experience. *IBERICA*, 24, 157-184. [http://www.aelfe.org/documents/13_24_Moreno.pdf]
- Moreno, A. I., Burgess, S., Sachdev, I., López-Navarro, I., & Rey-Rocha, J. (2013). The ENEIDA questionnaire: publication experiences in scientific journals in English and Spanish. Available from [<http://eneida.unileon.es/eneidaquestionnaire.php>]
- Nesi, H., & Gardner, S. (2012). *Genres across the disciplines: Student Writing in higher education*. Cambridge: Cambridge University Press.
- NVivo10 help (2014). Run a Coding Comparison query. http://help-nv10.qsrinternational.com/desktop/procedures/run_a_coding_comparison_query.htm [13/01/14]
- Nwogu, K. N. (1990). *Discourse variation in medical texts: Schema, theme and cohesion in professional and journalistic accounts*. Nottingham Monographs in Systemic Linguistics, 2.
- Nwogu, K. N. (1997). The Medical Research Paper: Structure and Functions. *English for Specific Purposes*, 16, 119-138.
- Orwin, R. G. (1994). Evaluating coding decisions. In H. Cooper, & L.V. Hedges (Eds.), *The Handbook of Research Synthesis* (pp. 139-162). New York: Russell Sage Foundation.
- Paltridge, B. (1994). Genre analysis and the identification of textual boundaries. *Applied Linguistics*, 15, 288-299.
- Paltridge, B. (1995). Working with genre: A pragmatic perspective. *Journal of Pragmatics*, 24, 393-406.
- Parkinson, J. (2011). The Discussion section as argument: the language used to prove knowledge claims. *English for Specific Purposes*, 30, 164-175.

- Peacock, M. (2002). Communicative moves in the discussion section of research articles. *System*, 30, 479-497.
- Perez-Llantada, C. (2012). *Scientific discourse and the rhetoric of globalization*. London: Continuum.
- Perez-Llantada, C., PLo, R., & Ferguson, G. R. (2012). “You don’t say what you know, only what you can”: The perceptions and practices of senior Spanish academics regarding research dissemination in English. *English for Specific Purposes*, 30, 18-30.
- Pho, P. D. (2008). Research article abstracts in applied linguistics and educational technology: A study of linguistic realizations of rhetorical structure and authorial stance. *Discourse Studies*, 10, 231-250.
- Sinclair, J. M. (1993). Written discourse structure. In J. M. Sinclair, M. Hoey, & G. Fox (Eds.), *Techniques of Description: Spoken and Written Discourse, a Festschrift for Malcolm Coulthard* (pp. 6-31). London: Routledge.
- Swales, J. (1981, 2011). *Aspects of Article Introductions*. Ann Arbor, MI: University of Michigan Press.
- Swales, J. (1990). *Genre Analysis: English in Academic and Research Settings*. Cambridge: Cambridge University Press.
- Swales, J. M. (2004). *Research Genres. Exploration and Applications*. Cambridge, UK: Cambridge University Press.
- Swales J.M. (2017) *Other Floors, Other Voices: A Textography of A Small University Building (Rhetoric, Knowledge, and Society Series)*. Ann Arbor, MI: University of Michigan Press.
- Tarone, E., Dwyer, S., Gillette, S., & Icke, V. (1981). On the use of the passive and active in two astrophysics journal papers. *The ESP Journal*, 1, 123-40.
- Tesch, R. (1990). *Qualitative research: Analysis types and software tools*. New York: Palmer Press.
- Teufel, S., Carletta, J., & Moens, M. (1999). An annotation scheme for discourse-level argumentation in research articles. Proceedings of EACL '99. (unpaginated).
- Vassileva, I. (1997). Hedging in English and Bulgarian academic writing. In A. Duszak, (Ed.), *Culture and Styles of Academic Discourse*, (pp. 203–221). Berlin: Mouton de Gruyter.
- Williams, I. A. (2011). Factors Affecting Discourse Structure and Style in Biomedical Discussion Sections, *Biomedical Engineering, Trends, Research and Technologies*, Dr. Sylwia Olsztynska (Ed.), ISBN: 978-953-307-514-3, InTech, Available from: <http://www.intechopen.com/books/biomedical-engineering-trends-researchand->

[technologies/factors-affecting-discourse-structure-and-style-in-biomedical-discussion-sections](#). Last accessed 14.02.17.

Yakhontova, T. (1997). The signs of a new time: academic writing in ESP curricula of Ukrainian universities. In A. Duszak (Ed.), *Culture and Styles of Academic Discourse* (pp. 103–112). Berlin: Mouton de Gruyter.

Yakhontova, T. (2006). Cultural and disciplinary variation in academic discourse: the issue of influencing factors. *Journal of English for Academic Purposes*, 5, 153-167. <http://dx.doi.org/10.1016/j.jeap.2006.03.002>

Yang, R., & Allison, D. (2003). Research articles in applied linguistics: Moving from results to conclusions. *Journal of English for Specific Purposes*, 22, 365-385.

Appendix A.

Bibliographical details of the pilot sample from the *EXEMPRAES Corpus* and their impact factor

N.	Bibliographical details of articles ordered by pairs	Quality indexes	
		JCR	SJR
LFS2ENG	Sawalha, R. M., Brotherstone, S., Conington, J., & Villanueva, B. (2007). Lambs with scrapie susceptible genotypes have higher postnatal survival. <i>PLoS One</i> , <i>11</i> , 1-5.	(2009-)	1.285
LFS2SP	Traoré, A., Royo, L. J., Álvarez, I., Fernández, I., Gutiérrez, J. P., Rincón, C., Goyache, F. (2008). Mutaciones adyacentes a los codones 136 y 171 del gen PrnP ovino afectan al protocolo diagnóstico basado en RT-PCR acoplado a las sondas fluorescentes. <i>ITEA</i> , <i>104</i> , 106-109.	(2009-)	0.118
LFS3ENG	Ollerton, J., Johnson, S. D., Cranmer, L., & Kellie, S. (2003). The pollination ecology of an assemblage of grassland asclepiads in south africa. <i>Annals of Botany</i> , <i>92</i> , 807-834.	1.370	0.892
LFS3SP	Jordano, P. (1990) ² . Biología de la Reproducción de tres especies del género <i>Lonicera</i> (Caprifoliaceae) en la Sierra de Cazorla. <i>Anales Jard. Bot. Madrid</i> , <i>48</i> , 31-52.	(2011-)	(2009-)
LFS4ENG	Jump, A. S., Hunt, J. M., & Peñuelas, J. (2006). Rapid climate change-related growth decline at the southern range edge of <i>fagus sylvatica</i> . <i>Global Change Biology</i> , <i>12</i> , 2163-2174. doi:10.1111/j.1365-2486.2006.01250.x	4.339	3.159
LFS4SP	Boada Junca, M., & Saurí Pujol, D. (2003). Evolución socioambiental del hayedo en el macizo del montseny (1945-1955). <i>Ería</i> , <i>62</i> , 317-324.	–	–
LFS6ENG	Pesic, V., Yam, R. S. W., Chan, B. K. K., & Chatterjee, T. (2012). Water mites (acari, hydrachnidia) from baishih river drainage in northern taiwan, with description of two new species. <i>Zookeys</i> , <i>203</i> , 65-83.	0.864	0.574
LFS6SP	Subías, L. S., & Shtanchaeva, U. (2012). Oribátidos (acari, oribatida) de las loreras (<i>prunus lusitanicus</i> L.) de extremadura (suroeste de españa) y descripción de una nueva especie de <i>cosmochthonius berlese</i> ,	–	0.203

² This article from the 20th century was allowed in the corpus because the informant explained that it would have been impossible to find two comparable pairs otherwise.

	1910 (cosmochthoniidae). <i>Graellsia</i> , 68(1), 7-16. doi:10.3989/graellsia.2012.v68.049		
HCS2ENG	Javanbakht, M., Abolhasani, F., Mashayekhi, A., Baradaran, H. R., & Jahangiri, Y. (2012). Health related quality of life in patients with type 2 diabetes mellitus in iran: A national survey. <i>PLoS One</i> , 7(8), e44526. doi:10.1371/journal.pone.0044526.t001	3.730	1.931
HCS2SP	Mata Cases, M., Roset Gamisans, M., Badia Llach, X., Antoñanzas Villard, F., & Ragel Alcázar, J. (2003). Impacto de la diabetes mellitus tipo 2 en la calidad de vida de los 11pacientes tratados en las consultas de atención primaria en españa. <i>Atención Pr12imaria</i> , 31, 493-499.	–	0.199
HCS4ENG	Ngan, J. M. W., Chow, D. H. K., & Holmes, A. D. (2005). The kinematics and intra- and inter-therapist consistencies of lower cervical rotational manipulation. <i>Medical Engineering & Physics</i> , 27, 395-401. doi:10.1016/j.medengphy.2004.10.009	1.151	0.459
HCS4SP	Cuesta-Vargas, A. I., & William, J. M. (2011). Estudio de la cinemática y fiabilidad inter e intraterapeutas de la manipulacion vertebral cervical basada en sensores inerciales. <i>Fisioterapia</i> , 33, 25-30.	–	0.166
HCS7ENG	Aisenbrey, S., Gelisken, F., Szurman, P., & Bartz-Schmidt, K. U. (2007). Surgical treatment of peripapillary choroidal neovascularisation. <i>British Journal of Ophthalmology</i> , 91, 1027-1030.	2.689	1.651
HCS7SP	Ruíz- Moreno, J. M., Amat-Peral, P., Lugo, F. L., & Montero, J. A. (2009). Extracción quirúrgica de la neovascularización coroidea peripapilar en pacientes jóvenes. <i>Archivos De La Sociedad Española De Oftalmología</i> , 84, 39-42.	–	0.206
SCS2ENG	Ammermueller, A. (2007). Poor background or low returns? why immigrant students in germany perform so poorly in the programme for international student assessment. <i>Education Economics</i> , 15, 215-230.	–	0.524
SCS2ENG	Blanchard Giménez, M. (2011). Análisis de los resultados de aprendizaje obtenidos al inicio y al final de curso por el alumnado inmigrante en comparación con el alumnado autóctono. <i>Pedagogía Social. Revista Interuniversitaria</i> , 18, 73-85.	–	–

SCS3ENG	Campbell, J. C., Webster, D. W., & Glass, N. (2009). The danger assessment validation of a lethality risk assessment instrument for intimate partner femicide. <i>Journal of Interpersonal Violence</i> , 24, 653-674.	1.332	1.118
SCS3SP	Echeburúa, E., Fernández-Montalvo, J., & de Corral, P. (2008). ¿Hay diferencias entre la violencia grave y la violencia menos grave contra la pareja?: Un análisis comparativo. <i>International Journal of Clinical and Health Psychology</i> , 8, 355-382.	1.649	0.397
SCS4ENG	Kloosterman, R., Notten, N., Tolsma, J., & Kraaykamp, G. (2011). The effects of parental reading socialization and early school involvement on Children's academic performance: A panel study of primary school pupils in the netherlands. <i>European Sociological Review</i> , 27, 291-306.	1.935	2.099
SCS4SP	Cervini Iturre, R. (2003). Relaciones entre composición estudiantil, proceso escolar y el logro en matemáticas en la educación secundaria en argentina. [Relationships Among School Composition, School Process and Mathematics Achievement in Secondary Education in Argentina] <i>Revista Electrónica De Investigación Educativa</i> , 5, 1-27.	-	(2007-)
SCS8ENG	Laursen, K., Masciarelli, F., & Prencipe, A. (2012). Regions matter: How localized social capital affects innovation and external knowledge acquisition. <i>Organization Science</i> , 23, 177-193.	3.351	7.511
SCS8SP	Delgado, M., Martín, G., Navas, J. E., & Cruz, J. (2011). Capital social, capital relacional e innovación tecnológica. Una aplicación al sector manufacturero español de alta tecnología y media-alta tecnología. <i>Cuadernos De Economía y Dirección De La Empresa</i> , 14, 207-221.	-	-
SCS11ENG	Dawson, C., Fuller-Love, N., Sinnott, E., & O'Gorman, B. (2011). Entrepreneurs' perceptions of business networks: Does gender matter? <i>Entrepreneurship and Innovation</i> , 12, 271-281. doi:10.5367/ijei.2011.0047	-	-
SCS11SP	González Álvarez, N., & Nieto Antolín, M. (2012). Descubrimiento y explotación de oportunidades empresariales. Un análisis desde la perspectiva del capital social y el género. <i>Economía Industrial</i> , 383, 65-74.	-	-

ONS2ENG	Gao, X., Jehng, J., & Wachs, I. E. (2002). In situ UV-vis-NIR diffuse reflectance and raman spectroscopic studies of propane oxidation over ZrO ₂ -supported vanadium oxide catalysts. <i>Journal of Catalysis</i> , 209, 43-50. doi:10.1006/jcat.2002.3635	3.118	2.669
ONS2SP	Herrera, C., Larrubia, M. A., Blasco, J. M., & Alemany, L. J. (2005). Estudio de la catálisis de la deshidrogenación oxidativa del propano. <i>Ingeniería Química</i> , 424, 132-144	–	0.151
ONS4ENG	Dold, B. (2003). Speciation of the most soluble phases in a sequential extraction procedure adapted for geochemical studies of copper sulfide mine waste. <i>Journal of Geochemical Exploration</i> , 80, 55-68.	0.931	0.500
ONS4SP	Delgado, J., Pérez-López, R., Álvarez-Valero, A. M., Nieto, J. M., Sáez, R., & Matos, J. X. (2007). Movilidad de metales en residuos mineros y valoración del impacto ambiental del distrito minero de são domingos (faja pirítica ibérica), SW de la península ibérica. <i>Geogaceta</i> , 43, 59-62	–	–
ONS5ENG	Tilkat, E., Onay, A., Yildirim, H., & Ayaz, E. (2009). Direct plant regeneration from mature leaf explants of pistachio, pistacia vera L. <i>Scientia Horticulturae</i> , 121, 361-365.	1.197	0.785
ONS5SP	García, E., Lorente, P., Marín, J. A., Arbeloa, A., & Andreu, P. (2010). Micropropagación e injerto in vitro de pistacho. <i>Información Técnica Económica Agraria</i> , 106, 294-302.	–	0.122

Bibliographical details of the Spanish ERA from the *EXEMPRAES Corpus* used for validating the coding scheme after the pilot study and its impact factor

N.	Bibliographical details	Quality indexes	
		JCR	SJR
SCS6SP	Jiménez, J. E., Artilés, C., Rodríguez, C., García, E., Camacho, J., & Moraes, J. (2008). Creatividad e inteligencia: ¿dos hermanas gemelas inseparables? <i>Revista Española De Pedagogía (REP)</i> , 240, 261-282.	–	0.188

Appendix B. Extract from codebook for the Discussion (and/or Closing) section

CODEBOOK FOR THE ERA DISCUSSION (AND/OR OTHER CLOSING) SECTION		
Communicative function label (CODE)	Definition	Examples in English and Spanish
Discussion (and/or other closing) section (D/C)	In the D/C section, writers announce what they are going to write about, remind readers of key features and results, draw on background knowledge, establish the meaning of key results, make comparisons with previous research, explain key results, discuss effects, make predictions, react to results or other features, evaluate their own study positively or negatively, state their contribution, evaluate other research or practice, assess the state in which the discipline or practice is/remains, make recommendations for future practice or research, suggest the applicability of their results or usability of their outcomes, hypothesize, and/or elaborate on some of these ideas.	
Move...	...	
Commenting on key results or other features (COMM)	In COMM, writers comment on their results in various ways: establishing the meaning of key results, comparing them (or other features) with those of previous research, explaining them or discussing effects, making predictions and/or reacting to results or other features. There are five possible steps:	
	Step...	...
Making predictions (COMM_PRED)	In PRED ⁱ , writers state more or less tentatively what phenomenon will happen in future considering the results or other features of the study. Predictions can be derived from implicit or explicit hypotheses by means of deduction. When this happens, the implication is that, if a hypothesis is true, the predicted future observations derived from it should occur.	Eng.: [In other words, out of 1000 lambs born alive, about 22 more lambs are expected to die if they are of the ARR/ARR genotype than if they are of the ARR / ARQ genotype]. (COMM_PRED) (LFS2ENG) Sp.: [De esta manera, las empresas conseguirán ventajas en cuanto a..., siendo las empresas externas una fuente de ... y, por lo tanto, un posible determinante de ...]. (COMM_PRED) (SCS8SP)
Drawing implications (IMP)	In IMP, writers draw implications of various kinds: they make recommendations for future practice and/or research, suggest the applicability of the current study results or the usability of its outcomes, or hypothesize for future studies. There are three possible steps:	

	Step...
	Hypothesizing for future research (IMP_HYP)	In HYP, writers speculate about a relationship between variables (e.g. by stating that some phenomenon might influence or may be related to something else; that it could make something else happen; or that it may be an indirect cause of it) to be tested in future research.	Eng.: [In this paper, we focus on...], [an important element in ...], but regional social capital might influence the effectiveness of ..., including ... (IMP_HYP). [Another avenue for future research would be to ...] (SCS8ENG) Sp.: [Habría que valorar, por tanto, el papel de (...) en (...) y su asociación a ...] [Cabría entonces barajar la hipótesis alternativa de que ... sea una consecuencia indirecta de ..., sin jugar un papel importante en ...] (IMP_HYP) (LFS3SP)

(Adapted with permission from Moreno, forthcoming)

Supplementary material:

Table 1E:

https://www.dropbox.com/s/xzoxo829zxx086c/Table%201E%20with%20analysis%20of%20English%20C_D%2007_12_17%20to%20share.pdf?dl=0

Table 2S:

https://www.dropbox.com/s/484s9uedxzqhd9/Table%202S%20with%20analysis%20of%20Spanish%20D%2007_12_17%20to%20share.pdf?dl=0

ⁱ To simplify, just the final parts of the code for each specific function are provided as required (thus, both PRED and COMM_PRED imply D/C_COMM_PRED).