Encoding of Information in Titles: Academic Practices across Four Genres in Linguistics

SARA GESUATO
University of Padua

1. Introduction

The title of an academic publication names, introduces, metonymically represents and advertises the content it labels, circulating from one text to the next. It is meant to be useful (informative and precise), logical (relevant to the work it names, easily classifiable and storable in databases) and reader-friendly (concise, understandable and appealing; Huth 1987; Swales and Feak 1994; Yitzhaki 1994; Busch-Lauer 2000; Dressler and Eckkramer 2001; Goodman et al. 2001; Yakhontova 2002; Hartley 2005; Lewinson and Hartley 2005; Hartley 2008). Yet, its encoding varies depending on editorial policies, individual authors' stylistic preferences, and awareness of the role and expected impact of the title on the communicative situation.

The following titles, which identify a book, a dissertation, a journal article and a proceedings paper, respectively, illustrate this variation:

(1) "The languages of a bilingual community"
(2) "The Acoustic-Phonetic Characteristics of Infant-Directed Speech in Mandarin Chinese and Their Relation to Infant Speech Perception in the First Year of Life"
(3) "On the Status of Infixation and Circumfixation in English Morphology"
(4) "Is This My Position? Teenagers' Response to Mass Media Discourse".

Titles (1), (3) and (4) are shorter than (2); (2) and (3) are richer in vocabulary and content than (1) and (4); (2) and (3) contain more denotationally precise terms than (1) and (4); (1) and (4) are easy to process, consisting of simple syntactic constituents, while (2) and (3) are more syntactically elaborate; finally, only (4) is partly formulated so as to directly involve the reader in the interaction. This linguistic heterogeneity can be attributed to the different content addressed, communicative goals and target readerships of the genres the titles are relevant to (Soler 2007: 91-92).

Books need to attract prospective readers and buyers; clear and appealing titles serve this marketing goal. Dissertations are expected to contain detailed, focused content about research carried out by fledgling academics, whose adequacy is evaluated by specially appointed examiners; their titles have to make a good first impression on these gate-keepers. Journal articles are meant to convey accurate information to interested expert peer readers, who are likely to use titles as guidelines when looking for data relevant to their work; their titles will be of help if they are non-misleading, and precise. Finally, proceedings papers typically originate as conference presentations meant to attract a large, interested expert audience; their titles contribute to this purpose by arousing the intended addressees' attention, and later on, they may retain the original formulation (cf. Busch-Lauer 2000). The different contextual expectations may call for different communicative strategies, or linguistic-textual realisations of functionally equivalent, communicative products.

More specifically, book titles that are short, divided into syntactically simple, easy-to-decode information units may be perceived as reader-friendly and thus appropriately label publications seeking commercial success. Titles of dissertations may achieve descriptive adequacy if they are exhaustive and to the point, although not necessarily appealing, easy to remember or easy to understand for the layperson: syntactic complexity, high level of information packaging and the use of jargon are appropriate choices. Journal article titles can be of use to the readers if they are highly informative and non-misleading; mild elaborateness can be tolerated for the sake of accuracy, and is unlikely to cause decoding problems to the expert, self-motivated readership. And titles of conference papers can arouse interest if they are structurally simple (easy to process), stylistically catchy (entertaining), but also lexically focused (informative).

If such linguistic-textual strategies actually recur in titles and lead to systematic differences across genres, experienced readers of academic discourse may become sensitive to them, and learn to recognize which genres given titles are associated with. In this paper I explore which systematic differences in structure, content and wording, if any, can be identified in titles of academic works in linguistics representing different genres. To this end, I overview the literature on titles of academic publications. Next I illustrate the data used for this study, and describe the analytical approach adopted. I then detail the findings. Finally, I summarise and comment on the results obtained.

2. Literature review

There is a vast literature on titlelogy, including titlelogy in academic discourse. Here I overview findings from recent linguistic and, partly, scientometric publications in English on titles of scientific publications.[1]

Buxton and Meadows (1977) measured the information content of research paper titles by looking at content words. They examined hundreds of journal article titles from English, French and German periodicals in several disciplines. The authors found that titles in natural sciences (especially chemistry and botany) had higher information content values than those in the social sciences (with philosophy having the lowest values). They also found that in most disciplines there had been an increase in the number of content words from 1947 to 1973, including in the English translations of the German and French titles, and that in chemistry these content words mostly identified new techniques and aspects studied. The authors attributed this finding to a widespread attributive use of nouns, but also to readers' need for easy retrievability of information as a result of ever-increasing numbers of papers being published. Finally, the authors pointed out that, independently of their information content, titles in the social sciences were less suitable to retrieval due to a lack of semi-systematic nomenclature, which is typical of the natural/hard sciences.

Dillon (1981) compared the use of colon in the titles of journal articles differing in degree of scholarship (defined in terms of dimensions abstraction of thought, protraction of endeavour and relation of entailment). His examination of 804 titles from journals in education revealed that the presence of colons in titles was the strongest in theoretical research journals (34%), less marked in empirical research journals (18%) and quite low in pedagogical journals (10%).

Dillon (1982) explored the use of colons in titles of 1,150 journal articles in education, psychology and literary criticism over a 100-year period (roughly, 1880-1980). The author noticed a steady increase in the use of colons across the disciplines, recording its first occurrence in a literary criticism journal.

Michelson (1994) examined over 2,000 titles in industrial relations journal articles. About 38% contained colons, and the presence of colons was inversely proportional to the status of the journals, as determined by their age.

Yitzhaki (1994) analysed the titles of research articles from 14 sciences, social sciences and humanities journals over a 60-year period. The author measured title informativeness as defined on the basis of number of content words, and its possible correlation with the number of authors. In the scientific fields there was a moderate positive correlation between number of authors and number of content words for most of the periods considered. However, in the social sciences the correlation was low and relevant to a minority of the titles considered. Finally, the humanities mainly displayed a negative correlation. The author attributed the positive correlation identified in the sciences to the high rate of multiple authorship in scientific papers.

Berkenkotter and Huckin (1995) examined 350 journal biology articles published between 1944 and 1989, and showed how, over time, the titles of these works had become more informative, syntactically more complex, semantically richer and more promotional in style.

Nord (1995) examined 12,000 titles and headings of various publications – including scholarly journals articles – in English, French, German and Spanish in order to describe their communicative functions, generic features, cultural conventions, structural patterns, and similarities and differences in source vs target texts in translation. She identified three essential functions of titles – relevant to the general communicative situation – and three optional functional of titles – determined by specific interactional circumstances. The first set was found to include the distinctive function, which ensures the unmistakable distinctiveness of the publication the title names, the metatextual function, meant to ensure the recognisability of the genre the publication exemplifies, and the phatic function, which involves arousing and maintaining attention in the culture-specific target audience. The second set was said to comprise the referential function, which corresponds to the understandability of the information conveyed, the expressive function, which has to do with the author's attitude toward the text labelled by the title, and the appellative function, which means the potential, envisaged attractiveness of the title for the prospective readers. Nord found that the four language- and culture-specific corpus components considered displayed the same frequency hierarchy of optional functions – in an average relationship of 100 : 30 : 6 for the referential, appellative and expressive function, respectively. But she also noticed intra-corpus differences, for instance, the highest and lowest degree of expressivity, respectively, in the German and the Spanish titles. The author also pointed out the lack of culture-specific differentiations in genres like poems and scholarly articles, the latter displaying a low proportion of appellative and expressive elements. Moreover, Nord exemplified the macrostructural types tiles realise: simple title, title-subtitle combinations, duplex titles with "or" and series titles (sequences of titles belonging to different texts), and their syntactic forms: nominal, verbal, sentence, adverbial, attributive and interjection titles. Finally, the author discussed the implications of her findings for translation practice, especially the need to be loyal both to the authors' intentions in the source text and the recipients' expectations in the target culture.

Fortanet et al. (1997) analysed the structure and content of 200 titles research articles in computer science, applied linguistics, business and economics, and chemistry. The chemistry and the linguistics titles turned out to contain, respectively, the highest and the lowest number of words. Three punctuation marks (colon, semicolon, and full stop) were the most frequent in business and economics titles, and the least common in the computer science titles. Their occurrence marked the presence of two information units expressing, respectively, the general framework of the article and the specific topic, or alternatively, the topic and the method. As for content, most titles (especially in chemistry, and least of all in linguistics) conveyed the general topic and specific focus of the studies, while one third mentioned the nature of the research conducted.

Fortanet et al. (1998) examined specific linguistic features of the above-mentioned corpus of titles. The most common syntactic structure was 'premodifier + head + postmodifier', although combinations of heads were more frequent in linguistics and business and economics, while combinations of pre- and postmodifiers were more frequent in chemistry and computer science. Both linguistics

and business and economics displayed a majority of *–ing* forms – functioning as either nouns or verbs – while chemistry and computer science showed a balanced distribution between *–ing* and *–ed* forms – functioning as either verbs or adjectives. Finally, the linguistics and the economics and business titles favoured the use of definite articles, while the chemistry and computer science titles showed an even distribution of definite and indefinite articles.

Whissell (1999) examined over 3,000 abstracts of psychology articles from highly cited journals. Their titles turned out to be, on average, 12 words long, and to contain punctuation marks marking strong conceptual pauses about 18% of the time, which contributed to the complexity of the abstracts, together with other textual strategies.

Busch-Lauer (2000) explored the appropriateness (i.e. length, structure and communicative effectiveness) of 150 German and English titles in linguistics and medicine collected from journals articles and conference papers, and 25 English titles written by German researchers. The linguistics titles were shorter than the medicine titles (8.4 vs 9.9 words), and the German titles were shorter than the English ones. Also, the medicine titles in English (whether L1 or L2) preferred a mono-structure format, while the titles in German and in linguistics preferred a title-subtitle structure. The sequencing of syntactic constructions in title-subtitle structures was mostly nominal-nominal, although the linguistics titles more frequently instantiated verbal and clausal constructions. The most common semantic relationship between the titles and relevant subtitles was 'general-to-specific topic' across disciplines and languages; moreover, the linguistics subtitles were often the only comprehensible and informative component of the title-subtitle combinations. As for content, the medicine titles turned out to be precise and informative about the sub-genre of the relevant papers (e.g. case study), and about the purpose and/or results of the studies. The linguistics titles, instead, often mentioned the process of the research carried out rather than the findings, were vague and unspecified, but also creative, richer in the use of rhetorical devices, and reflective of the writers' stylistic preferences. Overall, the medicine titles were long, precise, informative and helpful to researchers and bibliographers for their research and documentation purposes, respectively. The linguistics titles, instead, were short, vague, abstract, catchy and stylistically individualised, but less effective in helping readers trace recent research developments.

Anthony (2001) examined the length, word frequency, and preposition and punctuation usage of 600 titles representing various computer science sub-disciplines. The average title length was 9 words, most titles clustering around 6-to-12 words. On average, two-unit titles, with the colon separating them, made up about 13% of the data. The two most frequent semantic relationships holding between the two parts of a title were 'name : description' and 'topic : scope', but with considerable variation across journals. Other semantic categories identified for the title units of the corpus were 'name of approach/algorithm/application', 'description of approach/algorithm/application', 'topic of research', 'scope of research' and 'method of research'. Statistically significant high-frequency words varied from journal to journal, reflecting their sub-disciplinary content specificity.

Goodman et al. (2001) examined the content of the titles of 420 peer-reviewed medical journal articles and queried the journals' editors on title-specific editorial policies and practices. About 40% of the titles provided information only on the topics discussed, 33% on the studies' topics and methods, 18% on their topics and results, and 2% on their topics and conclusions, while others were ambiguous. The journal editors occasionally modified titles to increase their clarity and informativity, and the only journal having a policy on titles merely addressed the issue of title length.

Yakhontova (2002) compared structural, syntactic, semantic and pragmatic characteristics of 400 titles of conference presentation abstracts representative of the disciplines of linguistics and applied mathematics in English vs Ukrainian and Russian. Various realisation patterns were instantiated in the two language-specific sub-corpora, but to different degrees. Nominative constructions dominated in the corpus, especially in the Slavic component and in linguistics; titles consisting of two parts separated by a colon were more frequent in the English component and in mathematics; and titles realised as incomplete sentences were more common in the Slavic sub-corpus, especially in linguistics. The English titles were more self-promotional from a structural, lexical and rhetorical point of view. Also, the English and Slavic titles in mathematics were more similar to each other than those in linguistics. Finally, the Slavic titles more often than the English ones highlighted the aspects of the research themes investigated and focused on the theoretical aspects of the investigations carried out.

Haggan (2004) compared the syntactic encoding, structure and content of over 700 journal article titles in literature, linguistics and education. In the three disciplines considered, similar syntactic and structural choices were made, but to different degrees: titles could consist of single information units (especially in science and linguistics) or compound ones (especially in literature), and could be formulated as complete sentences (especially in science) or phrases (especially nominal), the latter being characterised by the coordination of heads (especially in literature and linguistics) and/or their post-modification (especially in science). However, Haggan also identified important disciplinary differences in terms of content and rhetorical effects, which were independent of the structural-syntactic encoding of the titles: the science titles presented straightforward information on the findings or topics of the papers; this was probably meant to quickly and efficiently orientate the reader. The literature titles, instead, offered titillating and enigmatic hints of the content to follow, which revealed an attempt to seduce and attract the reader. The linguistics titles fell in between, showing formal features typical of science titles, but content features more similar to literature titles.

Hartley (2005, 2007a, 2007b, 2007c, 2008) and Lewinson and Hartley (2005) conducted a series of studies on thousands of titles of academic publications in various disciplines. According to their findings, scholars and students preferred titles with colons; preference for colonic titles was strong among the Arts and Social Sciences, single-authored papers and keynote address speeches; the number of titles with colons had been increasing over time; titles with colons were longer and more informative than those without; there was an even distribution be-

tween colonic titles with a longer first part (opening statement) and those with a longer second part (qualification); preference for colonic titles was similar among both highly-cited and infrequently cited papers; titles differed widely across disciplines in length (8-15 words), structure (with or without colons), and content (e.g. they could introduce a general subject, specify a precise theme, indicate a question, express the author's argument, emphasise the methodology, suggest guidelines, and attract readers through allusions, alliteration or vagueness); the nationality of the authors had little effect on title length or structure; finally, titles differed across genres, book titles being shorter and more to the point than article titles, and to prefer full sentences over sentences split by colons.

Soler (2007) examined the structure, lexicon and syntax of the titles of 480 journal review papers and 90 journal research papers in biological and social sciences. The author identified four main structural constructions, which, however, reveal a conflation of syntactic, semantic and textual classificatory parameters: nominal group, compound, full-sentence and question titles. The most common construction across disciplines and genres was the nominal group, consisting of nominal or verbal heads, possibly with pre- and/or post-modifiers. This structure was used to name, classify and describe the phenomena studied. The full-sentence construction, instead, was a generic and disciplinary peculiarity of biology research papers, and was used to present findings of experiments in a conclusive and synthetic way. The compound construction, common in the research papers and the social sciences, was used to focus attention on specific aspects of the objects of study, the first part introducing a general topic, and the second a specific one. The question construction was infrequently used, but more common among the review papers. The length of the titles varied across disciplines: the average number of words was 10.89 in the social sciences (and 7.98 in linguistics), and 14.98 in the natural sciences (and 15.48 in medicine).

Wang and Bai (2007) analysed the structure and encoding of 417 titles in medical research articles. The average length of the titles was 10.9 words, most of them being realised as nominal groups (99%), with no subtitles (98%), and characterised by the presence of single heads (75%) accompanied by post-modifying prepositional phrases (68%).

Mungra (2007) looked at the use of metaphors in the titles of the 1,426 articles published in one medical journal over a one-year period. Only 62 titles (4%) were metaphorical, and these mostly labelled editorials or opinion articles. Two main types of metaphors occurred in the titles and the body of the texts: primary ones, having a straightforward reference to the tenor or vehicle, and complex ones, formed by the blending of two input domains.

Hyland (2002) and Campagna (2008) also briefly addressed the topic of titles in academic language. In his 1.8 million word corpus of research articles, textbooks and students' essays, Hyland noticed that questions in titles only occurred in research articles in the soft fields. Campagna, who examined conference handouts reproducing PowerPoint slides, found that while the typographic layout of titles tended to become more varied, its content and structure was similar to that of article titles, whose information flow moved from the generic to the specific.

In conclusion, research has revealed that titles in academic discourse vary

across disciplines, sub-disciplines, languages and cultures along several dimensions: length in words, punctuation strategies, structural organisation, syntactic encoding, lexical choices, content conveyed and the semantic relationships between structural components. Titles of scholarly publications have also been found to change over time, becoming more complex syntactically, more informative semantically, and more promotional stylistically. The focus of the research on journal article titles – except for a few studies examining conference papers and conference presentation abstracts – has left unaddressed the question of whether titles of academic works differ across the genres of the publications they name. Marginal exceptions to this trend are Hartley's, Dillon's and Soler's studies, reviewed above. Hartley noticed differences in length and content between book titles and article titles. Dillon noticed that the use of colons in titles correlates with degrees of publication (i.e. colons are more common in the titles of published rather than unpublished works, and in the titles of books and articles rather than dissertation abstracts). He also observed that the presence of colons correlates with degrees of scholarship among published works (i.e. it is the strongest in the titles of theoretical publications). Soler noticed generic preferences in the structure of titles, the full-sentence and compound titles being common among research papers, and the question construction among review papers. However, no study has investigated whether titles of publications representing different genres inherit the constraints and options of those genres, that is, whether the different communicative goals that specific genres satisfy call for different communicative strategies also in the titles metonymically representing them (see section 1.). This paper addresses the issue of the possible inter-generic differences among titles within the discipline of linguistics.

3. Data and approach

I describe 1,000 English titles of publications in linguistics, dated between 1970 and 2004, grouped into four 250-title sets. Each set exemplifies one publication type (i.e. books (BOOK), dissertations (DISS), journal articles (JOURN) and proceedings papers (PROC)) and 10 keywords (i.e. *bilingual(s); discourse; learning; morphology; phonetic/phonological; pragmatic(s); semantic; sociolinguistic; speech act(s)/lexical; syntactic/syntax*)[2] representing different areas and topics of investigation in linguistics.

I collected the titles from the MLA bibliography by using on-line queries. The query interface allowed me to specify the keywords in the *Title* box, the publication years in the *Years* box, and the publication type in the *Publication* box, except for *proceedings paper*. Therefore, to retrieve proceedings paper titles, I typed the search word *proceedings* in the *Anywhere* box. From each query output, I selected the first 25 relevant records, but excluding six types of titles:

(i) titles not (completely) in English; e.g.

> (5) "Svenska partikelverb med in, ut, upp och ner: En semantisk "Svenska partikelverb med in, ut, upp och ner: En semantisk studie ur kognitivt

perspektiv/Swedish Phrasal Verbs with in, ut, upp och ner: A Semantic Study from a Cognitive Perspective",

unless the non-English words or expressions identified the object of study; e.g.:

(6) "Declaring Speech Act in Conversation: A Study of Japanese Connective Datte" (2.sp-act/lex-proc);

(ii) titles not unambiguously about linguistics; e.g.:

(7) "Plastic Glasses and Church Fathers: Semantic Extension from the Ethnoscience Tradition";

(iii) titles containing a segment not unambiguously classifiable as a subtitle; e.g.:

(8) "Knowledge of Meaning: An Introduction to Semantic Theory; Bradford Book";

(iv) titles of atlases, bibliographies, reviews, dictionaries, anthologies, readers, encyclopaedias and multi-volume works; e.g.:

(9) "Teaching and Learning a Second Language: A Review of Recent Research" (review)

(10) "Ultra Lingua auf Deutsch: German-English Bilingual Dictionary" (dictionary)

(11) "Routledge Encyclopedia of Language Teaching and Learning" (encyclopedia);

(v) titles whose keyword was joined to another word by means of a hyphen or slash; e.g.:

(12) "Morphology-Driven Syntax: A Theory of V to I Raising and Pro-Drop" (keyword: *morphology*),

unless the query keyword itself consisted of two graphic units, possibly occurring hyphenated; e.g.:

(13) "Evidence for the Imperative as a Speech-Act Category" (16.sp-act/lex-journ; keyword: *speech act*);

and (vi) titles already included in another data set; e.g.:

> (14) "Sociolinguistic Constructs of Ethnic Identity: The Syntactic Delineation of an American Indian English" (2.socio-book; possible keyword: *syntactic*, but title already chosen for the socio-book corpus component).

I examined the length, textual organization, syntactic realizations and content of the titles. To measure length, I considered the number of words and syntactic constituents, both phrasal and clausal, making up the titles. To describe the textual organization of the titles, I examined their constituent information units, as identifiable by the presence of specific punctuation marks. The analysis of the syntactic realization of titles comprised, at a macro level, the classification of information units into phrasal vs clausal structures, and at a micro level, the identification of expansion strategies (i.e. embedding, complementation, apposition, and pre- and post-modification structures) in phrasal constituents. The examination of content involved the measurement of lexical density, and the classification of the semantic relationships between the main title and the subtitle in two-unit titles.

4. Findings

4.1. Length in words

I measured the length of titles in number of words. I defined words typographically, as strings of letters preceded and/or followed by spaces or punctuation marks. I thus regarded non-hyphenated compounds, capitalised abbreviations and numerical sequences as single words. Moreover, I resorted to syntactic-semantic criteria in case of hyphenated words: I regarded hyphenated sequences of strings of letters as instances of multiple words, if their constituents could function as independent units within a clause, and as instances of single words if their constituents functioned as bound morphemes; e.g.:

> (15) "You Know My Steez: An Ethnographic and Sociolinguistic Study of Styleshifting in a Black American Speech Community" (3.socio-diss; 17 words; *styleshifting*: 1 word)
> (16) "Implicatures in Discourse: The Case of Spanish NP Anaphora" (14. disc-book; 9 words; *NP*: 1 word)
> (17) "Acquisition of Spanish Verb Morphology by Bilingual Children: A Longitudinal Study between the Ages of 2;9 and 3;3" (18.bil-diss; 18 words; *2;9 and 3;3*: 3 words)
> (18) "English-Learning Toddlers' Sensitivity to Agreement Morphology in Receptive Grammar" (2.morph-proc; 9 words; *English-learning*: 2 words)
> (19) "The Effect of Character Structure on Children's Learning of Chinese Pseudo-Characters" (2.learn-journ; 11 words; *pseudo-characters*: 1 word).

Table 1 shows the total number of title words in the title-genres considered, and the average number of words per title across title-genres. The longest titles occur in DISS (about 13 words each, on average), followed by JOURN (about 11 words), PROC (about 10 words) and BOOK (about 9 words).[3]

Title-genre	Total words	Average words
BOOK	2,307	9.2
DISS	3,232	12.9
JOURN	2,701	10.8
PROC	2,496	9.9
Global	10,736	10.7

Table 1: Total words in sub-corpora and average words per title

4.2. Structural organisation

To examine the global structure of the titles, I distinguished between single-unit titles, consisting of one information unit, and multi-unit titles, comprising two or more. The presence of structural units within titles was determined by the occurrence of specific punctuation marks (i.e. colons, semicolons, full stops, question marks or dashes) marking strong internal pauses (cf. Anthony 2001: 189). However, I did not count other types of punctuation marks (e.g. commas, parentheses), or punctuation marks inside other punctuation marks, or linkers and prepositions as markers of title-internal boundaries;[4] e.g.:

(20) "Teaching Language, Learning Culture" (22.learn-book; 1 unit)
(21) "The Sociolinguistic Situation of the Polish Language of the Slavic-Lithuanian Borderlands (the Region of the Present-Day Countries: Byelorussia, Lithuania and Latvia)" (24.socio-journ; 1 unit)
(22) "Genuine Training in Academic Discourse or an Artificial Construct? Reconsidering the Past, Present, and Future of the College Research Paper" (11.disc-diss; 2 units)
(23) "Vocal Communication in the Small-Eared Bushbaby (Otolemur Garnettii): Morphology, Sound Structure, and Social Context" (20. morph-diss; 2 units)
(24) "Pragmatic Particles – Polite but Powerless? Tone-Group Terminal hein and quoi in Contemporary Spoken French" (3.pragm-journ; 3 units)
(25) "'I Lost the Bus: Can You Give Me a Ride Home?': Native and Nonnative English Speakers' Speech Act Production and Metapragmatic Judgments: A Study of Apologies, Complaints and Requests" (19.sp-act-diss; 3 units)
(26) "A War of Words: From Lod to Twin Towers: Defining Terrorism in Arab and Israeli Newspapers 1972-1996 (2001): A Study in Propaganda, Semantics and Pragmatics" (23.pragm-book; 4 units).

Title-genre	1-unit titles	2-unit titles	3-unit titles	4-unit titles	Total units
BOOK	96 (38.4%)	151 (60.4%)	2 (0.8%)	1 (0.4%)	408
DISS	128 (51.2%)	120 (48.0%)	2 (0.8%)	0 (0.0%)	374
JOURN	144 (57.6%)	104 (41.6%)	2 (0.8%)	0 (0.0%)	358
PROC	134 (53.6%)	116 (46.4%)	0 (0.0%)	0 (0.0%)	366
Total titles	502 (50.2%)	491 (49.1%)	6 (0.6%)	1 (0.1%)	1,506

Table 2: Structure of titles

Table 2 shows the frequency and distribution of one-, two-, three- and four-unit titles in the corpus, and the total number of units per corpus component. One- and two-unit titles account for most the data. They are fairly equally distributed across the sub-corpora. BOOK prefers one-unit titles (about 60%), while the other three sub-corpora slightly favour two-unit titles. Table 3 shows that the average number of words per unit varies from a minimum of 5.6 in BOOK to a maximum of 8.6 in DISS, while JOURN and PROC have intermediate values (7.5 and 6.7, respectively).

Title-genre	Total words	Total units	Average words per unit
BOOK	2,307	408	5.6
DISS	3,232	374	8.6
JOURN	2,701	358	7.5
PROC	2,496	366	6.7
Global	10,736	1,506	7.1

Table 3: Total words, total units and average words per unit across sub-corpora

4.3. Syntactic encoding

Title units in the corpus are realised as noun phrases, prepositional phrases, adjectival phrases, verb phrases, clauses and combinations of the above.[5] Additionally, these basic structures can be combined through coordination, or alternatively, expanded through embedding and/or pre- or post-modification of the heads of phrasal units.

The title units realised as noun phrases mostly have nominal, but occasionally verbal or even adverbial heads (cf. Wang and Bai 2007). They can consist of one or a series of coordinated phrases,[6] each of which can be enriched by modification or other forms of expansion; e.g.:

(27) "But Still a Yet" (12.sem-diss; single NP with an adverbial head)
(28) "Bilingual conversation" (10.bil-book; single NP with adjectival pre-modification)
(29) "The Filipino Bilingual's Competence" (15.bil-book; single NP with multiple pre-modification)
(30) "The Development of Past Tense Morphology in L2 Spanish" (13. morph-book; single NP with post-modification)
(31) "Rhetoric as Discourse" (19.disc-diss; single NP with complement expansion)
(32) "The Reasons We Speak" (10.disc-book; single NP with relative clause expansion)
(33) "A Pragmatic Logic for Commands" (20.pragm-book; single NP with pre- and post-modification)
(34) "Hellenistic and Roman Greece as a Sociolinguistic Area" (18.socio-book; single NP with pre-modification and complement expansion)
(35) "The Semantic Structure of Roget's, a Whole-Language Thesaurus" (14.sem-diss; single NP with pre- and post-modification and appositional expansion)
(36) "Control vs. Cooperation" (12.bil-diss; coordinated, unmodified NPs)
(37) "One System or Two?" (15-bil-proc; coordinated NPs, of which one with pre-modification and the other with ellipsis of the head)
(38) "Semitic and Indo-European, Volume II" (4.morph-book; coordinated NPs with appositional expansion)
(39) "Emotive Signs in Language and Semantic Functioning of Derived Nouns in Russian" (20.sem-book; coordinated NPs with nominal and verbal heads, the first NP with pre-modification and the second with pre- and post-modification)
(40) "A Pragmatic Logic for Commands" (20.pragm-book; NP with adjectival pre-modification and prepositional post-modification).

Other title units are realised as prepositional phrases, either single or coordinated; the latter are optionally characterised by the deletion of the prepositional head; e.g.:

(41) "On Predicting Pragmatic Relations" (18.pragm-proc; single PP)
(42) "Not by Perception Alone" (2.sem-proc; single PP)
(43) "From the Viewpoint of 'Perception' and 'Cognition'" (22.sem-journ; single PP)
(44) "From UG to Universals" (3.learn-journ; coordinated PPs)
(45) "On the Placement and Morphology of Clitics" (10.morph-book; coordinated PPs)
(46) "On Learning and Teaching a Second Language" (20.learn-book; coordinated PPs).

Title units are occasionally realized as adjectival phrases; these can be single or coordinated, and may be accompanied by pre- and/or post-modification; e.g.

(47) "Utterly Content in Each Other's Company" (19.sem-journ; single AP with adverbial pre-modification and embedded post-modifying PP)
(48) "Phonetic or phonological?" (8.phon-proc; coordinated APs)
(49) "Semantic and Syntactic" (2.sem-book; coordinated APs).

Title units are often realized as verb phrases. Single verb phrases always come with some form of nominal, prepositional and/or verbal expansion; e.g.

(50) "Making Semantic Interpretation Parser-Independent" (13.sem-proc; single VP with object NP and object complement)
(51) "Exploring the Role of Morphology in the Evolution of Spanish" (11. morph-book; single VP with post-modified object NP and an adverbial)
(52) "Based on Phonological and Morphological Principles" (17.phon-book; single VP with object PPs)
(53) "Learning to read" (3.learn-diss; single VP with object VP).

In the case of coordinated verb phrases, instead, post-verbal expansion or complementation is an option; e.g.:

(54) "Pretending and Meaning" (7.pragm-book; coordinated VPs with no expansion)
(55) "Teaching Language, Learning Culture" (22.learn-book; coordinated VPs with object NPs)
(56) "To Be an Actor or to Be an Observer?" (14.disc-journ; coordinated VPs with NP complementation).

The presence of a V-*ing* form as the head of a phrase required part-of-speech classification based on co-textual information. When a V-*ing* form was associated with the syntax typical of nouns (e.g. preceded by a determiner and/or adjective, and/or followed by an embedded prepositional phrase), I regarded it as the head of a noun phrase. When it co-occurred with the arguments typically required by the verb used as a predicate, I counted it as the head of a verb phrase; e.g.:

(57) "Meaning and Time" (12.disc-diss; V-*ing* coordinated with a noun: NP)
(58) "Speech-Language Pathologists' Training and Confidence in Serving Spanish-English Bilingual Children" (20.bil-journ; V-*ing* with nominal pre-modification: NP)
(59) "The Representation and Processing of Verbal Morphology in the First and Second Language" (14.morph-diss; V-*ing* with the syntax typical of nouns and coordinated with a noun: NP)
(60) "Choosing the Right Spelling in Greek" (2.morph-journ; V-*ing* + arguments: VP).

I found only a few examples of clausal title units; e.g.

(61) "Where Lexicon and Syntax Meet" (21.syn-book; declarative clause)
(62) "Does Latent Semantic Analysis Actually Have a Latent Structure?" (18.sem-diss; interrogative clause).

Certain syntactically complex title units required more elaborate classification procedures. I classified units displaying two or more different syntactic encoding strategies as a combination of syntactic constituents; instead, I classified units compatible with more than one syntactic interpretation as ambiguous; e.g.

(63) "Metaphoring as One Kind of Speech Act" (6.sp-act/lex-proc; double coding: VP + NP)
(64) "Changing Economy, Changing Markets" (16.socio-diss; NPs or VPs?: ambiguous).

Syntax of title units	BOOK (%)	DISS (%)	JOURN (%)	PROC (%)	Average %
NP	357 (87.5)	325 (86.9)	299 (83.5)	300 (82.0)	84.9
VP	27 (6.7)	32 (8.6)	18 (5.0)	31 (8.5)	7.2
PP	13 (3.2)	5 (1.3)	13 (3.6)	11 (3.0)	2.7
AP	1 (0.2)	0 (0.0)	2 (0.6)	1 (0.3)	0.2
Clause	5 (1.2)	9 (2.4)	20 (5.6)	21 (5.7)	3.7
Unclear	5 (1.2)	3 (0.8)	6 (1.7)	2 (0.5)	1.0
Global	408 (100)	374 (100)	358 (100)	476 (100)	

Table 4: Frequency and distribution of units across syntactic types

Table 4 shows the frequency and distribution of units over their syntactic realizations. NP title units account for about 85% of the data, the other encoding options being less frequent. This corroborates previous findings (e.g. Haggan 2004; Soler 2007; Wang and Bai 2007).

The frequency hierarchy of the syntactic options for encoding title units is NP > VP > Clause > PP both for the corpus as a whole and for the BOOK, DISS and PROC sub-corpora. The frequency hierarchy in JOURN is slightly different: NP > Clause > VP > PP. These main syntactic structures show different distributional preferences: noun phrases are the most common in BOOK and the least in PROC; verb phrases are the most frequent in DISS and the least in JOURN; prepositional phrases are the most common in JOURN and the least in DISS; and clauses are mostly found in PROC and the least frequently in BOOK.

In two-unit titles, various combinations are attested: AP/NP, NP/NP, NP/AP, NP/PP, NP/VP, NP/Clause, PP/PP, PP/NP, PP/VP, VP/VP, VP/NP, VP/PP, VP/Clause, Clause/Clause, Clause/NP and Clause/VP. However, only NP/NP accounts for most of the data, as shown in the following frequency hierarchy:

NP/NP (71.8%) > VP/NP (9.1%) > Clause/NP (4.6%) > NP/VP (3.6%).

Table 5 shows that the distribution preferences of these combinations are heterogeneous across the sub-corpora: NP/NP is the most common in BOOK (76.8%) and the least common in PROC (64.6%); PP/NP is the most common in PROC (3.4%) and totally absent from DISS (0%); and both VP/NP and Clause/NP are the most common in PROC (12% and 6.2%, respectively) and the least common in BOOK (6.2% and 2.6%, respectively).

Syntax of 2-title units	BOOK (%)	DISS (%)	JOURN (%)	PROC (%)	Average (%)
AP/NP	0 (0.00)	0 (0.00)	1 (0.96)	0 (0.00)	1 (0.20)
NP/NP	116 (76.82)	88 (73.33)	74 (71.15)	75 (64.65)	353 (71.89)
NP/AP	1 (0.66)	0 (0.00)	0 (0.00)	1 (0.86)	2 (0.40)
NP/PP	4 (2.64)	3 (2.50)	3 (2.88)	2 (1.72)	12 (2.44)
NP/VP	6 (3.97)	7 (5.83)	2 (1.92)	3 (2.58)	18 (3.66)
NP/Clause	0 (0.00)	2 (1.66)	3 (2.88)	2 (1.72)	7 (1.42)
NP/Unclear	1 (0.66)	0 (0.00)	1 (0.96)	0 (0.00)	2 (0.40)
PP/PP	1 (0.66)	0 (0.00)	1 (0.96)	0 (0.00)	2 (0.40)
PP/NP	1 (0.66)	0 (0.00)	2 (1.92)	4 (3.44)	7 (1.42)
PP/VP	1 (0.66)	1 (0.83)	0 (0.00)	0 (0.00)	2 (0.40)
NP/Unclear	1 (0.66)	0 (0.00)	1 (0.96)	0 (0.00)	2 (0.40)
PP/PP	1 (0.66)	0 (0.00)	1 (0.96)	0 (0.00)	2 (0.40)
PP/NP	1 (0.66)	0 (0.00)	2 (1.92)	4 (3.44)	7 (1.42)
PP/VP	1 (0.66)	1 (0.83)	0 (0.00)	0 (0.00)	2 (0.40)
VP/VP	1 (0.66)	0 (0.00)	0 (0.00)	1 (0.86)	2 (0.40)
VP/NP	10 (6.62)	13 (10.83)	8 (7.69)	14 (12.06)	45 (9.16)
VP/PP	1 (0.66)	0 (0.00)	0 (0.00)	0 (0.00)	1 (0.20)
VP/Clause	0 (0.00)	0 (0.00)	1 (0.96)	2 (1.72)	3 (0.61)
Clause/Clause	0 (0.00)	0 (0.00)	0 (0.00)	1 (0.86)	1 (0.20)
Clause/NP	4 (2.64)	4 (3.33)	6 (5.76)	9 (7.75)	23 (4.68)
Clause/VP	0 (0.00)	0 (0.00)	0 (0.00)	2 (1.72)	2 (0.40)
Unclear/NP	3 (1.98)	1 (0.83)	2 (1.92)	0 (0.00)	6 (1.22)
Unclear/PP	1 (0.66)	0 (0.00)	0 (0.00)	0 (0.00)	1 (0.20)
Unclear/VP	0 (0.00)	1 (0.83)	0 (0.00)	0 (0.00)	1 (0.20)
Total	151 (100)	120 (100)	104 (100)	116 (100)	491 (100)

Table 5: Syntactic encoding of two-unit titles

4.4. Sub-phrasal syntax

An examination of phrase-internal syntax in the title units reveals that pre-modification is a common strategy. Given the prominence of noun phrases, pre-modification is often applied to nominal heads, which can be qualified by nouns, verbs and/or adjectives in various combinations.[7] The frequency hierarchy of nominal pre-modification is DISS (611) > JOURN (525) > PROC (505) > BOOK (426); e.g.:

> (65) "Karelian-Russian Language Alternation" (4.bil-book; pre-modifier: compound adjective)
> (66) "Semantically Ambiguous Words" (19.sem-diss; pre-modifiers: adverb + adjective)
> (67) "Arab and American Panel News Interviews" (25.disc-diss; pre-modifiers: coordinated adjectives and compound noun)
> (68) "Chinese Indonesian Mother-Daughter Pairs" (24.socio-diss; pre-modifiers: adjectives + compound noun)
> (69) "Web-Based Language Learning System" (1.learn-proc; pre-modifiers: noun + verb form (adjectival compound))
> (70) "ESL Korean Learners' Decision-Making Processes" (17.sp-act/lex-journ; combination of pre-modifiers)
> (71) "The Changing Sociolinguistic Status" (18.socio-journ; combination of pre-modifiers)
> (72) "Mixed Language Varieties" (25.disc-journ; combination of pre-modifiers).

Similar forms of pre-modification apply to the verbal heads of noun phrases. Their frequency hierarchy is DISS (27) > JOURN/PROC (26) > BOOK (18); e.g.:

> (73) "Saudi Arabic-English Intrasentential Codeswitching" (17.syn-diss; pre-modifier: adjective)
> (74) "Language Switching" (22.bil-journ; pre-modifier: noun)
> (75) "Foreign Language Learning" (11.learn-book; pre-modifiers: adjective + noun)
> (76) "Data-Driven Learning" (3.learn-proc; pre-modifiers: noun + verb form).

The most frequent pre-modification types are 'single adjective' (55.5%), 'single noun' (18.8%), 'sequence of adjectives' (11.8%), 'single adjective and single noun' (8.2%) and 'sequence of nouns' (5.7%). Their distribution varies across the sub-corpora: BOOK displays the most limited occurrences of all pre-modification types; DISS has the highest number of three types, while JOURN and PROC have intermediate frequency values for two pre-modification types (i.e. 'single adjective' and 'single adjective + single noun'). Table 6 visually summarises these findings.

Pre-modification	BOOK	DISS	JOURN	PROC	Global (%)
1Adj+N	239	262	245	242	988 (55.5)
≥ 2Adj+N	44	61	48	57	210 (11.8)
1Adj+1N+N	28	50	39	29	146 (8.2)
1N+N	51	97	82	105	335 (18.8)
≥ 2N+N	19	27	30	25	101 (5.7)
Total	381	497	444	458	1780 (100)
Percentage	21.4%	27.9%	24.9%	25.8%	100%

Table 6: Most frequent types of pre-modification (heads are underlined)

Like pre-modification, post-modification typically qualifies the nominal and verbal heads of noun phrase title units. Post-nominal modification comprises the use of single or coordinated nouns, prepositional phrases or equivalents, and reduced relative clauses headed by V-*ing* or V-*ed* forms.[8] Moreover, pre- and post-modification can co-occur; e.g.:

(77) "The Pronouns rswo and rsaccaw" (4.socio-proc; post-modifier: coordinated nouns)
(78) "the Understanding of Verbal Irony" (15.pragm-journ; post-modifier: PP)
(79) "Language Learning as Social Modeling in the Northwest Amazon" (4.learn-journ; post-modifier: PP equivalent)
(80) "A Case Study Comparing Quebecois in Montreal and Texas Spanish in San Antonio" (17.bil-diss; post-modifier: V-*ing*-headed reduced relative clause)
(81) "The Modifying Strategies Used by Deaf Students in the Speech Act of Apologizing" (23.sp-act/lex-diss; post-modifier: V-*ed*-headed reduced relative clause)
(82) "English Spatial Prepositions by, on and into" (22.bil-diss; pre- and post-modification).

Title units with V-*ing* heads having a predicative function can be accompanied by nominal or clausal complementation. There are 165 occurrences of such complementation; they are twice as frequent in DISS (54) as in BOOK (23), while JOURN and PROC display intermediate frequency values (47 and 41, respectively); e.g.:

(83) "Understanding What Is Said and What Is Implicated" (24.pragm-diss; post-verbal clausal complementation)
(84) "Positioning Gender in Discourse" (7.disc-book; post-verbal nominal complementation)
(85) "Learning How to Do Things with Words in a Study Abroad Context" (5.learn-book; post-verbal clausal complementation).

Another form of expansion of phrasal title units is coordination. This usually combines phrases with heads of the same word class; e.g.:

(86) "Polite but Powerless?" (3.pragm-journ; coordinated APs)
(87) "Syntax, Information Structure and Intonation" (7.syn-book; coordinated NPs)
(88) "From the Japanese Case to a General Sociolinguistic Perspective" (17.socio-book; coordinated PPs)
(89) "Emblematizing or Stereotyping?" (25.disc-proc; coordinated VPs)
(90) "Phonological Phrasing and Syntactic Derivation" (19.syn-diss; coordinated NPs, of which one with a V-*ing* head).

Post-modification and coordination	**BOOK**	**DISS**	**JOURN**	**PROC**	**Total**
NP+1PP	129	90	116	112	447
NP+ ≥ 2PP	64	137	84	89	374
NP 1NP	42	36	27	42	147
NP 1NP+1PP	12	10	16	16	54

Table 7: Most frequent types of post-modification and coordination (heads are underlined)

Table 7 shows that the most common forms of post-modification are single prepositional phrases or sequences of prepositional phrases (cf. Wang and Bai 2007: 395), and that the most common forms of co-ordination are sequences of two noun phrases, optionally accompanied by a prepositional phrase. The former are much more frequent than the latter (821 vs 201 occurrences, respectively). Their distribution patterns vary, with one form (i.e. multiple prepositional post-modification) being the most and the least frequent, respectively, in DISS and BOOK.

Expansion form / Head	BOOK	DISS	JOURN	PROC	Total
Pre-modification / Adj	0	0	0	1	1
Pre-modification / N	426	611	529	505	2,071
Pre-modification / V-*ing*	18	27	26	26	97
Pre-modification / Other	0	1	0	0	1
Post-modification / Adj	0	0	2	2	4
Post-modification / N	220	267	242	233	962
Post-modification / O	7	5	16	6	34
Complementation / V-*ing*	23	54	47	41	165
Coordination / Adjs	1	0	1	1	3
Coordination / Ns	70	75	64	68	277
Coordination / PPs	7	1	4	1	0
Coordination / V-*ings*	4	3	0	1	8
Coordination / Other	4	7	16	3	30
Other	2	2	7	0	11
Total	782	1,053	954	888	3,677
Average per title	3.0	4.0	3.6	3.4	

Table 8: Distribution of expansion strategies

Table 8 summarises the details of the distribution of expansion forms (i.e. pre-modification, post-modification, coordination and complementation strategies) over the main word classes. Pre-modification and post-modification of nominal heads are the most frequent and the second most frequent expansion strategies, respectively, in the sub-corpora. The frequency hierarchy of four of the most frequent expansion strategies (i.e. pre-modification of nouns and V-*ing*s, post-modification of nouns, and complementation of V-*ing*s) is DISS > JOURN > PROC > BOOK. However, the frequency hierarchy of coordinated nouns is DISS > BOOK > PROC > JOURN.

4.5. Lexical density

Lexical density is the amount of information conveyed as a function of the number of content words employed; the assumption is that the higher the density of a discourse excerpt, the higher its informativity. Lexical density can be measured as the overall ratio of content words to function words, or as the frequency of content words per ranking (i.e. non-embedded) clause. In a corpus of titles, however, which mostly consists of sub-clausal units, it makes more sense to consider title units as the genre-specific equivalent of ranking clauses.

Assuming that, as written texts, the titles would be richer in content words than function words, I counted the latter. I regarded as function words occurrences of articles, conjunctions, prepositions, pronouns, possessives, *wh*-ques-

tion words, demonstratives and auxiliaries, unless these were mentioned as topics of research. I applied the same principle when similar reference was made to bound morphemes and phonemes; e.g.:

> (91) "Universal Semantic Primes, and Their Application to French Monolingual and English-French Bilingual Lexicography: English Spatial Prepositions by, on and into in French Translations of 'Alice's Adventures in Wonderland'" (22.bil-diss; 8 function words).[9]

Word counts	BOOK	DISS	JOURN	PROC	Global
Content words	1,574	2,207	1,852	1,731	7,364
Function words	733	1,025	849	765	3,372
Percentage of content words	68.2%	68.3%	68.6%	69.4%	68.6%
Average content words per title	6.3	8.8	7.4	6.9	7.4
Average function words per title	2.9	4.1	3.4	3.1	3.4
Average content words per title unit	3.9	5.9	5.2	3.6	4.9
Average function words per title unit	1.8	2.7	2.4	1.6	2.2
Total words / content words	1.5	1.5	1.5	1.4	1.5
Total words / function words	3.1	3.2	3.2	3.3	3.2
Content words / function words	2.1	2.2	2.2	2.3	2.2

Table 9: Measures of lexical density

Table 9 shows the total number of content and function words and various measures of lexical density. Content words outnumber function words both in the whole corpus and the sub-corpora. On average, there are about 7.3 content words per title and 4.8 content words per title unit. DISS and JOURN, however, have higher than average lexical density values per title and title unit.

4.6. Content analysis

To approach a description of the content of the titles, I used an intuitive, bottom-up approach. That is, I repeatedly read the titles until I started to notice a general pattern, namely that the information conveyed was relevant to two elements of the studies identified by the titles: the context of the research, including the phenomena investigated, and the research itself, or the way it had been carried out. After classifying the content of the titles into these two broad categories, I noticed that contextual information comprised reference to the spatio-temporal

setting of a linguistic phenomenon, the relevant language, the texts (or other data sources) examined, and the speakers involved, including their characteristics (e.g. problems, developmental stages, interactional scenarios and communities of practice). In coding the data, I classified each contextual element separately, while I grouped together the above-mentioned speaker-specific characteristics under the label 'context: other'; e.g.:

> (92) "Punjabi, Urdu, English in Pakistan: A Sociolinguistic Survey" (12. socio-book; context: spatial setting)
> (93) "Triglossia and Pragmatic Variety Choice in Nineteenth-Century Bruges: A Case Study in Historical Sociolinguistics" (11.pragm-journ; context: temporal setting)
> (94) "Aspects of the Syntax, the Pragmatics, and the Production of Code-Switching Cantonese and English" (24.pragm-book; context: languages)
> (95) "Cross-Language Blending of /l/ Gestures by Korean-English Bilingual Children" (3.bil-proc; context: speakers)
> (96) "Genuine Training in Academic Discourse or an Artificial Construct? Reconsidering the Past, Present, and Future of the College Research Paper" (11.disc-diss; context: text)
> (97) "A Comparison of Three Interventions for Children with Co-Occurring Pragmatic Language and Behavior Problems" (2.pragm-diss; context: other: speakers' problems)
> (98) "Imitation as a Basis for Phonetic Learning after the Critical Period" (9.phon-proc; context: other: speakers' developmental stages)
> (99) "Exploring Linguistic Differences in Academic Cultures" (22.disc-proc; context: other: communities of practice)
> (100) "Discourse Analysis at the Interface of Politics and the Media: Election Night Coverage" (24.disc-proc; context: other: interactional scenario).

Table 10 shows the frequency of occurrence of the most common context-relevant information units across corpus components and category types. 'Language' and 'speaker' are the notions most frequently referred to in the titles, followed by 'text', and then 'place'. The distribution hierarchy of information units about context is DISS > JOURN > PROC > BOOK.

Information unit	BOOK	DISS	JOURN	PROC	Global
Place	30	39	32	15	116
Language	91	107	97	82	377
Speaker	29	77	62	64	232
Text	40	56	36	28	160
Other	4	58	27	20	109
Total	194	337	254	209	994

Table 10: Information units about context with > 10 occurrences in one or more sub-corpora

When the title units contain information about the studies they name, the content expressed can have an appealing or promotional function or be clouded in vagueness (cf. Busch-Lauer 2000; Yakhontova 2002; Haggan 2004); e.g.:

> (101) "Give Syntax a Chance" (18.syn-journ; study: alluring/advertising function)
> (102) "Negotiating Bilingual and Bicultural Identities: Japanese Returnees betwixt Two Worlds" (1.bil-book; study: vagueness)
> (103) "Complaining and Commiserating: A Speech Act View of Solidarity in Spoken American English" (4.sp-act/lex-book; study: vagueness).

Much more often, however, research-focused titles do orientate the reader, by providing information about the linguistic fields relevant to the studies carried out, the domains investigated, specific facets of the linguistic phenomenon examined, the methods adopted in carrying out the studies (e.g. type of investigation, description of the project, the process of research, approach, implications or results), the theories used as background frameworks for the studies, and/or the work of individual scholars (cf. Busch-Lauer 2000; Yakhontova 2002; Haggan 2004); e.g.:

> (104) "Creolistics and Sociolinguistic Theories" (19.socio-journ; study: field)
> (105) "Expressivity and a Pragmatic Constraint on Object Reduplication in Bulgarian" (21.pragm-journ; study: domain)
> (106) "Sex Differences in Voice Onset Time: A Developmental Study of Phonetic Context Effects in British English" (7.phon-journ; study: facet)
> (107) "Blue in Old English: An Interdisciplinary Semantic Study" (7.sem-book; study: method)
> (108) "Speech Act Taxonomy as a Tool for Ethnographic Description: An Analysis Based on Videotapes of Continuous Behavior in Two New York Households" (7.sp-act/lex-book; study: method)
> (109) "Derivations: Exploring the Dynamics of Syntax" (11.syn-book; study: method)
> (110) "Phonetic Features in Language Production: An Experimental Examination of Phonetic Feature Errors" (13.phon-diss; study: method)
> (111) "Investigating Semantic Inhibition Using a Modified Independent Cue Task" (8.sem-diss; study: method)
> (112) "Connectionism, Language Production and Adult Aphasia: Elaboration of a Connectionist Framework for Lexical Processing and a Hypothesis of Agrammatic Aphasia" (22.sp-act/lex-book; study: theory)
> (113) "A Short Introduction to X-Bar Syntax and Transformations" (4.syn-book; study: theory)
> (114) "Brilliance, Energy and Size in Vowels: A Cross-Linguistic Study of Phonetic Symbolism" (15.phon-diss; study: theory)
> (115) "Investigating the Neuropsychological Bases of Script Knowledge: Differential Effects of Executive Dysfunction and Semantic Impairment in Dementia" (1.sem-diss; study: theory)

(116) "Pace Panini: Towards a Word-Based Theory of Morphology" (19. morph-book; study: scholar)
(117) "Syntactic Theories and Syntactic Methodology: A Reply to Seuren" (2.syn-journ; study: scholar).

A specific type of information about the studies consists in reference to their technical aspects, or to components of the phenomena studied or of the relevant investigations; these technical elements are identified by means of nominal or adjectival expressions; e.g.
- nominal: *alignment, analogy, causative, ergative, evidentiality, factivity, grammaticalisation, grapheme, implicature, infixation, logophority, minimal pair, move-structure, onset, phonotactics, reduplication, register, schemata, semantic field, stem, transformation, umlaut, variable*;
- adjectival: *constraint-ranking, corpus-driven, derivational, diachronic, electrophysiological, paleographic, pragmalinguistic, productive, right-branching, scalar, typological*); e.g.

(118) "Speech Act Schemata and Discourse Type" (3.sp-act/lex-proc; technical term: nominal)
(119) "Modeling Syntactic Constraints on Anaphoric Binding" (17.syn-proc; technical terms: adjectival and nominal)
(120) "Designing a Corpus-Based Grammar for Pragmatic Terminographic Definitions" (9.pragm-journ; technical terms: adjectival).

Content units about the studies are fairly equally distributed across the sub-corpora (see Table 11). They are the most common in PROC (550 occurrences) and DISS (549), less so in BOOK (522) and the least in JOURN (483). The most frequent content unit is 'Technical'. The second and third most frequent ones are 'Method' and 'Facet' in BOOK and DISS, and 'Facet' and 'Domain' in JOURN and PROC. The overall frequency hierarchy of information units about research is given below:

'Technical' (793) > 'Facet' (384) > 'Method' (346) > 'Domain' (279) > 'Theory' (177) > 'Vague' (86) > 'Field' (20) > 'Advertising' (19).

Information unit	BOOK	DISS	JOURN	PROC	Global
Advertising	3	3	10	3	19
Vague	20	21	27	18	86
Field	2	24	1	2	20
Domain	44	51	84	100	279
Facet	71	100	108	105	384
Method	115	104	57	70	346
Theory	39	39	42	57	177
Technical	228	229	151	185	793
Total	522	549	583	550	2,104

Table 11: Information units about studies with >10 occurrences in one or more sub-corpora

4.6.1. Information sequencing

I examined the sequencing of information in two-unit titles, that is, the types of semantic relationships holding between the first and second part of such titles, by drawing on my own classification of content units reported above (see section 4.6.) and on findings from previous studies, especially Swales and Feak (1994) and Busch-Lauer (2000). I identified four main types of information suitable for classifying the global content of title units: on the one hand, 'topic', 'method' and 'theory', relevant to the content of the studies, and on the other, 'context', relevant to the context of the studies. I also found it necessary to resort to three more content categories to deal with titles conveying the same type of information in both their first and second part. I classified the relevant information units as 'paraphrase' (if the same content was conveyed through alternative expressions), and 'general topic' or 'specific topic' (if the same type of content was being presented in broader vs more specific terms, respectively).[10] In the following examples, the hyphen (-) signals 'sequencing of information units in the first and second part of the title', while the plus sign (+) signals 'combination of content units in the same title unit':

(121) "Japanese Pidgin English in Hawaii: A Bilingual Description" (25.bil-book; context - method)

(122) "Literacy Learning in a Bilingual Classroom for Deaf Students: Negotiating between New Zealand Sign Language and English" (20.bil-diss; topic - context)

(123) "Symbolic Values of Foreign Language Use: From the Japanese Case to a General Sociolinguistic Perspective" (17.sem-book; topic - context+method)

(124) "Bilingual Instruction of Immigrant Children: A Theoretical Overview and Results from Empirical Research" (8.bil-book; topic - method)

(125) "Language Choice, Language Attitudes and Ethnic Identity in Bilingual Speakers: A Case Study Comparing Quebecois in Montreal and Texas Spanish in San Antonio" (17.bil-diss; topic - method+context)

(126) "Semantic Structure of Spanish: Meaning and Grammatical Form" (13.sem-book; general topic - specific topic)

(127) "Bilingual Education: Theories and Issues" (14.bil-book; topic - theory)

(128) "The Pragmatic Nature of Theatrical Discourse: The Performance-Text as a Macro Speech Act" (23.pragm-journ; topic - paraphrase)

(129) "Sociolinguistic Analysis of the Language Preferences of Adolescent-Bilinguals: Shifting Allegiances and Developing Identities" (8.socio-journ; specific topic – general topic).

Table 12 shows the frequency and distribution of the five most frequent sequencing patterns relevant to two-unit titles, accounting for 79% of the data. The frequency hierarchy of such sequences in two-unit titles reveals a high incidence of 'general-specific' semantic relationships (cf. Busch-Lauer 2000; Haggan 2004), followed by 'topic – method' in BOOK (26), DISS (16) and JOURN (13), and 'topic - method+context' in PROC (13):

> 'general-specific' (212) > 'topic-method' (63) > 'topic-context / topic-method+context' (40) > 'specific-general' (33).

Information unit	BOOK	DISS	JOURN	PROC	Global
Topic – context	91	2	11	8	40
Topic – method	26	16	13	8	63
Topic – method+context	7	16	4	13	40
General – specific	57	57	51	47	212
Specific – general	8	4	10	11	33
Total	107	105	89	87	388

Table 12: Five most frequent sequencing patterns in two-unit titles

5. Discussion and conclusion

Titles of academic publications count as *texts* (Nord 1995: 280; Haggan 2004: 312): they are cohesive and coherent self-contained discourse units, which are meant to satisfy a specific communicative purpose (i.e. to label and succinctly describe publications), and which can be removed from their context of production to be used in a different context of reception (e.g. for citation purposes). Despite being short, titles are *important* texts content-wise: in summarising the content of the publications they name, they affect the reader's first understanding of those publications. Also, titles are *consequential* texts at the level of writer-reader interaction: being the first, and visually prominent, components of larger texts, they determine the reader's first impressions of those texts, and thus influence the reader's decision of whether to read more of them (Bazerman 1985; Busch-Lauer 2000: 77; Day 1994: 15; Swales 2003: 179; Wang and Bai 2007: 389; Hartley 2007c: 554). Moreover, from the perspective of documentation and information science, titles are the *major* means of cataloguing and retrieving scientific literature (Busch-Lauer 2000: 90; Soler 2007: 91; Wang and Bai 2007: 389). Finally, titles are *demanding* texts for their writers: they have to be concise (due to space limitations), informative and/or appealing (to orientate and attract the reader) and acceptable (i.e. to conform to cultural, linguistic, generic, disciplinary and audience expectations).

The situational options and constraints of titles determine their shared characteristics such as economy (necessary to summarise and label other texts), nominal encoding (fit for naming entities) and informativity (use of content words, and of pre- and post-modification; Haggan 2004; Soler 2007: 97-98; Wang and Bai 2007), which are instantiated independently of the titles' relevance to specific languages, cultures, disciplines or genres. Yet, as contextually situated texts, titles also reflect the communicative needs, goals and peculiarities of the disciplines, sub-disciplinary fields, studies and texts they are relevant to (Anthony 2001; Fortanet et al. 1998), and thus differ across languages, cultures, disciplines and genres (see section 2.).

The above analysis has revealed similarities and differences among titles in linguistics across publication categories. The similarities outweigh the differences, and include: high lexical density; preference for one or two information units; sequencing of information in two-unit titles as a transition from a general topic to a specific one; frequent syntactic encoding of title units as noun phrases; high frequency of coordinated noun phrases, of adjectival pre-modification of nominal heads, and of prepositional post-modification of nominal heads; and frequent reference to the languages and technical aspects of the studies named by the titles. Such similarities corroborate previous findings, and are attributable to the titles' common genre and discipline membership, linguistic encoding, and communicative goals (i.e. to inform and guide, clearly, precisely and economically).

Among the differences that have been detected, some are systematic, that is, characterised by consistently ordered frequency hierarchies (i.e. DISS > JOURN > PROC > BOOK) in the distribution of certain features: the length of the titles in number of words (per sub-corpus, title and title unit), and the occurrence of total expansions, pre-modification strategies, post-modification resources, function words and information units about context. Other features, however, are non-systematic, revealing variable cross-genre distributional preferences, as is the case with number of total units, frequency of one-title units, occurrence of coordination or post-modification strategies, and sequencing of information units in two-unit titles. The following frequency hierarchies succinctly illustrate these divergent patterns:

- Title units: BOOK > DISS > PROC > JOURN
- One-unit titles: JOURN > PROC > DISS > BOOK
- Syntactic encoding of title units as NPs: BOOK > DISS > JOURN > PROC
- Syntactic encoding of title units as VPs: DISS > PROC > BOOK > JOURN
- Syntactic encoding of title units as Clauses: PROC > JOURN > DISS > BOOK
- Post-modification strategies alone: DISS > PROC > BOOK > JOURN
- Coordination strategies alone: PROC > BOOK > DISS > JOURN
- NP/NP sequence: BOOK (76.8%) > DISS (73.3%) > JOURN (71.1%) > PROC (64.6%)
- VP/NP sequence: PROC (12.0%) > DISS (10.8%) > JOURN (7.6%) > BOOK (6.6%)
- Clause/NP sequence: PROC (7.7%) > JOURN (5.7%) > DISS (3.3.%) > BOOK (2.6%)
- NP/VP sequence: DISS (5.8%) > BOOK (3.9%) > PROC (2.5%) > JOURN (1.9%)

It appears that the sub-corpora display fairly distinctive traits, but not totally clear-cut differences. BOOK prefers relatively short titles, often divided into two parts, each encoded as a noun phrase, and characterised by limited pre- and post-modification of head words in title units. These titles convey short, easy-to-decode and reader-friendly chunks of information units.

At the other end of the continuum, DISS prefers relatively long titles, with either one or two precise, detailed, elaborate information units. DISS shows the highest frequency of title units encoded as verb phrases, a strong preference for multiple post-modificaton of head words and high frequency of function words.

In between BOOK and DISS are the JOURN and PROC titles. The JOURN titles are shorter than the DISS ones, but longer than the PROC ones. Like those in DISS, they tend to consist of single units. As for their content and sequencing of information units, the JOURN titles are similar to the PROC ones, but their expansions (i.e. internal syntactic elaboration) are less frequent than in PROC. Overall, they are highly informative and moderately elaborate.

Finally, PROC comprises titles that are shorter than in JOURN, but longer than in BOOK. They display a slight preference for one-unit titles, in this more closely resembling DISS than BOOK. They are similar to the JOURN titles regarding types of information units. Surprisingly, their expansions are more frequent than in JOURN, and more generally, they are not as informal as expected (cf. Busch-Lauer 2000; Yakhontova 2002; Haggan 2004, reviewed in section 2.). A reason for this may be that the titles in this group came from published proceedings, and not from conference oral presentations; it may be that they were modified from their original formulation (contra Busch-Lauer 2000). Overall, the PROC titles are structurally simple, but moderately elaborate from a syntactic point of view, that is, easy to process and informative.

This study has explored intra-disciplinary and cross-generic similarities and differences in the encoding of titles of academic publications. Part of the findings – shared characteristics and distributional differences of certain features – are in line with previous findings, and can be plausibly explained with reference to the common communicative goal and different audience expectations of the genres the titles are relevant to. Other findings, revealing non-systematic differences, may tentatively be accounted for with reference to the corpus compilation procedure adopted. The collection of data was not carried out randomly. It involved the choice of keywords for the on-line queries identifying sub-disciplinary linguistic fields. The technical nature of these terms may have contributed to the collection of *similar* titles, that is, focused on the studies named rather than representative of varied content (including the encoding of evocative or vague notions). Given that, in general, lexicon is inextricably tied to phraseology, the a-priori selection of the same lexemes across title-genres may have triggered similar co-textual associations among the titles collected, which may have obscured or distorted some inter-genre differences.

The study can therefore be enhanced in two ways. First, further aspects of the titles collected can be investigated, such as the semantic relationships holding between coordinated noun phrases within the same title unit; the frequency of

occurrence of terms denoting abstract vs concrete entities; the occurrence of rare words; the semantic fields represented by keywords identified by comparing the title corpus with a larger, general corpus; the use of rhetorical effects; and distributional preferences of linguistic features already identified across sub-disciplinary linguistic fields (i.e. phonology vs discourse analysis; cf. Anthony 2001). In addition, the same analytical procedure followed here can be applied to sets of titles selected through different, non-lexical means (e.g. the titles of linguistics dissertations produced in a given year at a specific institution; the titles of the linguistics books published within a given series; the titles of the articles published by given leading journals over a specific period; and the titles of the presentations given at a series of conferences held at various institutions). The two complementary directions of investigation, oriented towards breadth and depth of analysis, are likely to shed more light on the degree and nature of intra- and inter-generic variation among titles.

NOTES

1 For studies addressing issues relevant to translation practice, the pedagogy of academic writing, and information comprehension, recall, selection and appeal, see the references in the literature reviewed below.

2 Parentheses and slashes signal alternative forms of a keyword that were typed together as variants of a given query by using the Boolean operator *OR* (e.g. *bilingual* OR *bilinguals*; *speech act* OR *speech acts* OR *lexical*).

3 Nord (1995: 274), instead, identified optimum book title length across cultures in the range of 3-4 words. Also, Soler (2007) found that the linguistics research papers in her corpus were 7.98 words long, on average.

4 Nord (1995: 280), however, identified slightly different macrostructural types of titles: single title, title + subtitle, title + *or* + title, and sequence of titles (see section 2.).

5 This is in line with the findings reported in other studies, such as Nord (1995) and Yakhontova (2002). However, Nord (1995) also retrieved adverbial and interjection titles.

6 Wang and Bai (2007) also reported similar variation in the realization of NP heads, and found instances of bi-head titles realized as text units coordinated by the function words *and*, *vs* and *or*.

7 Wang and Bai (2007) also found that pre-modifiers included mostly nominal or adjectival classifiers and deictics, but did not provide any frequency data.

8 Wang and Bai (2007) also found only embedded constructions as post-modifiers in their corpus, including one type not attested in my data, namely *to*-infinitive clauses.

9 Here and elsewhere, the words underlined are those relevant to the specific linguistic aspect being examined.

10 Hamp-Lyons (1987) suggested an alternative categorisation of titles – in terms of topic, focus, comment and viewpoint – which Anthony (2001) was able to adapt to the computer science paper titles he analysed.

REFERENCES

Anthony L. (2001) "Characteristic Features of Research Article Titles in Computer Science", *IEEE Transactions on Professional Communication*, 44(3): 187-194.

Bazerman C. (1985) "Physicists Reading Physics. Schema-Laden Purposes and Purpose-Laden Schema", *Written Communication*, 2(1): 3-23.

Berkenkotter C., Huckin T. (1995) *Genre Knowledge in Disciplinary Communication: Cognition/Culture/Power*, Hillsdale, NJ: Lawrence Erlbaum.

Busch-Lauer I.-A. (2000) "Titles of English and German Research Papers in Medicine and Linguistics Theses and Research Articles", in Trosborg A. (ed.) *Analysing Professional Genres*, Amsterdam/Philadelphia: John Benjamins, 77-94.

Buxton A. B., Meadows A. J. (1977) "The Variation in the Information Content of Titles of Research Papers with Time and Discipline", *Journal of Documentation*, 33(1): 46-52.

Campagna S. (2008) "Projecting Visual Reasoning", paper presented at the *CERLIS Conference "TRADING IDENTITIES: Commonality and Individuality in English Academic Discourse"*, University of Bergamo, 19-21 June, 2008.

Day R. A. (1994) *How to Write and Publish a Scientific Paper*, Cambridge: CUP.

Dillon J. T. (1981) "Functions of the Colon: An Empirical Test of Scholarly Character", *Educational Research Quarterly*, 5: 71-75.

Dillon J. T. (1982) "In Pursuit of the Colon: A Century of Scholarly Progress: 1880-1980", *Journal of Higher Education*, 53(1): 93-99.

Dressler W. U., Eckkramer E. M. (2001) "Functional Explanation in Contrastive Textology", *Logos and Language*, 2(1): 25-43.

Fortanet I., Coll J. F., Palmer J. C., Posteguillo S. (1997) "The Writing of Titles in Academic Research Articles", in Chamorro R. M., Navarrete A. R. (eds.) *Lenguas aplicadas a las ciencias y la tecnologia. Approximaciones*, Caceres, Spain: Universidada de Extremadura, Servicio de Publicaciones, 155-158.

Fortanet I., Posteguillo S., Coll J. F., Palmer J. C. (1998) "Linguistic Analysis of Research Article Titles: Disciplinary Variations", in Vazquez I., Camilleu I. (eds.) *Perspectivas Praguietices en Linguistica Aplicada*, Zaragoza, Spain: Anubar, 443-447.

Goodman R. A., Thacker S. B., Siegel P. Z. (2001) "What's in a Title? A Descriptive Study of Article Titles in Peer-Reviewed Medical Journals", *Science Editor*, 24(3): 75-78.

Haggan M. (2004) "Research Paper Titles in Literature, Linguistics and Science: Dimensions of Attraction", *Journal of Pragmatics*, 36: 293-317.

Hamp-Lyons L. (1987) *Study Writing: A Course in Written English for Academic and Professional Purposes*, Cambridge: CUP.

Hartley J. (2005) "To Attract or to Inform: What are Titles for?" *Journal of Technical Writing and Communication*, 35: 203-213.

Hartley J. (2007a) "Colonic titles!", *The Write Stuff: The Journal for European Medical Writers*, 16(4): 147-149.

Hartley J. (2007b) "There's More to the Title than Meets the Eye: Exploring the Possibilities", *Journal of Technical Writing and Communication*, 37: 95-101.

Hartley J. (2007c) "Planning that Title: Practices and Preferences for Titles with Colons in Academic Articles", *Library and Information Science Research*, 29: 553-568.

Hartley J. (2008) *Academic Writing and Publishing: A Practical Handbook*, Abingdon: Routledge.

Huth E. (1987) *Medical Style & Format: An International Manual for Authors, Editors, and Publishers*, Baltimore: Williams & Wilkins.

Hyland K. (2002) "What do they Mean? Questions in Academic Writing", *Text*, 22(4): 529-557.

Lewinson G., Hartley J. (2005) "What's in a Title? Number of Words and the Presence of Colons", *Scientometrics*, 63(2): 341-356.

Michelson G. (1994) "Use of Colons in Titles and Journal Status in Industrial Relations Journals", *Psychological Reports*, 74: 657-658.

Mungra P. (2007) "Metaphors among Titles of Medical Publications: An Observational Study", *Ibérica*, 14: 99-122.

Nord C. (1995) "Text-Functions in Translation: Titles and Headings as a Case in Point", *Target*, 7(2): 261-284.

Soler V. (2007) "Writing Titles in Science: An Exploratory Study", *English for Specific Purposes*, 26: 90-102.

Swales J. (2003) *Genre analysis. English in Academic and Research Settings*, Shangai: Shangai Foreign Language Education Press.

Swales J., Feak C. B. (1994) *Academic Writing for Graduate Students. A Course for Nonnative Speakers of English*. Ann Arbor: The University of Michigan Press.

Wang Y., Bai Y. (2007) "A Corpus-based Syntactic Study of Medical Research Article Titles", *System*, 35: 388-399.

Whissell C. (1999) "Linguistic Complexity of Abstracts and Titles in Highly Cited Journals", *Perceptual and Motor Skills*, 88: 76-86.

Yakhontova T. (2002) "Titles of Conference Presentation Abstracts: A Cross-Cultural Perspective", in Ventola E., Shalom C., Thompson S. (eds.) *The Language of Conferencing*, Frankfurt: Peter Lang, 277-300.

Yitzhaki M. (1994) "Relation of Title Length of Journal Articles to Number of Authors", *Scientometrics*, 30: 321-332.