

Jan Kochanowski University Press

This is a contribution from *Token: A Journal of English Linguistics*
Volume 16/2023.

Edited by John G. Newman, Marina Dossena.

© 2023 Jan Kochanowski University Press.

PATTERNS OF
LANGUAGE VARIATION AND CHANGE
IN ACADEMIC WRITING

Language variation and change in academic writing: Recent trends through globalisation and digitalisation

Josef Schmied*, Marina Bondi**,
Olga Dontcheva-Navratilova*** and Carmen Pérez-Llantada****

** Chemnitz University of Technology*

*** University of Modena and Reggio Emilia*

**** Masaryk University*

***** University of Zaragoza*

ABSTRACT

This article discusses variation and change in academic writing, integrating different approaches, from English for academic purposes to lingua franca studies and from contrastive rhetoric to discourse analysis, and various comparative perspectives from national to genre/part genre (e.g. research article abstracts or conclusions) or career-specific writings (e.g. BA, MA and PhD theses). It focuses on the interrelated development of discourse as social interaction in the context of technological affordances and societal demands and on the specific applications of the well-known trends of globalisation and digitalisation to non-native academic writing. Of course, the impact of recent changes varies with (sub-) disciplines, genres, and even individual researchers in their construction of careers and identities. The general trends, however, can be observed independently of whether we see them as functional necessity or advancement or threats to established conventions individually. A great number of small-scale empirical corpus studies should be able to provide a detailed mosaic where researchers can collaborate to provide a background for individual academic writers to choose from. Global rhetorical features (like IMRaD) and small-scale usages of pronouns are just examples of current variation and changes that are worth tracing in the wide field of metadiscourse that shapes academic interaction today, for the advancement of science communication and thus of science as a whole.

Keywords: academic writing, technological affordances, societal demands, corpus studies, genre analysis.

1. Research backgrounds and developments

Over the past 50 years, scholarly research in the field of English for Academic Purposes has provided insightful analyses of the rhetorical, linguistic and grammatical features characterising academic prose. It has also offered empirically informed descriptions of the ways in which different disciplinary cultures engage in academic writing practices. If we look at the literature in retrospect, we find early accounts of language use for academic and research purposes in pioneering works such as those of Barber (1962), Connor (1996), Kaplan (1966) and Swales (1984). This seminal research has set the grounds for a large stock of literature that has examined different aspects of academic writing from multidisciplinary perspectives, such as genre, discourse and register, EAP (English for Academic Purposes), contrastive rhetoric and academic ELF (English as a Lingua Franca), to name just the most prolific areas of research. Seminal work in these areas can be found in, e.g. Connor (1996), Hyland (2012), Mauranen (2018) and Swales (2004). Taking a grammatical perspective, research has also provided an important number of corpus-based multidimensional analyses that have claimed that academic prose exhibits economy features at a phrase level that make its discourse style different from those of other written registers such as journalism or science popularisations (Biber – Egbert 2018; Biber – Gray 2012). In recent years, increasing attention has been paid to digital media and how new insights into language variation in emerging genres can be gained by comparing online practices with traditional forms of academic writing. This can be seen, for instance, in descriptions of genres such as online registered reports that Mehlenbacher (2019) defines as a rhetorical hybrid of the journal article or, as another example, open access peer reviews that Breeze (2019) describes as having language conventions distinctly different from those of their antecedent genre, the traditional occluded peer review. From a different standpoint, recent genre and rhetorical studies have foregrounded the impact of remediation (or the shift from print form to digital form) on traditional genres that have moved on the web, and the emergence of hybrid research-related genres such as open data articles, video abstracts or open research notebooks, to name a few, that exhibit distinct patterns of language use and writing conventions at the level of discourse pragmatics different from traditional genres such as journal articles, abstracts and laboratory notebooks in print form (e.g. Cavalieri 2020, Pérez-Llantada 2022, Wickman 2016). In sum, this rich stock of research has not only captured a comprehensive picture of diversity in academic writing conventions across English language discourses, academic genres, and disciplinary cultures but

also engaged in the investigation of aspects of language change in relation to genre evolution and innovation in digital environments.

2. Current comparative research attempts and objectives

In this Special Issue of *Token* we build upon this valuable knowledge base to offer critical and reflective views of language variation and change in academic writing conventions in relation to discipline, language, culture and genre-specific diversity. This is in line with previous *Token* volumes on healthcare communication (Bondi – Poppi 2019) or on specialised discourse in general (Cavalieri – Mocini – Turnbull 2020). We also aim to better understand whether traditional language and discourse features merge with new features when new situational and digital contexts arise.

This Special Issue is the outcome of a seminar held at the European Society for the Study of English (ESSE) online conference in Lyon (France) in post-pandemic 2021. The seminar, organised by the four co-editors of this issue, was intended to encourage lively participation on the part of both speakers and members of the audience with a view to examining critically the topic under discussion (language variation and change in academic writing conventions) and do so from an international (and cross-national) perspective, by this means aiming to bring together insights from as many different European (English language) departments and universities as possible. The call for contributions of the seminar included an invitation to submit empirically informed (i.e. corpus-based and/or discourse-analytical) studies of academic metalanguage usage (e.g. hedging/boosting resources, markers of modality, reader mentions, and writer self-mentions, etc.), and comparative analyses of argumentative structures, research questions/hypotheses, cohesion/coherence resources, referencing, and other metadiscourse elements. We also welcomed contributions from all sub-disciplines (linguistics, literature, methodology, cultural/area studies, digital humanities, etc.) in order to establish a comparative state-of-the-art evaluation that could also provide guidelines for postgraduate seminars, summer schools, or on-line teaching.

3. From variation to change

This Special Issue addresses two broad topics, language variation and diachronic language change, narrowing them down to the context of academic writing. Broadly speaking, language variation is the outcome of social interactions,

for example, within and across discourse communities that may or may not have the same linguacultural and/or disciplinary backgrounds and scholarly traditions (Connor 1996; Hyland 2012; Schmied 2016, 2018). Thus, variation in this Special Issue specifically refers to linguistic diversity represented by different scholarly writing conventions and academic writing styles. Differences across such writing conventions and styles can be traced from a range of perspectives, but here the two most salient ones are the intercultural rhetoric perspective and the cross-disciplinary perspective. Along with linguistic, culture-specific and disciplinary differences, further evidence of purported language variation in the context of academic writing has been mainly associated with the phenomena of languages in contact and socialisation into other academic practices and, as a result, the widespread use of English as a *Lingua Franca* (ELF hereafter), particularly in the context of the general expansion and development of English-medium scholarly publications. The ELF perspective assumes that there is no universal standard of ‘good academic writing’ and claims that Anglo-American rhetoric and recurring lexicogrammar represent just one scholarly tradition for textualising new knowledge and structuring academic texts (Mauranen – Pérez-Llantada – Swales 2010: 664).

On the other hand, diachronic change and, in particular, changes that affect the English language system itself, can be defined as changes that are shaped and constrained by the social and technological signs of the times. Here variation is inextricably related to language change and to the impact that the proliferation of English-medium publications has had on writing in English. Focusing on changes at the turn of the 21st century, Hyland and Jiang (2019: 227-230) have explored research articles by carrying out a corpus-based multidimensional analysis of three moments: 1965, 1990 and 2015. The study shows for example that the social sciences seem to be moving towards greater informational focus and a preference for empirical, experimental and data-informed investigations, while the hard sciences are increasing their use of involvement features (e.g. first- and second-person pronouns, modality, evaluative language and engagement markers). Changes may be related to the impact of a wider audience for science in recent years.

4. From multidimensional analyses to small-scale comparative case studies of non-native writing traditions

Using multidimensional analysis, studies with large-scale corpora such as the Longman Corpus of Spoken and Written English (Biber et al. 1999) and the FROWN/FLOB longitudinal corpora have reported “variety-internal

variation or diachronic change" (Leech et al. 2009: 180), specifically variation of American/British English at the level of grammatical and syntactic complexity (Baker 2017). These studies further report that while some written registers like fiction exhibit an increasing use of colloquial linguistic features, academic prose is a "tight-up register" resisting historical change. Diachronically, scientific writing proves to rely more heavily on economy linguistic features for informational purposes (Biber – Gray 2012: 326).

This Special Issue seeks to complement studies of language variation and diachronic change from genre and register perspectives by offering the perspectives of English for Academic Purposes and academic writing development. It focuses on writing practices in the non-native European traditions, which has an important potential for examining linguistically and culturally diverse academic writing conventions in this geographic region. Accordingly, all the selected contributions explore patterns of language variation and change involving different academic writing practices. They intend to shed light on aspects of language variation and change, and discipline and genre conventions over time, but all enquiring into the use of standardisation conventions cutting across academic genres and languages for scholarly communication. The contributions of this Special Issue cover a range of (sub)disciplines (e.g. linguistics, literature, cultural studies, philosophy, educational sciences and applied psycholinguistics) as well as both traditional and emerging forms of communication online (e.g. journal articles and abstracts, Masters/PhD theses, online newsletters and monograph chapters, research projects websites and corporate websites). Furthermore, they also enquire into academic writing conventions by putting the focus on language variation across linguacultural backgrounds (e.g. Czech, English, German, Italian, Polish and Spanish). While examining a range of linguistic and discourse features, covering rhetorical move/step organisation of the texts, discourse topicalisation, phraseological features, and interactive and interactional metadiscourse features such as stance and evaluation, all the contributions identify language features that indicate or signal continuity or change. They all report on the main tendencies observed in a period of increasingly international dimension of research and they reflect on how such variation and changes should be taken on board when teaching specialised writing across the disciplines and subdisciplines (as in the widely-used textbook by Swales – Feak 2012). Some of the reported findings further attest to the widespread use of English as an international lingua franca, no longer viewed as a language exclusively belonging to native English speakers, by this means contesting uniform standards in academic

writing practices in English academic writing and, therefore, advocating the diversity of particular academic writing cultures and scholarly traditions.

The corpus-based studies compiled in this Special Issue also showcase the synergies that can be created among interrelated interpretative theoretical frameworks and perspectives (English for Academic Purposes, genre analysis, discourse analysis, intercultural rhetoric and English as a Lingua Franca). In doing so, they offer complementary insights into convergent and divergent academic English language usage across various European traditions, cultural contexts and over time. In this respect, these mutually-informing perspectives may pave the way for cross-regional comparisons in future language descriptions and linguistics research in academic writing. Small-scale comparative corpus studies of relatively simple linguistic and rhetorical features may in the end contribute to a wide mosaic of linguacultural, genre- and discipline-specific analyses that are also presented here to tempt individual researchers to contribute either in their own research or their own writing.

As also claimed in other contributions of this issue, diachronic variation in the use of rhetorical conventions (move/step organisation) may be related to socio-cultural factors such as different intellectual styles and cultural patterns or to the influence (or lack) of formal academic writing instruction. It is also concluded that the observed cross-linguistic variation might indicate not only differences in cultural writing conventions, but also in the relationship between the writer and the discourse community s/he addresses. Most contributions report important findings regarding academic language and discourse variation in the use of recurring phraseological units, frame markers, labellers, self-mentions, stance markers and positive/negative words, among other features, in traditional genres such as articles, abstracts and theses that rely on well-established academic writing conventions. In turn, these findings also further invite reflection on the use of metadiscourse strategies in web-mediated writing practices supported by Internet affordances (e.g. journal websites and research group websites) that are also explored in the last contributions of this Special Issue. These studies could pave the way to future investigations into similar practices in relation to web-mediated genres and social media (Twitter, LinkedIn, etc.), where writers need to interact with diversified audiences.

5. Socio-technological changes: The example of academic reviews

The impact of the digital transition on academic discourse has been profound. On the one hand, this has coincided with the internationalisation

of discourse communities and it has certainly favoured it. International scientific communities thrive on evolving technological possibilities that potentially favour access to knowledge and allow for closer transnational contact. On the other hand, the digital transition has also diversified the forms of academic interaction through new platforms, forms of collaborative writing and multiplying the possibilities of interaction both within the discourse community and across different audiences. Academics now find themselves trying their hand at new communicative environments (such as websites or social media), and at the same time have to adjust to changes in more traditional forms of communication.

A particularly interesting example of the multi-faceted interplay between technological affordances and societal demands can be seen in recent trends in reviews (Schmied 2021). This genre has received particular importance because publications in peer-reviewed journals with high impact factors are considered central for an academic career nowadays. The spoken equivalent to the central journal publication in the written genres is the conference presentation, interestingly also called somewhat euphemistically a “paper”. Again, the main quality-assurance mechanism is the peer review. In this case, the review is an evaluation of a promise, since the reviewer does not have to comment on a complete paper but only on a conference abstract or proposal, which might be developed into a much longer presentation by the time of the conference – often months ahead. Here disciplinary conventions have to be distinguished: whereas in social sciences the abstract may be much longer and receive more qualitative than quantitative changes, the humanities may leave more room for review adaptations by the presenter.

Recent technological affordances have made the review process cloud-based. Many conferences use general submission systems (e.g. the free Easy Abstract from the Linguist list), which also include a “review facility”. Some conferences make all texts available (to conference participants) counting on inspiring constructive and fair academic online interaction even before and after the conference.

The social demands of open science have diversified the review process, as not only conference advisors but theoretically all interested self-appointed specialists can contribute openly online. This allows the author to choose from more diverse evaluations while at the same time often making the assessment of these evaluations more difficult, so that a self-critical confidence rating in addition to other standardised ratings may be helpful. The open review process may be expanded in related online genres like

rebuttals, that allow presenters to comment on reviewers' comments and appraisals, which allow authors to indicate how useful they find reviewers' comments.

Both technological and societal changes have made the review process more accessible and more transparent. Interactivity in the review process has thus created new opportunities and new conventions. Now new open online text types like "How to write constructive online reviews" or even "Criticising with Kindness" or "Mistakes Reviewers Make" can be found.¹

It remains to be seen whether open reviews are "the solution" to several issues in academic writing today: they may help solve the current problem of finding peers who are willing to sacrifice some of their precious research time to support the careers of their colleagues by evaluation and thus improving their academic output (especially in conference papers and journal publications) and they may also make academic evaluation more transparent, from power-conscious gate-keeping to the functional development of genres. However, it also remains to be seen whether academic writers will be sufficiently rewarded for contributing to improving others' academic texts instead of developing their own and whether they will always provide constructive criticism instead of fighting "Wiki wars" at a professional level. In any case, open science solutions have the potential to bring more social interaction between academic professionals and the general public, but at least between research novices and research experts – which is another important development worth following.

6. Diachronic perspectives through socio-technological and personal career developments

The diachronic comparative analyses or diachronic studies included in this Special Issue also shed important light on the potential of the language of stance to express authorial voice and to construct an academic identity in relation to researchers' professional development over time. All diachronically oriented contributions reflect on the impact of social factors on the construction of culturally and disciplinary defined academic identities, by considering socio-political changes (like the Bologna process introducing BA and MA studies everywhere to harmonise higher education in the EU or the spread of English-medium instruction within the process of

¹ See, for example, <https://iclr.cc/Conferences/2020/ReviewerGuide>.

internationalisation and marketisation of EU universities [Bowles – Murphy 2020]), the growing status of English as an international language and the size of the national and international discourse community typically correlating with an orientation towards competitive or collegial community dynamics.

The findings of several contributions strongly suggest that the process of integration of local European academic discourse communities in the global scholarly exchange has resulted in considerable idiosyncratic variation in the discourse of L2 scholars. Integration deepens gradually through the functional expansion of scientific writings from digesting science at the beginning of a career, to developing one's own identity through theses writing up to (post-)doctoral levels and contributing to the global advancement in the (sub-)discipline through international research journals and conferences at professional level. Throughout this process, tendencies emerging within the local academic discourse communities interact with changes taking place in international discourse communities, thus leading to multifarious academic identities and considerable diversity in the expression of stance and the construction of authorial voice. These divergences have gradually given rise to hybrid, 'glocal' academic discourses bearing the signs of the dominant Anglo-American academic discourse conventions and of local L2 academic literacies (Dontcheva-Navratilova 2014).

On a more fine-grained level, variation and change might be the outcome of changing scholarly practices, for example, a scholar's expertise, affiliation with local or foreign universities, preferred context of publication, changes in research methodologies or increasing use of open access. Yet all diachronic studies comprised in this volume point to a movement from diversity in rhetorical choices and language means for the expression of stance and voice towards gradual conventionalisation and standardisation over the last three decades. This tendency towards convergence seems to respond to the requirements of the highly competitive research marketplace (Cheek – Øby 2019), where scholars strive to boost their credibility by enhancing the replicability of their research and the informativeness and surveyability of their academic texts. Together with this general trend towards a greater role of promotional elements in academic writing, the studies included in this Special Issue also show convergence towards a more stable generic structure in traditional research genres in the humanities, along the lines of the standards developed in the sciences (Bondi 2022). These findings could further pave the way to carrying out future research on writing within and across the boundaries of academic conventions.

7. Globalisation and the debate of English native-speakerism

It is widely accepted that academia is one of the domains in which English has expanded and become the main lingua franca of international scholarly communication. The predominant role of English was initially viewed as a source of tension with scholars having to choose either English or their national language for publishing their research work. Such a role was triggered by national and supranational research and internationalisation policies in higher education and research institutions, which made non-Anglophone scholars in Europe and elsewhere gradually shift to English-only for research publication purposes. In other words, publishing in English was seen as a policy-imposed necessity rather than as a personal opportunity to share their work in the globalising research world (Ferguson – Pérez-Llantada – Plo 2011; Pérez-Llantada – Plo – Ferguson 2011). Notwithstanding this imposition, research has concluded that academic English is perceived not as a ‘language for identification’ but as a ‘language for communication’ insofar as it provides them global access, greater visibility of research and possibilities for international collaboration.

The shift to English for research publication purposes brought about an important language phenomenon that has conceptualised academic discourse as a linguistically diverse discourse. Intercultural rhetoric and linguistic research over the past decades have provided compelling evidence that while the syntax of the L1 English texts is very simplified, with short sentences and straight argumentation, the syntax of the texts written in English by non-native English scholars may be syntactically dense, with greater use of coordination, subordination and complementation constructions. This suggests that the English of L2 writers maintains conventions of the native language and culture of the scholars (Berns 1995: 6). Academic language variation has been traced through comparisons of academic texts written in English by non-native English scholars from different linguacultural backgrounds. At present, there is substantial empirical evidence of language variation in L2 English academic texts, for example regarding aspects of authorial identity, stance and audience engagement. Pragmatic features of the texts also differ significantly across academic English varieties. At the same time, different academic Englishes are shown to share common trends. This Special Issue further attests to the fact that in academic writing there is room for shared normative conventions, whether based on Anglo-American conventions or determined by general evolving features of an international discourse community, as well as for writing conventions that

are more typical of the scholars' L1 epistemological and rhetorical traditions. Academic Englishes have been described in the European context and elsewhere (e.g. South America, Africa and Asia) and research concludes that academic English needs to be dissociated from its native linguacultural roots, and decentralised from the Anglo-American English (core) variety (Mauranen – Pérez-Llantada – Swales 2010). Academic English as a lingua franca contests standardisation, it observes no national boundaries and it has no definite centres. It is part of transcultural flows, with scholars using English in their own ways, constructing their own identities and conveying authorial positioning in very diverse ways.

It is also worth recalling that while these varieties of academic English were initially considered to be “defective forms of English” (Greenbaum 1996: 17) from the perspective of peer reviewers and journal gatekeepers (cf. Gosden 2003), current ELF research very convincingly argues that they are not. They are an entirely natural development of the widespread use of English in academia, a parallel phenomenon to that of World Englishes (Mauranen 2018). Supporting this claim, empirical research has shown that academic communication relies on the use of ‘good English’, not necessarily Anglo-American English, and that non-canonical grammar usage in journal articles, as long as it does not impede clarity and intelligibility of content, does not hamper acceptance for publication in journals (Rozycki – Johnson 2013; Hynninen – Kuteeva 2017). In other words, from the initial views of academic Englishes as variants or deviations of Anglo-American standards as a result of first language influences and incomplete English language competence set against English native-speakerism, the current view of academic English diversity in scholarly writing practices is one that acknowledges the plurilingual and multicultural diversity of scholarly communication in the age of globalisation (Sano 2002: 49). Today, increasing digitalisation of scholarly communication practices leads to increasing societal demands (and pressure) to reach non-specialised, diversified audiences through open science policies. Open science² has become a key driver, aiming to promote public communication of science so that research has a *de facto* impact on society and other science stakeholders (Bartling – Friesike 2014). Considering

² This article is a contribution to the project Digital genres and Open Science “Géneros digitales y ciencia en abierto” (project code PID2019-105655RB-I00 MCIN/AEI 10.13039/501100011033) funded by the Spanish Ministry of Science and Innovation and the Spanish Agency for Research. It is also a contribution to the research group CIRES (Comunicación Académica y Retos Sociales) supported by the Gobierno de Aragon (H16_20R).

this, it seems reasonable to predict further English language variation and change over time. In fact, this has already been traced in emerging genres that support social accountability of science by relying on dissemination in social media, such as homepages, podcasts, tweets and blogs.

8. Linguistic and discourse variables in national, disciplinary, and genre contexts

The analysis of patterns in academic genres in this Special Issue focuses on research articles, because this is seen as the central genre for research development today. Four articles follow the “national” development over the last 30 years by Czech (**Dontcheva-Navratilova**), Italian (**Bondi – Nocella and Diani**), and German (**Ivanova**) linguists in English. They concentrate on global rhetorical structure and interactional features, such as self-mention and evaluative markers, as they are seen as most important for impact in readers. These diachronic comparisons are complemented by studies on the personal career development of an individual linguist from Poland as reflected in the structure and stance marker choices in conclusions of research articles and book chapters (**Warchał**), on the parallel development of the early career genre of English MA theses in the fields of Linguistics and Cultural Studies from Germany with a focus on evaluative that-complement clauses (**Schmied – Ivanova**) and on the comparison of the use of frame markers by non-native (ELF) writers with a comparable corpus by Anglophone writers (**Guziurová**). Finally, the latest developments in identity construction in European research project websites (with parallel company websites) adds a new multimodal dimension to the discussion of variation and change in academic writing (**Lafuente Millán**). The articles pay special attention to the variation and change in texts written by non-native linguists in English, with occasional comparisons to other languages (like Italian; see **Diani**), other subdisciplines (like cultural studies; **Schmied – Ivanova**) and native English writings (**Guziurová**). Of course, even as a group the articles in this Special Issue are only a small contribution to the wide multidimensional genre mosaic of academic writing today.

The variables analysed in these genres often include a wide range of linguistic and discourse features, especially global rhetorical structures, such as steps and moves, and various metadiscourse features, such as writer-reader interaction through pronouns, frames and evaluative markers. Our focus is on explicitness and transparency, or addressivity

and interactivity, which may be increased or reduced, depending on many individual choices that come together as varying patterns in the academic communities studied in the small-scale case studies here. The focus is not on the “accurate” distinction between specific and generic usage of determiners or the flexibility of idiomatic expressions found in many non-native, SciELF contributions, even in this Special Issue. Whether the few cases where such variation has an impact on readers are more relevant than pragmatic metadiscourse choices is arguable. The patterns observed are interpreted as legitimate diversity in international academic discourse rather than as deviation from a traditional or native norm (cf. Schmied – Hofmann – Esimaje 2018). The patterns found often display predictable variation; some changes point towards more uniformity, which may indicate new disciplinary genre conventions, others may still be within the usual fluidity or hybridity of genres. However, diversity and change are a decisive feature of academic writing and only constant, critical, empirical analysis may show new patterns in the constantly changing technological and societal contexts.

9. Further research and developments

In putting together this selection of contributions, with different explorations of discourses pertaining to a specialised domain of language, we also invite readers to reflect on the important implications for teaching academic writing across languages, writing cultures, modes and media that they pose. Both the empirical (corpus-based) data on language variation and the diachronic analyses of academic writing can provide better instructional support to researchers across the disciplines previously mentioned. The attested linguistic variation deems it necessary to acknowledge the role of ELF as a language variety or rather, varieties or “similects”, as Mauranen (2018) conceptualises them, that are closely tied to the rich repertoire of linguistic and cultural identities reflected in English academic writing practices. At the same time, the outcome of the diachronic studies strongly suggests that EAP practitioners need to take on board the increasing use of digital resources for text composing, by this means taking a proactive stand and raising students’ awareness of emerging digital writing practices. Indeed, the fast-changing technological scenario brings in new language and communication needs, and therefore, new learning needs. This should prompt reflection on possible ways to address emerging and evolving forms of communication that will very likely reflect language variation and change.

We hope that readers find this Special Issue³ inspiring and relevant in order to better understand current trends in the rhetorical and linguistic analysis of academic writing and, more broadly, aspects of language variation and change from a diachronic perspective. After all, even though technological affordances and sociocultural changes may be drivers of change, in the end it is up to academic writers and their identities to decide whether they find trends functionally convenient in their discourse communities and follow them or whether they see such trends as unnecessary universals and reject them. In any case, it is important to be aware of variation and change in academic styles and their implications, and we hope to make an informed empirical contribution to this debate to raise critical awareness in general and to empower non-native English, multilingual academic writers in particular to find their place in their local and international discourse communities.

REFERENCES

- Baker, P.
 2017 *American and British English. Divided by a Common Language?*
 Cambridge: Cambridge University Press.
<https://doi.org/10.1017/9781316105313>.
- Barber, C. L.
 1962 "Some measurable characteristics of modern scientific prose".
 Reprinted in J. M. Swales (ed.) (1988) *Episodes in ESP: A Source and Reference Book on the Development of English for Science and Technology*.
 New York: Prentice Hall, 1-15.
- Bartling, S. – S. Friesike (eds.)
 2014 *Opening Science*. Cham: Springer.
<https://doi.org/10.1007/978-3-319-00026-8>.

³ We gratefully acknowledge the scholars who supported the quality of the articles included in this Special Issue by reviewing them. Their time and expertise is very much appreciated. Their names are listed in alphabetical order: Martin Adam (Masaryk University, Czech Republic), Mirela Bardi (Bucharest University of Economic Studies, Romania), Maria Luisa Caiazzo (University of Basilicata, Italy), Silvia Cavalieri (University of Verona, Italy), Sara Gesuato (University of Padova, Italy), Marketa Mala (Charles University, Prague, Czech Republic), Silvie Murillo Ornat (University of Zaragoza, Spain), Pilar Mur Dueñas, (University of Zaragoza, Spain), Sandra Petroni (University of Rome 2, Italy), Renata Povolná (Masaryk University, Czech Republic), Michele Sala (University of Bergamo, Italy), Annalisa Sezzi (University of Modena and Reggio Emilia, Italy), Jolanta Sinkuniene (Vilnius University, Lithuania).

- Berns, M.
1995 "English in the European Union", *English Today* 11 (3), 3-11.
<https://doi.org/10.1017/S0266078400008348>.
- Biber, D. et al. (eds).
1999 *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education Limited.
- Biber, D. – J. Egbert
2018 *Register Variation Online*. Cambridge: Cambridge University Press.
<https://doi.org/10.1017/9781316388228>.
- Biber, D. – B. Gray
2012 "The competing demands of popularization vs. economy: Written language in the age of mass literacy". In: T. Nevalainen – E. C. Traugott (eds.) *The Oxford Handbook of the History of English*. Oxford: Oxford University Press, 314-328.
<https://doi.org/10.1093/oxfordhb/9780199922765.013.0028>.
- Bondi, M.
2022 "The scientific article: Variation and change in knowledge communication Practices". In: C. Hanganu-Bresch et al. (eds.) *The Routledge Handbook of Scientific Communication*. London: Routledge, 159-169.
<https://doi.org/10.4324/9781003043782>.
- Bondi, M. – F. Poppi (eds.)
2019 *Exploring Healthcare Communication and Health Literacy: Cognitive and Social Community Action in an Online World*. Special Issue of *Token: A Journal of English Linguistics* 9.
<https://dx.doi.org/10.25951/3288>.
- Bowles, H. – A. C. Murphy (eds.)
2020 *English-Medium Instruction and the Internationalization of Universities*. International and Development Education. Cham: Palgrave Macmillan.
https://doi.org/10.1007/978-3-030-47860-5_1.
- Breeze, R.
2019 "Continuity and change. Negotiating relationships in traditional and online peer review genres". In: M. J. Luzón – C. Pérez-Llantada (eds.) *Science Communication on the Internet: Old Genres Meet New Genres*. Amsterdam/Philadelphia: John Benjamins, 107-129.
<https://doi.org/10.1075/pbns.308.06bre>.
- Cavalieri, S.
2020 "From abstracts to video-abstracts: Academic argumentation in genre hybridization", *Token: A Journal of English Linguistics* 11, 41-58. Special Issue *Argumentation and Ideology in the Dissemination of Specialized Discourse*, edited by S. Cavalieri – R. Mocini – J. Turnbull.
<https://doi.org/10.25951/4306>.

- Cavalieri, S. – R. Mocini – J. Turnbull (eds.)
 2020 *Argumentation and Ideology in the Dissemination of Specialized Discourse*. Special Issue of *Token: A Journal of English Linguistics* 11. <https://doi.org/10.25951/4304>.
- Cheek, J. – E. Øby
 2019 “‘Getting attention’. Creating and presenting the visible, online, researcher self”, *Qualitative Inquiry* 25 (6), 571-582. <https://doi.org/10.1177/1077800418806593>.
- Connor, U.
 1996 *Contrastive Rhetoric. Cross-Cultural Aspects of Second Language Writing*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139524599>.
- Dontcheva-Navratilova, O.
 2014 “The changing face of Czech academic discourse”. In: K. Bennett (ed.) *The Semiperiphery of Academic Writing*. Basingstoke: Palgrave Macmillan, 39-61. https://doi.org/10.1057/9781137351197_3.
- Ferguson, G. – C. Pérez-Llantada – R. Plo
 2011 “English as an international language of scientific publication: A study of attitudes”, *World Englishes* 30 (1) (February), 41-59. <https://doi.org/10.1111/j.1467-971X.2010.01656.x>.
- Gosden, H.
 2003 “‘Why not give us the full story?’: Functions of referees’ comments in peer reviews of scientific research papers”, *Journal of English for Academic Purposes* 2 (2), 87-101. [https://doi.org/10.1016/S1475-1585\(02\)00037-1](https://doi.org/10.1016/S1475-1585(02)00037-1).
- Greenbaum, S.
 1996 *The Oxford English Grammar*. Oxford: Oxford University Press.
- Hyland, K.
 2012 *Disciplinary Identities. Individuality and Community in Academic Discourse*. Cambridge: Cambridge University Press. <https://doi.org/10.1093/applin/ams070>.
- Hyland, K. – F. (K.) Jiang
 2019 *Academic Discourse and Global Publishing: Disciplinary Persuasion in Changing Times*. London: Routledge. <https://doi.org/10.4324/9780429433962>.
- Hynninen, N. – M. Kuteeva
 2017 “‘Good’ and ‘acceptable’ English in L2 research writing: Ideals and realities in history and computer science”, *Journal of English for Academic Purposes* 30, 53-65. <https://doi.org/10.1016/j.jeap.2017.10.009>.
- Kaplan, R.
 1966 “Cultural thought patterns in intercultural education”, *Language Learning* 16 (1), 1-20. <https://doi.org/10.1111/j.1467-1770.1966.tb00804.x>.

Leech, G. et al.

- 2009 *Change in Contemporary English: A Grammatical Study*. Cambridge: Cambridge University Press.
<https://doi.org/10.1017/CBO9780511642210>.

Mauranen, A.

- 2012 *Exploring ELF: Academic English Shaped by Non-Native Speakers*. Cambridge: Cambridge University Press.
- 2018 "Second Language Acquisition, World Englishes, and English as a Lingua Franca (ELF)", *World Englishes* 37 (1), 106-119.
<https://doi.org/10.1111/weng.12306>.

Mauranen, A. – C. Pérez-Llantada – J. M. Swales

- 2010 "Academic Englishes: A standardised knowledge?". In: A. Kirkpatrick (ed.) *The Routledge Handbook of World Englishes*. London/ New York: Routledge, 659-676.

Mehlenbacher, A. R.

- 2019 "Registered reports: Genre evolution and the research article", *Written Communication* 36 (1), 38-67.
<https://doi.org/10.1177/0741088318804534>.

Pérez-Llantada, C.

- 2022 "Online data articles: The language of intersubjective stance in a rhetorical hybrid", *Written Communication* 39 (3), 400-425.
<https://doi.org/10.1177/07410883221087486>.

Pérez-Llantada, C. – R. Plo – G. Ferguson

- 2011 "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 (1), 18-30.
<https://doi.org/10.1016/j.esp.2010.05.001>.

Rozycki, W. – N. H. Johnson

- 2013 "Non-canonical grammar in Best Paper award winners in engineering", *English for Specific Purposes* 32 (3), 157-169.
<https://doi.org/10.1016/j.esp.2013.04.002>.

Sano, H.

- 2002 "The world's lingua franca of science", *English Today* 18 (4), 45-49.
<https://doi.org/10.1017/S0266078402004078>.

Schmied, J.

- 2016 "Academic writing world-wide: Comparing metadiscourse conventions, credibility, and new functional standards". In: J. Schmied – D. A. Nkemleke (eds.) *Academic Writing and Research across Disciplines in Africa*. Göttingen: Cuvillier, 9-34.
- 2018 "A global view on writing research articles for international journals: Principles and practices". In: J. Schmied – M. Hofmann – A. Esimaje (eds.) *Academic Writing for Africa: The Journal Article*. Göttingen: Cuvillier, 1-18.

- 2021 "Applying Appraisal Theory to conference reviews: Systemic functional perspectives in corpus-based genre-specific discourse analyses". In: J. Schmied – M. Hofmann – C. Chang (eds.) *Critical Discourse and Corpus Approaches to Systemic Functional Grammar*. Göttingen: Cuvillier, 1-22.
- Schmied, J. – M. Hofmann – A. Esimaje (eds.).
- 2018 *Academic Writing for Africa: The Journal Article*. Göttingen: Cuvillier.
- Swales, J. M.
- 1984 "ESP comes of age? – 21 years after 'Some Measurable Characteristics of Modern Scientific Prose'", *Unesco Alsed-LSP Newsletter (1977-2000)* 7 (2), 9-20.
- 2004 *Research Genres: Explorations and Applications*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139524827>.
- Swales, J. M. – C. B. Feak
- 2012 *Academic Writing for Graduate Students: Essential Tasks and Skills* (3rd edn.). Ann Arbor: University of Michigan Press. <https://doi.org/10.3998/mpub.2173936>.
- Wickman, C.
- 2016 "Learning to 'share your science': The open notebook as textual object and dynamic rhetorical space". In: A. G. Gross – J. E. Buehl (eds.) *Science and the Internet: Communicating Knowledge in a Digital Age*. Amityville, NY: Baywood's Technical Communications Series, 11-32.

Address: JOSEF SCHMIED, "English and American Studies, Chemnitz University of Technology, 09107 Chemnitz, Germany. josef.schmied@phil.tu-chemnitz.de.
ORCID code: <https://orcid.org/0000-0001-8499-3158>.

Address: MARINA BONDI, Department of Studies on Language and Culture, University of Modena and Reggio Emilia, Largo S. Eufemia 19, 41121 Modena, Italy. marina.bondi@unimore.it.
ORCID code: <https://orcid.org/0000-0002-3462-8387>.

Address: OLGA DONTCHEVA-NAVRATILOVA, Department of English Language and Literature, Faculty of Education, Masaryk University, Poříčí 9, Brno 602 00, Czech Republic. navratilova@ped.muni.cz.
ORCID code: <https://orcid.org/0000-0002-0378-7975>.

Address: CARMEN PÉREZ-LLANTADA, Department of English and German Studies / Institute for Biocomputation and Physics of Complex Systems (BIFI), University of Zaragoza, Spain. llantada@unizar.es.
ORCID code: <https://orcid.org/0000-0003-4052-1321>.

Academic writing conventions in Czech English-medium linguistics journals: Continuity and change over the last 30 years¹

Olga Dontcheva-Navratilova
Masaryk University

ABSTRACT

This paper studies the development of academic writing conventions in English-medium research articles (RAs) by Czech linguists published in two national journals (*Brno Studies in English* and *Linguistica Pragensia*) over the last 30 years. Drawing on the genre analysis framework, the study investigates possible changes in the titles, rhetorical structure, statement of aims, research questions and hypotheses, and personal and locational metadiscourse markers for writer and reader reference in a small corpus of 20 RAs. The comparative diachronic analysis aims to identify continuity and change in the evolution of academic writing conventions and the factors influencing them. The findings indicate that Czech English-medium RAs have gradually adopted a more transparent rhetorical structure close to the IMRAD model, their titles have gained in informativeness, and researcher visibility has been enhanced due to an increase in locational and exclusive personal self-mention. These tendencies point to hybridity in the present-day English-medium discourse of Czech linguists which stems from the adaptation of diverging academic writing traditions to meet the publication needs of the authors.

Keywords: research article, rhetorical structure, diachronic analysis, reader reference, writer reference.

¹ This work was supported by the Czech Science Foundation grant 21-12150S *Intercultural variation in writer-reader interaction in English-medium academic discourse by Czech and Anglophone novice writers*.

1. Introduction

Over the last 30 years, English has been firmly established as the language of international communication in the globalised academic world. The “academic lingua franca” status of English has facilitated rapid communication of new scientific knowledge while forcing non-Anglophone scholars to make their claims, report their results and interact with a large and varied international readership in an international language. Since writing in a language is typically associated with specific disciplinary, epistemological and cultural conventions, non-Anglophone writers may be seen as standing at a crossroads as they need to decide whether to accommodate the dominant Anglophone norms or continue to abide by the research writing conventions established in their original academic literacy. While English may be seen as a *Tyrannosaurus Rex* “gobbling up the other denizens of the academic linguistic grazing grounds” (Swales 1997: 347), the fate of the non-Anglophone academic literacies has gradually emerged as somewhat more complex (Tardy 2004).

Several studies have shown (e.g. Bondi 2009; Dontcheva-Navratilova 2014, 2021; Lafuente-Millán 2014; Lorés-Sanz 2011; Mur-Dueñas 2007; Šinkūnienė 2018; Shaw – Vassileva 2009) that the English-medium discourse of non-Anglophone scholars is somewhat similar to, but also different from, the writing of native Anglophone authors. This suggests that non-Anglophone scholars strive to resist stigmatization when aspiring to publish in international journals (cf. Flowerdew 2008) by preserving some of the discursive and rhetorical traditions of their original academic literacy, while at the same time adjusting to some extent to Anglophone academic discourse conventions. This simultaneous process of convergence with and divergence from Anglophone academic writing norms leads to hybridization (Pérez-Llantada 2013: 263) or ‘glocalization’ (Swales 2004:11; Sancho Guinda 2015: 29) of the discourse of non-Anglophone scholars. The diversity of culture-specific variants of academic English seems to be aptly grasped by the term ‘alternative academic written Englishes’ (Mauranen – Pérez-Llantada – Swales 2010: 647). Since the development of alternative academic Englishes has not yet attracted much scholarly attention from a diachronic perspective, the present study focuses on the development of one of these culture-specific variants of academic English (Czech) over the last 30 years.

The period 1990-2020 is characterised by major political and economic changes in Czech society which have impacted on the academic discourse community. The most significant change was the fall of the Iron Curtain (1989), which resulted in the opening of borders and led to more intensive

interaction with the global academic community, and gradually increased access to more recent publications. In this period, the local status of the Czech Anglicist community changed dramatically. Since before 1989 all relations with English-speaking countries were regarded as undesirable, and Russian was the second compulsory language taught in all schools, English studies at universities had a limited number of staff and students; consequently, the Czech Anglicist community was very small. However, in the 1990s Russian was replaced by English and German as the prevailing second languages taught at school, and by the end of the millennium English was clearly the preferred choice (*Framework Education Programme for Basic Education* 2007). In addition, as a result of the establishment of English as an international language in practically all domains, there was an urgent need for translators. In response to this demand for teachers and translators, English studies at universities raised the number of staff and students and included academic writing in English in their programmes to enhance the internationalisation of scientific research. At the beginning of the 21st century, the enlarged Czech Anglicist community is aspiring to be part of global academia and already feeling the impact of globalisation and commodification (Swales 2004; Sanchez Guinda 2015) marked by increased competitiveness and pressure to publish, especially in English. When looking for publication opportunities, Czech Anglicists typically submit their work to national English-medium journals and, only recently and less frequently, to international journals.

This paper studies the development of academic writing conventions in the emblematic academic genre of the research article (RA) with the aim of finding out how English-medium RAs by Czech linguists published in two national journals have evolved dynamically in response to disciplinary, sociocultural and broader socioeconomic changes in the local and global context over the period 1990-2020. The analysis focuses on the form and function of titles, rhetorical structure, the form and location of research aims, hypotheses and research questions, and the functions of metadiscourse markers for writer and reader reference. I begin by comparing the Czech and the Anglophone academic writing traditions and their development over the last three decades. Then I present my corpus and describe the analytical approach taken to the study of rhetorical features included in the bounds of this investigation. In the following sections, I present the findings of the quantitative and qualitative analyses of the selected rhetorical features and highlight the markers of continuity and change in the discourse of Czech linguists in relation to factors influencing the development of academic writing conventions as reflected by the two journals.

2. The Czech and the Anglophone academic writing traditions

The Czech and the Anglophone academic writing traditions differ in many respects, the most prominent of which are primarily the way they approach discourse organization and writer-reader interaction (e.g. Čmejrková – Daneš 1997). The differences seem to stem primarily from divergences in their epistemologies, literacy traditions, the size and type of audience they address, and their contact with other academic writing norms.

The Anglophone academic tradition is associated with the Saxonian intellectual style, which is essentially empirically oriented and democratic. In contrast, the Czech academic tradition has been influenced by Teutonic epistemology, and, due to common Slavonic origins and historical circumstances, the Russian academic literacy (Čmejrková – Daneš 1997; Kozubíková-Šandová 2019), both of which tend towards theorizing, deductive reasoning, and elitism (Clyne 1987). According to previous research and style manuals, Anglophone academic texts are typically characterised by explicit discourse organisation indicated by section headings, a reader-friendly attitude and a high level of interactivity conveyed by metadiscourse markers, as it is the writer who takes responsibility for making the text understandable when striving to persuade the implied heterogeneous audience to accept the suggested views and claims (Bennett 2009; Hyland 2002; Thompson 2001). Czech academic discourse, in contrast, typically lacks explicit discourse structure, and when addressing the homogenous Czech disciplinary discourse community it shows a preference for a reader-oriented, depersonalized style associated with the use of impersonal constructions and the exclusive 'editorial we' (Čmejrková 1996; Čmejrková – Daneš 1997; Vassileva 1998; Yakhontova 2006).

However, the Czech and Anglophone academic traditions are undergoing various modifications. Recent research has shown that Anglophone academic writing conventions have evolved dynamically over the past 30 years. Several studies have evidenced changes in the generic structure of RAs (Bondi 2022), the form and function of titles (e.g. Busch-Lauer 2000; Li – Xu 2019; Xiang – Li 2020), the explicit formulation of research questions and hypotheses (Thewell – Bas-Bleda 2020), citation patterns (Hyland – Jiang 2017b), level of formality (Hyland – Jiang 2017a) and expression of stance (Hyland – Jiang 2016a), engagement (Hyland – Jiang 2016b) and metadiscourse in general (Hyland – Jiang 2018). These changes indicate that there is a tendency towards an increase of informative value and standardization of article components, a significant rise in the number

of citations, a slight decrease in formality, and, in the soft knowledge field, a less visible conveyance of stance and a decline in the extent to which authors engage with the reader. Apart from genre and disciplinary development incentives, these transformations may also suggest that Anglophone discourse is influenced by the growing number of second language writers who publish in international journals (Hyland – Jiang 2018).

Czech academic writing conventions have also evolved dynamically over the years. As Kozubíková Šandová's (2019) diachronic study of Czech RAs in linguistics indicates, several changes occurred in the 1990s concerning the form of titles and the overall organisation of the text, which gradually moved from a spiral with numerous digressions towards greater linearity and transparency, as indicated by the inclusion of section headings and/or numbering. The use of interactive metadiscourse increased after the beginning of the new millennium, thus facilitating text processing for the reader. Interactional metadiscourse, however, is marked by considerable idiosyncratic variation; it seems to show an increase in the use of self-mentions and a certain decrease in the use of hedges. Kozubíková Šandová (2019) attributes most of these changes to the impact of globalization and the influence of Anglophone academic literacy. Apart from this, the evolution of Czech academic discourse might be attributed to digitalisation, as computer-mediated text processing favours clearer text structure, and online access to numerous publications fosters the need to make texts easily surveyable to face the challenge of increased competition in academia.

It is obvious that differences between the Czech and the Anglophone academic traditions are considerable and their clash in the English-medium discourse of Czech linguists is likely to result in some degree of hybridization.

3. Data and method

3.1 Corpus

The study was carried out on a corpus of 20 single-authored linguistics RAs published in two Czech English-medium journals between 1990 and 2020. The choice of journals was limited to the only two peer-reviewed English-medium linguistics journals published in the Czech Republic (including the period before the split of Czechoslovakia in 1993) which have a sufficiently long history to provide data for this diachronic research and as such may be regarded as representative of the development of the Czech culture-specific variant of English. Both journals have now adopted an open-access policy.

The first journal, *Linguistica Pragensia* (LP; published by Charles University, Prague), was founded in 1991 as a successor to the journal *Philologica Pragensia* (1958-1990). The second journal, *Brno Studies in English* (BSE; published by Masaryk University, Brno), was founded in 1959; it also includes a literature and cultural studies section, which is outside the scope of this investigation. For the purposes of the diachronic comparative analysis, the corpus was sub-divided into two sub-corpora, each comprising ten RAs (five from each journal), representing the periods 1990-1995 and 2015-2020 respectively.

Table 1 provides an overview of the size and composition of the corpus.

Table 1. Composition of the corpus

Period	Journal	No. of Articles	Wordcount
Period A – 1990-1995	<i>Linguistica Pragensia</i>	5	18,444
	<i>Brno Studies in English</i>	5	28,860
	Total	10	47,304
Period B – 2015-2020	<i>Linguistica Pragensia</i>	5	23,863
	<i>Brno Studies in English</i>	5	39,935
	Total	10	63,978
Total		20	111,282

The articles included in the corpus were selected to represent equally both journals in the two periods compared (1990-1995 and 2015-2020) and the spectrum of authors publishing in them. All articles are single-authored by different scholars, which reflects the dominant authorship pattern in the journals and minimises the impact of author idiosyncrasies. The selection of authors in each of the two periods includes highly experienced authors (with more than 25 years in academia) who have published extensively, authors who have substantial expertise (with more than 15 years in academia) but have begun to publish more recently, and authors who are at the beginning of their academic career. These three levels of expertise are represented relatively evenly across the two periods, although in the 1991-1995 period highly experienced authors represent the most numerous category (four) and the other two categories are each represented by three authors, while the 2015-2020 author sample comprises two highly experienced authors, four experienced and four early-career researchers. This difference reflects changes in the size and dynamics of the Czech Anglicist discourse community, which over the last 30 years has increased in membership and now comprises relatively experienced scholars and early-career researchers in considerable proportions.

The corpus was compiled using the software tool *SketchEngine* (Kilgariff et al. 2014), which was also used to search the corpus for personal expressions for writer and reader reference. In agreement with the common procedure in contrastive corpus-based research, the difference in word-count between the sub-corpora was neutralized by normalization to occurrences per 10,000 words.

3.2 Method

The diachronic comparative corpus-based analysis draws on the genre analysis framework (Swales 1990; Bhatia 1993) to investigate the rhetorical structure of RAs and realizations and functions of metadiscourse markers for writer and reader reference. The rhetorical structure of RAs is explored drawing on the IMRAD model (Swales 1990; Cotos et al. 2015; Gray et al. 2020; Schmied 2011, 2015), while taking into consideration the dynamic evolution of genres in response to sociocultural changes (Hyland 2000; Sancho Guinda 2015). As typical of social sciences, the rhetorical sections are coded as 'Introduction', 'Data and method', 'Results and discussion' and 'Conclusion'. The comparison between the articles published in Period A and Period B is carried out at the macrolevel of their overall organisation, while a more detailed moves analysis is applied to introduction, method and conclusion sections. Also investigated are specific rhetorical elements, namely the form and function of RA titles, the location and form of research aims, research questions, and hypotheses and functions of metadiscourse markers for writer and reader reference, so as to trace possible diachronic changes across the two periods.

The analysis of RA titles focuses on two formal features, i.e. length and syntactic structure, and two content-related features, i.e. information types and functions. Following Cheng et al. (2012) and Xiang and Li (2020), five types of syntactic structures are considered: nominal constructions, prepositional phrases, *V-ing* phrases, full sentences, and compound constructions (titles containing two parts separated by punctuation marks, such as a colon, dash or period). The taxonomy of information types included in titles used in this study draws on Sahragard and Mejhami (2016), Li and Xu (2019) and Xiang and Li (2020), and comprises five categories: topic only, method/design, dataset, result and conclusion. Finally, to identify the interactional potential of titles the analysis applies Busch-Lauer's (2000) three functions of titles: nominative, aiming at identifying a piece of work (which in this study is regarded as subsumed in the other two functions); designating, focusing on

indicating the content of the piece of work; and advertising, appealing to the reader.

The significance of statements of research aims, research questions and hypotheses in the rhetorical structure of RAs is considered with regard to frequency of occurrence, location and linguistic realisation (cf. Lim 2014; Jalilifar 2010).

Realizations and functions of expressions for writer and reader reference are explored within Hyland's (2005) interpersonal metadiscourse framework. The functional classification of personal expressions used in this study is informed by several taxonomies as suggested by previous research (e.g. Dontcheva-Navratilova 2013; Fløttum et al. 2006; Ivanič 1998; Harwood 2005; Hyland 2002; Sheldon 2009; Tang – John 1999). My classification takes into consideration the context and location of the personal structure within the genre of the RA and the semantic-pragmatic meaning of the verb with which the personal expressions occur (cf. Halliday 1994; Hyland 1999) as indicative of their role in writer-reader interaction.

The five roles of self-mention considered in this study are typically realised by the exclusive pronouns *I* and *we*; they include: (1) the discourse organiser role, which at the macro-level outlines the text structure, and at the micro-level indicates rhetorical moves, transition points and intra-textual connections, often expressed by the co-occurrence of *I/we* with discourse verbs (e.g. *discuss, present, summarise, focus on*); (2) the researcher role, which describes research and data collection processes and procedures, and tends to be signalled by the co-occurrence of *I/we* with research and cognition verbs (e.g. *analyse, find, use, understand*); (3) the arguer role, which is used to elaborate arguments, put forward claims and comment on findings, typically manifested by the co-occurrence of *I/we* with position verbs (e.g. *argue, claim, dispute, show*); (4) the evaluator role, which expresses agreement or disagreement with an issue, position or belief, and is commonly realised by the co-occurrence of *I/we* with evaluative and emotive verbs (e.g. *feel, be sceptical about, find something + evaluative adjective*), and (5) the reflexive-self role, conveying the situatedness of the researcher and casting his/her autobiographical self into the text by providing personal comments related to the design and realisation of the study or the teaching or research experience of the author (cf. Starfield – Ravelli 2006). Finally, personal expressions for writer reference are compared to 'locational self-mention' (Bondi 2014), i.e. the use of discourse (*paper, article, section*), research (*analysis, study*) and cognitive (*assumption, interpretation*) nouns to indirectly refer to the author, so as to explore the interplay of personality and impersonality across the two periods.

While previous studies assign a single function to reader reference, this investigation considers two roles of reader reference, typically expressed by the inclusive pronoun *we*: (1) discourse community member, associated with reference to established practices and shared knowledge, and (2) the ‘reader-in-the text’ (Thompson – Thetela 1995), conveyed by direct address to the reader positioned as an interested colleague following the argument presented in the text.

4. Results and discussion

Before focusing on analysis of the rhetorical structure of the RAs representing the two periods, it is worth noting that there is considerable difference in the wordcount of the samples – 47,304 in Period A (average article length 4,730 words) and 63,978 in Period B (average article length 6,397 words). Moreover, one of the BSE articles in Period A is an outlier in many respects, including wordcount (10,916); the reason for its inclusion in the dataset is that it is authored by probably the most prominent Brno linguist of the period (Jan Firbas), whose influence on the Czech Anglicist community is considerable. Without this article, the average wordcount per text in Period A would be 4,043 words. This difference in article wordcount between the two periods already suggests a certain change in the way Czech authors conceive their texts and the level of informativeness the articles are intended to convey (a similar tendency is also found by Bondi (this issue) in her study of Italian research articles). Another factor leading to the increase in wordcount of articles concerns requirements imposed by journal editors; BSE, for instance, set 5,000 words as the minimum article length in 2008 (personal communication of the editor-in-chief J. Chovanec).

4.1 The rhetorical structure of RAs

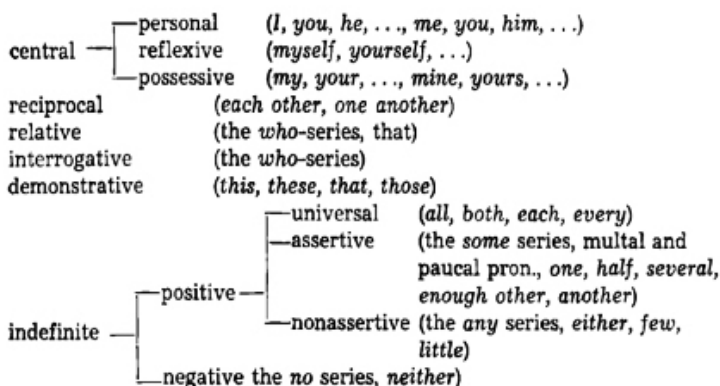
The overall organisation of the RAs in the two periods is clearly different. As Example 1 shows, Period A RAs typically lack explicit indication of organisation in sections by headings and numbering; only three RAs use topic-based section headings (one of them only ‘Conclusions’), while two use numbering of paragraphs (one with no section headings). Period B RAs, on the other hand, generally display the standard numbered section headings pertaining to the IMAD framework, although, as is typical of applied linguistics articles, there is some variation in section headings (Yang – Allison 2003). This

suggests that while in Period A Czech authors to a large extent abide by Czech academic writing conventions, in Period B they opt for a different article structure. The reason for this change seems to be twofold. Firstly, the way Czech linguists write their academic texts is clearly influenced by Anglophone academic conventions due to training in academic writing courses, access to international journals, and pressure to publish in English-medium local and international journals. The second reason stems from a certain change in article typology and methods used, as not all RAs included in the Period A sub-corpus are empirical. Drawing on Gray (2015), four of them may be categorised as hybrid, or mostly general theoretical, i.e. discussing a linguistic feature illustrated by examples taken from previous research or provided by the author without identification of a source. These hybrid articles are generally characterised by a fuzzy structure, which may have also affected the organisation of the Period A texts. The RAs in Period B, however, are empirical and use quantitative, qualitative or mixed methodology for data analysis; only one may be regarded as hybrid of the theoretical and empirical types.

(1) Period A (06_BSE_1991)

This paper is a sequel to the analysis of pronouns published in BSE 18 (Chamonikolasová, 1990). The previous paper dealt with the communicative dynamism (cf. Firbas 1990) and prosodic prominence (cf. Firbas 1990) of personal pronouns. The present analysis covers all pronominal categories and provides their comparison.

The formal classification of pronouns that has been applied in this paper is based on the conception of pronominal categories given in Quirk et al. 1985:



Syntactically, the analysis is limited to pronouns implemented as simple noun phrases and functioning as independent sentence elements; it does not deal with pronouns inside a complex noun phrase (premodifiers or headwords), which are only part of a sentence element. The former and the latter show certain differences in functional sentence perspective and

Period B (12_LP_2016)

1. INTRODUCTION

This paper deals with the analysis of the representation of the different functions of the subject *it* in two text types: academic prose and fiction. It examines four basic functions of subject *it* — anaphoric *it*, empty *it*, and *it* as the component of extraposition (anticipatory *it*) and the cleft sentence (focusing *it*). Focusing not only on the distribution and frequency of the different functions but also on their correlations with other features of the two registers, it aims to discover in which respects and to what extent the two text types differ and whether such a basic, common word as *it* can play a role as a style marker.

2. THEORETICAL BACKGROUND

The pronoun *it* has the same basic functions as the other third person pronouns, serving deictic and anaphoric uses. Aside from those, it can also have several “special uses” whose exact number varies in different interpretations. The *Longman Grammar* identifies the following three:

- Empty subject/object
- Anticipatory subject/object¹
- Subject in cleft constructions (Biber et al., 1999: 332).

3. MATERIAL AND METHOD

The analysis presented in this article is based on 200 excerpts obtained from the British National Corpus (BNC), a corpus of 100 million words of British English from the latter part of the 20th century, accessed using the web-based interface *BNCweb* (CQP-Edition) available on-line from <http://bncweb.lancs.ac.uk/>. Two subcorpora of similar sizes were created from this corpus, one of academic prose and one of fiction.

The same search for *it* in the subject position was then carried out in both subcorpora, using the query “*it* (_ {ADV})? _ V* [D,Z]”. The result of this query should be it followed by an optional adverb and any verb in either past tense form or third person singular present tense form.

The occurrence of the individual moves and steps in the Introduction, Method and Conclusion sections in the corpus is summarised in Table 2. (Only steps occurring in the corpus are listed; steps occurring only in Period B are highlighted in italics.) Comparison of the data for Period A and Period B shows a considerable change in terms of the quantity and range of moves and steps represented in the two sub-corpora. From a quantitative perspective, there is an increase in the occurrence of almost all moves and steps. Moreover, two moves, i.e. M3 ‘Establishing credibility’ in the Method section and M3 ‘Reshaping the research territory’ in the Conclusions, and several steps in other moves, i.e. ‘Claiming centrality’, ‘Reviewing previous research’ and ‘Clarifying definition’ in the Introduction, ‘Introducing participants’, ‘Delineating study procedures’ and ‘Identifying variables’ in the Method section, and ‘Explicating results’, appear only in Period B. It is significant that these newly adopted moves and steps contribute decisively to the replicability of the research and enhance its persuasive force.

In the Introduction, the M1 'Establishing territory' move is typically realised by S1 'Providing general information', which occurs in most articles across the two periods. In Period A, this is sometimes combined with presentation of the study as a sequel to a previous study by the same author (Example 6, Period A text), while in Period B, this is usually done by indicating the topic of the study (Example 6, Period B text). The most significant change in Move 1 is the frequent occurrence of the S3 'Reviewing previous research' step, which not only allows writers to contextualise their research but also highlights its importance (Bhatia 1993; Samraj 2002). The absence of this step in Period A reflects an assumption on the part of the authors that their readership is acquainted with the main theoretical tenets and methodology to which they refer (cf. Dontcheva-Navratilova 2013), while its presence in Period B may be attributed to the influence of Anglophone rhetorical conventions and the efforts of Czech linguists to reach an international audience. The increased prominence of M2 'Identifying a niche' (from 9.52% to 23.07%) shows the greater argumentative effort invested by Czech linguists to point to the need for their research, while the more frequent use of the 'Outlining the structure of the paper' step in Move 3 indicates an attempt to help the reader with text processing.

Table 2. Rhetorical structure of the Introduction, Method and Conclusion sections

Rhetorical section	Moves/steps	Period A		Period B	
		N	%	N	%
Introduction	M1 Establishing territory	9	42.86	10	38.46
	S1 Providing general background	9		9	
	S2 Claiming centrality	0		2	
	S3 Reviewing previous research	0		8	
	M2 Identifying a niche	2	9.52	6	23.08
	Raising general questions	1		6	
	Highlighting a problem	1		1	
	Indicating a gap	0		3	
	M3 Addressing the niche	10	47.62	10	38.46
	S1 Introducing present research descriptively	8		2	
	S2 Announcing research aims/ purposes	5		10	
	S3 Presenting research questions	2		2	
	S4 Presenting research hypotheses	1		1	
	S5 Clarifying definitions	0		1	
	S5 Outlining the structure of the paper	1		3	
	Total moves	21	100	26	100

Method	M1 Contextualizing methods	10	71.43	9	42.86
	S1 Providing general background	9		5	
	S2 Identifying the methodological approach	7		7	
	S3 <i>Introducing participants</i>	0		1	
	M2 Describing the study	4	28.57	9	42.86
	S1 Describing data	4		8	
	S2 <i>Delineating study procedures</i>	0		7	
	S3 <i>Identifying variables</i>	0		2	
	M3 Establishing credibility	0	0	3	14.28
	S1 <i>Describing the data analysis</i>	0		3	
	Total moves	14	100	21	100
Conclusions	M1 Re-establishing the territory	9	60.00	9	36.00
	S1 Highlighting principal findings	9		9	
	M2 Framing new knowledge	1	6.66	6	24.00
	S1 <i>Explicating results</i>	0		6	
	S2 Addressing limitations	1		0	
	M3 Reshaping the territory	0	0	2	8.00
	S1 <i>Reshaping the field</i>	0		1	
	S2 <i>Supporting with evidence</i>	0		1	
	M4 Expanding the niche	5	33.33	8	32.00
	S1 Generalising results	4		6	
	S2 Proposing directions	2		2	
	S3 Claiming value	1		3	
	Total moves	15	100	19	100

(Steps occurring only in Period B are highlighted in italics; percentage shows the ratio for the specific rhetorical section, i.e. introduction, method and conclusions.)

Changes in the Method section concern primarily M2 ‘Describing the study’ (marking a rise from 28.57% to 42.86%), which becomes a conventional move in Period B, i.e. it occurs in more than 60 per cent of texts (Kanoksilapatham 2015), and the appearance of M3 ‘Establishing credibility’ via description of the data analysis (14.28%). These developments indicate that Czech linguists endeavour to enhance the replicability and reliability of their research, which may be due to the wider variety of topics under investigation, more frequent use of quantitative methods and an increased competitiveness when seeking to reach a wider readership.

The Conclusion section also appears to have gained in rhetorical complexity. The range of rhetorical moves is extended by the appearance of the new M3 ‘Reshaping the territory’ (8%). While several of the Period A articles comprise very brief conclusions (Example 2) and only the most experienced authors point to limitations and implications of the study, in

Period B M2 ‘Framing new knowledge’ (up from 6.66% to 24.00%) and M4 ‘Expanding the niche’ (32.00%) may be regarded as conventional moves. For instance, in Example 3 the author emphasises the value of the approach used in the study in terms of contribution to knowledge and reliability, which can be interpreted as a ‘promotional’ feature; since promotional features are generally more frequent in Anglophone than in L2 discourse (Moreno 2021), their increased occurrence may be attributed to the influence of Anglophone rhetorical conventions.

- (2) Summing up the answers to the three questions raised at the beginning of the present paper, the following statement can be made: the most frequent rheme in both English and Czech is an object in the final sentence position implemented by a modified noun. (09_BSE_1995)
- (3) The analysis of corpus data, based on the activation of paradigmatic relation of alternation and the syntagmatic relation of co-occurrence of PHS with other constituents of the utterance, proved to be a reliable basis for context-sensitive interpretation of emergent communicative strategies and the dual, i.e. therapeutic and diplomatic usage of PHS in current communicative situations. (19_BSE_2019)

All in all, the evolution of the rhetorical structure of Czech English-medium RAs across the two periods clearly indicates a move towards the IMRAD framework associated with a prevalence of clearly empirical research and greater rhetorical complexity.

4.2 Titles of RAs

The results of the analysis of the form features of RA titles presented in Table 3 indicate obvious changes between the two periods.

Table 3. Formal features of RA titles

Titles	Title Length	Surface form				
		Nominal	Prep. phrase	V-ing phrase	Compound	Sentence
Period A	8.0	8	2	0	0	0
Period B	12.9	3	2	0	5	0

There is a substantial increase in title length between the two periods, from 8.0 to 12.9 words per title. This might be attributed to the influence of Anglophone academic discourse, since as Busch-Lauer (2000) suggests, L2 writers tend to use shorter titles than Anglophone linguists (8.5 vs 9.08 respectively in 2000), and to a general tendency over the last 30 years towards use of longer and thus more informative titles (Xiang – Li 2020). For instance, Li and Xu's (2019) diachronic research into titles of articles published in *Journal of Pragmatics* show an increase of title length from 8.02 in 1978 to 12.01 in 2018.

The syntactic structure of titles used by Czech linguists in both periods shows only three of the five structures available in the taxonomy of title structure, although the two non-represented types (sentence and V-ing phrase) account for only 10 per cent of all title structures in Anglophone texts (cf. Xiang – Li 2020). Period A displays only two structural title types: nominal (80%) and prepositional phrase (20%) (Examples 4 and 5). While there is no change in the occurrence of prepositional titles across the two periods, the nominal type has more restricted use in Period B (30%). The most striking diachronic change concerns the appearance of compound titles (Example 6), which do not occur at all in Period A, but represent 50 percent of all titles in Period B. This change may be associated with the influence of Anglophone conventions, since, as Busch-Lauer's (2000) findings indicate, the occurrence of compound titles is less frequent in RAs by L2 writers than in RAs by Anglophone writers (28% vs 64%) (see also Bondi in this issue). The compound structural title type has also been on the increase in international journals over the last 30 years, most likely in response to the need to provide the reader with more initial information amid the exponential surge in the number of RAs published. This is in line with previous diachronic research (Li – Xu 2019; Xiang – Li 2020) reporting that compound titles nearly doubled their rate in the period between 1988 and 2018.

- (4) Some thoughts on contrastive grammar (03_LP_1994)
- (5) On the discourse functions of the English language of conversation (07_BSE_1991)
- (6) "Now is the time to root out evil": The role of natural world metaphors in the construction of the US and THEM dichotomy (18_BSE_2019)

The changes in title length and structure are also reflected in changes in content-related features. Table 4 summarizes the distribution of information types and functions of titles.

Table 4. Content-related features of RA titles

Titles	Information type					Function	
	Topic	Topic + dataset	Topic + Method	Topic + Results	Topic + Conclusion	Designate	Advertise
Period A	8	2	0	0	0	10	0
Period B	1	3	2	3	1	7	3

In the first period, titles only have an informative function. They typically indicate the topic of the research. Only occasionally (20%) do they indicate the type of dataset the study is performed on (Example 7). In the second period, however, the titles are considerably richer in information content: the predominant types of titles indicate Topic and Dataset and Topic and Results (Example 8). Titles designating Topic and Method (Example 5) are less frequent, while those indicating the Title only or Title and Conclusion are rare. The increased occurrences of the Results, Dataset and Method components may be attributed to their higher degree of explicitness in comparison with Conclusion and Topic (Haggan 2004). My results differ to some extent from those reported by Xiang and Li (2020); while their results for linguistics RAs also indicate a decrease of Topic type titles (from 74% in 1988 to 53% in 2018), in their data, the Method information type is considerably more frequent than the Dataset type. This difference may stem from the small sample used in this study, as well as from the somewhat narrower range of methods used by the members of Czech Anglicist community.

- (7) The vagaries of subject *it*: can it serve as a style marker? (12_LP_2016)
- (8) Boring as hell: a corpus study of intensifying post-modification of predicative adjectives in the 'ADJ as NOUN' frame (15_LP_2020)

Another important evolution is visible in the function of titles, which in Period B are not confined to designating the content of the study, as 30 per cent of them have also an eye-catching or advertising function (Example 3). This seems to suggest an effort on the part of Czech linguists to become more competitive, as they seem to wish to make their titles more creative, stylistically rich and appealing by including quotes (Example 3) and expressive language (Example 5; cf. Busch-Lauer 2000: 90).

4.3 Aims of the research, research questions and hypothesis

Research aims, questions and hypotheses are part of the key move of the Introduction section M3 'Addressing the niche', which occurs in all RAs in the corpus. While in Period A, research presentation tends to be performed more frequently by S1 'Introducing present research descriptively', in Period B, this is realised primarily by S2 'Announcing research aims/purposes' (Table 4). In both periods, these steps may occur at the very beginning of the Introduction; in such cases, M1 'Establishing research territory' and M2 'Identifying a niche' may follow, or M3 'Addressing the niche' may constitute the only move in the Introduction (Example 6). When indicating the aims of their research, most Period A authors frame their study as an 'attempt', using set phrases such *in the present paper I will attempt to, this study attempts, the paper attempts*, which may be associated with the Czech convention of showing authorial modesty (Čmejrková – Daneš 1997). In Period B, all RAs include a clear indication of the specific aims of the study, typically using the set phrases *the aim of this paper is* and *this paper aims to* collocating with the verbs *examine, discover* and *show*, which have been categorised as lexical bundles typical of Anglophone academic discourse (Cortes 2013). This change seems to stem from the influence of Anglophone academic conventions as well as from the need to specify more clearly the aims of the study in a research territory packed with an ever higher number of publications.

Explicit research questions and hypotheses are not frequent in the corpus, which confirms the findings of previous research indicating that they are more common in RAs by Anglophone authors (Sheldon 2011). The frequency of occurrence of hypotheses and research questions does not differ across the two periods, as there is one RA including a hypothesis and two RAs comprising research questions in each period. Both hypotheses take the form of a statement, using the verb form *be* in Period A and *will* in Period B. Although the low number of research questions does not allow for generalisations, it might be relevant to note that they tend to be factual (Lim 2014) rather than polar, i.e. in the case of research questions (Example 9) they include a *wh*-pronoun and a present tense copular verb in an SVC structure. This kind of research question is typically associated with qualitative research (the prevailing kind of method used by Czech linguists) and seeks interpretation of situations, meanings and experiences (Lim 2014:84).

- (9) What are the parts of speech that implement rheme most often? (09_BSE_1995)

It is somewhat surprising that the occurrence of hypotheses and research questions is restricted to only one of the journals (*BSE*), although the Instructions for authors in both journals do not mention any requirements concerning these steps. While it is difficult to find a satisfactory explanation for this situation, it might be associated with the topic and method applied, as both studies using research questions in Period A study Theme-Rheme articulation, and both studies in Period B combine discourse analysis and pragmatics.

4.4 Expressions for writer and reader reference

A quantitative analysis of self-mention and reader reference (Table 5a and Table 5b) shows a considerable decrease in the frequency of use of both features across the two periods: from 48.6 to 11.7 for self-mention and from 24.6 to 13.8 for reader reference. The relatively frequent use of self-mentions is typical of Czech-medium academic discourse (Dontcheva-Navratilova 2020: 132), yet the rate of author-reference may be seen as somewhat skewed by the outlier (10_BSE_1995) mentioned in Section 3, which comprises about 35 per cent of all uses of *I*, *me* and *my*. However, even without this article the occurrence of self-mentions in Period A is 28.8, which is more than two times higher than in Period B. There is also a change in the distribution of occurrences across the articles – in Period A all ten articles use self-mentions, while three articles comprise no reader reference expressions. In contrast, in Period B, all articles employ reader reference; however, two articles show no instances of self-mention. This may indicate that, while trying to avoid explicit indication of human agency, Czech linguists show a growing tendency to use reader-oriented features more typical of Anglophone academic discourse.

Table 5a. Self-mention across the two periods (per 10,000 words)

Periods	<i>I</i>	<i>me</i>	<i>my</i>	<i>we</i>	<i>us</i>	<i>our</i>	<i>the author</i>	Total
Period A	16.9	5.2	12.3	7.5	0	6.5	0	48.4
Period B	6.3	0.4	2.6	0.8	0	1.6	0	11.7

Table 5b. Reader reference across the two periods (per 10,000 words)

Periods	<i>we</i>	<i>us</i>	<i>our</i>	<i>you</i>	<i>one</i>	<i>the reader</i>	Total
Period A	5.7	5.7	3.7	0	8.0	1.5	24.6
Period B	8.9	1.4	1.2	0	1.3	1.0	13.8

The decreasing tendency in the incidence of expressions for writer and reader reference may be attributed to three factors. Firstly, it may reflect instructions provided by academic style manuals (Bennett 2009) and academic writing courses, which typically advise authors to maintain an impersonal tone. A second factor may be the impact of Anglophone writing conventions on applied linguistics and soft sciences in general, which, as Hyland and Jiang (2016a, 2016b) show, are characterised by a decline in self-mention and reader reference over the last 50 years. Finally, the change might reflect the expertise of authors represented in the corpus; Period A includes a higher proportion of highly experienced authors, who generally seem to be more prone to the use of writer and reader reference expressions (cf. Dontcheva-Navratilova 2014).

Realisations of self-mention and reader reference have changed to some extent. In Period A, both singular and plural first person pronouns are used to indicate self-mention (16.9 and 7.5 respectively); only two authors indicate self-mention exclusively by *I/me/my*, three use exclusively *we/our* and five use both singular and plural forms. This is most likely a transference from the Czech academic writing style where the use of self-mention *we/our* is standard practice intended to indicate author modesty. In addition, as Hyland and Jiang (2016a) note, over the same period *I* has replaced *we* as the preferred marker of stance in Anglophone applied linguistics discourse, which might also have impacted Czech linguists' writing habits. In Period B, the use of exclusive plural forms is rare (0.8) and always occurs in combination with singular forms, as in Example 10, where the author adopts the researcher role to highlight his/her agentive role in the research process (*my analysis*) but, when commenting on the results, opts for reducing personal attribution by using *we* in combination with several hedging devices (*should, tentative*).

- (10) As follows from Table 2, the results of *my analysis* imply that the presentation scale is confined to the second passive only (...). Nevertheless, *our results* should be viewed as tentative; it would require analysis of more examples to verify whether the presentation scale may also be implemented in the first passive. (11_LP_2015)

Another significant change is the decrease in incidence of *one* and *the reader* (from 8.0 and 1.5 in Period A to 1.3 and 1.0 in Period B), which are less direct in their appeal to the reader. On the other hand, there is an increase in the use of *we* (from 5.7 in Period A to 8.9 in Period B), which is more explicitly dialogic and thus more likely to involve readers in the argument as colleagues and persuade them to accept the author's claims.

Tendencies in the use of rhetorical functions of metadiscourse markers for writer and reader reference have also undergone some changes, which are summarised in Tables 6a and 6b.

Table 6a. Rhetorical functions of self-mention (%)

Self-mention	Period A	Period B
Discourse organiser	32.9	16.9
Researcher	39.6	64.4
Opinion-holder	9.8	3.4
Arguer	8.6	10.2
Reflexive self	9.1	5.1

Table 6b. Rhetorical functions of reader reference (%)

Reader reference	Period A	Period B
Disc. community member	16.9	20
'Reader-in-the-text'	83.1	80

In both periods, the most prominent role of self-mention is that of researcher, which appears in all RAs in the corpus. The researcher role is typically realised by the personal pronoun *I* in the agentive subject position collocating with research verbs, such as *find*, *analyse*, *examine*, and the possessive *my* collocating with research nouns such as *analysis*, *research*, *data*, *material*. The researcher role shows a striking increase in Period B (39.6% in Period A vs 64.4% in Period B), which may be associated with an attempt to increase the visibility of the researcher and to gain credit for decisions that scholars make while carrying out their study (Example 11), as well as with an effort to enhance research replicability and credibility.

- (11) (...) after careful consideration *I decided* to winsorize the data, i.e. replace the outlier participants' means with the next highest value which is not an outlier (Field 2014), so as not to influence the control group's overall means. (20_BSE_2020)

The role of discourse organiser is the second most frequent in both periods; however, it indicates a decreasing tendency (32.9% in Period A vs 19.9% in Period B). Overall, self-mentions outlining the text structure are rather rare, since, as mentioned above, research aims are typically formulated in an impersonal way and the step 'Outlining the structure of the paper' in the

introduction is rather rare. Thus, most instances of the discourse organiser role operate at the micro-level to indicate transition points and intra-textual connections. In Period A, this role is most visible in the outlier article (10_BSE_1995), which comprises all 13 uses of *let me* (26% of all occurrences of the discourse organiser role), indicating transition points in discourse (Example 12), which may account to a large extent for the less frequent occurrence of this role in Period B.

(12) *Let me* now turn to the analysis of the other text. (10_BSE_1995)

The more authoritative roles of opinion-holder, arguer and reflexive self are considerably less frequent. Only the arguer role shows an increase in use (from 8.6% in Period A to 10.2% in Period B); this seems to indicate a transition from a descriptive presentation of results (typical of the Czech academic literacy) towards a more argumentative reasoning aimed at persuading the reader to accept the writer's claims. It is interesting to note that while in Period A the arguer role is realised primarily by *I* and *we* in the agentive subject position collocating with position verbs, such as *argue*, *propose*, *show* (Example 13), in Period B, it is also conveyed by the possessive *my* collocating with nouns, such as *proposal*, *understanding* and *hypothesis*.

(13) *I will argue* now that it is one of the substantial tasks of CGR to find out which of these differences can be ascribed to the translator's imagination and creativity only, and which; on the contrary, must be regarded as differences in the language structure. (03_LP_1995)

The opinion-holder and reflexive-self roles show a decreasing tendency, perhaps because most of the instances occur in the outlier 10_BSE_1995. In Period A, the opinion holder role is typically expressed by the collocations *I believe* and *I trust*, which might be interpreted as boosters. The considerably lower frequency of this role (9.8% in Period A vs 3.4% in Period B) may stem from a tendency to reduce authorial intrusion in the text. The self-reflexive role in Period A is associated with personal comments concerning the realisation of the study and the teaching or research experience of the author (Example 14), while in Period B it refers primarily to personal communications of colleagues that the author has used in the study.

(14) With one exception, the analyses were carried out *under my direction* by students who attended *my seminars* on FSP. (10_BSE_1995)

Reader reference conveys predominantly the ‘reader-in-the text’ role (83.1% in Period A and 80% in Period B), which shows a slight decrease in Period B, complemented by an increase in instances representing the reader and the writer as in-group members of the same community (e.g. *our country, our talking habits*). The ‘reader-in-the text’ role typically endeavours to position the reader as a co-researcher who follows the writer’s chain of arguments and thus may be induced to accept the writer’s views and claims (Example 15).

- (15) *If we extracted* this purported modificant from the verb’s meaning, *we would be left* with a core describable as the very general ‘move’.
(17_BSE_2016)

A comparison of personal self-mention (Table 5a) and locational self-mention (Table 7) shows that while self-mention expressed by personal pronouns has considerably decreased (from 48.4 in Period A to 11.7 in Period B), locational self-mention has strikingly increased over the two periods (from 31.3 in Period A to 64.2 in Period B), as also found in Bondi’s (this issue) case study. This indicates that Czech scholars tend to adopt a more impersonal style of writing which may be associated with the scientific paradigm dominant at the end of the 20th century in Anglophone academic writing related to “clarity, economy, rational argument supported by evidence, caution and restraint” (Bennett 2009: 52).

Table 7. Locational self-mention (per 10,000 words)

Periods	Discourse nouns	Research nouns	Cognitive nouns	Total
Period A	9.5	19.4	2.5	31.3
Period B	22.5	38.7	3.2	64.2

The most frequent forms of locational self-mention across both periods are research nouns (e.g. *analysis, study, sample, material, results*: Example 16), followed by discourse nouns (e.g. *paper, article, section, table*: Example 17) and cognitive nouns (e.g. *conclusion, interpretation*), which are considerably less frequent. The increase in research nouns may be associated with the tendency towards explicit statement of research aims, description of data and orientation towards empirical research in Period B. The raise in frequency of discourse nouns may be attributed to an awareness of the importance of structuring the text more transparently to help readers with text processing.

- (16) *The present study* is to be regarded as a specific inquiry into the relationship between the immediately relevant context and the retrievability span. (10_BSE_1995)
- (17) Following the analysis, *the paper* works with two basic topic management evasion techniques: raising a new safe topic and perspective reprojected. (13_LP_2017)

Overall, the use of writer and reader reference expressions is marked by considerable variation across the two periods and across individual texts. The changes observed seem to indicate that the influence of Czech academic conventions in English-medium texts has diminished, while the influence of Anglophone conventions has increased.

5. Conclusion

The diachronic comparative analysis in this study has explored the evolution of Czech linguists' writing practices in the genre of Czech English-medium RAs in response to the emergence of new sociocultural and publication situations over the last 30 years. Since the fall of the Iron Curtain, Czech Anglicists have established active interaction with the international linguistics community and aspire to make their work available to an international readership in response to the pressure to publish in order to advance in their academic career. My analysis has shown that the English-medium discourse of Czech linguists has undergone several changes indicating a transition from writing conventions strongly influenced by the Czech academic literacy to hybrid-in-nature writing practices accommodating the Czech culture-specific tradition and Anglophone academic discourse conventions dominant in international academic publications, thus enhancing Czech linguists' chances of having their work published in an international context.

The hybridity of the present-day English-medium discourse of Czech linguists is manifested by the persistence of several features typical of Czech academic writing conventions. These show considerable variation in section labelling, infrequent occurrence of certain rhetorical moves and steps, such as *Indicating a gap*, *Proposing directions for further research* and *Reshaping the field*, and rare but still present self-mention by means of *we* and *our*. The process of merging Anglophone and local rhetorical norms is also indicated by the presence of considerable variation across individual texts. However, despite these signs of continuity, it can be argued that all rhetorical features

of English-medium RAs by Czech linguists have been affected by change. Over the 30-year period, Czech linguistics English-medium RAs have become more empirically oriented, their length has increased, their organisation has become more explicit, and they have adopted mostly the IMRAD rhetorical structure, thus enhancing the informativeness and surveyability of the text. A similar tendency is noticeable in the development of titles, which, apart from becoming longer and more informative, may display an eye-catching function indicating an increase in competitiveness on the part of Czech linguists. The rhetorical structure of RAs has also gained in complexity by comprising more moves and steps which make the discourse more argumentative, clearly state the research aims, help the reader with text processing, contribute to the replicability of the research and enhance its persuasive force. The frequency of use and functions of personal expressions for writer and reader reference have also changed: the exclusive *we* has been gradually replaced by the use of *I* and impersonal forms, while the researcher role of self-mentions has shown a striking increase aimed at enhancing the visibility and credibility of the researcher. These changes appear to reflect the influence of the Anglophone academic literacy on the English-medium discourse of Czech linguists; however, they could also have been impacted by changes taking place in Anglophone academic writing conventions, which are characterised by increased informativeness and decreased conveyance of stance and reader engagement (cf. Hyland – Jiang 2016a, 2016b, 2018).

The findings of this study may raise our awareness of the ways in which culture-specific variants of English adapt diverging academic writing traditions in order to enhance L2 academic writers' chances of publishing in a national and international context. Obviously, the results of this study cannot be overgeneralized, as they are limited by the small size of the corpus and the limited number of features included in the research. Nevertheless, I believe that this study has shown the potential of diachronic comparative analysis for exploration of the evolution of English-medium academic discourse in different linguacultural settings and for helping us understand the reasons for the changes that we observe.

REFERENCES

Primary sources

Brno Studies in English (BSE)

<https://journals.phil.muni.cz/bse/about>, accessed November 2023.

Linguistica Pragensia

<https://linguisticapragensia.ff.cuni.cz/en/about-journal/>, accessed November 2023.

Secondary sources

- Bennett, K.
2009 "English academic style manuals: A survey", *Journal of English for Academic Purposes* 8, 43-54. <https://doi.org/10.1016/j.jeap.2008.12.003>.
- Bhatia, V.
1993 *Analysing Genre: Language Use in Professional Settings*. Longman. <https://doi.org/10.4324/9781315844992>.
- Bondi, M.
2009 "Polyphony in academic discourse: A cross-cultural perspective?" In: E. Suomela-Salmi – F. Dervin (eds.) *Cross-Cultural and Cross-Linguistic Perspectives on Academic Discourse*. Amsterdam: John Benjamins, 83-108. <https://doi.org/10.1075/pbns.193.06bon>.
2014 "Changing voices: Authorial voice in abstracts". In: M. Bondi – R. Lorés-Sanz (eds.) *Abstracts in Academic Discourse: Variation and Change*. Bern: Peter Lang, 243-270. <https://doi.org/10.3726/978-3-0351-0701-2>.
2022 "The scientific article: Variation and change in knowledge communication practices". In: C. Hanganu-Bresch et al. (eds.) *The Routledge Handbook of Scientific Communication*. London: Routledge, 159-169. <https://doi.org/10.4324/9781003043782>.
- Busch-Lauer, I.
2000 "Titles in English and German research papers in medicine and linguistics". In: A. Trosborg (ed.) *Analysing Professional Genres*. Amsterdam/Philadelphia: John Benjamins, 77-97. <https://doi.org/10.1075/pbns.74.08bus>.
- Cheng, S. W. – C. W. Kuo – C. H. Kuo
2012 "Research article titles in applied linguistics". *Journal of Academic Language and Learning*, 6 (1), A1-A14.
- Clyne, M.
1987 "Cultural differences in the organisation of academic texts: English and German", *Journal of Pragmatics* 11 (2), 211-247. [https://doi.org/10.1016/0378-2166\(87\)90196-2](https://doi.org/10.1016/0378-2166(87)90196-2).
- Čmejrková, S.
1996 "Academic writing in Czech and in English". In: E. Ventola – A. Mauranen (eds.) *Academic Writing*. Amsterdam/Philadelphia: John Benjamins, 137-152. <https://doi.org/10.1075/pbns.41.11cme>.
- Čmejrková, S. – F. Daneš
1997 "Academic writing and cultural identity: The case of Czech academic writing". In: A. Duzsak (ed.) *Culture and Styles of Academic Discourse*.

- Berlin/New York: Mouton de Gruyter, 40-62.
<https://doi.org/10.1515/9783110821048.41>.
- 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. <https://doi.org/10.1016/j.jeap.2012.11.002>.
- Cotos, E. – S. Huffman – S. Link
 2015 "Furthering and applying move/step constructs: Technology-driven marshalling of Swalesian genre theory for EAP pedagogy", *Journal of English for Academic Purposes* 19, 52-72.
<https://doi.org/10.1016/j.jeap.2015.05.004>.
- Dontcheva-Navratilova, O.
 2014 "The changing face of Czech academic discourse". In: K. Bennett (ed.) *The Semiperiphery of Academic Writing*. Basingstoke: Palgrave Macmillan, 39-61. https://doi.org/10.1057/9781137351197_3.
 2020 "Persuasion in academic discourse: Metadiscourse as a means of persuasion in Anglophone and Czech linguistics and economics research articles". In: O. Dontcheva-Navratilova et al. *Persuasion in Specialised Discourses*. Cham: Palgrave Macmillan, 121-158.
<https://doi.org/10.1007/978-3-030-58163-3>.
 2021 "Engaging with the reader in research articles in English: Variation across disciplines and linguacultural backgrounds", *English for Specific Purposes* 63, 18-32. <https://doi.org/10.1016/j.esp.2021.02.003>.
- Fløttum, K. – T. Dahl – T. Kinn
 2006 *Academic Voices: Across Languages and Disciplines*. Amsterdam: John Benjamins Publishing.
- Flowerdew, J.
 2008 "Scholarly writers who use English as an Additional Language: What can Goffman's 'Stigma' tell us?", *Journal of English for Academic Purposes* 7 (2), 77-86. <https://doi.org/10.1016/j.jeap.2008.03.002>.
- Framework Education Programme for Basic Education
 2007 Ministry of Education, Youth and Sports of the Czech Republic, <https://www.msmt.cz/vzdelavani/zakladni-vzdelavani/framework-education-programme-for-basic-education>, accessed November 2023.
- Gray, B.
 2015 *Linguistic Variation in Research Articles. When Discipline Tells only Part of the Story*. Amsterdam: John Benjamins. <https://doi.org/10.1075/scl.71>.
- Gray, B. – E. Cotos – J. Smith
 2020 "Combining rhetorical moves analysis with multi-dimensional analysis: Research writing across discipline". In: U. Römer – V. Cortes – E. Friginal (eds.) *Advanced Corpus-based Research on Academic Writing*. Amsterdam and Philadelphia: John Benjamins., 137-168.
<https://doi.org/10.1075/scl.95.06gra>.

- Haggan, M.
 2004 "Research paper titles in literature, linguistics and science: Dimensions of attractions", *Journal of Pragmatics* 36 (2), 293-317. [https://doi.org/10.1016/S0378-2166\(03\)00090-0](https://doi.org/10.1016/S0378-2166(03)00090-0).
- Halliday, M. A. K.
 1994 *An Introduction to Functional Grammar* (2nd edn.). London: Edward Arnold.
- Harwood, N.
 2005 "'We do not seem to have a theory... The theory I present here attempts to fill this gap': Inclusive and exclusive pronouns in academic writing", *Applied Linguistics* 26, 343-375. <https://doi.org/10.1093/applin/ami012>.
- Hyland, K.
 1999 "Academic attribution: Citation and the construction of disciplinary knowledge", *Applied Linguistics* 20 (3), 341-367. <https://doi.org/10.1093/applin/20.3.341>.
 2000 *Disciplinary Discourses: Social Interactions in Academic Writing*. Harlow, UK: Longman.
 2002 "Authority and invisibility: Authorial identity in academic writing", *Journal of Pragmatics* 34 (8), 1091-1112. [https://doi.org/10.1016/S0378-2166\(02\)00035-8](https://doi.org/10.1016/S0378-2166(02)00035-8).
- Hyland, K. – F. (K.) Jiang
 2016a "Change of attitude? A diachronic study of stance", *Written Communication* 33, 251-274. <https://doi.org/10.1177/0741088316650399>.
 2016b "'We must conclude that...': A diachronic study of academic engagement", *Journal of English for Academic Purposes* 24, 29-42. <https://doi.org/10.1016/j.jeap.2016.09.003>.
 2017a "Is academic writing becoming more informal?", *English for Specific Purposes* 45, 40-51. <https://doi.org/10.1016/j.esp.2016.09.001>.
 2017b "Points of reference: Changing patterns of academic citation", *Applied Linguistics* 40 (1), 64-85. <https://doi.org/10.1093/applin/amx012>.
 2018 "'In this paper we suggest': Changing patterns of disciplinary metadiscourse", *English for Specific Purposes* 51, 18-30. <https://doi.org/10.1016/j.esp.2018.02.001>.
- Ivanič, R.
 1998 *Writing and Identity*. Amsterdam: John Benjamins.
- Jalilifar, A. R.
 2010 "Research article introductions: Sub-disciplinary variation in applied linguistics", *Journal of Teaching Language Skills* 2 (2), 29-55.
- Kanoksilapatham, B.
 2015 "Distinguishing textual features characterizing structural variation in research articles across three engineering sub-discipline corpora", *English for Specific Purposes* 37, 74-86. <https://doi.org/10.1016/j.esp.2014.06.008>.

Kilgariff, A. et al.

2014 "The Sketch Engine: Ten years on", *Lexicography* 1, 7-36.

Kozubíková-Šandová, J.

2019 "Proměny českého akademického diskurzu během posledních padesáti let" [Changing Czech academic discourse in the past 50 years], *ČMF* 101, 54-71.
<https://doi.org/10.14712/23366591.2019.1.4>.

Lafuente-Millán, E.

2014 "Reader engagement across cultures, languages and contexts of publication in business research articles", *International Journal of Applied Linguistics* 24 (2), 201-223. <https://doi.org/10.1111/ijal.12019>.

Li, Z. – J. Xu

2019 "The evolution of research article titles: The case of *Journal of Pragmatics*, 1978-2018", *Scientometrics* 121, 1619-1634. <https://doi.org/10.1007/s11192-019-03244-3>.

Lim, J. M-H.

2014 "Formulating research questions in experimental doctoral dissertations on Applied Linguistics", *English for Specific Purposes* 35, 66-88. <https://doi.org/10.1016/j.esp.2014.02.003>.

Lorés-Sanz, R.

2011 "The construction of the author's voice in academic writing: The interplay of cultural and disciplinary factors", *Text & Talk* 31 (2), 173-193. <https://doi.org/10.1515/text.2011.008>.

Mauranen, A. – C. Pérez-Llantada – J. M. Swales

2010 "Academic Englishes: A standardised knowledge?" In: A. Kirkpatrick (ed.) *The Routledge Handbook of World Englishes*. London/New York: Routledge, 659-676.

Moreno, A. I.

2021 "Selling research in RA discussion sections through English and Spanish: An intercultural rhetoric approach", *English for Specific Purposes* 63, 1-17. <https://doi.org/10.1016/j.esp.2021.02.002>.

Mur-Dueñas, P.

2007 "'I/we focus on...': A cross-cultural analysis of self-mentions in business management research articles", *Journal of English for Academic Purposes* 6, 143-162. <https://doi.org/10.1016/j.jeap.2007.05.002>.

Pérez-Llantada, C.

2013 "'Glocal' rhetorical practices in academic writing", *European Journal of English Studies* 17 (3), 251-268.
<https://doi.org/10.1080/13825577.2013.867180>.

Sahragard, R. – H. Meihami

2016 "A diachronic study on the information provided by the research titles of applied linguistics journals", *Scientometrics* 108, 1315-1331
<https://doi.org/10.1007/s11192-016-2049-4>.

- Samraj, B.
 2002 "Introductions in research articles: Variations across disciplines", *English for specific purposes* 21 (1), 1-17.
[https://doi.org/10.1016/S0889-4906\(00\)00023-5](https://doi.org/10.1016/S0889-4906(00)00023-5).
- Sancho Guinda, C.
 2015 "Genres on the move: Currency and erosion of the genre moves construct", *Journal of English for Academic Purposes* 19, 73-87.
<https://doi.org/10.1016/j.jeap.2015.07.001>.
- Schmied, J.
 2011 "Academic writing in Europe: A survey of approaches and problems". In: J. Schmied (ed.) *Academic Writing in Europe: Empirical Perspectives*. Göttingen: Cuvillier Verlag, 1-22.
 2015 "Graduate writing in Europe in comparison: A research-based approach to metalanguage and genre". In: J. Schmied (ed.) *Academic Writing for South Eastern Europe: Practical and Theoretical Perspectives*. Göttingen: Cuvillier Verlag, 1-24.
- Shaw, P. – I. Vassileva
 2009 "Co-evolving academic rhetoric across culture; Britain, Bulgaria, Denmark, Germany in the 20th century", *Journal of Pragmatics* 41 (2), 290-305. <https://doi.org/10.1016/j.pragma.2008.07.009>.
- Sheldon, E.
 2011 "Rhetorical differences in RA introductions written by English L1 and L2 and Castilian Spanish L1 writers", *Journal of English for Academic Purposes* 10, 238-251.
<https://doi.org/10.1016/j.jeap.2011.08.004>.
- Šinkūnienė, J.
 2018 "The power of English: I and we in Lithuanian, Lithuanian English and British English research writing". In: P. Mur-Dueñas – J. Šinkūnienė (eds.) *Intercultural Perspectives on Research Writing*. Amsterdam/Philadelphia: John Benjamins, 15-37.
<https://doi.org/10.1075/aals.18.03sin>.
- Starfield, S. – J. Ravelli
 2009 "'The writing of this thesis was a process that I could not explore with the positivistic detachment of the classical sociologist': Self and structure in New Humanities research theses", *Journal of English for Academic Purposes* 5, 222-243. <https://doi.org/10.1016/j.jeap.2006.07.004>.
- Swales, J. M.
 1990 *Genre analysis: English in Academic and Research Settings*. Cambridge: Cambridge University Press.
 1997 "English as *Tyrannosaurus rex*", *World Englishes* 16 (3), 373-382.
<https://doi.org/10.1111/1467-971X.00071>.
 2004 *Research Genres. Explorations and Applications*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139524827>.

Swales, J. – C. B. Feak

- 1994 *Academic Writing for Graduate Students: Essential Tasks and Skills*.
Ann Arbor: University of Michigan Press.
<https://doi.org/10.3998/mpub.2173936>.

Tang, R. – S. John

- 1999 "The 'I' in identity: Exploring writer identity in student academic writing through the first person pronoun", *English for Specific Purposes* 18, 23-39. [https://doi.org/10.1016/S0889-4906\(99\)00009-5](https://doi.org/10.1016/S0889-4906(99)00009-5).

Tardy, C.

- 2004 "The role of English in scientific communication: *Lingua franca* or *Tyrannosaurus rex*?", *Journal of English for Academic Purposes* 3, 247-269.
<https://doi.org/10.1016/j.jeap.2003.10.001>.

Thewell, M. – A. Bas-Bleda

- 2020 "How common are explicit research questions in journal articles?", *Quantitative Science Studies* 1 (2), 730-748.

Thompson, G.

- 2001 "Interaction in academic writing: Learning to argue with the reader", *Applied Linguistics* 22 (1), 58-78. <https://doi.org/10.1093/applin/22.1.58>.

Thompson, G. – P. Thetela

- 1995 "The sound of one hand clapping: The management of interaction in written discourse", *Text & Talk* 15, 103-127.

Vassileva, I.

- 1998 "Who am I/who are we in academic writing?" *International Journal of Applied Linguistics* 8 (2), 163-192.
<https://doi.org/10.1111/j.1473-4192.1998.tb00128.x>.

Xiang, X. – J. Li

- 2020 "A diachronic comparative study of research article titles in linguistics and literature journals", *Scientometrics* 122, 847-866.
<https://doi.org/10.1007/s11192-019-03329-z>.

Yang, R. – D. Allison

- 2004 "Research articles in applied linguistics: Structures from a functional perspective", *English for Specific Purposes* 23, 264-279.
[https://doi.org/10.1016/S0889-4906\(03\)00005-X](https://doi.org/10.1016/S0889-4906(03)00005-X).

Yakhontova, T.

- 2006 "Cultural and disciplinary variation in academic discourse: The issue of influencing factors", *Journal of English for Academic Purposes* 5, 153-167.
<https://doi.org/10.1016/j.jeap.2006.03.002>.

Address: OLGA DONTCHEVA-NAVRATILOVA, Department of English Language and Literature, Faculty of Education, Masaryk University, Poříčí 9, Brno 602 00, Czech Republic. navratilova@ped.muni.cz.

ORCID code: <https://orcid.org/0000-0002-0378-7975>.

Academic writing conventions in English-medium linguistics journals in Italy: Continuity and change over the last 30 years

Marina Bondi and Jessica Jane Nocella

University of Modena and Reggio Emilia

ABSTRACT

Against the background of studies on “academic Englishes”, this paper is a study parallel to Dontcheva-Navratilova (this issue). Focusing on the use of English in Italian academic publishing and on English linguistics in particular, we look at the development of academic writing conventions in research articles written by Italian scholars over the last 30 years. The study is based on a small corpus of 20 single-authored English-medium research articles – ten representing the period from 1990 to 1995 and ten from between 2014 and 2019 – published in the official journal of the Italian association of Anglicists (*Textus*) and in the applied linguistics journal *Rassegna Italiana di Linguistica Italiana* (RILA). The study draws on genre analysis to explore possible changes in rhetorical structure and on corpus analysis to study forms of self-mention. Special attention is paid to introductions, methodology, and conclusions. At a macrolevel, results show diachronic changes in rhetorical structure with a clearer IMRAD structure and a more empirical methodology in the second phase, while at a microlevel forms of self-mention show a marked increase in non-personal and implicit (locational) self-mention. This seems to respond to the tension between personal and impersonal forms that has largely characterized the development of the genre in English as well as to the contact between different academic cultures.

Keywords: academic English, research articles, corpus linguistics, genre analysis.

1. Introduction

Studying academic discourse means looking at ways of thinking and using language within the context of academia. The research article (RA) certainly takes pride of place within this broad realm, as it represents the key communicative practice of research communication, the primary genre

used by scholars to share the results of their research. Rhetorical structures and linguistic patterns of RAs have been widely examined in discourse and genre studies from both a diachronic and a synchronic perspective.

Following groundbreaking work in rhetoric by Bazerman (1988) and Gross et al. (2002), diachronic perspectives have often focused on changes from an author-centred perspective in the 17th century to a rather more impersonal and object-centred perspective in the 20th century. Indeed, the development of generic structures has been increasingly oriented to the structure of the scientific experimental paper, often identified by the acronym IMRAD (Introduction, Methods, Results and Discussion) (see Bondi 2022 for a diachronic overview). Academic discourse studies have confirmed these trends while also drawing attention to elements of variation. Rhetorical structures and features of author involvement are arguably influenced by disciplinary variation (e.g. Hyland 2000; Hyland – Bondi 2006), methodological choices (e.g. Gray 2015 on theoretical, quantitative and qualitative research), and types of interdisciplinary research (Thompson – Hunston 2019). In a diachronic perspective, Hyland and Jiang (2016, 2019) have explored language changes at the turn of the century, highlighting opposing trends in the hard sciences vs the social sciences: the social sciences seem to be moving towards greater informational focus and a preference for empirical, experimental and data-informed investigations, while the hard sciences are increasing their use of involvement features such as first- and second-person pronouns or modality, interpersonal and evaluative meanings, self-mention and engagement markers.

The recent dramatic expansion of English as the language of the global scholarly exchange has added another dimension of variation, by giving prominence to cross-cultural and intercultural issues. If cross-cultural studies have a long tradition, the perspective adopted more recently is often one of looking at English-medium writing as an interesting part of English in academic publishing (Hyland 2016). While drawing attention to issues of communicative inequality and potential discrimination between native and non-native speakers (Ferguson et al. 2011; Lillis – Curry 2010; Flowerdew 2019), the present-day role of English in scholarly communicative practices has also highlighted the role of non-native speakers as contributing a greater richness of perspectives and elements of variation to the language (Flowerdew 2001; Pérez-Llantada 2013, 2014; Heng Hartse – Kubota 2014). Scholarly attention has been increasingly paid to the tension between Anglophone norms and the rhetorical traditions of the original academic discourse communities of non-Anglophone writers, a tension which can be

seen in fact to stimulate attention to forms of ‘alternative academic written Englishes’ (Mauranen et al. 2010: 647). Within this framework, it becomes interesting to look at both elements of variation and hybridity in the English that gets published (e.g. Pérez-Llantada 2013; Martinez 2018) and at the features of English-medium discourse of different non-Anglophone scholars (e.g. Lillis – Curry 2006; Mur-Dueñas 2007, 2011; Bondi 2009; Shaw – Vassileva 2009; Hewings et al. 2010; Lorés-Sanz 2011; Dontcheva-Navratilova 2012, 2014, 2016, 2020; Luzón 2018; Moreno 2021), in a perspective that often weaves together the fields of Intercultural Rhetoric, English for Academic Purposes and English as a Lingua Franca (Mur-Dueñas – Šinkūnienė 2018).

The combination of a diachronic approach with a focus on English-medium publications is understandably less explored, as the phenomenon is relatively recent. It is our contention, however, that it is important to look at micro-diachronic changes that may help us trace the way local forms of the global language have developed, to see where they diverge or converge. That is why our work concentrates on the specific context of English-medium publications in Italy, with a view to changes in rhetorical structure and authorial involvement that have been observed by others on a global scale.

The aim of this paper is thus to explore how academic language has changed in the context of Italian academia over the time span of thirty years. We do this through a case study of a single discipline, i.e. English linguistics, using a small corpus of twenty RAs in the field. Drawing upon Dontcheva-Navratilova’s parallel diachronic study (this issue), we aim to explore elements of continuity and change in the generic structure of RAs (looking at titles, abstracts, introductions, methodology and conclusions) and in involvement features (looking at authorial self-mention).

The rest of the article is structured as follows. Section 2 provides the reader with a methodological background on RAs and self-mention, followed by a description of corpus design and collection, analytical tools, and methods (Section 3). The outcomes from the macro-analysis (Section 4) are divided into four sub-sections: the first one (4.1) focuses on the general structure of RAs, looking at changes in their IMRAD structure and headings’ patterns, the other sections focus on rhetorical changes in introductions (4.2), methods (4.3) and conclusions (4.4). Results from the microanalysis (Section 5) focus on personal pronouns dealing with their frequency and distribution (5.1), their collocates and the types of identity they construct (5.2), and the different forms of self-mention attested (5.3). The study then closes with conclusions (6), which sum up the main results of our analysis and take into account discussions on language and cultural variation.

2. Methodological background: Rhetorical structure and self-mention in RAs

2.1 Move analysis

Academic discourse studies have long investigated the rhetorical structures and linguistic patterns of RAs as a genre, largely on the basis of seminal work on genre by Swales (1990, 2004; Swales – Feak 2012) and Bhatia (1993). The genre analytical approach has led to a vast literature focusing on the major sections of the RA (see Samraj 2016 for an overview).

Studies on Introductions, for example, have been greatly influenced by Swales's Create-a-Research-Space model (CARS) (see Swales 1990 and 2004), identifying the moves and steps that characterize them:

- Move 1, 'Establishing a territory' (via 'Topic generalizations of increasing specificity');
- Move 2, 'Establishing a niche' (via Step 1A 'Indicating a gap' or Step 1B 'Adding to what is known', with Step 2, optional, 'Presenting positive justification');
- Move 3, 'Presenting the present work' (via Step 1 'Announcing present research descriptively and/or purposively', and a range of other options like Step 2 'Presenting RQs or hypotheses', Step 3 (optional) 'Definitional clarifications', Step 4 'Summarizing methods', Step 5 'Announcing principal outcomes', Step 6 'Stating the value of the present research', Step 7 'Outlining the structure of the paper' (Swales 2004: 230-232).

Successive models have introduced terminological variations, such as 'Claiming relevance of field', 'Establishing the gap present research is meant to fill', and 'Previewing authors' new accomplishment' (Lewin et al. 2001). The basic organizational structure, however, has been widely accepted and the three main moves occupy pride of place in many successive adaptations or studies working at the interface of rhetorical structure and linguistic features (e.g. Cortes 2013; Gray et al. 2020).

The rhetorical composition of other sections has only attracted attention more recently (Swales – Feak 2012 [1994]: 291). Methods sections, in particular, have undergone various classifications. The most recent models, combining move analysis with multidimensional study of linguistic features (Cotos et al. 2015; Gray et al. 2020), have centred on three basic moves which would be applicable to a range of disciplines:

- Move 1, 'Contextualizing methods' (with the following steps: 'Referencing previous works', 'Providing general background', 'Identifying

- the methodological approach', 'Describing the setting', 'Introducing the subjects/participants' and 'Rationalizing pre-experiment decisions');
- Move 2, 'Describing the study' ('Acquiring the data', 'Describing the data', 'Identifying variables', 'Delineating experimental/ study procedure', 'Describing tools/ instruments/ materials/ equipment', 'Rationalising experiment decisions', 'Reporting incrementals');
 - Move 3, 'Establishing credibility' ('Preparing the data', 'Describing the data analysis', 'Rationalising data processing/ analysis') (Gray et al. 2020: 144)

As clearly shown by Gray et al. (2020: 143), the process of move analysis usually involves developing a framework and validating it for the corpus under examination. Models such as these offer an interesting complex of categories validated on a large multidisciplinary corpus, which can then be adapted to the corpus under investigation. In the case of our small corpus, for example, as there was no experimental research, the range of steps to be considered could be limited to a shorter list that excluded the steps typically referred to as experimental research.

Similarly, when it comes to discussion/conclusions, the models available are numerous (e.g., Swales 1990, 2004; Holmes 1997). Corpus-based work has produced both more complex and comprehensive multidisciplinary models (Gray et al. 2020) and more specific models, such as the very influential Yang and Allison's (2003) model, based on applied linguistics RAs. Gray et al.'s model (2020) has the advantage of adopting the CARS metaphors of the territory and the niche, thus reflecting the close relationship between conclusions and introductions, with its four moves:

- Move 1, 'Re-establishing territory' ('Drawing on theoretical/general background', 'Drawing on study-specific background', 'Highlighting principal findings', 'Previewing the discussion roadmap');
- Move 2, 'Framing the new knowledge' ('Explicating and accounting for results', 'Clarifying expectations' and 'Addressing limitations');
- Move 3, 'Reshaping the territory' ('Supporting' or 'Countering with evidence');
- Move 4, 'Expanding the niche' ('Generalizing results', 'Claiming the value', 'Noting implications' and 'Proposing directions') (Gray et al. 2020: 145).

Yang and Allison's (2003) model, on the other hand, reflects the reality of applied linguistics RAs: Move 1, 'Summarizing the study', corresponds to 'Re-establishing the territory', whereas Move 2, 'Evaluating the study' ('Indicating

significance/advantage', 'Indicating limitations', 'Evaluating methodology') and Move 3, 'Deductions from the research' ('Recommending further study' and 'Drawing pedagogic implications') (Yang – Allison 2003: 379) combine freely elements belonging to Gray et al.'s Moves 2, 3 and 4.

The context of our study (English linguistics in Italy) somehow accompanies English linguistics from its first stages of recognition as a discipline in Italian academia to the present: at the end of the '80s English linguistics was acknowledged in the Italian University system as a separate discipline, independent from both English literature and language teaching per se. Because of this contextual change, we were particularly interested in the distinction between recommending further research and drawing pedagogic implications, which is only present in Yang and Allison's (2003) model, and thus decided to focus on those particular moves and steps.

2.2 Authorial voice: Self-mention

Involvement features analysed in multidimensional studies of language change (e.g. Hyland – Jiang 2019) obviously encompass a wide range of elements that contribute to the creation of authorial voice (Thompson 2001; Matsuda – Tardy 2007; Hyland – Guinda 2012; Bondi 2014). These elements include evaluative language, modality, markers of positioning in general. While acknowledging that a specific focus on authorial voice would require consideration of all these features, we limited our focus to markers of self-mention in RAs because these have repeatedly been highlighted as distinctive of different academic cultures (Mur-Dueñas 2007; Lorés-Sanz 2011; Dontcheva-Navratilova 2012, 2016; Mur-Dueñas – Sinkuniene 2018; Ädel 2022). Cross-cultural analysis of personal and impersonal constructions, in particular, has shown that Italian writers of linguistics RAs tend to use personal forms less frequently than native speakers of English (Molino 2010).

Studies of self-mention often consider both tools of self-reference and semantic elements of writer identity. Writer identity is commonly investigated by looking at first-person pronouns. Tang and John's (1999) influential study established a typology of these identities in students' essays, rating authorial presence from the least to the most powerful. The semantic taxonomy they propose identifies different roles: "the representative", when a generic first-person pronoun is used (e. *as we know today*); "the guide through the essay" (e.g. *let us look*), accompanying the reader through the various parts of the writing; "the architect" (e.g. *I will outline*), where the

first person *I* highlights the person who writes, structures, and outlines the essay; the “recounters of the research process”, describing the various steps of research (e.g. *I have worked, read...*); “the opinion-holder”, usually co-occurring with mental-process verbs, and “*I* the originator”, signalling what is new in an essay (Tang – John 1999: 26-29). Building on this work, Vladimirova (2007) classifies personal pronouns according to their pragmatic functions in a corpus of linguistics RAs. The main categories are reduced to four: “*I* as researcher”, which includes Tang and John’s categories of “*I* as guide/architect/opinion-holder/originator”; “*I* indefinite”, which expands the use of the “indefinite *we*” to the first person singular; “*I* biographical”, where the researcher presents him/herself as a person in the body of the article, while “acknowledgements” where writers express their gratefulness to people that have supported them during their research are considered separately (Vladimirova 2007: 143-145).

When looking at forms of self-mention, however, first-person pronouns are not the only markers that should be considered. Self-mention can also be realized by nominal forms. These may identify writers with a generic “author(s)” or metonymically through their own research activity, i.e. referring to the research design, the actual research publication or even elements of the publication. Self-mention can thus be realized by forms such as *the study shows*, *the paper argues* or *the table illustrates*. Building upon Dahl’s “locational metatext” (2004: 1811), these could be referred to as forms of “locational self-mention” (Bondi 2014), where self-mention is realized by reference to discourse units of different levels (e.g. *article*, *results*, *table*) and by cognitive constructs (e.g. *approach*, *analysis*), acting as textual substitutes for the self. Writers are not as fully visible as in first-person reference, but they are adequately represented by the text itself as Actor, especially in organizing and framing discourse.

3. Materials and methods

In order to carry out our study, we created a small diachronic corpus which consists of twenty single-authored RAs, for a total of 122,964 tokens. This is divided into two small sub-corpora, each corresponding to a different period. The first, which we will call Phase A, consists of ten English-medium RAs published in the time span of 5 years between 1990 to 1995, while the second, which we will call Phase B, contains ten RAs published between 2014 to 2019.

Texts were collected from the official journal of the Italian association of Anglicists *Textus*, founded in 1988, and from *RILA* (*Rassegna Italiana di Linguistica Applicata*) which is an Italian applied linguistics journal founded in 1969, publishing articles in a number of different languages. Both journals are edited and published in Italy and were chosen because of the high number of English-medium articles on English Linguistics by Italian scholars. They are indexed in international databases and are largely available outside Italy, but they clearly have a national audience in mind as a primary target.

Articles were selected on the basis of two criteria: first, they had to be written by a single author who is a native speaker of Italian working abroad or in Italy, second, their area of investigation and topic had to be English Linguistics.

Our initial aim was to collect one article per year from each journal, each one of them written by a different author. However, concerning Phase A, only a few authors were publishing in English at the time. Given such obstacles in Phase A, we selected all RAs that met the two main criteria of collection, even if that meant having no materials for certain years (1993–1994 – see Table 1) or two articles written by the same author.

Table 1. Details of Phase A sub-corpus

Samples Year	<i>RILA</i>			<i>Textus</i>		
	Number of Texts	Tokens	Affiliation	Number of Texts	Tokens	Affiliation
1990	–	–	–	–	–	–
1991	3	11,994	<ul style="list-style-type: none"> • University of Southern California • Lancaster University (2) 	–	–	–
1992	–	–	–	2	12,468	University of Bologna
1993	–	–	–	–	–	–
1994	–	–	–	–	–	–
1995	1	2,318	Lancaster University	4	34,446	University of Bologna (1); University of Milan (3)
Total	4	14,312		6	46,914	
Total number of tokens Phase A: 63988						

Similarly, in Phase B, not all *RILA* articles met the criteria of selection, which is why the corpus includes a higher number of texts from *Textus* (Table 2). However, whenever it was possible, articles were selected according to the chronological order in which the first article that met the two criteria of selection appeared in the Table of Contents of each year's journal.

Table 2. Details of Phase B sub-corpus

Samples Year	<i>RILA</i>			<i>Textus</i>		
	Number of Texts	Tokens	Affiliation	Number of Texts	Tokens	Affiliation
2014	1	4,954	University of Bergamo	1	4,492	University of Modena e Reggio Emilia
2015	–	–	–	1	7,277	University of Campania
2016	1	6,472	University of Bolzano	1	7,011	University of Catania
2017	1	6,472	Ca' Foscari	1	4,624	University di Genova
2018	–	–	–	1	6,243	University of Bologna
2019	1	5,147	University of Rome	1	6,644	University of Turin
Total	4	22,583		6	36,393	
Total number of tokens Phase B: 58976						

In order to analyse our data, we carried out a corpus-based study in combination with a genre analysis of the main sections. First, we looked at the general IMRAD structure of the papers and how this has changed diachronically.

The analysis of abstracts was limited to their presence/absence as elements of the RA, since the presence of this section was too limited to allow for an analysis of change in the structure or language. Italian academic publishing seems to have been slow in understanding the key role of abstracts in providing readers with access to the text as well as to the rationale behind the research (Bondi 2014).

The study of titles, headings and subheading patterns was oriented to understanding whether they reflected more the themes or the research

structure of the article. Titles were analysed according to the taxonomy presented by Cheng et al. (2012) and Xiang and Li (2020), looking first at changes in their syntactic structure (nominal constructions; prepositional phrases; V-ing phrases, full sentences, and compound constructions – divided by a punctuation mark) and then at the type of information presented. Information types were divided according to five categories: topic only, method/design, dataset, result and conclusion (Sahragard – Mejhami 2016; Li – Xu 2019; Xiang – Li 2020). Lastly, adapting Busch-Lauer (2000) and following Dontcheva-Navratilova (this issue), we classified title functions into the designating function, which shows the content of the text, and the advertising function, which attracts and appeals to readers.

Particular attention was then paid to introductions, methods sections, and conclusions, exploring possible changes in rhetorical structures. We decided to focus on these sections not only because of word-limit constraints, but also to carry out a specular analysis to Dontcheva-Navratilova's work (this issue), which focuses on the same research sections.

Moves and specific steps were manually tagged using *UAM Corpus Tool* (see O'Donnell 2014 for an overview), which allowed us to semantically annotate sentences of our corpus according to each annotation scheme. More specifically, our analysis of introductions centred on Swales's CARS model, also known as 'Establishing Territory' (Move 1), 'Identifying a Niche' (Move 2), and 'Addressing a Niche' (Move 3) (See Section 2 for more detail). Similarly, rhetorical methodological moves and conclusions were analysed and tagged according to their respective move model. Methods were analyzed according to the model provided by Cotos et al. (2015) and Gray et al. (2020), while conclusions were tagged according to Yang and Allison's (2003) taxonomy (see Section 2).

At the microlevel of analysis, we looked first at personal structures for writer and reader references availing ourselves of *WordSmith Tools* 8.0 (Scott 2020) for the quantitative part. We focused on the use of personal pronouns in the texts – excluding references, footnotes and acknowledgments – and checked their frequency and collocations, with special attention to their semantic and pragmatic features. When looking at writer identity in self-mentions, the simplified model we adopted (adapted from Tang – John [1999] and Vladimirou [2007]) was meant to identify how often writer identity coincided with the Writer as Researcher/Writer/Interpreter and how often it rather referred to the Biographical self of the author, or to an Indefinite self. Finally, we looked at diachronic trends in nominal and pronominal forms of self-reference; in particular we studied personal and

depersonalized “locational” forms of self-mention (Bondi 2014) involving different types of verbs that represent argumentative procedures across the two sub-corpora.

4. Results: Macro-analysis of RAs

This section looks at the general structure of the RAs (abstracts, titles and paragraph headings, IMRAD structure) and at the rhetorical moves adopted in their introduction, methods sections, and conclusion.

4.1 Structure

Before going into further detail, it is worth mentioning that the total wordcount of the articles in Phase A is 66,134 (an average of 6,613 tokens per article), while there are 61,945 tokens in Phase B (an average of 6,194). However, due to the presence of an appendix, one RA in Phase A is an outlier, with a wordcount of 18,173 tokens, which is almost three times the general average. Excluding this outlier from our data, the average wordcount per article becomes 4,796 tokens, leading to a significant difference with the second phase. This change in length over the years might be due not only to editorial reasons of the journals (i.e. requiring a minimum length in words), but also to a rise in the informativeness of RAs (see Dontcheva-Navratilova, this issue).

Abstract

From a close examination of the ten RAs of Phase A it emerged that papers are never introduced by abstracts, and that their general IMRAD structure is not always consistent. In three of the RAs, for example, the methodology is not clearly stated. In articles belonging to the Phase B sub-corpus, abstracts are consistently present in all RAs, with a difference in language choice, probably depending on each journal's editorial line: in *Textus* these are always in English, whereas in *RILA*, until 2018 they are in Italian and start to appear in English from 2019.

Titles

The general length of titles increases from Phase A (11.9) to Phase B (12.7). Again, this could be due to the need of authors to be more informative, but also to the rise in the use of acronyms (e.g. EMI, CLIL, EU), which have been counted as separate words.

Table 3. Titles’ syntactic structure in Phase A and Phase B

Titles	Title Length	Surface form				
		Nominal	Prep. phrase	V-ing phrase	Compound	Sentence
Phase A	11.9	6	0	0	4	0
Phase B	12.7	2	0	0	8	0

Table 4. Titles’ information type and function in Phase A and Phase B

Titles	Information type					Function	
	Topic	Topic + dataset	Topic + Method	Topic + Results	Topic + Conclusion	Designate	Advertise
Phase A	4	1	1	2	2	10	0
Phase B	1	5	1	3	0	5	5

Phase A shows a predominance in nominal structures, which represent 60% of the small sub-corpus (Table 3), while the remaining 40% consists of compound titles. As also found in Dontcheva-Navratilova’s case study (this issue), Phase B sees an increase in compound titles, which have doubled since Phase A (80%), while the remaining 20% is made up of nominal structures. Once more, this might be related to the need of providing more information and detail about the article.

With regards to the information type of the titles (Table 4), in Phase A 40% of cases are topic-oriented (Example 1), while the remaining 60% are divided between mentioning dataset (10%), methods (10%), results (20%) and conclusions (20%; see Example 2). In Phase B, strictly topic-oriented titles decrease to 10%, and most of them hint at their dataset (Example 3), methods, and results (Example 4).¹ Titles in Phase B of our small corpus do not hint at information on conclusions.

- (1) “Need” As Modal Auxiliary and As Lexical Verb in Present-Day English (PA-MB)
- (2) Italian and the English s-genitive: a contrastive analysis and its pedagogical implications for the teaching of Italian as a foreign language. (PA-ES)

¹ Examples are identified by phase and author initials (see Table 1, 2 and 8).

- (3) On the phraseological dimension of legal discourse: the case of English and Italian contracts (PB-GD)
- (4) Invested effort for learning in CLIL and student motivation: How much are they related? Answers from the Italian context. (PB-AB)

The function of titles has also seen a significant change over the years (Table 4). In Phase A, all titles focus on the content of a piece of work, making the designating function predominant (Example 5). In Phase B, 50% of the titles show a designating function, while 50% have an advertising one (Example 6).

- (5) Semantic Syntax. English Phrasal Verbs. (PA-GA)
- (6) Tastes We've Lived By. Taste Metaphors in English. (PB-MB)

The increase in the advertising function might be linked to higher competitiveness in research, leading authors to find strategies that might encourage readers to read their work (i.e., puns, question marks, evaluative language, etc.). Therefore, the advertising function of titles is a shop window, whose purpose is to attract and appeal to editors, reviewers, and readers.

Paragraph headings

Paragraph heading patterns belonging to Phase A are irregular: three of the selected articles have no headings at all; in other words, there were no numbers or spaces dividing the different sections of the paper. Four papers present their work in a so-called research or hybrid structure, which means that some sections, for example background, would be introduced by more thematic nomenclatures, focusing on the nature of its content, while other parts, such as results, follow the IMRAD denomination (e.g. *Introduction – Newspaper Language – The Corpus – Results*, etc.). The remaining three articles instead are thematic in that they were structured with a content nomenclature for each heading rather than a research one (e.g. *I. Introduction – II. The Data – III. Theoretical approaches to the S-Genitive – IV. Genitive of measure – V. Genitives referring to geographical entities – VI Names of countries and continent.*, etc.).

With regards to Phase B, while articles have a regular IMRAD structure, headings do not show fixed patterns in their nomenclature: as a matter of fact, most of them (7/10) present a hybrid structure, with some sections introduced by more thematic headings (e.g. 1. *Theoretical background: Ideology, groupness and security discourse*, 3. *Dissemination and ideology in Europol's Annual Reports*) and other ones by those resembling the IMRAD structure

of the paper (e.g. 2. *Materials and Methods*). However, even when following the IMRAD research structure, headings can present variation in their nomenclature (Swales 2004: 219), such as *Quantitative analysis*, or *Qualitative analysis* to refer to the results' section. The remaining three articles, instead, are more thematic (e.g. 1. *Introduction* – 2. *Comparing newspaper data* – 3. *Labelling pronunciation in The Times*, etc.).

4.2 Introductions

A close examination of introductions shows minor but interesting diachronic changes in moves and steps at a quantitative level of analysis.

Table 5. Introductions' Rhetorical Moves present in Phase A and Phase B RAs

FEATURE	Phase A				Phase B			
	NT	%M	NS	%(S)	NT	%M	NS	%(S)
Move 1: Establishing territory	10	50			10	43		
Step 1: Providing general background	9		9	52.94	9		10	37.04
Step 2: Claiming centrality	2		3	17.65	4		5	18.52
Step 3: Reviewing previous research	5		5	29.41	7		12	44.44
Move 2: Identifying a niche	3	15			4	17		
Step 1: Raising general questions	2		2	28.57	0		0	
Step 2: Highlighting a problem	1		1	14.29	2		2	33.33
Step 3: Indicating a gap	2		4	57.14	4		4	66.67
Move 3: Addressing the niche	7	35			9	39		
Step 1: Introducing present research descriptively	5		5	31.25	6		7	26.92
Step 2: Announcing research aims/ purposes	2		4	25	6		10	38.46
Step 3: Presenting research questions	2		2	12.5	3		3	11.54
Step 4: Presenting research hypotheses	1		1	6.25	1		1	3.85
Step 5: Clarifying definitions	2		2	12.5	1		1	3.85
Step 6: Outlining the structure of the paper	2		2	12.5	4		4	15.38

Table 5 shows the number of texts in which each move type is found (NT), the ratio of each move type to the total number of moves (%M), the number of sentences involved in the realization of each step (NS) and the percentage of all steps that each step represents (%S).

The data shows that, while the number of texts containing Move 1 remains the same, there is a slight increase in Move 2 and a more marked increase for Move 3. The stability of Move 1 in terms of number of texts involved, however, means that its role has proportionately reduced in Phase B, as the importance of Move 2 and especially 3 grows accordingly.

The data on steps shows a general tendency to produce slightly more elaborate texts in Phase B (with more sentences or occurrences for the same step), especially for 'Reviewing previous research' in Move 1 and 'Announcing purposes' in Move 3. When looking at percentage values for each step, we may see that Move 1, dominated by 'General background' in Phase A is rather dominated by 'Reviewing previous research' in Phase B. Move 2 shows a constant centrality of 'Indicating a gap', while Move 3 sees a shift from 'Introducing present research descriptively' to 'Announcing research purposes'.

Overall then, not only does 'Addressing the niche' gain ground over 'Establishing territory', but also the territory and the niche provide increasing attention to placing the study within the context of the scientific debate, by reviewing previous research, indicating a gap and announcing purpose.

It is also interesting to notice that 'Indicating a gap' shows a slightly wider distribution and variable position. While in Phase A such step appears in the first paragraph following the aims, in Phase B it also appears either in the following paragraphs (3rd, 4th or 5th) or preceding the aims. Similarly, the statement of purpose ('Announcing research purposes') appears in the first two paragraphs of the introduction in Phase A, while in Phase B not only does it tend to appear in the following paragraphs (3rd or 4th), but is also repeated in other sections of the paper, such as in the methodology. Qualitatively, on the other hand, the two phases show very similar patterns (Examples 7-8), as in both of them the statement of purpose achieves its realization by highlighting elements of novelty and linking them to the existing literature and the gap identified (Samraj 2002).

- (7) The present paper *aims at complementing these studies by* (...). (PA-GI)
- (8) *This article, therefore, complements the wide and constantly growing range of studies centred on the Frog Stories, and adds novel information on using English* (...) (PB-MI).

The lexico-grammatical patterns are the same in both phases: *the present paper; in this paper/study; the aim of this project is to...; the aim is to...; the article is aimed at...; in order to have some evidence...; this article deals with...; etc.*).

4.3 Methods section

As already noticed, not all RAs in Phase A present a clear regular IMRAD structure; in other words, for some papers the methods section is not clearly outlined, whereas it is much more obvious in Phase B. This is most likely due to the influence of publications in international peer reviewed journals, where authors had adhered to the IMRAD structure for some time; Italian academics were thus beginning to follow suit. By the time we enter Phase B, there appears to have been a general increase in the explicit signalling of rhetorical moves and steps, again probably due to the fact that research has become more competitive, and it has become necessary to provide details on both the data and the methods applied to one’s work in order to establish credibility and significance to one’s work.

Table 6 illustrates that Phase B is characterized by greater attention to the rhetorical moves of ‘Contextualizing methods’ (especially focusing on participants) and ‘Establishing credibility’, which are found in a higher number of texts, represent a higher percentage of the moves, and are characterized by more complex steps. The data also shows that the increase in complexity of the steps represented by a higher number of sentences also characterizes the descriptive move which decreases in relative importance.

Table 6. Diachronic changes in Methodology’s Moves and Steps

FEATURE	Phase A				Phase B			
	NT	%M	NS	%(S)	NT	%M	NS	%(S)
M1 Contextualizing methods	5	35.71			8	38.07		
S1 Providing general background	4		4	18.18	6		6	16.22
S2 Identifying the methodological approach	3		3	13.64	5		6	16.22
S3 Introducing participants	1		1	4.55	2		4	10.81
M2 Describing the study	8	57.14			9	42.87		
S1 Describing data	7		7	31.82	6		8	21.62

S2 Delineating study procedures	5		5	22.71	6		6	16.22
S3 Identifying variables	1		1	4.55	3		3	8.10
M3 Establishing credibility	1	7.15			4	19.06	4	10.81
S1 Describing the data analysis	1		1	4.55	4		4	10.81

In general, throughout Phase A, there is no clear order in the way moves and steps appear, while in Phase B, authors seem to follow a distinct rhetorical pattern.

4.4 Conclusions section

With regards to Yang and Allison's (2003) rhetorical moves for conclusions, Phase A shows no clear order in which they appear, hence leading to more variety in style (e.g. Move 1 → Move 2 → Move 1 → Move 3). However, despite not appearing in the expected order, moves are quantitatively more balanced throughout conclusions (Table 7). In Phase B instead, not all moves are equally present even though they appear in the expected order (e.g. Move 1 → Move 2 → Move 3 – Step 1 – Step 2).

Table 7. Diachronic changes in Conclusions' Moves and Steps

FEATURE	Phase A				Phase B			
	NT	%M	NS	%(S)	NT	%M	NS	%(S)
Move 1: Summarizing	9	40.9	10	31.26	10	45.45	13	38.24
Move 2: Evaluating study	7	31.8			8	36.36		
Step 1: Indicating significance	3		3	9.37	5		6	17.65
Step 2: Indicating limitations	4		6	18.75	3		4	11.76
Step 3: Evaluating methodology	2		2	6.25	5		5	14.71
Move 3: Deduction from the research	6	27.3			4	18.19		
Step 1: Pedagogical implications	4		5	15.62	4		4	11.76
Step 2: Recommendations	5		6	18.75	2		2	5.88

In particular, Move 1 has remained diachronically stable, while Move 3 sees a significant drop, which is mainly determined by a decrease in the writing of recommendations for further studies (Step 2), which fell from 18% to 6%, accompanied by a less marked reduction in the pedagogical implications. This decrease – in line with a generalized growing attention to methodology – might reflect the national context, where English linguistics was establishing itself as an academic discipline, with an interest in implications but not necessarily in direct professional applications. Table 7 also shows an increase in the use of Move 2, which is the one used to evaluate the research study. Specifically, Step 1, used to indicate the significance of a study, and Step 3, adopted to evaluate its methodology, have significantly increased since Phase A, while there has been a decrease in the description of one study's limitations (Step 2). Such decrease might indicate a hybridization between the Anglo-Saxon model and the Italian one, which tends to avoid highlighting limitations, as these might be seen as a weakness. In general, however, the respective increase and decrease in the use of such steps clearly indicate a rise in the self-promotion of authors' research, in line with international trends (Hyland – Jiang 2019).

5. Self-mention: Personal and locational references

This section looks at forms of self-reference. We first deal with the presence of personal pronouns and adjectives manifesting forms of self-mention, looking at their frequency and distribution (5.1), as well as their collocates and the types of identity they construct (5.2). We then contrast personal (*we argue*) and depersonalized ("locational") (*the article argues*) forms of self-mention attested (5.3).

5.1 Personal pronouns: Frequency and distribution

A closer look at first-person pronouns and adjectives in our corpus shows a sharp decrease in the use of personal references (*I, my, me, mine, we, us, our, ours*) from Phase A with a normalized frequency of 5.52 occurrences per thousand words (ptw) to one of 2.59 ptw in Phase B. Neither of the two sub-corpora shows traces of the second-person pronoun *you*.

Starting from Phase A, we can see that the use of pronouns mainly varies according to the individual style of each author (Table 8). For example, two of the authors (GI and VP) use few personal references. This preference

might be related to sub-disciplinary trends, as they both carry out research in lexicography, but it may also be influenced by the impersonality of Italian academic culture noted for example by Molino (2010).

Table 8. Personal pronouns and possessive adjectives per author in Phase A and B

Phase A					Phase B				
author	singular	plural	total	standard deviation	author	singular	plural	total	standard deviation
MB*	25	–	25		AM	6	2	8	
LL*	14	–	14		CB	6	4	10	
ES* (1)	13	12	25		AB	3	51	54	
ES* (2) ¹	16	1	17		MS	7	2	9	
GA	12	45	57		GT	7	11	18	
MBP	8	24	32		MG	1	1	2	
GI	2	1	3		GD	–	18	18	
VP	1	1	2		MI	–	2	2	
GM	–	56	56		CPP	–	–	–	
ETB* ²	26	96	122		MB	24	8	32	
Total	117	236	353	34.02	Total	54	99	153	15.81
Normalized frequency (ptw)	1.83	3.69	5.52		Normalized frequency (ptw)	0.91	1.68	2.59	

¹ ES is the author for whom we collected two research articles.

² Authors marked with an asterisk were carrying out their research in Anglophone Universities at the time of publication.

Another important observation is that the use of the first-person singular pronoun *I* is reduced in authors developing their research in Italy, while higher in those carrying out their research abroad (marked with an asterisk in Table 8). This might reflect the impact of closer contact with different writing norms. Italian academic writing tends to avoid personalization, while Anglophone writing norms seem to be more open: Hyland (2002), for example, shows that personal pronouns followed by lexical verb phrases represent 60% of his four-word bundles of academic English in the British National Corpus Baby edition.

However, we should also notice high individual variation in the use of plural personal references (*we/us/our/ours*), which vary from 0 (LL and MB – See Table 8) to 96 (ETB – See Table 8). This may be related to the fact that plural pronouns can carry different meanings, such as the use of the *inclusive we* (Example 9), adopted to include the reader in the writing, and *exclusive we* (Example 10), which is mainly used to describe the author's meta-discursive or research process.

- (9) (...) is acquisition of ideological meaning which *we all take for granted* but is usually difficult. (PA-ETB)
- (10) The other type of movement *we are concerned with* here in extraposition (...). (PA-GA)

The choice of plural personal references over singular ones might also depend on the different argumentative logic adopted by each single author. For example, those using a higher number of plural personal references (e.g. ETB, ES, GA, GM, MBP) carry out a more empirical kind of study (e.g. corpus linguistics, use of prepositions, syntax, etc.), often hinting at a research group even with single-author texts, while most of the authors using predominantly singular personal references (GI, LL, VP) adopt a more analytical and philosophical approach to research, hence with less reference to data (i.e. macro observations on approaches adopted on second language acquisition or contemporary receptions of authors' addenda and corrigenda to Johnson's Dictionary).

In Phase B, there is less individual variation in the use personal references, with a preference for impersonality, and a sharp decrease in the use of singular personal pronouns and adjectives which drop from 1.83 occurrences ptw in Phase A to 0.91 ptw in Phase B. Individual variation is at its highest in the use of plural personal references (*we/us/our/ours*), which vary from 0 (CPP) to 51 (AB) and which remain dominant over the use of singular ones in three quarters of the RAs.

5.2 Self-mentions and self-representation

A closer qualitative analysis of the use of singular personal references shows a reduction in the range of collocates in favour of a more consolidated phraseology which confirms that authors are increasingly paying more attention to methodological statements rather than to meta-discursive

frames. Singular person pronoun *I*, which is the most frequent pronoun adopted in RAs, generally collocates with verbs expressing verbal and cognitive processes (e.g. *I deal with, analyze, think, concentrate on, give examples, reconsider*, etc.), which frame the discursive presentation of the research (Table 9). This is also visible and confirmed in collocates of the personal adjective *my*, which is again followed by items related to research methodology and results (e.g. *corpus, example, data*, etc.)

Table 9. Collocates of singular personal references in Phase A and Phase B

Form	Phase A		Phase B	
	Collocates	Frequency (ptw)	Collocates	Frequency (ptw)
I	ask, analyze, answer, compare, consider, describe, discuss, evaluate, find, have the space, introduce, list, look at, mention, proceed, report, point out, quote, regard, say, show, think, trace,	88 (1.37)	argue, classify, consider, don't know, elaborate, further extend, identify, include, report, think, will focus/ argue/ compare	41 (0.69)
Me	Seems to me, let me, available to me	5 (0.08)	–	–
My	analysis, book, data, dictionary, discussions, example, investigation, knowledge, only hope/ possibility, own, opinion, study, theory, view, work	21 (0.33)	corpus, contention, disposition, months, research	13 (0.22)
Mine	publication, work, analysis	3 (0.05)	–	–

By taking a step further and looking at the semantic referents of the first-person singular pronouns of both phases, we notice that the notion of “biographical self” is present only in Phase A with 17 cases (Example 11). The author as researcher, writer, and interpreter, predominant in Phase A (88 hits), is the only type of referent attested in Phase B, though much more limited in number (41 hits).

- (11) (...) by exploring its possible applications in an environment *I am familiar with* (...). (PA-ES2)

When looking more specifically at the representation of the writer, in Phase A the representation of *I* as a researcher is that of a writer acting as guide, leading the reader through the RA, framing once more the structure of the paper (e.g. *shall/will report, introduce..., I have found, drawn examples..., I would like to discuss/consider, etc.*). In Phase B instead, the writer as a researcher takes up mostly the role of a recounter reporting the steps of his or her research, as it is mainly followed by material process verbs (e.g. *I compiled, I identified, I included, etc.*).

With regard to plural person referents in both Phases (Table 10), despite the quantitative decrease over the years, collocations remain similar. The pronoun *we* is the most frequent one (136 hits in PA and 54 in PB) and, similarly to *I*, it collocates with discourse and cognitive process verbs (i.e. *investigate, question, generalize, etc.*). This also happens with the pronoun *us* which is preceded by verbs of permission (*let, allow, bring, help*) and followed meta-discursive and research verbs (i.e. *compare, examine, sum up, etc.*). Possessive adjectives and pronouns (*our, ours*) are followed or anticipated by research process nouns (i.e. *our analysis, our study, etc.*) and their object and focus of analysis (i.e. *our corpus, our language*).

Table 10. Collocates of plural personal references in Phase A and Phase B

Form	Phase A		Phase B	
	Collocates	Frequency (ptw)	Collocates	Frequency (ptw)
We	accept, adopt, analyze*, apply, are concerned/ created/ faced with/ going to/ dealing, assume, can think/ assume/ apply/ turn..., cannot label/ take up..., decide, do, examine, found, have, identify, include, may find/ wonder..., need, notice, see, will consider/ deal/ notice...	136 (2.12)	conclude, decide, describe, find, hypothesize, we would like to comment/ highlight	54 (0.91)

Our	analysis, volume, bread and butter, choice, corpus, culture, discussion, endeavor, knowledge, language, sample, students, undertaking, worse fear	65 (1.01)	analysis, corpora/ corpus, findings, hypothesis, knowledge, perception...	29 (0.49)
Ours	(course)	1 (0.02)	–	–
Us	let/ bring/ allow/ help + compare, examine, sum up, claim, affirm, explain, understand, consider, question, query, assume	34 (0.53)	tell, provide + results, evidence, let/ allow + affirm, claim, consider	16 (0.27)

Looking at the collocates of *we* and *our*, we can notice that the emphasis is on the research process (*our analysis*) and its object (*our corpus/our language*). The focus is mostly on the ongoing research process (e.g. *we studied clusters...*, *our analysis has confirmed...*, *we would like to highlight*, etc.) and therefore on the writer as researcher. However, *we* can be more inclusive and refer to both the writer and reader (*If we now turn...*, *our students...*), thus representing the author as writer. There are also a few occurrences of references to an indefinite self or more generally to humankind (*we* as humans, e.g. *our perception*).

5.3 Personal and locational self-mention

Together with forms of personal self-mention, in both sub-corpora we find traces of “locational self-mentions” (Bondi 2014), where authors indirectly mention themselves through references to their article (Example 12), study (Example 13), sections (Example 14), and cognitive components of the article (Example 15).

(12) The article is divided into four parts: (...) (PA-GA)

(13) The findings of this study show (PB-AM)

(14) Data Analysis (PB-GT – Section heading)

(15) This paper reports the analysis of the lexical overlap (...) (PA-ETB)

Locational self-mentions are mostly used in Phase B (76% of all cases), with a reverse proportion of the general trend of the use of personal pronouns, which were more frequent in Phase A.

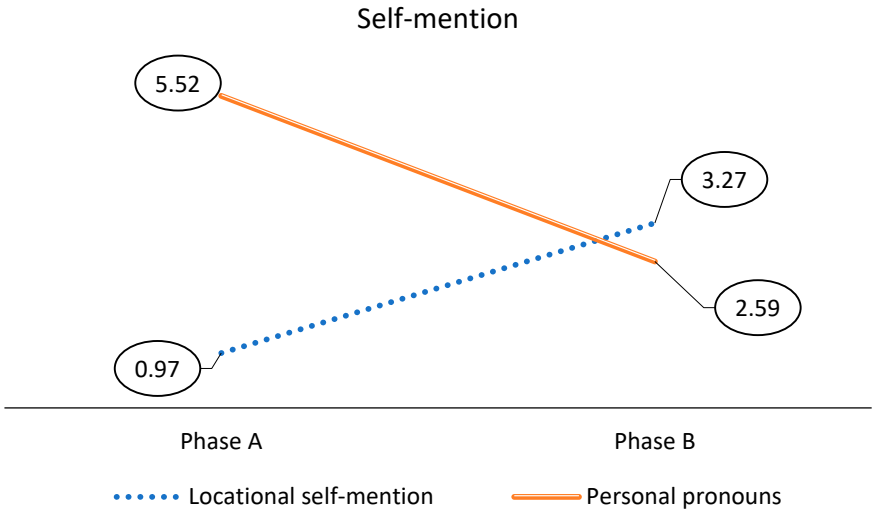


Figure 1. General diachronic trend in self-mentions (normalized frequency ptw)

Overall, as is also visible from Fig. 1, our corpus of RAs shows a slight diachronic decrease in the use of self-references and a marked increase in the use locational self-mentions (with normalized frequency going from 0.97 ptw to 3.27 ptw), along with a general decrease in personal references (from 5.52 ptw to 2.59 ptw) in Phase B. The increasing preference for locational forms of self-mention is in line with the general trend observed in the study of abstracts (Bondi 2014), leading towards what may be seen as a process of depersonalization, but also as a form of self-presentation which lies somewhere between the traditional scientific poles of subjectivity and objectivity.

6. Conclusions

This small case study has shown diachronic changes and patterns in RAs written in English by Italian scholars (in English Linguistics) over the last thirty years. A first analysis of our material revealed how English has become

a well-established medium of publication in Phase B, while in the early nineties most authors of such a subfield of research published in Italian. As a matter of fact, in Phase A almost half of the authors publishing in English were living in English-speaking countries, so they were probably used to English as their main professional language.

Delving into our corpus, we noticed that over the years the IMRAD structure of empirical research papers has become clearer, with more precise sections, titles, and headings, which, as previously seen, may also depend on the sampled sub-disciplines and their respective approach to their work. Moreover, the theoretical and methodological background have continued to occupy a conspicuous section of articles over the years, with a slight increase in the mentioning of research gaps and in the description of elements of novelty in the introductions. Phase B conclusions have also shown an increase in the use of Yang and Allison's (2003) Move 2, adopted to evaluate a study, with, however, a decrease in the description of the limitations of a study. Titles and headings show a parallel path towards a clearer structure and more complex attempts to reflect both the content and the structure of the article, while also adding an "advertising" element in titles.

Such trends – which seem to match those by Czech authors noticed by Dontcheva-Navratilova in the parallel study (this volume) – might be related to a number of factors that characterize current international publishing. The current proliferation and competition of academic publications seem to favour formats that make reading more efficient (such as the IMRAD structure of RAs or more informative titles and more structured abstracts) and at the same time determine the increasing need of authors to self-promote their work and their research in a competitive job sector.

Microanalysis has also revealed a marked diachronic increase of "locational self-mention" (Bondi 2014) together with a steady decrease in personal self-mention. This reversed trend might be due to the increase of a more empirical methodology in linguistic subfields, which may partly explain the decline in first person singular pronouns. The general decrease in first person personal pronouns seems to be in line with the paradigm of objectivity set by the hard sciences and recently adopted by the humanities, and with the need for increased credibility. The choice of depersonalized locational forms of self-mention, however, also seems to reflect the need for a "third way" between direct involvement and full impersonality. The combination of personal, impersonal and depersonalized forms of reference goes along with the growing awareness of the nature of research in the humanities, which, despite its empirical features, remains interpretative.

It is a choice that emphasizes the role played by interpretation in the act of writing, while at the same time presenting the text as Actor in academic communication.

The changes seen in the frequency of use and functions of personal structures for self-reference do not seem to match the results in the parallel study by Dontcheva-Navratilova, notwithstanding the great convergence observed in the standardization of the format of the empirical RA. The visibility and credibility of the Italian researcher is enhanced in this corpus rather more by the emphasis on the author as researcher than on the actual manifestation of personal presence. Whether this is a matter of stylistic preference of the individual authors or a truly generalizable trend in English-medium RAs remains to be checked against a larger corpus.

Admittedly, it is difficult to generalize on the basis of such a small corpus. It might be necessary to study all these aspects individually and more thoroughly on a much wider corpus. The limits of the corpus, moreover, are not only quantitative; they are also qualitative, as it is limited to papers by Italian English linguists publishing in Italian journals with some international distribution and any substantial hypothesis should be checked against other disciplines and lingua cultural backgrounds. It might even be useful to compare different publication contexts by following the same authors when publishing in English in contexts ranging from international to local or “glocalized”, much in the same way as it would be interesting to compare their writing in Italian and in English, or to explore how change is taking place in writing styles of authors from different countries. Combining the focus on Italian authors and Italian publishing may also be a limitation, as it may be worth investigating the two perspectives separately. The use of English as the publishing language in non-Anglophone countries is a relatively recent phenomenon that certainly deserves to be explored further in its diachronic dimension, but it should be seen in the context of the general expansion and development of English-medium publications.

More generally, in fact, we hope to have contributed to showing how diachronic comparative analysis can help trace change in academic discourse by mapping an area – however small – of the general radical change at work in the publishing industry. Technological, economic and distributional factors have greatly influenced academic publishing over the past thirty years, while the global dimension of discourse communities and knowledge communication seems to be able to embrace both elements of standardization and elements of variation that are worth studying.

REFERENCES

Primary sources

RILA [RASSEGNA ITALIANA DI LINGUISTICA APPLICATA]

<https://www.bulzoni.it/it/riviste/rasseгна-italiana-di-linguistica-applicata/results,1-0>, accessed November 2023.

Textus. English Studies in Italy

<https://www.carocci.it/prodotto/textus-2>, accessed November 2023.

Secondary sources

Ädel, A.

- 2022 "Writer and reader visibility in humanities research articles: Variation across language, regional variety and discipline", *English for Specific Purposes* 65, 49-62.
<https://doi.org/10.1016/j.esp.2021.09.001>.

Bazerman, C.

- 1988 *Shaping Written Knowledge: The Genre and Activity of the Experimental Article in Science*. Madison: University of Wisconsin Press.

Bhatia, V.

- 1993 *Analysing Genre. Language Use in Professional Settings*. London: Longman. <https://doi.org/10.4324/9781315844992>.

Bondi, M.

- 2009 "Polyphony in academic discourse: A cross-cultural perspective?" In: E. Suomela-Salmi – F. Dervin (eds.) *Cross-Cultural and Cross-Linguistic Perspectives on Academic Discourse*. Amsterdam: John Benjamins, 83-108. <https://doi.org/10.1075/pbns.193.06bon>.
- 2014 "Changing voices: Authorial voice in abstracts." In: M. Bondi – R. Lorés-Sanz (eds.) *Abstracts in Academic Discourse: Variation and Change*. Bern: Peter Lang, 243-270.
<https://doi.org/10.3726/978-3-0351-0701-2>.
- 2022 "The scientific article: Variation and change in knowledge communication practices". In: C. Hanganu-Bresch et al. (eds.) *The Routledge Handbook of Scientific Communication*. London: Routledge, 159-169. <https://doi.org/10.4324/9781003043782>.

Busch-Lauer, I.

- 2000 "Titles in English and German research papers in medicine and linguistics". In: A. Trosborg (ed.) *Analysing Professional Genres*. Amsterdam/Philadelphia: John Benjamins, 77-97.
<https://doi.org/10.1075/pbns.74.08bus>.

Cheng, S. W. – C. W. Kuo – C. H. Kuo

- 2012 "Research article titles in applied linguistics", *Journal of Academic Language and Learning*, 6 (1), A1–A14.

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. <https://doi.org/10.1016/j.jeap.2012.11.002>.

Cotos, E. – S. Huffman – S. Link

- 2015 "Furthering and applying move/step constructs: Technology-driven marshalling of Swalesian genre theory for EAP pedagogy", *Journal of English for Academic Purposes* 19, 52–72. <https://doi.org/10.1016/j.jeap.2015.05.004>.

Dahl, T.

- 2004 "Textual metadiscourse in research articles: A marker of national culture or of academic discipline?", *Journal of Pragmatics* 36, 1807–1825. <https://doi.org/10.1016/j.pragma.2004.05.004>.

Dontcheva-Navratilova, O.

- 2012 "Cross-cultural differences in the construal of authorial voice in the genre of diploma theses". In: C. Berkenkotter – V. Bhatia – M. Gotti (eds.) *Insights into Academic Genres*. Bern: Peter Lang, 301–328. <https://doi.org/10.3726/978-3-0351-0410-3>.
- 2014 "The changing face of Czech academic discourse". In: K. Bennett (ed.) *The Semiperiphery of Academic Writing*. Basingstoke: Palgrave Macmillan, 39–61. https://doi.org/10.1057/9781137351197_3.
- 2016 "Cross-cultural variation in the use of hedges and boosters in academic discourse", *Prague Journal of English Studies* 5 (1), 163–184. <https://doi.org/10.1515/pjes-2016-0009>.
- 2020 "Persuasion in academic discourse: Metadiscourse as a means of persuasion in Anglophone and Czech linguistics and economics research articles". In: O. Dontcheva-Navratilova et al. *Persuasion in Specialised Discourses*. Cham: Palgrave Macmillan, 121–158. <https://doi.org/10.1007/978-3-030-58163-3>.

Ferguson, G. – C. Pérez-Llantada – R. Plo

- 2011 "English as an international language of scientific publication: A study of attitudes", *World Englishes* 30 (1) (February), 41–59. <https://doi.org/10.1111/j.1467-971X.2010.01656.x>.

Flowerdew, J.

- 2001 "Attitudes of journal editors towards non-native speaker contributions", *TESOL Quarterly*, 35 (1) (Spring), 121–150. <https://doi.org/10.2307/3587862>.

- 2019 "The linguistic disadvantage of scholars who write in English as an additional language: Myth or reality", *Language Teaching* 52 (2), 249-260. <https://doi.org/10.2307/3587862>.
- Gray, B.
- 2015 *Linguistic Variation in Research Articles. When Discipline Tells only Part of the Story*. Amsterdam: John Benjamins. <https://doi.org/10.1075/scl.71>.
- Gray, B. – E. Cotos – J. Smith
- 2020 "Combining rhetorical moves analysis with multi-dimensional analysis: Research writing across discipline". In: U. Römer – V. Cortes – E. Friginal (eds.) *Advanced Corpus-based Research on Academic Writing*. John Benjamins, 137-168. <https://doi.org/10.1075/scl.95.06gra>.
- Gross, A. – J. Harmon – M. Reidy
- 2002 *Communicating Science: The Scientific Article from the Seventeenth Century to the Present*. Oxford: Oxford University Press.
- Heng Hartse, J. – R. Kubota
- 2014 "Pluralizing English? Variation in high-stakes academic texts and challenges of copyediting", *Journal of Second Language Writing* 24, 71-82. <https://doi.org/10.1016/j.jslw.2014.04.001>.
- Hewings, A. – T. Lillis – D. Vladimirov
- 2010 "Who's citing whose writings? A corpus-based study of citations as interpersonal resource in English medium national and English medium international journals", *Journal of English for Academic Purposes* 9 (2), 102-115. <https://doi.org/10.1016/j.jeap.2010.02.005>.
- 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 (4), 321-337. [https://doi.org/10.1016/S0889-4906\(96\)00038-5](https://doi.org/10.1016/S0889-4906(96)00038-5).
- Hyland, K.
- 2000 *Disciplinary Discourses: Social Interactions in Academic Writing*. Harlow, UK: Longman.
- 2002 "Authority and invisibility: Authorial identity in academic writing", *Journal of Pragmatics* 34 (8), 1091-1112. [https://doi.org/10.1016/S0378-2166\(02\)00035-8](https://doi.org/10.1016/S0378-2166(02)00035-8).
- 2016 *Academic Publishing. Issues and Challenges in the Construction of Knowledge*. Oxford: Oxford University Press.
- Hyland, K. – M. Bondi (eds.)
- 2006 *Academic Discourse across Disciplines*. Bern: Peter Lang. <https://doi.org/10.3726/978-3-0351-0446-2>.

Hyland, K. – C. S. Guinda (eds.)

- 2012 *Stance and Voice in Written Academic Genres*. Basingstoke: Palgrave Macmillan. <https://doi.org/10.1057/9781137030825>.

Hyland, K. – F. (K.) Jiang

- 2016 "Change of attitude? A diachronic study of stance", *Written Communication* 33, 251-274. <https://doi.org/10.1177/0741088316650399>.
- 2019 *Academic Discourse and Global Publishing: Disciplinary Persuasion in Changing Times*. London: Routledge. <https://doi.org/10.4324/9780429433962>.

Lewin, B. – J. Fine – L. Young

- 2001 *Expository Discourse: A Genre-Based Approach to Social Research Text*. London/New York: Continuum.

Li, Z. – J. Xu

- 2019 "The evolution of research article titles: The case of *Journal of Pragmatics*, 1978-2018", *Scientometrics* 121, 1619-1634. <https://doi.org/10.1007/s11192-019-03244-3>.

Lillis, T. – M. J. Curry

- 2006 "Professional academic writing by multilingual scholars: Interactions with literacy brokers in the production of English-medium texts", *Written Communication* 23 (1), 3-35. <https://doi.org/10.1177/0741088305283754>.
- 2010 *Academic Writing in a Global Context. The Politics and Practices of Publishing in English*. London/New York: Routledge.

Lorés-Sanz, R.

- 2011 "The construction of the author's voice in academic writing: The interplay of cultural and disciplinary factors", *Text & Talk* 31 (2), 173-193. <https://doi.org/10.1515/text.2011.008>.

Luzón, M.-J.

- 2018 "Variation in academic writing practices: The case of reporting verbs in English-medium research articles", *IBERICA* 36, 171-194.

Martinez, R.

- 2018 "'Specially in the last years...': Evidence of ELF and non-native English forms in international journals", *Journal of English for Academic Purposes* 33, 40-52. <https://doi.org/10.1016/j.jeap.2018.01.007>.

Matsuda, P.K. – C. M. Tardy

- 2007 "Voice in academic writing: The rhetorical construction of author identity in blind manuscript review", *English for Specific Purposes* 26 (2), 235-249. <https://doi.org/10.1016/j.esp.2006.10.001>.

Mauranen, A. – C. Pérez-Llantada – J. M. Swales

- 2010 "Academic Englishes: A standardised knowledge?". In: A. Kirkpatrick (ed.) *The Routledge Handbook of World Englishes*. London/New York: Routledge, 659-676.

Molino, A.

- 2010 "Personal and impersonal authorial references: A contrastive study of English and Italian Linguistics research articles", *Journal of English for Academic Purposes* 9 (2), 86-101.
<https://doi.org/10.1016/j.jeap.2010.02.007>.

Moreno, A.

- 2021 "Selling research in RA discussion sections through English and Spanish: An intercultural rhetoric approach", *English for Specific Purposes* 63, 1-17. <https://doi.org/10.1016/j.esp.2021.02.002>.

Mur-Dueñas, P.

- 2007 "'I/we focus on...': A cross-cultural analysis of self-mentions in business management research articles", *Journal of English for Academic Purposes* 6, 143-162. <https://doi.org/10.1016/j.jeap.2007.05.002>.
- 2011 "An intercultural analysis of metadiscourse features in research articles written in English and in Spanish", *Journal of Pragmatics* 43 (12), 3068-3079. <https://doi.org/10.1016/j.pragma.2011.05.002>.

Mur-Dueñas P. – J. Šinkūnienė (eds.)

- 2018 *Intercultural Perspectives on Research Writing*. Amsterdam/Philadelphia: John Benjamins.

O'Donnell, M.

- 2008 "Demonstration of the UAM CorpusTool for text and image annotation". In: *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Demo Session*, 13-16. <https://doi.org/10.5555/1564144.1564148>.

Pérez-Llantada, C.

- 2013 "'Glocal' rhetorical practices in academic writing", *European Journal of English Studies* 17 (3), 251-268.
<https://doi.org/10.1080/13825577.2013.867180>.
- 2014 "Formulaic language in L1 and L2 expert academic writing: Convergent and divergent usage", *Journal of English for Academic Purposes* 14, 84-94. <https://doi.org/10.1016/j.jeap.2014.01.002>.

Sahragard, R. – H. Meihami

- 2016 "A diachronic study on the information provided by the research titles of applied linguistics journals", *Scientometrics* 108, 1315-1331.
<https://doi.org/10.1007/s11192-016-2049-4>.

Samraj, B.

- 2002 "Introductions in research articles: Variations across disciplines", *English for specific purposes* 21 (1), 1-17.
[https://doi.org/10.1016/S0889-4906\(00\)00023-5](https://doi.org/10.1016/S0889-4906(00)00023-5).

- 2016 "Research Articles". In: K. Hyland – P. Shaw (eds.) *The Routledge Handbook of English for Academic Purposes*. Abingdon/New York: Routledge, 405-415. <https://doi.org/10.4324/9781315657455-37>.

Scott, M.

- 2020 *WordSmith Tools version 8*. Stroud: Lexical Analysis Software.

Shaw, P. – I. Vassileva

- 2009 "Co-evolving academic rhetoric across culture; Britain, Bulgaria, Denmark, Germany in the 20th century", *Journal of Pragmatics* 41 (2), 290-305. <https://doi.org/10.1016/j.pragma.2008.07.009>.

Swales, J. M.

- 1990 *Genre Analysis: English in Academic and Research Settings*. Cambridge: Cambridge University Press.
- 2004 *Research Genres. Explorations and Applications*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139524827>.

Swales, J. M. – C. B. Feak

- 2012 *Academic Writing for Graduate Students: Essential Tasks and Skills* (3rd edn.). Ann Arbor: University of Michigan Press.
<http://doi.org/10.3998/mpub.2173936>.

Tang, R. – S. John

- 1999 "The 'I' in identity: Exploring writer identity in student academic writing through the first person pronoun", *English for Specific Purposes* 18, 23-39. [https://doi.org/10.1016/S0889-4906\(99\)00009-5](https://doi.org/10.1016/S0889-4906(99)00009-5).

Thompson, G.

- 2001 "Interaction in academic writing: Learning to argue with the reader", *Applied Linguistics* 22 (1), 58-78. <https://doi.org/10.1093/applin/22.1.58>.

Thompson, P. – S. Hunston

- 2019 *Interdisciplinary Research Discourse. Corpus Investigations into Environment Journals*. London: Routledge.
<https://doi.org/10.4324/9781003009344>.

Vladimirou, D.

- 2007 "'I suggest that we need more research': Personal reference in linguistics journal articles". In: C. Gabrielatos – R. Slessor – J. Unger (eds.) *Papers from the Lancaster University Postgraduate Conference in Linguistics and Language Teaching*. Vol. 1. Lancaster: Lancaster University, 139-157.

Xiang, X. – J. Li

- 2020 "A diachronic comparative study of research article titles in linguistics and literature journals", *Scientometrics* 122, 847-866.
<https://doi.org/10.1007/s11192-019-03329-z>.

Yang, R. – D. Allison

- 2003 "Research articles in applied linguistics: Moving from results to conclusions", *English for Specific Purposes* 22 (4), 365-385.
[https://doi.org/10.1016/S0889-4906\(02\)00026-1](https://doi.org/10.1016/S0889-4906(02)00026-1).

Address: MARINA BONDI, Department of Studies on Language and Culture, University of Modena and Reggio Emilia, Largo S. Eufemia 19, 41121 Modena, Italy.
marina.bondi@unimore.it.

ORCID code: <https://orcid.org/0000-0002-3462-8387>.

Address: JESSICA JANE NOCELLA, Department of Studies on Language and Culture, University of Modena and Reggio Emilia, Largo S. Eufemia 19, 41121 Modena, Italy.
ORCID code: <https://orcid.org/0000-0001-8614-2664>.

German English-medium linguistics journal abstracts over the last 30 years: Quantitative and qualitative structural developments

Marina Ivanova

Chemnitz University of Technology

ABSTRACT

Abstracts are central to academic writing as they summarise and promote publications – this paper shows that the widespread use of abstracts started in the 90s and increased rapidly, becoming a standard. It analyses 593 articles with 555 abstracts from nine linguistics journals, affiliated with the Web of Science to German institutions. The focus is on global rhetorical structures adopted – from introduction/research questions to methods, results, discussions, and conclusions. Additionally, I identify trends in writers' stance expression through selected metadiscourse features as expressed in subjects and verbs. The analyses demonstrate that abstracts from Germany have become more unified towards the scientific IMRAD model. This model, however, has been adapted to the advertising function of the abstract with a stronger emphasis on the authors' contributions and article's importance (especially in introductions, methods, and results) and rare discussions of conclusions and limitations. Thus, general academic writing structures have been adapted to genre-specific functional practices over the last 30 years.

Keywords: journal abstract, metadiscourse, stance, rhetorical structures, IMRAD.

1. Introduction

Abstracts are an indispensable part of journal publications today as they are used to summarise publications and promote them to a specific research community (see Hyland 2000: 64). However, this has not always been the case – abstracts have changed both quantitatively and qualitatively in the

last 30 years. This article aims to show that the use of abstracts in linguistics journals started in the 90s, becoming a standard today. It also explores the genre developments in terms of rhetorical structure and word length.

Although the journal abstract is attached to a research article, it is traditionally viewed as a genre of its own due to its different structure and purpose (Gillaerts – van de Velde 2010: 135; Hyland 2000: 64). The genre characteristics have been systematized in different typologies (Bhatia 1993; Gillaerts 2013; Hyland 2000; Swales – Feak 2009), mostly through the classification into rhetorical moves. Still, notable disciplinary, cultural and diachronic differences have been recognized (see among others Bondi – Lorés Sanz 2014; Friginal – Mustafa 2017). This paper explores abstracts' developments in terms of global rhetorical structures, in particular in the use of the common IMRAD structure (Introduction, Method, Results and Discussion) together with Research Questions, Conclusions and Limitations. IMRAD is a common structure for research articles, which has also been applied to abstracts (e.g., Lorés Sanz 2004).

In order to avoid confounding variables and potential cultural differences, the focus is on abstracts by scholars affiliated with German institutions. The texts are compiled from the Web of Science (WoS) (2022) database and comprise 593 entries from nine high-impact international English-medium linguistics journals. The academic writing discourse is therefore explored in a context where authors want to reach a relatively broad audience and attract readers from their research communities.

After an overview of the general diachronic development (1969-2021), the paper turns to a quantitative and qualitative analysis of moves in 100 randomly sampled abstracts. The distribution of rhetorical structures is viewed in terms of the overall sample, its development throughout the last 30 years, and its distribution in the selected journals. Then, the qualitative analysis highlights small-grain tendencies in move structure. This prompts a discussion of the extent to which the tendencies are explained by the time articles were published and by the article approach.

The paper also turns to the most frequent subjects and verbs and explores how they are used as metadiscourse markers to express stance and attitude. It tests whether authors prefer personal and positive markers, which would be the logical hypothesis for the advertising function of abstracts. Finally, it combines the insights into the global rhetorical structure and the local metadiscourse expressions to draw conclusions on the genre developments of abstracts throughout the last 30 years.

2. Global rhetorical structure of journal abstracts

The global rhetorical structure of genres has most commonly been analysed through a classification into moves. A move can be seen as a rhetorical stage in a genre. It has a minor communicative purpose to fulfill, which in turn serves the major communicative purpose of the genre (Dos Santos 1996: 485).

Several authors have classified the moves of the journal abstract with a different level of detail. Based on Bhatia's summary, Move 1 'Introducing purpose' is used to express aims and (hypo)theses; Move 2 'Describing methodology' indicates the experimental design and scope; Move 3 'Summarizing results' comprises the findings and problem solutions and Move 4 'Presenting Conclusions' interprets the results and provides applications (Bhatia 1993: 148-149).

Dos Santos provides options for submoves and even substeps to these submoves: Move 1 'Situating the research'; Submove 1A 'Stating current knowledge' and/or Submove 1B 'Citing previous research' and/or Submove 1C 'Extended previous research' and/or Submove 2 'Stating a problem' (Dos Santos 1996: 485).

Hyland (2000) presents 'Purpose' as a separate move and frames the results as a 'Product' in his classification of abstract moves: 'Introduction', 'Purpose', 'Method', 'Product' and 'Conclusion' (Hyland 2000: 67). This terminology suits the advertising function of abstracts stressed in his study, which presents the use of moves as strategies for rhetorical persuasion (2000: 68), claiming significance (2000: 75) and insider credibility (2000: 78).

Gillaerts (2013: 52) summarizes abstract moves and intermediate steps as follows: Move 1 'Situating the research' (1a 'Current knowledge', 1b 'Specific references', 1c 'Indicating the gap'); Move 2 'Presenting the research' (2a 'Purpose', 2b 'Research question or hypothesis'); Move 3 'Describing the methodology', Move 4 'Summarizing the findings' and Move 5 'Discussing the research' (5a 'Conclusions', 5b 'Recommendations'). He recognizes several 'methodological issues with the categorization of Moves both in terms of separating overlapping moves and in terms of identifying them in the abstracts.

There are several instructional works on abstract Moves. For instance, Salager-Meyer gives a critical dimension to abstract structure and argues that the move structure is one of the most important features of a well-written abstract and it should include "purpose, methods, results (or data synthesis) and conclusions (optional in case reports)" (Salager-Meyer 1990: 370). These Moves are set in correlation with the overall paper structure. Swales and

Feak's (2009) textbook summarises the abstract moves in a similar manner: Move 1 'Background/ introduction/ situation'; Move 2 'Present research/ purpose'; Move 3 'Methods/ materials/ subjects/ procedures'; Move 4 'Results/ findings'; Move 5 'Discussion/ conclusion/ implications/ recommendations' (Swales – Feak 2009: 5).

Abstracts have also been compared to other genres, for instance by applying the CARS model (Creating A Research Space) designed for introductions (Swales 1990, 2004) to abstracts (Gillaerts 2013; Lorés Sanz 2004). Lorés Sanz (2004) found both examples of CARS and IMRAD structure in linguistics journal abstracts. Comparing abstracts to the research articles, Gillaerts and van de Velde discuss how abstracts have started functioning as "mini articles" (2010: 136).

Overall, the common IMRAD structure seems to be reflected in all studies listed above. Although it compensates with some detail of the move substeps, this is a good trade-off for a qualitative and quantitative analysis of the structure of 100 abstracts. Therefore, for this study, I will focus on the IMRAD structure and investigate how it is used in linguistics abstracts.

Still, it should be noted that it is difficult to make generalisations on the structure of abstracts. Broadly speaking, the hard and soft field scientists are shown to suggest credibility in their abstracts in different ways. While the soft sciences reference the community and discuss background knowledge from the literature, the hard sciences focus more on the methods and procedures (Hyland 2000: 83). The edited volume by Bondi and Lorés Sanz (2014) has provided evidence for other prominent cultural, disciplinary and diachronic variations of abstracts. For instance, comparing English and Italian abstracts, Diani (2014: 83) discerns a correlation between the size of the community and the use of moves – writers to larger audiences with more competition (English) focus more on the results whereas those are often omitted in the Italian sample. The personalisation differences observed by Diani (2014) are also evident in the study by Friginal and Mustafa (2017), who compare US and Iraqi PhD students' research article abstracts. They suggest that non-native speakers with Iraqi background express more distance through passive and non-personal constructions (Friginal – Mustafa 2017). In terms of discipline, Cavalieri (2014) shows that medicine abstracts focus more on the background (situating research move) than linguistics abstracts. Medicine researchers also use more personal patterns than linguists, putting the writer in the foreground (Cavalieri 2014: 174). In terms of diachronic changes in Economics, Linguistics and History journal article abstracts from 1990 to 2010, Bondi (2014) shows an increase in personal markers putting an accent on the author (*we*). The study also demonstrates the growing need to

prove significance and novelty by using modals and evaluative adjectives. All these “voice markers” are shown to underscore the individuality and authority in writers’ contributions (Bondi 2014: 268).

This paper will therefore also limit its focus in terms of culture, discipline and time. It examines the changes in abstract writing by authors with German university affiliations in English-medium linguistic journals during the last 30 years.

3. Corpus and method

3.1 The Web of Science database

The corpus used for this study was compiled from the Web of Science (WoS) online database. The database was filtered for the following journals: *Linguistics*, *Journal of Phonetics*, *Cognitive Linguistics*, *Applied Psycholinguistics*, *Language and Speech*, *Phonetica*, *Linguistic Review*, *Applied Linguistics*, and *Language Learning & Technology (LLT)*. The 15,130 entries were batch exported in a rich CSV format, which also included the available abstracts. The distribution of paper entries per journal correlated with the age of the journal – the journals with the oldest first entry in the WoS database logically had more entries (see Table 1).

The author information section was filtered to include “Germany” as part of the author affiliations. Although this does not exclude authors from non-German background and international multiple-author papers, it can be expected that national academic standards are reflected in the papers and their abstracts. Since this research focuses on broader rhetorical structure, which is likely to be influenced by national conventions, and does not focus on foreign language-specific features like grammar, this limitation should not largely interfere with the validity of the results. It is unfortunately impossible to isolate a true national German style because most scholars have international experience, so research in academic writing is often influenced by the interference of these confounding variables. Still, the 555 abstracts in this study should provide a relatively comprehensive overview of the rhetorical tendencies of researchers from Linguistics departments in Germany. Future studies can attempt to replicate these results by manually filtering the data to include a sample only with works by scholars affiliated with German institutions.

Table 1 presents the distribution of German papers per journal and the proportion of papers to abstracts. The table also describes the whole corpus

for comparison and shows that the German sample accounts for only 5.66% of all entries in the linguistics WoS database. Thus, the WoS database has great potential for future research on journal abstracts from other countries. The WoS database is a representative example of the new technological affordances and societal demands in academic publishing (Schmied et al., this issue) – the database links a vast number of publications from the numerous individual journal portals and thereby allows researchers to get a quick overview of the rapidly growing field.

Table 1. Journals and German abstracts in the collected WoS corpus

Journal	since	N D	% journals D	N abstracts D	% abstracts D	N all	% all	% D / all
<i>Linguistics</i>	1969	200	33.73	171	85.5	2627	25.08	1.91
<i>Journal of Phonetics</i>	1993	111	18.72	111	100	1576	15.05	1.06
<i>Cognitive Linguistics</i>	1998	61	10.29	61	100	1461	13.95	0.58
<i>Language and Speech</i>	1991	59	9.95	59	100	1283	12.25	0.56
<i>Applied Psycho-linguistics</i>	1992	54	9.11	54	100	1127	10.76	0.52
<i>Phonetica</i>	1972	48	8.09	40	83.33	977	9.33	0.46
<i>Linguistic Review</i>	1995	41	6.91	40	97.56	561	5.36	0.39
<i>Applied Linguistics</i>	1992	10	1.69	10	100	473	4.52	0.10
<i>Language Learning & Technology</i>	2003	9	1.52	9	100	389	3.71	0.09
		593		555		10474		5.66

However, one limitation of the WoS database should be noted – a substantial part of existing abstracts has not been compiled, in particular in the journals *Applied Psycholinguistics*, *Language and Speech*, *LLT*, and *Journal of Phonetics*. Some journals like *Phonetics* have included abstracts from their first issue release, but many of them are not included in the database. A lot of these problems arise when publishers provide only PDFs as this complicates

the automatic creation of paper entries. In order to provide a truthful representation of the abstract distribution throughout the years, I manually checked if the missing abstracts exist online and noted this in the corpus. Future studies can use web scraping and character recognition techniques to complete the WoS database.

In order to increase the representativeness of the abstract distribution overview (Section 4.1), I filtered out genres which typically do not have an abstract: Bibliography, Biographical Item, Book Review, Correction, Item About an Individual, Note, Meeting Abstract, Editorial Material, Letter, News Item, Software and Hardware Review. For the whole corpus, these were 10,474 out of 15,130 papers (31%). For the German section, the 137 filtered papers accounted for 19% of the corpus. They had only 5 abstracts, which proves that these genres typically do not have abstracts. The remaining 81% (583 articles and 10 review articles) constituted the 593 texts in the 100,266-word corpus of journal abstracts written by German authors in English (WoS-D).

3.2 Methodology

The whole WoS-D corpus was tagged with Part of Speech (PoS) tags and parsed in terms of dependencies using spaCy (Honnibal et al. 2020) in Python (Van Rossum – Drake 2009). This was done in order to easily determine the most frequent subjects and verbs in the corpus and analyse the distribution of these metadiscourse markers. The `en_core_web_trf` language model used has high accuracy in PoS-tagging (98%) and dependency parsing (94%) (spaCy 2022) which is sufficient for an initial overview of the most widespread subjects and verbs. The tagged data was lemmatised with the *textstem* package (Rinkler 2018) in R (R Core Team 2020) for corpus analysis. Again, the scope of this paper does not allow an analysis of many potentially interesting tendencies in the distributions of parts of speech and dependencies. These can be explored in future studies focusing more on metadiscourse features.

For the manual analysis of moves, a random sample of abstracts ($n = 100$, see Appendix) was drawn with R and manually tagged with INCEpTION (Klie et al. 2018) with active learning assistance from the built-in Sentence Classifier (OpenNLP Document Categorizer). The tags used corresponded to the IMRAD categories: Introduction, Method, Results, and Discussion. In addition, the tags Research Questions (RQ), Conclusion, and Limitations were also used in order to test whether they are part of the abstract genre. The output was converted from XML/XMI to TSV in Python

(Van Rossum – Drake 2009) and analysed with R (R Core Team 2020) and the *tidyverse* packages (Wickham et al. 2019). Finally, a qualitative analysis of the abstracts explored the move distribution in a selection of cases in more detail.

4. Results and discussion

4.1 Overview

The raw distribution of German English papers with and without abstracts in Fig. 1 clearly shows the rise of the abstract in journal articles. The abstract gained popularity in the 90s and has become a standard part of publications. This is also confirmed by Fig. 2, which shows the percentage of German English papers with and without abstracts per year. The error bars show the standard error of each bar and depict how uncertain its information is – the smaller the error bar, the smaller the uncertainty and the higher the reliability. Before the 90s, papers without abstracts have smaller error bars than those with abstracts, which used to be the exception. Then, papers from 1990-1996 all have abstracts but also large error bars because the visualisation is made based on little data. Finally, from 2000 to 2021, the few papers without an abstract have large error bars and the papers with an abstract have small error bars. This shows that the papers with abstracts have become a natural part of journals.

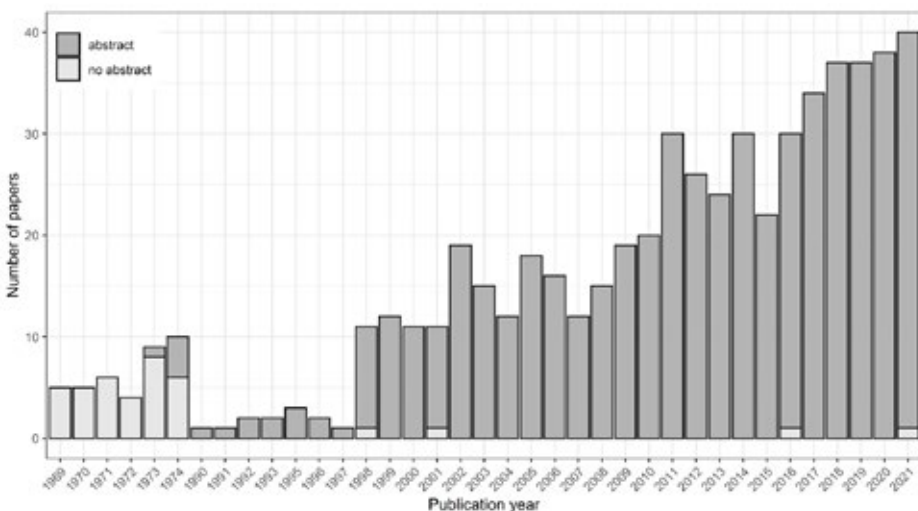


Figure 1. Number of papers with abstracts per year

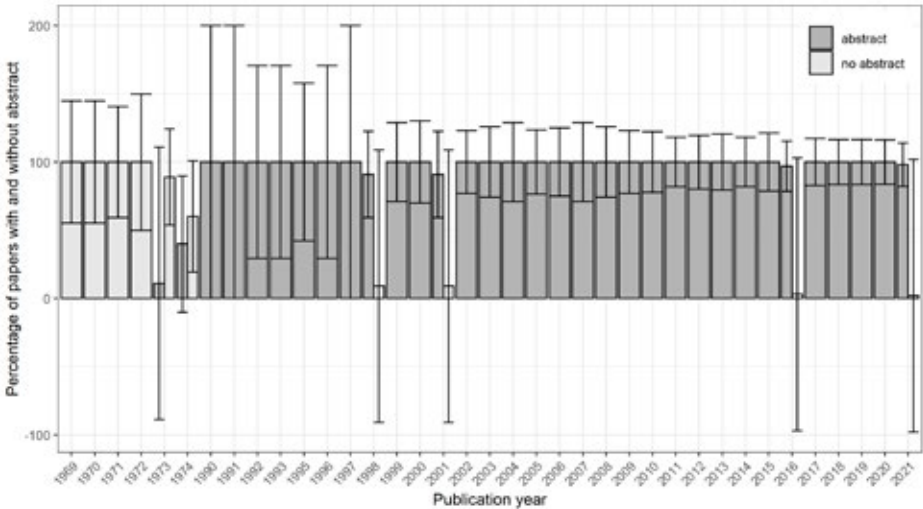


Figure 2. Percentage of papers with abstracts per year

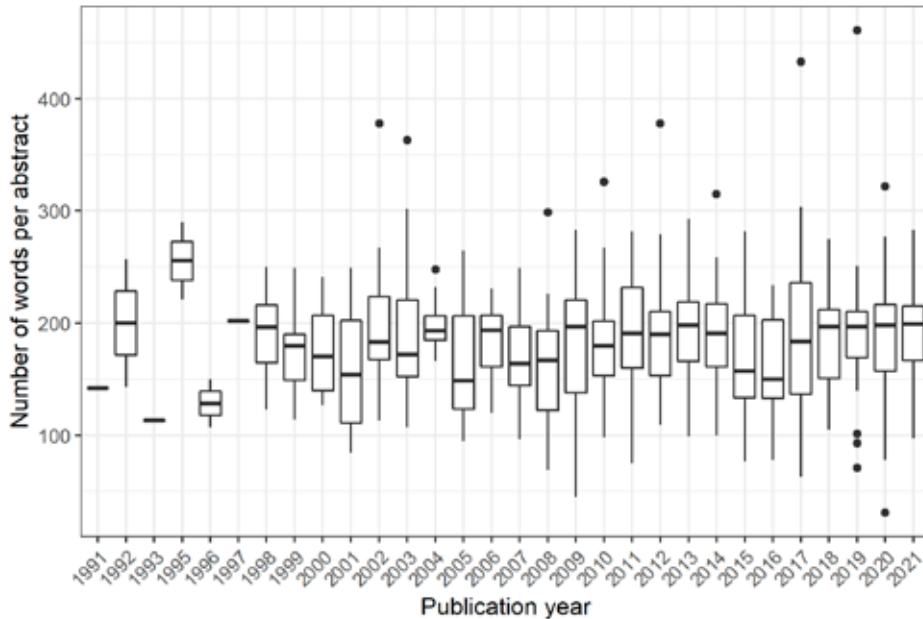


Figure 3. Word count of German English abstracts per year

Fig. 3 visualises the changes in word count of German English abstracts per year. Here it is again obvious that abstracts have become more uniform throughout the years. While in the 90s the median abstract length could

differ between 100 and 200 words, in the last four years, abstracts have had the same median worth length of 200 words. The data also partly reflect the established “increase in the average length of the RA abstracts” (Gillaerts – van de Velde 2010: 136). However, the increasing standardisation in word length is a more prominent observation here.

4.2 Rhetorical structures

4.2.1 Quantitative results

The manual annotation of 100 abstracts with one move per sentence gives a comprehensive overview of the tendencies in abstract global rhetorical structure. Fig. 4 presents the raw frequency distribution of IMRAD moves in the annotated sample (full overview of move annotations in the Appendix).

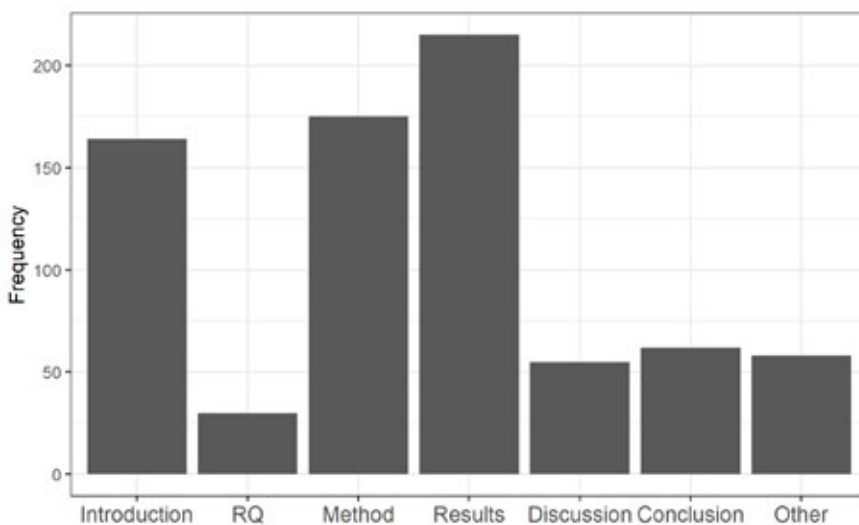


Figure 4. Distribution of Moves in the manually annotated sample (n = 100 abstracts)

The most frequent structure is Results, followed by Methodology and Introduction. This shows that the authors aim to create a research space and address a specific scientific community with the details described in their methodology. Authors summarise the results in the results section but leave limited discussion and conclusion of the findings. There is a nearly equal share of the Discussion and Conclusion moves, which are two of the most infrequent categories. This goes along with the advertising function

of abstracts, i.e. authors use this genre as a teaser to promote their full publications. The category “Other” contains irrelevant text like journal copyright and citations. Research questions or aims have the smallest share on the one hand, because they are mostly implied in the introduction and on the other hand, because of the short abstract length. There were no Limitations sections, which shows that this critical category is reserved for the article. There, authors have an opportunity to clarify the rationale for their decisions, to justify the limitations, and to provide suggestions for future research. The economic size of the abstract does not leave space for these considerations, which would considerably increase the face threat towards the authors and reduce their credibility. Thus, abstracts are left with the summarizing and advertising functions.

Looking at the percentage of IMRAD moves per year in the annotated abstracts (Fig. 5), it becomes evident that the structure of abstracts has become more uniform throughout the years.

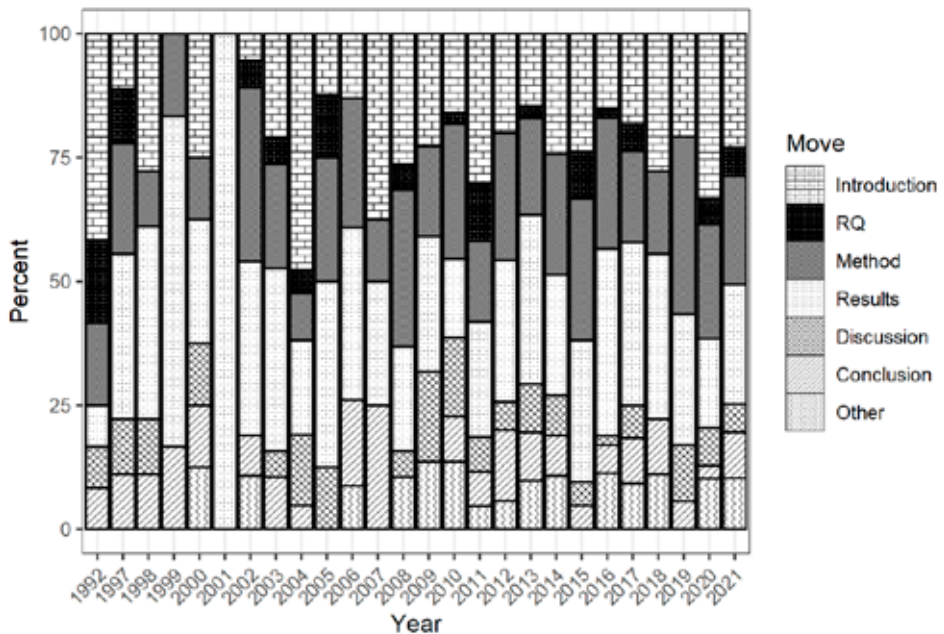


Figure 5. Percentage of Moves per year in the manually annotated sample (n = 100 abstracts)

In the 90s, there are several outliers, e.g. the abstracts from 1999 and 2001 show mostly just Results. However, in the last five years, the moves have had more uniform shares. These results mirror the findings in Hyland

(2000), where moves are shown to become more frequent and more diverse between 1980 and 1997 (Hyland 2000: 82). Similarly, Dontcheva-Navratilova (this issue) also observes a rise of moves in Czech journal articles, showing that academic writing has generally become more complex but also with a standardised, replicable, and thereby easy-to-follow structure. The strong increase in Move 3 “Describing the methodology” in English-medium journal abstracts from *Applied Linguistics* and *English for Specific Purposes* (ESP) observed by Diani (this issue) is not that prominent here, but rather a more uniform distribution of moves throughout time. Still, *Applied Linguistics* accounts for only 10 abstracts in my corpus and two abstracts analyzed qualitatively, and *ESP* was not featured, thus a replication study with more samples could provide interesting insights on journal-specific effects.

Fig. 6 shows a comparison of the moves across journals. Although some of the journals are not strongly represented in the database (e.g. *Applied Linguistics* has only two abstracts in the sample), the charts provide a preliminary overview of some of the rhetorical tendencies in the different journals.

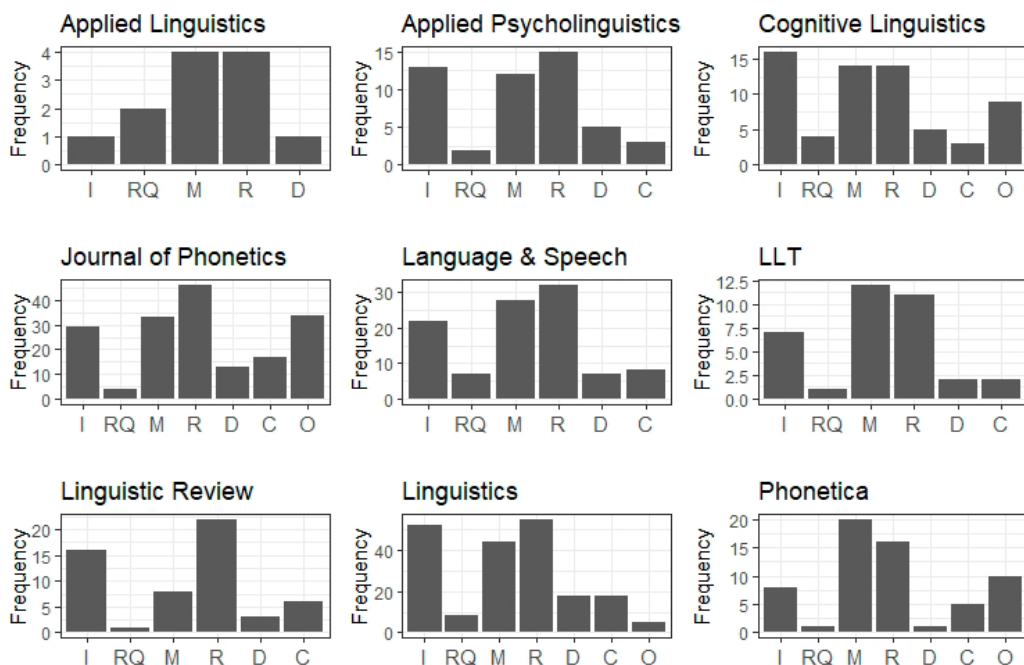


Figure 6. Distribution of moves per journals

For instance, *Linguistic Review* abstracts have the fewest Method moves, perhaps because the journal advertises a focus on theoretical and Generative Grammar articles (*The Linguistic Review* 2022), which already implies some of the methods. Meanwhile, the journals with explicit focus on phonetics (*Journal of Phonetics*, *Language and Speech*, *Phonetica*) share a similar move distribution with few Introduction moves and many Method and Results moves. A similar distribution is evident in *Applied Psycholinguistics* and *LLT* too. *Cognitive Linguistics* and *Linguistics* have the highest shares of Introduction moves, which reveals that their studies need to establish more background knowledge in the field.

4.2.2 Qualitative results

This section looks at qualitative differences between the abstracts throughout the years. As shown in the previous section, Moves have become uniform. For instance, in 2001 there is a *Linguistics* paper abstract which has only Results Moves in its two sentences:

(1)

- a) [Results]: This paper argues that the common denominator of topic-comment (TC) constructions in natural languages is not a single functional feature (e.g. aboutness) but rather the fact that they share some salient semantic attributes with prototypical examples of TC.
- b) [Results]: The paper tries to disentangle these prototypical semantic attributes of TC and then shows, mainly referring to examples from German, that different TC constructions are characterized by different combinations of these attributes and therefore cannot be analyzed properly by unitary theories of TC function.

In contrast, a 2020 abstract from the same journal incorporates all IMRAD moves in its nine sentences:

(2)

- a) [Introduction]: Numerous crosslinguistic studies on motion events have been carried out in investigating the scope of the two-fold typology “path versus manner” (Talmy 1985, 2000) and its possible implications.
- b) [Introduction]: This typological contrast is too narrow as it stands, however, to account for the diversity found both within and across types.

- c) [Method]: The present study is based on what can be termed a process-oriented perspective.
- d) [Method]: It includes the analyses of all relevant conceptual domains notably the domain of temporality, in addition to space, and thus goes beyond previous studies.
- e) [Method]: The languages studied differ typologically as follows: path is typically expressed in the verb in French and Tunisian Arabic in contrast to manner of motion in English and German, while in the temporal domain aspect is expressed grammatically in English and Tunisian Arabic but not in German and French.
- f) [Method]: The study compares the representations which speakers construct when forming a reportable event as a response to video clips showing a series of naturalistic scenes in which an entity moves through space.
- g) [Discussion]: We assume that each of these three cognitive categories is shaped specifically by language structure (both system and repertoire) and language use (frequency of constructions).
- h) [Results]: The findings reveal systematic differences both across, as well as within, typologically related languages with respect to (1) the basic event type encoded, (2) the changes in quality expressed, (3) the total number of path segments encoded per situation, and (4) the number of path segments packaged into one utterance.
- i) [Results]: The findings reveal what can be termed language-specific default settings along each of the conceptual dimensions and their interrelations which function as language specific attentional templates.

One limitation of this annotation method becomes evident from this comparison – annotating with one move per sentence is difficult because there can be several overlapping moves in a sentence. In example 2, the move is labelled “Results”, but it combines information both on the results and the methodology. Move e) clarifies the methodology but also the theoretical background behind it. The methodological issues of move repetition, embedding and identification in genre analysis have also been recognised by Gillaerts (2013: 53).

As can also be seen from the examples above, moves span across several sentences. Many abstracts do not show the linear move structure typical for the sections of papers, but have a shifted order, as in f), g), and h) – Method-

Discussion-Results. There are some examples of abstracts with dynamic moves where the results are interrupted by introduction and method statements for clarification, especially when there are multiple experiments with different methods and results. Dos Santos (1996: 497) similarly identifies differences in the move size (move balance), move blending (move embedding), and move sequence reversal as the main features of abstracts that are used for emphasis of relevant parts and cohesion (Dos Santos 1996: 497). Hyland also observes articles “recycling” moves when several applied methods generate results (Hyland 2000: 69).

As in Gillaerts (2013: 53) considerations, I found it difficult to put some of the moves in definite and unambiguous categories. The Discussion move g) is actually an assumption, but since it is provided with relation to the methods and the results, it was classified as a discussion item.

So far, I have focused on the diachronic changes in structure. However, it should be noted that analyses of global rhetorical structure should also consider the article’s approach. Some recent theoretical, methodological or typological studies like the following 2021 *Linguistics* abstract continue using limited moves:

(3)

- a) [Introduction]: Many researchers seem to think that Construction Grammar posits the existence of only wholly idiosyncratic constructions.
- b) [Introduction]: However, this misconception betrays a deep misunderstanding of the approach because it glosses over the fact that constructions rarely if ever emerge *sui generis*.
- c) [Introduction]: Rather, Construction Grammar aims to balance the fact that some linguistic uses cannot be fully predicted from other well-established uses with the fact that extensions of a construction, while not predictable, are motivated by other senses in the constructional network.
- d) [Method]: This paper illustrates this idea by providing an analysis of the Spanish completive reflexive marker *se*.

The abstract builds up the background and concludes with one move on the case study. Again, d) in example 3 contains overlapping moves and presents a dilemma to the annotator. It combines method and results with a reference to the background (*this idea*), so both labels could fit the sentence. However, it has been labelled “Method” due to the emphasis on the approach. There

are several other recent abstracts with homogeneous moves in the database, which shows that the approach of the article should also be considered. This idea was also discussed with relation to the hard and soft sciences in Hyland (2000: 70). It can be generalized that empirical articles tend to have a clear IMRAD structure while theoretical, methodological and typological articles often omit moves and focus on those rhetorical structures which suit their purpose.

4.3 Selected metadiscourse features

The choice of subjects and verbs can reveal a lot about the communicative goals of the abstracts. Table 2 shows the top 10 subjects and verbs (roots) in the corpus. No diachronic change in the use of these features was observed, which indicates they are core academic discourse features of linguistics abstracts regardless of the publication period. The most frequent subject is the personal pronoun *we*, which expresses authorial agency. The language model has also coded relative clauses (*that*, *which*) as subjects and they are the second most frequent item, showing that abstract sentences can be relatively complex. The expletive construction or personal pronoun *it* is also popular as it allows researchers to use passive voice or to refer to non-human subjects like the frequent nouns *results*, *paper* and *study*. These nouns put agency on an inanimate object and distance the author from the contribution. The first-person pronoun *I* is relatively low on the list with 97 occurrences, showing 1) the predominantly collaborative nature of journal article writing and 2) single authors' preference to deemphasize their individual contribution. These results go along with Bondi's (2014) findings of the growing personalization through the use of self-referential *we* in Linguistics and Economics articles (Bondi 2014: 254, 257). The locational self-referential nouns like *study* and *paper* were also shown to be prominent (Bondi 2014: 254, 257f).

In terms of the verbs, in addition to the verb *be*, the most frequent verbs are *show*, *argue*, and *suggest*. These verbs are typically carriers of *that*-complement clauses, which are a widespread structure in academic writing (Hyland – Tse 2005; Pho 2013). They incorporate extraposition (Hyland – Tse 2005: 42), which allows the author to provide evaluation of the statement in the main clause. For instance, *suggest that* is weaker than *show that* and *claim that* is more negative than *argue that* (Pho 2013). The most frequent verbs here are all positive or strong: *show* (positive strong), *argue* (ambiguous strong), *suggest* (positive weak), and *find* (positive strong). This reveals that authors focus on the advantages of their work and do not address their or other

studies’ limitations in the abstract. This goes along the “linguistic positivity” trend, which suggests that academic writing uses more positively connoted words (Wen – Lei 2021: 20). The top verbs also confirm Bondi’s observation of the widespread use of verbs as acts of topic-setting (*investigate*, here on place 5) and claim-making with growing claim-making verbs (*show*, *argue*, *suggest*, here on place 2-4) (Bondi 2014: 262). Similarly, Schmied (this issue) observed a rise in positive strong verbs like *show* in the recent Master’s theses, confirming the linguistic positivity trend.

Table 2. Top 10 subjects and root in all abstracts of the WoS corpus (100,266 words)

	Subject	n	Verb	n
1	<i>we</i>	539	<i>be</i>	477
2	<i>that</i>	379	<i>show</i>	264
3	<i>it</i>	346	<i>argue</i>	125
4	<i>which</i>	254	<i>suggest</i>	107
5	<i>results</i>	197	<i>investigate</i>	102
6	<i>they</i>	194	<i>find</i>	96
7	<i>paper</i>	159	<i>reserve</i>	92
8	<i>study</i>	155	<i>discuss</i>	90
9	<i>I</i>	97	<i>present</i>	75
10	<i>speakers</i>	93	<i>examine</i>	73

All in all, the subjects and verbs show that authors resort to metadiscourse markers to put or take away focus from their individuality and to express their stance and attitude. Together with the rhetorical moves, authors combine global and local structures to promote their work.

5. Conclusion

Abstracts use their global and local structure to condense and advertise the article they introduce. This paper took entries from linguistics journals in the Web of Science database authored by scholars with affiliations to German research institutions and examined 593 papers with 555 abstracts as well as a sample of 100 abstracts quantitatively and qualitatively. It showed that abstracts have become a standard part of articles after the 90s. Abstracts have also become more uniform both in terms of word length and of IMRAD

structure. This goes along with the findings of Busch-Lauer, who noted that German and English native speakers show more awareness of genre norms and have “adapted to international standards” (Busch-Lauer 2014: 60).

In the field of linguistics, abstracts focus on background knowledge (introduction) but also specify their sample and analysis procedures in the methodology and highlight their results. The move distribution usually follows the IMRAD article structure, but authors often switch, combine, and omit moves in order to suit the purposes of their article or journal approach.

All these strategies aim to promote the study, following the “linguistic positivity” tendency identified in abstracts by previous corpus studies (Cao et al. 2021; Vinkers et al. 2015; Wen – Lei 2021). In this paper, linguistic positivity is demonstrated on two levels: the structural level, where there were no Limitations moves, and the metadiscourse level, where there are no negatively connoted verbs. Limitations are almost obligatory in journal articles but absent in abstracts, which also shows that the IMRAD structure of articles is not completely compatible with the structure of abstracts.

The observed developments in abstract length and structure raise the question of the extent to which they are a result of authorial choices and editorial prescriptions. This is difficult to address due to the occluded nature of peer review. Hyland judges from the variety of patterns in his sample that “how writers use such practices is not *determined* by editorial prescription or genre constraints” (Hyland 2000: 75, emphasis in original). Here it can also be claimed that since abstracts have become uniformly structured across nine journals, each with their own requirements, their structure is more indirectly influenced. Still, future studies can compare the abstracts with the official journal requirements to determine how thoroughly authors comply to them. The study can also be extended to abstracts written by scholars affiliated with research institutions in other countries. Such analyses can gain insights in the impact of editorial and cultural conventions on one of the most popular genres in academic writing.

Overall, German abstracts have become more unified towards the scientific Anglo-American IMRAD model. This general model, however, has some major differences from that of the articles. It has been adapted to the advertising function of the abstract where stronger emphasis is placed on the authors’ contributions and article’s importance (through the introduction, method and results) and limitations are rarely discussed. Thus, over the last 30 years, common academic writing structures like IMRAD have been adapted to fulfil genre-specific functions.

REFERENCES

Primary sources

Web of Science

- 2022 <https://www.webofscience.com/wos/woscc/basic-search>, accessed June 2022.

Secondary sources

Bhatia, V.

- 1993 *Analysing Genre. Language Use in Professional Settings*. London: Longman. <https://doi.org/10.4324/9781315844992>.

Bondi, M. – R. Lorés Sanz

- 2014 “Changing voices: Authorial voice in abstracts”. In: M. Bondi – R. Lorés Sanz (eds.) *Abstracts in Academic Discourse: Variation and Change*. Bern: Peter Lang, 243-269.

Busch-Lauer, I.

- 2014 “Abstracts: Cross-linguistic, disciplinary and intercultural perspectives”. In: M. Bondi– R. Lorés Sanz (eds.) *Abstracts in Academic Discourse: Variation and Change*. Bern: Peter Lang, 43-63.

Cao, X. et al.

- 2021 “Promoting science with linguistic devices: A large-scale study of positive and negative words in academic writing”, *Learned Publishing* 34 (2), 82-88. <https://doi.org/10.1002/leap.1322>.

Cavalieri, S.

- 2014 “Variation across disciplines. The case of Applied Linguistics and Medicine”. In: M. Bondi– R. Lorés Sanz (eds.) *Abstracts in Academic Discourse: Variation and Change*. Bern: Peter Lang, 161-174.

Diani, G.

- 2014 “On English and Italian research article abstracts: Genre variation across cultures”. In: M. Bondi– R. Lorés Sanz (eds.) *Abstracts in Academic Discourse: Variation and Change*. Bern: Peter Lang, 65-83.

Dos Santos, M. B.

- 1996 “The textual organization of research paper abstracts in applied linguistics”, *Text & Talk* 16 (4), 481-499.

Friginal, E. – S. S. Mustafa

- 2017 “A comparison of U.S.-based and Iraqi English research article abstracts using corpora”, *Journal of English for Academic Purposes* 25, 45-57. <https://doi.org/10.1016/j.jeap.2016.11.004>.

Gillaerts, P.

- 2013 “Move analysis of abstracts from a diachronic perspective: A case study”. In: N. – L. Johannesson – G. Melchers – B. Björkman (eds.)

- Of Butterflies and Birds, of Dialects and Genres: Essays in Honour of Philip Shaw*. Stockholm: Acta Universitatis Stockholmiensis, 49-60.
- Gillaerts, P. – F. van de Velde
 2010 "Interactional metadiscourse in research article abstracts," *Journal of English for Academic Purposes* 9 (2), 128-139.
<https://doi.org/10.1016/j.jeap.2010.02.004>.
- Honnibal, M. – I. Montani
 2017 *spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing*.
<https://spacy.io/>, accessed June 2022.
- Hyland, K.
 2000 *Disciplinary Discourses: Social interactions in Academic Writing*. Harlow, UK: Longman.
- Hyland, K. – P. Tse
 2005 "Evaluative *that* constructions: Signalling stance in research abstracts", *Functions of Language* 12 (1), 39-63.
- Klie, J.-C. et al.
 2018 *The INCEpTION Platform: Machine-Assisted and Knowledge-Oriented Interactive Annotation*. In: *Proceedings of System Demonstrations of the 27th International Conference on Computational Linguistics (COLING 2018)*, Santa Fe, New Mexico, USA.
<https://inception-project.github.io/>, accessed June 2022.
- Lorés Sanz, R.
 2004 "On RA abstracts: From rhetorical structure to thematic organisation", *English for Specific Purposes* 23 (3), 280-302.
<https://doi.org/10.1016/j.esp.2003.06.001>.
- Pho, P. D.
 2013 *Authorial Stance in Research Articles: Examples from Applied Linguistics and Educational Technology*. Basingstoke: Palgrave Macmillan.
- R Core Team
 2020 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
<https://www.R-project.org/>, accessed June 2022.
- Rinkler, T.
 2018 *textstem: Tools for Stemming and Lemmatizing Text*. [Version 0.1.4].
<https://cran.r-project.org/web/packages/textstem/index.html>,
 accessed June 2022.
- Salager-Meyer, F.
 1990 "Discoursal flaws in medical English abstracts: A genre analysis per research- and text-type", *Text* 10 (4), 365-384.
<https://doi.org/10.1515/text.1.1990.10.4.365>.
- spaCy
 2022 *Available Trained Pipelines for English*. <https://spacy.io/models/en>,
 accessed June 2022.

Swales, J. M.

1990 *Genre Analysis: English in Academic and Research Settings*. Cambridge: Cambridge University Press.

2004 *Research Genres. Exploration and Applications*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139524827>.

Swales, J. M. – C. B. Feak

2009 *Abstracts and the Writing of Abstracts*. Michigan: University of Michigan Press.

The Linguistic Review

2022 <https://www.degruyter.com/journal/key/tlir/html?lang=en>, accessed June 2022.

Van Rossum, G. – F. L. Drake

2009 *Python 3 Reference Manual*. CreateSpace. <https://dl.acm.org/doi/book/10.5555/1593511>, accessed June 2022.

Vinkers, C. H., et al.

2015 “Use of positive and negative words in scientific PubMed abstracts between 1974 and 2014: Retrospective analysis”, *BMJ (Clinical Research Ed.)* 351, h6467. <https://doi.org/10.1136/bmj.h6467>.

Wen, J. – L. Lei

2021 “Linguistic positivity bias in academic writing: A large-scale diachronic study in life sciences across 50 years”, *Applied Linguistics* 43 (2), 340-364. <https://doi.org/10.1093/applin/amab037>.

Wickham, H. et al.

2019 “Welcome to the Tidyverse”, *Journal of Open Source Software* 4 (43), 1686. <https://doi.org/10.21105/joss.01686>.

APPENDIX

Overview of the random sample for manual analysis (n = 100)

Year	Journal	N	all D	%
1	2	3	4	5
1992	<i>Language and Speech</i>	1	2	50
1997	<i>Language and Speech</i>	1	1	100
1998	<i>Linguistics</i>	1	22	5
1999	<i>Linguistics</i>	1	20	5
2000	<i>Linguistics</i>	1	22	5
2001	<i>Linguistics</i>	1	18	6
2002	<i>Journal of Phonetics, Language and Speech, Phonetica</i>	3	26	12

1	2	3	4	5
2003	<i>Language Learning & Technology, Linguistics</i>	2	22	9
2004	<i>Language Learning & Technology, Linguistics</i>	2	18	11
2005	<i>Cognitive Linguistics, Linguistics</i>	2	22	9
2006	<i>Journal of Phonetics, Language and Speech, Linguistics</i>	3	20	15
2007	<i>Cognitive Linguistics, Linguistic Review</i>	2	13	15
2008	<i>Cognitive Linguistics, Linguistics, Phonetica</i>	3	19	16
2009	<i>Cognitive Linguistics, Journal of Phonetics, Linguistic Review</i>	3	23	13
2010	<i>Journal of Phonetics, Language and Speech, Linguistics, Phonetica</i>	4	22	18
2011	<i>Cognitive Linguistics, Journal of Phonetics, Language Learning & Technology, Linguistic Review, Linguistics, Phonetica</i>	6	30	20
2012	<i>Cognitive Linguistics, Language Learning & Technology, Linguistic Review, Linguistics, Phonetica</i>	5	28	18
2013	<i>Journal of Phonetics, Linguistic Review, Linguistics</i>	3	24	12
2014	<i>Applied Psycholinguistics, Cognitive Linguistics, Journal of Phonetics, Linguistic Review</i>	4	33	12
2015	<i>Language and Speech, Linguistic Review, Linguistics</i>	3	29	10
2016	<i>Applied Linguistics, Journal of Phonetics, Language and Speech, Linguistics, Phonetica</i>	5	34	15
2017	<i>Applied Psycholinguistics, Journal of Phonetics, Language and Speech, Linguistic Review, Linguistics</i>	5	36	14
2018	<i>Journal of Phonetics, Linguistics</i>	2	37	5
2019	<i>Applied Psycholinguistics, Language and Speech, Language Learning & Technology, Linguistics</i>	4	37	11
2020	<i>Applied Linguistics, Cognitive Linguistics, Journal of Phonetics, Linguistics</i>	4	40	10
2021	<i>Applied Psycholinguistics, Cognitive Linguistics, Journal of Phonetics, Language & Speech, Linguistic Review, Linguistics, Phonetica</i>	7	42	17

Overview of the manually annotated moves per abstract (n = 100)

Abs ID	I	M	RQ	R	D	C	Other
1	2	3	4	5	6	7	8
1	1	1	NA	2	NA	NA	NA
10	2	1	NA	3	NA	1	NA

1	2	3	4	5	6	7	8
100	2	4	NA	3	NA	NA	2
11	1	5	NA	2	1	NA	NA
12	2	NA	NA	2	NA	1	NA
13	NA	3	1	3	NA	NA	NA
14	1	2	NA	1	1	NA	7
15	1	2	NA	3	NA	1	NA
16	5	NA	1	1	NA	NA	NA
17	2	NA	NA	1	NA	1	NA
18	2	2	1	1	2	NA	NA
19	1	1	2	1	2	3	2
2	2	1	NA	3	2	NA	2
20	2	1	1	NA	1	1	NA
21	1	2	1	3	1	1	NA
22	2	2	NA	4	NA	1	2
23	1	NA	3	3	NA	1	3
24	1	3	NA	2	NA	1	2
25	1	NA	NA	3	NA	1	NA
26	2	2	1	1	NA	NA	NA
27	3	NA	2	1	NA	NA	NA
28	2	1	1	1	NA	1	NA
29	2	3	NA	2	NA	1	2
3	2	NA	NA	1	1	NA	NA
30	2	1	NA	3	NA	1	2
31	1	NA	NA	1	1	NA	NA
32	2	2	NA	3	1	1	NA
33	2	4	NA	3	NA	2	2
34	5	2	2	1	1	1	NA
35	4	1	NA	4	NA	1	NA
36	1	NA	1	4	NA	NA	NA
37	2	3	NA	1	NA	NA	NA
38	3	1	NA	NA	NA	NA	NA
39	4	1	NA	NA	NA	1	NA
4	2	2	1	1	NA	NA	NA
40	2	2	1	2	1	NA	NA
41	1	1	NA	1	NA	NA	NA

1	2	3	4	5	6	7	8
42	NA	1	NA	4	NA	1	NA
43	1	3	NA	1	NA	1	NA
44	NA	1	NA	4	1	2	NA
45	3	1	NA	2	NA	1	NA
46	NA	2	NA	3	NA	NA	2
47	3	4	1	1	NA	NA	NA
48	NA	2	NA	2	NA	NA	2
49	2	4	NA	2	1	NA	NA
5	1	3	NA	1	1	NA	NA
50	2	2	NA	1	1	NA	NA
51	NA	1	NA	1	NA	NA	2
52	NA	4	NA	1	2	NA	NA
53	1	3	1	4	NA	1	2
54	2	1	NA	1	1	1	2
55	3	2	NA	7	NA	NA	NA
56	2	2	NA	3	NA	NA	NA
57	1	2	NA	3	2	NA	2
58	3	1	NA	2	1	NA	1
59	1	2	NA	2	1	NA	NA
6	NA	NA	NA	2	NA	NA	NA
60	2	1	NA	1	1	1	NA
61	1	1	NA	1	1	NA	NA
62	1	1	1	4	NA	1	NA
63	2	4	NA	2	NA	2	NA
64	3	2	NA	1	1	1	NA
65	4	1	NA	1	NA	1	1
66	NA	2	NA	4	NA	1	NA
67	2	NA	NA	6	1	1	NA
68	2	1	NA	3	NA	2	2
69	NA	2	NA	2	1	NA	NA
7	2	2	NA	1	NA	2	NA
70	1	1	NA	2	NA	NA	NA
71	NA	1	NA	4	NA	1	NA
72	NA	4	NA	2	3	1	NA
73	2	2	NA	3	2	NA	NA

1	2	3	4	5	6	7	8
74	4	1	NA	3	NA	1	NA
75	4	2	NA	NA	3	NA	NA
76	1	3	NA	1	1	1	NA
77	NA	2	NA	2	1	1	NA
78	2	2	NA	1	1	NA	NA
79	NA	4	NA	6	NA	2	2
8	NA	4	1	4	NA	NA	2
80	1	6	NA	5	NA	2	NA
81	NA	3	1	3	1	1	NA
82	2	1	NA	2	1	1	1
83	3	2	NA	2	1	1	2
84	NA	1	NA	3	1	1	NA
85	1	1	NA	1	NA	1	NA
86	2	2	NA	5	NA	2	2
87	NA	1	1	1	1	NA	NA
88	3	1	1	2	1	NA	2
89	2	1	NA	1	NA	NA	NA
9	3	3	NA	3	1	NA	NA
90	1	3	NA	1	NA	NA	NA
91	4	1	NA	5	NA	NA	NA
92	1	1	1	1	1	NA	NA
93	NA	NA	NA	NA	NA	NA	1
94	1	2	NA	1	NA	1	NA
95	4	1	NA	1	NA	NA	2
96	2	1	NA	1	NA	1	NA
97	1	2	NA	1	4	2	2
98	1	1	NA	1	1	NA	NA
99	2	1	3	3	NA	1	NA

Address: MARINA IVANOVA, English and American Studies, Chemnitz University of Technology, 09107 Germany. marina.ivanova@phil.tu-chemnitz.de.

ORCID code: <https://orcid.org/0000-0001-7414-6668>.

Research article abstracts in English and Italian: Generic and cross-linguistic variation over the last 20 years

Giuliana Diani

University of Modena and Reggio Emilia

ABSTRACT

The paper examines comparatively and diachronically the generic structure of research article abstracts written in English for international scientific journals and abstracts written in Italian and published in Italian journals in the field of linguistics, with the aim of exploring generic and cross-linguistic variation over the last 20 years. The data consist of two small corpora of English and Italian linguistics research article abstracts spaced at ten-year intervals, namely 1997, 2007 and 2017. The analysis shows that the generic structure and rhetorical organisation of abstracts written in Italian conforms to the international conventions based on the norms of the English academic discourse community. However, they are rhetorically less complex than English abstracts. Diachronic variation has also been observed in the frequency and distribution of every single move across the two language corpora over the decades.

Keywords: research article abstracts, genre analysis, cross-linguistic (English and Italian), diachronic variation.

1. Introduction

The importance of research article (RA) abstracts in academic writing is widely acknowledged and has generated a significant number of studies over the last 20 years. Early scholarly research focused on its textual structure. The most representative contributor in this field is Bhatia (1993: 78-79), who identifies four moves, reflecting Swales' (1990) Introduction-Method-Results-Discussion (IMRD) research article structure: 1) "Introducing purpose";

2) “Describing methodology”; 3) “Summarizing results”; 4) “Presenting conclusions”. Compared to the four-move model, Dos Santos (1996: 484-490) postulates a fifth move, preceding all the others: “Situating the research”. Similar to Dos Santos, Hyland (2000) suggests that an ‘Introduction’ move is needed, where the context of the paper and research motivation are provided. The move referred to as “Introduction” and “Situating the research” by Hyland and Dos Santos, respectively, provides “a justificatory context for the research”, to be differentiated from the “Purpose” move, which “outlines the intention behind the paper” (Hyland 2000: 67).

Considerable research interest has been invested in tracking disciplinary variation in the textual structure of RA abstracts. Attention has been paid to abstracts written in English within single disciplines. Salager-Meyer (1990), for example, investigated medical abstracts and found that only 62% contained Bhatia’s four-move model, whereas Anderson and Maclean (1997) identified five moves in the medical abstract: Background, Purpose, Methods, Results and Conclusion. Huckin (2001) found that biomedical abstracts often exclude the Purpose move. Lorés Sanz’s (2004) analysis of abstracts from linguistics journals revealed two major types of rhetorical organisation: the IMRD structure and the CARS pattern (Create a Research Space, Swales 1990). Samraj’s (2005: 152) study of abstracts from biology showed that they “share a number of moves, which have been traditionally ascribed to this genre, namely, a statement of purpose, results, and conclusions”, but that they “fail to mention the methods employed in the study, implying that this genre is not a simple summary of the full-length article”. In her study of academic law abstracts, Breeze (2009: 13-14) pointed out that their macrostructure differs from “abstracts in other areas, reflecting the priorities of legal academic culture”. For example, she noted that the prominence of the issue statement, as she calls it, “which acts as the principal ‘hook’ to attract the reader, appears to be peculiar to legal academia, appealing to a discourse community in which the ‘issue’ is the crucial focus of any discussion”. Tankó (2017) found that RA abstracts in the field of literature have a non-hierarchical eight-move structure (Outcome, Purpose, Background, Method, Topic, Conclusion, Niche, Implications) with four stable moves, whose functions are to present the background, purpose, methodology and outcomes of the research.

Research interest has also been extended to cross-disciplinary variation. Pho (2008), for example, focused on abstracts written in English in the fields of applied linguistics and educational technology. Cavalieri (2014) investigated the move structure of abstracts in English across applied linguistics and medicine. Hatzitheodorou (2014) compared law with business. Further

research has been devoted to cross-linguistic variation. There are papers comparing English with Spanish (Martín-Martín 2003, 2005; Lorés Sanz 2009; Sandoval Cruz 2015), French (Van Bonn – Swales 2007; Cavalieri – Preite 2017), Italian (Diani 2014; Cavalieri – Diani 2018), Portuguese (Johns 1992; Feltrim et al. 2003), German (Busch-Lauer 1995), Swedish (Melander et al. 1997), Arabic (Alharbi – Swales 2011), and Chinese (Duan – Wei 2021).

Parallel to the synchronic studies discussed above, generic variation in RA abstracts has also been investigated from a diachronic perspective. Scholarly investigation has mainly focused on changes in the textual structure of abstracts in English within single disciplines. Atkinson (1992) offered a detailed analysis of the evolution of medical papers, including the abstract, published in the *Edinburgh Medical Journal* between 1735 and 1985. Bondi's (1997) study traced the development of abstracts in economics in the 20th century. Bondi and Cavalieri (2012) investigated the changes in the generic structure of the abstract in the field of applied linguistics over a time span of twenty years (1988-2008). Similarly, Gillaerts (2013) examined the move structure of abstracts from the *Applied Linguistics* journal from 1987 to 2007. Although these previous studies offer invaluable insights into diachronic variation across disciplines, the RA abstract may also display diachronic changes in its rhetorical strategies across languages. Yet, relevant research is still lacking.

This paper is meant to contribute to filling this gap, by examining comparatively and diachronically the generic structure of RA abstracts written in English for international scientific journals and abstracts written in Italian and published in Italian journals in the field of linguistics, with the aim of exploring generic and cross-linguistic variation in the rhetorical preferences that characterise the members of the international and Italian scientific communities in this domain over the last 20 years. The context for this analysis is provided by a previous study by the author (Diani 2014), which looked at the rhetorical organisation of English and Italian linguistics RA abstracts published in 2007 in two leading journals of Italian and English linguistics. The purpose now is to extend the analysis to English and Italian RA abstracts published in their 1997 and 2017 issues. To this end, the paper aims to answer the following research questions:

1. What generic variation can be observed in the English and Italian linguistics RA abstracts over the last 20 years?
2. What cross-linguistic variation can be identified in the moves of the English and Italian abstracts over the decades?

The next Section provides a short description of the materials used for the study as well as the method adopted. A diachronic overview of the move structure of English and Italian linguistics RA abstracts is provided in Section 3. Section 4 presents the results emerging from the comparative analysis of each single move across the two language corpora over the decades under investigation. The paper concludes with a summary of the findings (Section 5).

2. Materials and methods

The present study is carried out on two small corpora of English and Italian RA abstracts taken from two international and three Italian journals in the field of linguistics spaced at ten-year intervals over the past 20 years: 1997, 2007 and 2017. The 2007 English and Italian corpora include abstracts of a previous work (Diani 2014). The compilation of the 2007 Italian corpus was conditioned by the fact that most Italian linguistics journals publish abstracts written in English, although the associated papers are written in Italian. This is a policy practised by many journal editorials who require abstracts in English for papers in other languages (Ventola 1994a, 1994b). Of the twelve leading journals of Italian linguistics I consulted, only two were found to publish abstracts in Italian fairly regularly and the associated papers are written in English, French or German: *Linguistica e Filologia* and *Rassegna Italiana di Linguistica Applicata*.¹ From these, I took all the abstracts available from the 2007 issue, for a total of 34 (about 5,700 words). The same criteria were followed for the compilation of the 2017 Italian corpus. Since neither of the two journals published abstracts in Italian in their 1997 issues, this decade was excluded from the analysis. Since the 2017 *Linguistica e Filologia* issue only published one abstract in Italian, I created the 2017 Italian corpus by taking abstracts from *Rassegna Italiana di Linguistica Applicata* and *Italiano LinguaDue*, founded only in 2009. The corpus consists of 35 abstracts (5,861 words). To compile the English corpus, two leading international journals were selected: *Applied Linguistics* and *English for Specific Purposes*. The English corpus consists of 148 abstracts (28,937 words), divided into three subsets, comprising all the abstracts available in the 1997,

¹ The Italian linguistics journals considered are: *Annali d'Italianistica*, *Archivio Glottologico Italiano*, *Bollettino d'Italianistica*, *Itals/Didattica e Linguistica dell'Italiano come Lingua Straniera*, *Lingue e Linguaggio*, *Linguistica e Filologia*, *Rassegna Italiana di Linguistica Applicata*, *Rivista Italiana di Dialettologia*, *Rivista di Linguistica/Italian Journal of Linguistics*, *Studi e Saggi Linguistici*, *Studi di Lessicografia Italiana*, *Studi di Grammatica Italiana*.

2007 and 2017 issues: 39, 47, 62, respectively. Tables 1 and 2 give an overview of the English and Italian corpus characteristics.

Table 1. The English corpus characteristics over the last 20 years

English Journals	1997	2007	2017	Total
<i>Applied Linguistics</i>	20	22	35	
<i>English for Specific Purposes</i>	19	25	27	
Total number of abstracts	39	47	62	
Total number of words	7,067	10,500	11,370	28,937

Table 2. The Italian corpus characteristics over the last 10 years

Italian Journals	2007		2017	Total
<i>Linguistica e Filologia</i>	14	<i>Italiano LinguaDue</i>	29	
<i>Rassegna Italiana di Linguistica Applicata</i>	20	<i>Rassegna Italiana di Linguistica Applicata</i>	6	
Total number of abstracts	34	Total number of abstracts	35	
Total number of words	5,700	Total number of words	5,861	11,561

An analysis of the move structure of the English and Italian linguistics abstracts was carried out by following previous work that has examined English RA abstracts in applied linguistics based on a modified version of Bhatia's (1993) four-move model for abstracts (Purpose-Method-Results-Conclusion). More specifically I adopted Dos Santos's (1996: 481) five-move model for the structure of abstracts in applied linguistics, which includes:

- Move 1, 'Situating the research', realised by 'Stating current knowledge' (Submove 1A), and/or 'Citing previous research' (Submove 1B), and/or 'Extending previous research' (Submove 1C), and/or 'Stating a problem' (Submove 2);
- Move 2, 'Presenting the research', realised by 'Indicating main features' (Submove 1A), and/or 'Indicating main purpose' (Submove 1B), and/or outlining hypotheses or questions, 'Hypothesis raising' (Submove 2);
- Move 3, 'Describing the methodology' ("describes the study design");
- Move 4, 'Summarizing the results' ("states the major findings");
- Move 5, 'Discussing the research' realised by 'Drawing conclusions' (Submove 1) and/or 'Giving recommendations' (Submove 2).

In order to conduct a move-structure identification and classification of the abstracts, I first manually tried to associate each sentence with the

communicative functions corresponding to the moves and submoves characterising Dos Santos's model. I then proceeded to see, in a preliminary analysis, whether these moves were present in both the English and Italian subsets and track cross-linguistic variation and change over the decades.

3. Move analysis in English and Italian linguistics RA abstracts: A diachronic overview

The move analysis reveals that the five-move abstract model, as described by Dos Santos (1996), is rare in the two diachronic language corpora analysed. In fact, only two abstracts contain all five moves in the Italian corpus: one in 2007 and one in 2017, respectively. This is also the case for the English subsets, although there is a slightly greater frequency in 2017, in which 11 abstracts contain the five moves compared to 2007 and 1997, when only one and three used all five moves, respectively. Overall, the results here echo those of Hyland (2000), who found that the presence of all five moves is quite rare in humanities abstracts.

The findings show similarities in frequency distribution of these moves across the diachronic corpora, although statistical differences appear to exist. If we take an overview to the distribution of the moves in the two diachronic language corpora (see Tables 3 and 4), undoubtedly the most striking (albeit predictable) feature is the predominance of the 'Presenting the research' move in both languages over the periods under investigation. As for the English data, the overall occurrence of the move demonstrates a steady increase in its use over the last 20 years, with a 6% rise between 1997 and 2017. Similarly, in the Italian subsets, the percentage rises between 2007 and 2017 (94.11% vs 100%). Another level of similarity can be observed in the frequency distribution of the 'Describing the methodology' move in both languages in the 2007 subsets, occurring in 44.68% of the English abstracts and in 38.23% of the Italian abstracts. Diachronically, on the contrary, there is a substantial increase of 37% in the use of the move from 2007 to 2017 in the English data, rising from 44.68% to 81.45% in 2017, whereas the 2017 Italian subset shows no significant variation: the frequency amounts to 40% compared to 38.23% in 2007.

However, the analysis points to diachronic variation with respect to the other moves. A significant difference exists in the frequency of occurrence of the 'Summarizing the results' move in both languages over the decades. The Italian abstracts show a much lower frequency of this move with 52.94%

and 51.42% in the 2007 and 2017 subsets compared to 80.85% and 100% in the 2007 and 2017 abstracts in English. Interestingly, when focusing on the English data, we see a substantial increase in the use of the move over the last 20 years, with a massive 31% rise between 1997 and 2017. Yet variation is particularly evident in the ‘Discussing the research’ move. This move only appears in 8.82% and 11.42% of the 2007 and 2017 Italian subsets in contrast to 48.93% and 35.48% of the 2007 and 2017 English abstracts. Even though a discussion of the research is an important aspect of a RA, it is omitted in the Italian linguistics abstract, suggesting that Italian linguists are less inclined to anticipate what conclusions can be drawn from their study. As regards the ‘Situating the research’ move, its frequency of occurrence in the three English subsets does not emerge as being significantly different over the last 20 years (56.41% in 1997, 53.19% in 2007, 56.45% in 2017). On the contrary, we see a decrease in the use of the move in the Italian abstracts between 2007 and 2017 (61.76% in 2007 as opposed to 48.57% in 2017).

Table 3. Frequency distribution of moves in English abstracts over the last 20 years

Moves	EN 1997 (n = 39) %	EN 2007 (n = 47) %	EN 2017 (n = 62) %
Move 1 Situating the research	56.41	53.19	56.45
Move 2 Presenting the research	92.30	93.61	98.38
Move 3 Describing the methodology	41.02	44.68	81.45
Move 4 Summarizing the results	69.23	80.85	100.00
Move 5 Discussing the research	43.58	48.93	35.48

Table 4. Frequency distribution of moves in Italian abstracts over the last 10 years

Moves	IT 2007 (n = 34) %	IT 2017 (n = 35) %
Move 1 Situating the research	61.76	48.57
Move 2 Presenting the research	94.11	100.00
Move 3 Describing the methodology	38.23	40.00
Move 4 Summarizing the results	52.94	51.42
Move 5 Discussing the research	8.82	11.42

The analysis of data suggests that there are differences in the preferences in move sequences across the two diachronic language corpora.

Table 5. Frequency distribution of move sequences in English abstracts over the last 20 years

Move sequences EN 1997	(n = 39) %	Move sequences EN 2007	(n = 47) %	Move sequences EN 2017	(n = 62) %
Pr-M-R	12.82	Pr-M-R	17.02	Pr-M-R	22.58
Sr-Pr-R-D	5.12	Sr-Pr-R-D	14.89	Sr-Pr-R-D	3.22
Sr-Pr-R	12.82	Sr-Pr-R	14.89	Sr-Pr-R	4.83
Pr-R-D	10.25	Pr-R-D	10.63	Pr-R-D	3.22
Pr-M-R-D	12.82	Pr-M-R-D	10.63	Pr-M-R-D	11.29
Sr-Pr-M-R	7.69	Sr-Pr-M-R	8.51	Sr-Pr-M-R	25.80
Sr-Pr-D	7.69	Sr-Pr-D	6.38	Sr-Pr-D	0.00
Sr-Pr	15.38	Sr-Pr	4.25	Sr-Pr	1.61
Pr-M-D	0.00	Pr-M-D	4.25	Pr-M-D	0.00
Pr-R	0.00	Pr-R	0.00	Pr-R	8.06
Sr-Pr-M-R-D	7.69	Sr-Pr-M-R-D	2.12	Sr-Pr-M-R-D	17.74
Sr-M-R	0.00	Sr-M-R	2.12	Sr-M-R	3.22

Sr = Situating the research; Pr = Presenting the research; M = Describing the methodology; R = Summarizing the results; D = Discussing the research

Table 6. Frequency distribution of move sequences in Italian abstracts over the last 10 years

Move sequences IT 2007	(n = 34) %	Move sequences IT 2017	(n = 35) %
Sr-Pr	32.35	Sr-Pr	31.42
Pr-R	14.70	Pr-R	8.57
Pr-M-R	11.76	Pr-M-R	20.00
Sr-Pr-R	11.76	Sr-Pr-R	11.42
Sr-M-R	5.88	Sr-M-R	0.00
Sr-Pr-M-R	5.88	Sr-Pr-M-R	2.85
Pr-M-D	5.88	Pr-M-D	0.00
Pr	2.94	Pr	5.71
Sr-Pr-M-R-D	2.94	Sr-Pr-M-R-D	2.85
Pr-M	2.94	Pr-M	8.57
Sr-Pr-M	2.94	Sr-Pr-M	0.00
Pr-R-D	0.00	Pr-R-D	2.85
Pr-M-R-D	0.00	Pr-M-R-D	5.71

Although the results in Tables 5 and 6 show that the move structure Presenting the research-Describing the methodology-Summarizing the results (Pr-M-R) shows a steady increase in both language corpora across the periods (EN 12.82% in 1997, 17.02% in 2007, 22.58% in 2017; IT 11.76% in 2007, 20% in 2017), the corpora display preferences for different move sequences. In the English abstracts a general preference is for Situating the research-Presenting the research-Describing the methodology-Summarizing the results (Sr-Pr-M-R), showing a substantial increase of 18% over the last 20 years, rising to 25.80% in the 2017 data, and also initially for Presenting the research-Describing the methodology-Summarizing the results-Discussing the research (Pr-M-R-D), despite the overall decline from 1997 to 2017 (from 12.82% in 1997 to 10.63% and 11.29% in 2007 and 2017, respectively). In the Italian data, on the contrary, the tendency is for a two-move abstract Situating the research-Presenting the research (Sr-Pr), with frequencies that have been fairly uniform over the last 20 years (32.35% and 31.42% in 2007 and 2017, respectively). Like the Italian data, the sequence Situating the research-Presenting the research is present in the English corpora, but it tends to be considerably lower in terms of frequency (4.25% in 2007 and 1.61% in 2017). This initial examination indicates that the Italian abstracts are rhetorically less complex than the English ones over the last 20 years. They only present some of the basic structural moves which constitute the different sections of the underlying RA, whereas the English abstracts reflect these moves more closely.

I now examine the five moves in order to explore whether their use and frequency have changed over the past 20 years.

4. Detailed analysis of moves in English and Italian linguistics RA abstracts: A diachronic comparison

4.1 Move 1 'Situating the research'

As Table 4 shows, this move is the second most frequent move in the 2007 Italian corpus (61.76%), while it ranks as the third in the 2017 subset (48.57%). As regards the three English subsets (see Table 3), in both the 1997 and 2007 English data it ranks as the third most frequent move (56.41% and 53.19%, respectively), while it ranks the fourth in the 2017 corpus (56.45%). In spite of these different distributions, the writers of the English and Italian abstracts use the same submoves connected to the move, as described by Dos Santos (1996), and no instances of submoves outside the parameters of this model are found in the samples analysed.

Table 7. Frequency distribution of submoves in the 'Situating the research' move in English abstracts over the last 20 years

Move 1 – Situating the research	EN 1997 (n = 39) %	EN 2007 (n = 47) %	EN 2017 (n = 62) %
Submove 1 A – Stating current knowledge and/or	68.18	44.00	57.14
Submove 1 B – Citing previous research and/or	9.09	4.00	5.71
Submove 1 C – Extended previous research and/or	0.00	0.00	0.00
Submove 2 – Stating a problem	9.09	16.00	8.57
Merging of Submove 1 with Submove 2	13.63	36.00	28.57

Table 8. Frequency distribution of submoves in the 'Situating the research' move in Italian abstracts over the last 10 years

Move 1 – Situating the research	IT 2007 (n = 34) %	IT 2017 (n = 35) %
Submove 1 A – Stating current knowledge and/or	76.19	88.23
Submove 1 B – Citing previous research and/or	0.00	0.00
Submove 1 C – Extended previous research and/or	0.00	0.00
Submove 2 – Stating a problem	0.00	0.00
Merging of Submove 1 with Submove 2	23.81	11.76

As can be seen in Tables 7 and 8, for both groups of linguists, the move is most commonly realised through submove 1A 'Stating current knowledge' to identify their research field and topic over the 20-year period. In Italian, this is present in 76.19% of the cases in 2007 and in 88.23% in 2017. In English, the overall occurrence of submove 1A demonstrates a fluctuating trend through the three subsets. It falls significantly from 1997 to 2007 (from 68.18% to 44%, respectively) and then rises in 2017, amounting to 57.14%. Examples from both language corpora are given in (1) and (2).

- (1) The validity of instruments appropriate for eliciting L1 (first language) and L2 (second language) speech act data has been widely debated.

Over the past decade, investigations devoted to establishing L2 learners' socio-pragmatic competence and performance, discourse completion tests (DCTs) have become increasingly common as instruments for comparative analyses of native speaker (NS) and non-native speaker (NNS) socio-pragmatic behaviors. (EN 1997)

- (2) La combinazione italiano-tedesco è stata nel 1983 al centro di uno degli studi più noti, ma in seguito più discussi nell'ambito della ricerca sul bilinguismo precoce. (IT 2007)

[In 1983 the Italian-German combination attracted much attention from scholars interested in research on bilingualism at an early stage.]

For the other submoves in the three English subsets, low frequencies of submove 1B 'Citing previous research' and submove 2 'Stating a problem' are found over the period as a means of realising Move 1, as (3) and (4) illustrate.²

- (3) The introductory it pattern, as in it is important to remember the differences, is a functionally diverse pattern of great importance to academic discourse. *However, previous studies have found that* using the pattern in an expert-like manner can prove challenging in particular for non-native-speaker students. (EN 2017)
- (4) *Despite the impact of the ESP genre-based framework of teaching discipline-specific writing to L2 learners, especially to L2 graduate students,* the writing performance of learners in such a framework is still not fully explored. (EN 2007)

The overall occurrences of both submoves show a fluctuating trend through the three English subsets. 1B falls from 1997 (9.09%) to 2007 (4%) and then rises in 2017 (5.71%). Submove 2, on the contrary, rises from 1997 (9.09%) to 2007 (16%) and then drops in 2017 (8.57%).

An interesting point to note is that there is a tendency in both languages to merge submove 1A 'Stating current knowledge' with submove 2 'Stating a problem', as a means of realising Move 1, in order to indicate that the research field under investigation has not been thoroughly successful or complete. In 2007, this rhetorical strategy is reported in 36% of the English cases and in 23.81% of the instances in Italian. But it records a decrease in its use in both languages in 2017, totalling 28.57% and 11.76%, respectively.

² For purposes of illustration, examples of the linguistic exponents are italicised.

All the examples of this move begin with an adversative sentence, mainly introduced by the connectors *however*, *but* and its Italian equivalent *tuttavia*. Two representative examples are given in (5) and (6).

- (5) Content and Language Integrated Learning (CLIL) programmes are burgeoning in European school contexts due to the widespread belief that they help to significantly improve foreign language learning while content learning is not negatively affected. *However, some of its purported benefits still need to be empirically confirmed.* (EN 2017)
- (6) Visto il suo ruolo centrale nella comunicazione, l'abilità interattiva nell'uso e apprendimento della lingua 2 è stata ampiamente rivalorizzata dal CEF (2001:14). *Tuttavia, lo spoken interaction spesso non ottiene il focus che merita nei curricula TESOL [...].* (IT 2007)
[Because of its fundamental role in communication, ability in second language acquisition has been widely reconsidered by CEF (2001:14). However, spoken interaction is often ignored in TESOL syllabuses.]

4.2 Move 2 'Presenting the research'

This move is the most frequent in both English and Italian abstracts, showing a steady increase across the periods (EN subsets: 92.30% in 1997, 93.61% in 2007, 98.38% in 2017; IT subsets: 94.11% in 2007, 100% in 2017). A common practice in both diachronic language corpora is that linguists describe the general purpose or an outline of their research immediately after the 'Situating the research' move (EN subsets: 56.41% in 1997, 54.35% in 2007, 56.46% in 2017; IT subsets: 59.37% in 2007, 48.57% in 2017). Although less frequent, another generalised tendency in both languages is to open directly an abstract with this move, thus omitting the 'Situating the research' move (EN subsets: 43.59% in 1997, 45.65% in 2007, 43.54% in 2017; IT subsets: 40.63% in 2007; 51.42% in 2017).

The move is realised in both languages predominantly through two submoves as described by Dos Santos (1996): submove 1A 'Indicating main features' or submove 1B 'Indicating main purpose'. The first option is favoured by writers in both English and Italian data: 77.42% of the abstracts in English over the last 20 years and 64.88% of the Italian abstracts over the last 10 years. While submove 1A remains the preferred option for the writers in both language corpora, the data indicate frequency changes across the periods. As regards the English subsets, there is an increase of 12% in its use from 1997 to 2007 (from 74.35% in 1997 to 86.95% in 2007), but a decrease of 16% in 2017,

falling to 70.96%. In the Italian subsets, on the contrary, its use increases by 17% between 2007 and 2017 (from 56.25% in 2007 to 73.52% in 2017).

As for submove 1A 'Indicating main features', in both the English and Italian abstracts there is a preference for using an inanimate noun referring to the present work (*paper, study, research, analisi* [analysis], *contributo* [paper], *saggio* [paper], *articolo* [article], *lavoro* [work]) followed by a verb signalling cognitive processes such as *examine, explore, investigate, analyse, report, present, describe, studiare* [study], *presentare* [present], *descrivere* [describe]. Writers of both English and Italian abstracts favour the practice of using those verbs in the present tense across the periods. There are only two instances of a verb in the future tense in the Italian 2007 subset (*in questo articolo verranno analizzate* [in this article X will be analysed]; *in questo intervento analizzerò* [in this paper I shall/will analyse]), and one instance in English in 2007 (*this paper will examine*). In addition, 13 instances of past tense are found in the English subsets over the last 20 years (1 occ. in 1997, 4 in 2007, 8 in 2017, respectively). Typical linguistic exponents of this submove in English and Italian are: *the study examines, this paper explores, il presente contributo studia* [this paper studies], *questo articolo describe* [this paper describes]. Such usage has been shown to be a typical feature of scientific language (Prelli 1989; Master 2001).

Although much less frequent (only 10 instances in English and three instances in Italian), there are cases in which the writers of English and Italian abstracts use first-person pronouns (*I/io* and *we/noi* for multi-authored abstracts) to introduce their research: *I/we explore, I analyse, we present, analizzerò* [I shall/will analyse]. Interestingly, this use has increased over the decades in English, from 1 instance in 1997 to 4 and 5 instances in 2007 and 2017, respectively. This result is in line with the general increasing trend in RA abstracts of the same discipline during 1990 and 2010 observed by Bondi (2014) and Li (2021). In Italian, on the contrary, the instances are only found in 2007.

As regards submove 1B 'Indicating main purpose', the writers of both English and Italian abstracts exhibit a clear preference for using a noun indicating the purpose of their studies across the periods (*the aim of this paper/article, the purpose of this paper, the main objective, l'intento di questo lavoro, scopo di questo articolo* [the aim of this work/article]), followed by verbs that typically mark intentions such as *aim, proporre* [intend]. Only two abstracts in the Italian data contain the subject form of the first-person pronoun in this submove (*ci proponiamo* [we intend]; *in questo contributo intendiamo* [in this paper we intend]). As for the English data, one abstract of the 2017

subset uses *our objective* and, interestingly, no instances of *I/we* pronouns are present. This finding shows that in both English and Italian abstracts writers tend to conceal their presence when stating their purposes. This behaviour is particularly evident in Italian, where the preference is for impersonal forms realised through the impersonal *si* constructions (Renzi et al. 1988), as shown in (7) and (8):

- (7) Nel presente contributo *si intende mettere in evidenza* la relazione che esiste tra genere filmico, strategie traduttive e acquisizione di una lingua straniera. (IT 2007)
[In the present article we intend to highlight the relationship between film, translation strategies and foreign language acquisition.]
- (8) In questo studio *si presenta* una proposta di intervento didattico mirato alla consapevolezza degli aspetti pragmatico-prosodici in apprendenti di italiano lingua seconda. (IT 2017)
[In this study we present a teaching proposal aiming at developing awareness of pragmatic-prosodic aspects in learners of Italian as second language.]

As can be seen in the examples, although the *si* constructions indicate an unspecified referent, it is clear that the voice coincides with that of the writer. In explaining their use, I cite evidence from a study by Molino (2010: 95) on the use of personal and impersonal authorial references in English and Italian RAs, which showed that these constructions “have to do with the need to background the agent in order to guarantee an impersonal and objective tone which is expected in Italian in formal registers such as scientific writing”. As observed by Molino, their use can be compared to the so-called *pluralis majestatis*. In terms of frequency changes over the decades under investigation, the use of *si* shows no variation: the overall occurrence is fairly similar and stable over the last 20 years: 10 instances in 2007 and 11 in 2017.

4.3 Move 3 ‘Describing the methodology’

This move shows a substantial increase of 38% in its use in English over the last 20 years, rising to 81.45% in the 2017 data (from 41.02% in 1997 to 44.68% and 81.45% in 2007 and 2017, respectively). This evolution may indicate that authors are more oriented towards a specialised readership. Although much less frequent, the move exhibits an increase of 2% in the Italian data: from 38.23% in 2007 to 40% in 2017.

Methodology is the most frequently occurring move after the 'Presenting the research' move in 78.56% of the three subsets analysed in English and in 74.18% in Italian over the decades. Two representative examples are given in (9) and (10):

- (9) *The study involved two business English courses for university students in Hong Kong. The data sources included questionnaire surveys and samples of learners' language produced during role-plays.* (EN 2017)
- (10) *La complessa problematica è esemplificata attraverso lo studio delle parole chiave in un corpus di testi di argomento turistico [...].* (IT 2007)
[This complex topic is investigated through the analysis of keywords using a corpus of texts on tourism.]

However, on other occasions, and as reported by other authors such as Dos Santos (1996) and Anderson and Maclean (1997), this move merges with the 'Presenting the research' move, into a single sentence, as the following examples illustrate:

- (11) *The present study, based on a corpus of simulated sales negotiations involving American and Japanese participants, focuses on differences in back channelling, repair, repetition, pausing, and private speech among the two groups of subjects, relating these features to norm differences in negotiating and to the management of face wants.* (EN 1997)
- (12) *Attraverso l'analisi di (n)ostalgiche pubblicità tedesco-orientali degli anni Novanta, ci si propone qui di esemplificare come la semiotica possa contribuire alla comprensione della relazione tra testi e cultura.* (IT 2007)
[Through the analysis of East-German nostalgic ads of the '90s, we aim at illustrating how semiotics can contribute to an understanding of the relation between text and culture.]

This merging is reported in 21.44% of the English abstracts over the decades and in 25.82% of the 2007 and 2017 Italian subsets, suggesting that this strategy is favoured by the linguists of both languages due, perhaps, to lack of space.

We also find that in three instances in Italian and one in English this move merges with the 'Summarizing the results' move into a single sentence in the 2007 data:

- (13) *Two hundred and fifty-one questionnaires completed by graduates from 1998 to 2000 revealed a distinct profile for competent language users compared with incompetent users, based on the six-level Common European Framework of Reference for Languages [...]* (EN 2007)
- (14) *Un campionamento da diversi corpora di italiano L2 mostra infatti un'alta percentuale di errori e di omissioni nella selezione dell'ausiliare anche per quanto riguarda i core verbs.* (IT 2007)
 [A pilot study of different corpora of Italian L2 does reveal a high percentage of mistakes and omission of auxiliary verb. This is also the case with core verbs.]

Just as in English applied linguistics abstracts, as described by Dos Santos (1996), linguistic features that typically characterise the 'Describing the methodology' move in the English abstracts analysed over the last 20 years are the past tense and passive voice, as shown in (15):

- (15) Intermediate to advanced adult ESL learners (N = 33) *participated* in the study. *They completed* four picture-based narrative tasks under the two planning conditions over a two-week period via text-chat. Two tasks *were used* as experimental tasks to gauge immediate planning effect on L2 production and the other two as new tasks to gauge L2 development (EN 2017).

In terms of frequency distribution, linguists appear to prefer the passive voice. However, it shows significant diachronic variation. It exhibits a remarkably higher presence in the earliest decade (86.6% in 1997) than in the two later decades, in which the percentage drops sharply (52.38% in 2007 and 46.15 in 2017). In 2007 and 2017, writers also use the past tense (9.6% and 30.75% respectively) and the present tense (38.09% and 17.94% respectively).

Italian linguists, on the contrary, opt for the use of the present tense, as shown in (16) and (17):

- (16) *I dati analizzati consistono* in produzioni testuali orali e comprendono sia testi in lingua materna (L1) sia testi in lingua seconda/straniera (L2). (IT 2007)
 [The data consist of oral texts in L1 as well as L2.]
- (17) *Il corpus su cui si basa il lavoro proviene* da un progetto di pragmatica transculturale dedicato all'atto linguistico del rifiuto a un invito. (IT 2017)
 [The corpus on which the paper is based derives from a pragmatic transcultural project devoted to the speech act of refusing an invitation.]

Only one instance in which Italian linguists use the *passato prossimo* (formally equivalent to the present perfect) is found in the 2007 subset, as (18) illustrates:

- (18) *Abbiamo allestito un corpus di titoli di giornali di economia inglesi e spagnoli al fine di esaminare il linguaggio figurato utilizzato in entrambe le culture. (IT 2007)*
 [We collected a corpus of English and Spanish newspapers headlines and aimed at exploring the use of figurative language across the two cultures.]

A high level of similarity can, therefore, be observed between this move in the English and Italian linguistics abstracts across the decades under investigation.

4.4 Move 4 ‘Summarizing the results’

This move exhibits significant diachronic variation in its occurrence across the two language corpora over the decades under investigation. The Italian abstracts show a much lower frequency with 52.94% and 51.42% in the 2007 and 2017 subsets compared to 80.85% and 100% in the 2007 and 2017 abstracts in English. Apart from this difference, the results are stated most frequently in both languages by means of a sentence beginning with a noun which makes reference to the function of this move (*results, analysis, risultati* [results], *analisi* [analysis]), followed by verbs signalling results such as *show, reveal, indicate, find, rivelare* [reveal], *mostrare* [show]. The only distinctive feature that is worth noting is that the prevalent practice of using such verbs in the present and past tense or passive voice in the English abstracts (*showed, reveal, indicate, it was found that, it is shown that*) is not so clearly favoured by the Italian linguists. These show a significantly higher incidence in the use of the present tense (*rivelano* [reveal], *mostra* [shows]), representing 94.44% of all the tenses in 2007 and 88.88% in 2017. The passive construction is used in only one instance in this move in 2007 (*viene mostrato che* [it is shown that]) and in three instances in 2017. In terms of the specific tenses in the three English subsets, the present tense has been predominant all over the decades. It accounts for 33% in 1997, but increases to 44% in the two later subsets (77.5% in 2007 and 77.04% in 2017, respectively). Over the last 20 years, frequencies of the past tense fall by 9% (from 25.92% in 1997 to 16.12% and 19.67 in 2007 and 2017, respectively). The passive voice exhibits a greater presence in the earliest decade (40.74% in 1997) than in the two later decades, in which it drops sharply (10% in 2007 and 3.27% in 2017).

Differences across the two groups of texts can also be noted in the type of personal references used by the linguists in this move. So far, we have seen that in both the English and Italian abstracts, first-person references are not the favoured choice. But this is not the case with this move in the three English subsets, where on some occasions the main findings are presented by means of a sentence beginning with personal pronouns so that writers promote themselves in the text (4 occ. of *I* and 4 occ. of *we* in co-authored abstracts in 2007; 1 occ. of *I*, 7 occ. of *we* and 5 occ. of *our* in 2017). These figures show that co-authors are now becoming more likely to intrude into their abstracts than single authors.

- (19) *We found that* these texts all have a three-part structure consisting of introduction, body, and conclusion. (EN 2007)
- (20) *Our main results indicate* that despite the fact that English majors are motivated to enhance their abilities in professional writing, only a third of them seem to possess the ability and willingness to control their writing processes despite the fact that self-regulatory strategy use is linked to an increased level of motivation and self-efficacy and to a decreased level of writing anxiety. (EN 2017)

On the contrary, in Italian no instances of personal pronouns are found in this move in the 2007 and 2017 subsets. As observed for Move 2 'Presenting the research', Italian linguists seem more eager to employ impersonal *si* constructions (5 occurrences in 2007 and 9 occurrences in 2017):

- (21) Dall'esame *si osserva* la coesistenza e l'interdipendenza degli elementi linguistici dell'indeterminatezza e della precisione linguistica nei contratti, nonostante questi ultimi non debbano rispondere espressamente alla necessità di garantire la più generale applicabilità. (IT 2007)
[The analysis shows the coexistence and inter-dependence of elements such as linguistic indeterminacy and precision in contracts, even though these kinds of texts do not have to meet the need of conveying general applicability to the extent that laws, rules and statutes must.]
- (22) A supporto e complemento della sezione operativa in cui si guida il lettore nell'analisi, *si presentano* i risultati di un lavoro di ricerca glottodidattica nel quale la scheda di osservazione proposta è stata applicata a una selezione di manuali di diverse L2. (IT 2017)

[To support and complement the Section in which the reader is guided in the analysis, the results of a language teaching study on a selection of different L2 textbooks using the proposed observation grid are presented.]

Again, the overall impression is that the preference is for a more objective and detached interpersonal style in Italian writing.

The ‘promotional’ aspect of the abstract as one of the most important features of the genre (Dahl 2000; Hyland 2000) is also signalled through overt evaluative lexical items related to the promotion of findings in the English corpus. The adjectives *significant*, *important*, *interesting*, *useful*, and the adverbs *significantly*, *interestingly* are the commonest. Their use culminates in 2007, accounting for 12 occurrences, compared with 3 instances in 1997 and 6 in 2017. The same result was noted in a study by Melander et al. (1997) on the use of persuasive strategies in RA abstracts, which showed that American linguists used promotional elements in such texts.

- (23) Analysis of the corpus reveals that elaboration is a complex and *important* rhetorical function in academic writing, and that both its use and meanings vary according to discipline. (EN 2007)
- (24) Results evidence an *interesting* finding regarding the full-sentence title construction which appears not only as a generic peculiarity of RPs but also as a disciplinary peculiarity of Biology RP titles. (EN 2007)

In the Italian corpus no instances of evaluative language are found.

4.5 Move 5 ‘Discussing the research’

A significant difference exists in the frequency of occurrence of this move in the Italian and English linguistics abstracts across the periods, as it is present in 8.82% and 11.42% of the 2007 and 2017 Italian subsets in contrast to 48.93% and 35.48% of the 2007 and 2017 English abstracts. The move is most commonly realised through submove 1 ‘Drawing conclusions’ as Dos Santos (1996) calls it. A common linguistic strategy used by the linguists in both languages to signal the initiation of this move consists in explicitly using a noun which makes reference to the function of this move, e.g. *the article/paper concludes*, *the paper suggests*, *lo studio suggerisce* [the study suggests]; *si conclude che* [we conclude that]. Typical verbs in this move in both languages are *suggest*, *conclude*, *discuss*, *suggerire* [suggest], *concludere* [conclude], used predominantly in the present tense. In English,

the present tense significantly rises from 1997 (45%) to 2007 (73.07%) and then slightly falls in the following decade (72.72% in 2017). The passive voice is also used and shows significant variation over the three periods: it displays a higher frequency in 1997 (50%) than the present tense, but then sharply drops in the following decades (26.92% and 27.27% in 2007 and 2017, respectively). The typical verb used is *suggest* (*it is suggested that*). As regards the Italian data, in the seven Italian 'Discussing the research' moves (3 in 2007 and 4 in 2017), there are only instances of verbs in the present tense.

Similar to the linguistic realisation of this move in the English samples analysed, there are cases (only 10 instances) in which linguists present their conclusions by means of a sentence beginning with personal pronouns like *I/we* followed by verbs such as *argue*, *suggest*, *conclude*, introducing an argumentative proposition. This gives a good indication of the writer's role in writing the abstract. His/her role is that of a writer who clearly assumes an 'arguer' role, to borrow Fløttum's (2003) terminology. By using these verbs, the writer indicates her/his position in relation to his/her own research findings.

- (25) Basing our assumptions on the reported pedagogical skills gained by the trainee teachers and the favorable reactions of the Precourse ESOL students, *we suggest* that such a joining of forces is of unequivocal benefit to both parties. (EN 1997)
- (26) Based on the analysis, *I argue that* the significance of genre-based learning can be captured more fully through observing how learners recontextualize their genre awareness in their writing. (EN 2007)
- (27) *We conclude* that intense engagement with a foreign language may put the native language at risk for temporary attrition, but studying with a foreign-language as the medium of instruction does not. (EN 2017)

The very clear signalling of author position in the English linguistics abstracts corroborates the general picture emerging from other studies that "the English linguists take on the arguer role in the abstracts to a great extent" (Dahl 2009: 131).

Although less frequent (6 instances over the decades), we also find that the English linguists use submove 2 'Giving recommendations', as described in Dos Santos's (1996) model, with the function of outlining suggestions for future research:

- (28) The results are explained in terms of the working memory constraint, and *directions of further research are discussed*. (EN 1997)
- (29) The paper concludes with a discussion of the utility of ESP work in medical discourse and *sets a research agenda for further work* in the communicative setting and tasks of IMGs. (EN 2007)

5. Conclusion

This diachronic study shows that the move structure of RA abstracts written in Italian in the field of linguistics conforms to the international conventions based on the norms of the English academic discourse community. However, they are rhetorically less complex than the English abstracts. Most of the Italian abstracts adopt a two-move model, whereas the English ones exhibit a tendency for a four-move pattern. We could account for these differences on the basis of several socio-cultural factors: different intellectual styles and cultural patterns, or the influence (or lack) of academic writing instruction.

More specifically, the analysis reveals that the generic structure of the English RA abstract has a diachronic evolution towards a “mini-article”, as Gillaerts and Van de Velde (2010: 136) call it, in the sense that there has been a steady increase in the use of the four-move model ‘Situating the research-Presenting the research-Describing the methodology-Summarizing the results’ over the last 20 years. This result tends to confirm Gillaerts’s (2013: 58) claim that “the abstract has become more important for information retrieval”. The tendency for the four-move model thus reflects the “informative” function of the abstract (Lorés Sanz 2004). In the Italian data, on the contrary, there is a tendency for a two-move abstract ‘Situating the research-Presenting the research’, with frequencies that have been fairly uniform over the last 10 years. The fact that Italian scholars give more importance to ‘Situating and Presenting the research’ rather than the other moves may respond to their intent of being more “indicative” (Lorés Sanz 2004). But this may depend on the context of publication and the relationship between the writer and the discourse community. Considering that abstracts in Italian address a small community and are included in local or national publications, Italian writers may not see themselves in competition with their peers, thus placing less emphasis on the need to negotiate their findings with their colleagues. They seem to be more interested in situating their research in context and outlining the purposes of the research. In contrast, as the abstracts for international publication have a much wider audience

than the Italian abstracts, writers feel the need to be persuasive as there is more competition to publish and consequently a greater need to justify their work and meet the expectations and norms of the international discourse community. Novel ideas and originality may have to be emphasised more strongly when scholars address a wider, international readership than when they write for a smaller, national peer group.

Finally, a cross-linguistic feature where we see the clearest difference between the two diachronic language corpora is the use of personal references. There is little use of personal pronouns in the Italian abstracts over the past 10 years. Only five instances are found in comparison with 42 instances in the abstracts in English in the 20-year span. The Italian linguists seem to prefer a more detached interpersonal style by opting for impersonal *si* constructions (41 instances). Hence it appears that, compared to their colleagues writing in English, they tend to be invisible. The choice not to appear as visible authors may be related to the Italian style of formal language or to the need to produce a text which conforms to the level of writer invisibility expected within a particular discourse community. Such trends match those of Bondi and Nocella (this issue), who found that Italian academic writing tends to avoid personalization, while Anglophone writing norms seem to be more open.

The specificity and small size of the corpora analysed here calls for more diachronic investigation across disciplines which could explore the cultural/linguistic differences identified so as to gain a better understanding of the complex interplay between “big” and “small” disciplinary cultures (Holliday 1999; Atkinson 2004). Nevertheless, it can be tentatively concluded that the generic variation observed through the two language corpora might reflect differences in cultural writing conventions, but also differences in the relationship between the writer and the discourse community s/he addresses. As rightly noted by Hyland (2000: 75), the variety of patterns represents the writer’s choice of “how best to convince others of their work, given the particular circumstances of their research, their individual goals and considerations of discipline membership”.

Editorial policies about abstract writing of the linguistics journals from which the abstracts have been sampled may influence our findings to some degree. Further work might take into consideration editorial guidelines that the RA abstract authors are asked to comply with from cross-disciplinary and diachronic perspectives so as to explore whether discipline-specific diachronic variations exist. The study might also be expanded by data from interviews conducted with a small sample of authors to better understand how they approach their writing within the boundaries of academic conventions. In

any case, as Schmied et al. (this issue) suggest, academic writers alone are responsible for deciding whether they find trends functionally convenient in their discourse communities and follow, or whether they see them as unnecessary universals and reject them.

REFERENCES

Primary sources

Applied Linguistics

<https://academic.oup.com/applij>, accessed November 2023.

English for Specific Purposes

<https://www.sciencedirect.com/journal/english-for-specific-purposes>, accessed November 2023.

Italiano LinguaDue

<https://riviste.unimi.it/index.php/promoitals/index/>, accessed November 2023.

Linguistica e Filologia

https://aisberg.unibg.it/sr/docs/LeF/LF_indici.html, accessed November 2023.

RILA [RASSEGNA ITALIANA DI LINGUISTICA APPLICATA]

<https://www.bulzoni.it/it/riviste/rasseгна-italiana-di-linguistica-applicata/results,1-0>, accessed November 2023.

Secondary sources

Alharbi, L. M. – J. M. Swales

- 2011 "Arabic and English abstracts in bilingual language science journals: Same or different?", *Languages in Contrast* 11 (1), 70-86.
<https://doi.org/10.1075/lic.11.1.06alh>.

Anderson, K. – J. Maclean

- 1997 "A genre analysis study of 80 medical abstracts", *Edinburgh Working Papers in Applied Linguistics* 8, 1-23.

Atkinson, D.

- 1992 "The evolution of medical research writing from 1735-1985: The case of the Edinburgh medical journal", *Applied Linguistics* 13 (4), 337-374.
<https://doi.org/10.1093/applin/13.4.337>.
- 2004 "Contrasting rhetorics/contrasting cultures: Why contrastive rhetoric needs a better conceptualization of culture", *Journal of English for Academic Purposes* 3 (4), 277-289.
<https://doi.org/10.1016/j.jeap.2004.07.002>.

- Bhatia, V. K.
1993 *Analysing Genre. Language Use in Professional Settings*. London: Longman. <https://doi.org/10.4324/9781315844992>.
- Bondi, M.
1997 "The rise of abstracts. Development of the genre in the discourse of economics", *Textus* 10, 395-418.
2014 "Changing voices: Authorial voice in abstracts". In: M. Bondi – R. Lorés Sanz (eds.) *Abstracts in Academic Discourse: Variation and Change*. Bern: Peter Lang, 243-270.
<https://doi.org/10.3726/978-3-0351-0701-2>.
- Bondi, M. – S. Cavalieri
2012 "The evolution of the abstract as a genre: 1988-2008. The case of applied linguistics". In: G. Garzone – P. Catenaccio – C. Degano (eds.) *Genre Change in the Contemporary World. Short-term Diachronic Perspectives*. Bern: Peter Lang, 43-54.
- Breeze, R.
2009 "Issues of persuasion in academic law abstracts", *Revista Alicantina de Estudios Ingleses* 22, 11-26.
<https://doi.org/10.14198/raei.2009.22.02>.
- Busch-Lauer, I.
1995 "Textual organization in English and German abstracts", *Anglicana Turkuensia* 14, 175-186.
- Cavalieri, S.
2014 "Variation across disciplines. The case of applied linguistics and medicine". In: M. Bondi – R. Lorés Sanz (eds.) *Abstracts in Academic Discourse. Variation and Change*. Bern: Peter Lang, 161-174.
<https://doi.org/10.3726/978-3-0351-0701-2>.
- Cavalieri, S. – G. Diani
2018 "Rhetorical variation in English and Italian law research article abstracts: A cross-linguistic analysis". In: G. Tessuto – V. K. Bhatia – J. Engberg (eds.) *Frameworks for Discursive Actions and Practices of the Law*. Newcastle upon Tyne: Cambridge Scholars Publishing, 89-113.
- Cavalieri, S. – C. Preite
2017 "Genre variation and changes in frame sequences across cultures: The case of criminology RA abstracts in English and French", *Studies in Logic, Grammar and Rhetoric* 49 (62), 37-53.
- Dahl, T.
2000 *Lexical Cohesion-Based Text Condensation. An Evaluation of Automatically Produced Summaries of Research Articles by Comparison with Author-Written Abstracts*. Dissertation. Bergen: University of Bergen.
2009 "Author identity in economics and linguistics abstracts". In: E. Suomela-Salmi – F. Dervin (eds.) *Cross-Linguistic and Cross-Cultural Perspectives on Academic Discourses*. Amsterdam: John Benjamins, 123-133.

- Diani, G.
 2014 "On English and Italian research article abstracts: Genre variation across cultures". In: M. Bondi – R. Lorés Sanz (eds.) *Abstracts in Academic Discourse. Variation and Change*. Bern: Peter Lang, 65-83.
<https://doi.org/10.3726/978-3-0351-0701-2>.
- Dos Santos, M. B.
 1996 "The textual organization of research paper abstracts in applied linguistics", *Text* 16 (4), 481-499.
<https://doi.org/10.1515/text.1.1996.16.4.481>.
- Duan, J. – J. Wei
 2021 "A genre analysis of English and Chinese legal research article abstracts: A corpus-based approach", *Journal of Language Teaching and Research* 12 (5), 810-821. <http://dx.doi.org/10.17507/jltr.1205.21>.
- Feltrim, V. D. – S. M. Aluísio – M. G. V. Nunes
 2003 "Analysis of the rhetorical structure of computer science abstracts in Portuguese". In: D. Archer et al. (eds.) *Proceedings of the Corpus Linguistics 2003 Conference*. Lancaster: Lancaster University Press, 212-218.
- Fløttum, K.
 2003 "'Je' et le verbe", *TRIBUNE* 14, 7-14.
- Gillaerts, P.
 2013 "Move analysis of abstracts from a diachronic perspective: A case study". In: N. – L. Johannesson – G. Melchers – B. Björkman (eds.) *Of Butterflies and Birds, of Dialects and Genres. Essays in Honour of Philip Shaw*. Stockholm: Acta Universitatis Stockholmiensis, 49-60.
- Gillaerts, P. – F. Van de Velde
 2010 "Interactional metadiscourse in research article abstracts", *Journal of English for Academic Purposes* 9 (2), 128-139.
<https://doi.org/10.1016/j.jeap.2010.02.004>.
- Hatzitheodorou, A.-M.
 2014 "A genre-oriented analysis of research article: Abstracts in law and business journals". In: M. Bondi – R. Lorés Sanz (eds.) *Abstracts in Academic Discourse. Variation and Change*. Bern: Peter Lang, 175-198.
<https://doi.org/10.3726/978-3-0351-0701-2>.
- Holliday, A. R.
 1999 "Small Cultures", *Applied Linguistics* 20 (2), 237-264.
<https://doi.org/10.1093/applin/20.2.237>.
- Huckin, T.
 2001 "Abstracting from abstracts". In: M. Hewings (ed.) *Academic Writing in Context: Implications and Applications. Papers in Honour of Tony Dudley-Evans*. Birmingham: The University of Birmingham Press, 93-103.
- Hyland, K.
 2000 *Disciplinary Discourses: Social Interactions in Academic Writing*. Harlow, UK: Longman.

- Johns, T.
1992 "‘It is presented initially’: Linear dislocation and interlanguage strategies in Brazilian academic abstracts in English and Portuguese", *Ilha do Desterro* 27 (1), 9-32.
- Li, Z.
2021 "Authorial presence in research article abstracts: A diachronic investigation of the use of first person pronouns", *Journal of English for Academic Purposes* 51, 1-13.
<https://doi.org/10.1016/j.jeap.2021.100977>.
- Lorés Sanz, R.
2004 "On RA abstracts: From rhetorical structure to thematic organisation", *English for Specific Purposes* 23 (3), 280-302.
<https://doi.org/10.1016/j.esp.2003.06.001>.
2009 "Different worlds, different audiences: A contrastive analysis of research article abstracts". In: E. Suomela-Salmi – F. Dervin (eds.) *Cross-Linguistic and Cross-Cultural Perspectives on Academic Discourses*. Amsterdam: John Benjamins, 187-197.
<https://doi.org/10.1075/pbns.193.12lor>.
- Martín-Martín, P.
2003 "A genre analysis of English and Spanish research paper abstracts in experimental social sciences", *English for Specific Purposes* 22 (1), 25-43.
[https://doi.org/10.1016/S0889-4906\(01\)00033-3](https://doi.org/10.1016/S0889-4906(01)00033-3).
2005 *The Rhetoric of the Abstract in English and Spanish Scientific Discourse. A Cross-Cultural Genre-Analytic Approach*. Bern: Peter Lang.
- Master, P.
2001 "Active verbs with inanimate subjects in scientific research articles". In: M. Hewings (ed.) *Academic Writing in Context. Implications and Applications. Papers in Honour of Tony Dudley-Evans*. Birmingham: University of Birmingham Press, 169-181.
- Melander, B. – J. M. Swales – K. M. Friedrickson
1997 "Journal abstracts from three academic fields in the United States and Sweden: National or disciplinary proclivities?". In: A. Duszak (ed.) *Culture and Styles of Academic Discourse*. Berlin: Mouton De Gruyter, 251-272.
- Molino, A.
2010 "Personal and impersonal authorial references: A contrastive study of English and Italian linguistics research articles", *Journal of English for Academic Purposes* 9 (2), 86-101.
<https://doi.org/10.1016/j.jeap.2010.02.007>.
- 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 (2), 231-250.
<https://doi.org/10.1177/1461445607087010>.

- Prelli, L. J.
1989 *A Rhetoric of Science: Inventing Scientific Discourse*. Columbia: University of South Carolina Press.
- Renzi, L. – G. Salvi – A. Cardinaletti
1988 *Grande Grammatica Italiana di Consultazione*. Vol. I: *La Frase. I Sintagmi Nominale e Preposizionale*. Bologna: Il Mulino.
- Salager-Meyer, F.
1990 "Discoursal flaws in medical English abstracts: A genre analysis per research – and text-type", *Text* 10 (4), 365-384. <https://doi.org/10.1515/text.1.1990.10.4.365>.
- Samraj, B.
2005 "An exploration of a genre set: Research article abstracts and introductions in two disciplines", *English for Specific Purposes* 24 (2), 141-156. <https://doi.org/10.1016/j.esp.2002.10.001>.
- Sandoval Cruz, R. I.
2015 "Análisis contrastivo de movimientos retóricos en resúmenes de lingüística aplicada escritos en español mexicano e inglés norteamericano". In: M. D. Perales-Escudero – M. G. Méndez López (eds.) *Experiencias de Docencia e Investigación en Lenguas Modernas*. Chetumal, Mexico: Universidad de Quintana Roo, 20-31.
- Swales, J. M.
1990 *Genre Analysis: English in Academic and Research Settings*. Cambridge: Cambridge University Press.
- Tankó, G.
2017 "Literary research article abstracts: An analysis of rhetorical moves and their linguistic realizations", *Journal of English for Academic Purposes* 27, 42-55. <https://doi.org/10.1016/j.jeap.2017.04.003>.
- Van Bonn, S. – J. M. Swales
2007 "English and French journal abstracts in the language sciences: Three exploratory studies", *Journal of English for Academic Purposes* 6 (2), 93-108. <https://doi.org/10.1016/j.jeap.2007.04.001>.
- Ventola, E.
1994a "Abstracts as an object of linguistic study". In: S. Čmejrková – F. Daneš – E. Havlová (eds.) *Writing vs. Speaking. Language, Text, Discourse, Communication*. Tübingen: Gunter Narr, 333-352.
1994b "Problems in producing scientific abstracts in L2". In: S. Čmejrková – F. Štícha (eds.) *The Syntax of Sentence and Text*. Amsterdam: John Benjamins, 283-303.

Address: GIULIANA DIANI, Department of Studies on Language and Culture, University of Modena and Reggio Emilia, Largo Sant'Eufemia 19, 41121 Modena, Italy.
ORCID code: <http://orcid.org/0000-0001-7044-9039>.

Concluding sections over 30 years of research writing: The case of a Polish scholar

Krystyna Warchał

University of Silesia

ABSTRACT

This paper offers an individual perspective on the evolution of genre standards by looking into a collection of texts published throughout 30 years of a research career by a scholar whose main field is applied psycholinguistics and whose main language of publication is English, her second language. The material is limited to concluding sections of English-language single-authored monograph chapters and journal articles published in the years 1990-2019, beginning with early, pre-doctoral publications, to full professorship contributions. The analysis focuses on the relative prominence of concluding moves and the changes in the use of first-person pronouns and epistemic markers in texts representing different stages of academic career. Apart from documenting the development of genre competence and the growth of the second language writer, the results may be indicative of an evolution of expectations towards the final text section and shed some light on the development of academic identity.

Keywords: academic writing, genre analysis, genre evolution, moves, academic identity, second language writing.

1. Introduction

Genre analysis, pioneered by Swales (1981, 1990) and developed by Bhatia (1993), has been applied to a variety of academic and occupational genres beyond the research article, such as lectures (Thompson 1994), conference presentations (Rowley-Jolivet – Carter-Thomas 2005), abstracts (Cross – Oppenheim 2006), research grant proposals (Connor – Mauranen 1999), corporate home pages (Luzón Marco 2002), and business letters of negotiation (Pinto dos Santos 2002). These studies have provided insights

into rhetorical practices of various discourse communities, added to the understanding of specific goals accomplished through these practices, and offered teaching tools to help novice members to develop their competence in producing texts that are central for the community. Among these studies, diachronic analyses, aimed at identifying changes in rhetorical structures, have been comparatively rare.

One such attempt has been undertaken by Dressen-Hammouda (2008), who examines emerging genre competence in the context of disciplinary identity. She traces the early academic development of a geology student, from his third undergraduate year to the final stage of his PhD project, showing how his growing ability to identify and relate to his field's central concerns and to recognise disciplinary practices is reflected in the use of appropriate discursual forms. Not much, however, is known about textual signals of academic development in longer time frames, from novice to expert community member. This paper offers an individual perspective on the development of genre competence in an extended time-frame by looking into a collection of academic texts published throughout 30 years of an active research career by a scholar whose main field is applied psycholinguistics, and whose main language of publication is English, her second language in terms of the order of acquisition. By focusing on selected aspects of text structure and rhetoric, it traces the development from a novice to an expert scholar and the growth of a second language academic writer.

The analysis is limited to concluding sections, text segments where English-language authors typically summarise main results, relate them to previous research, point out their significance, and identify their implications. While "there is no one 'correct' way to conclude" (Arnaudet – Barrett 1984: 88), concluding sections are usually expected to go beyond a simple restatement of the findings; rather, they are expected to situate them in a broader context and to make them relevant in view of previous studies and social needs (Hewings – Thaine 2012; Swales – Feak 2012; Wallace – Wray 2016). In other words, it is often at this stage that writers demonstrate that their research makes sense and what sense it makes. This turns conclusions into rhetorically complex and demanding sections. In view of that, it can be expected that the task of concluding will be approached differently by novice and experienced academic writers. This paper is an attempt to check these intuitions by looking into concluding sections produced at three different stages of the subject's academic career in order to establish whether there are any differences in the concluding moves taken by the author, the relative prominence of the moves, the expression of authorial presence, and the use

of epistemic markers modifying the degree of certainty, self-confidence, and authority that she invests in the text.

The next section provides the background for the study by discussing concluding sections and introducing the rhetorical model applied in the analysis. Section 3 introduces the methodology and describes the procedures. Section 4 presents the results of the analysis, followed by discussion and concluding remarks.

2. Concluding sections

Initiated by Swales in the early 1980s and originally applied to research article introductions, move analysis has since been used to examine the rhetorical structure of a variety of academic genres and their sections (see, e.g., Carter-Thomas – Rowley-Jolivet 2003, for conference presentations; Feng – Shi 2004, for research grant proposals; Lorés Sanz 2004, for research article abstracts; Shamsudin – Ebrahimi 2013, for lecture introductions; Bruce 2008, for methods sections of research articles). Among these text segments, conclusions have received less attention, but important contributions have been made by, among others, Hopkins and Dudley-Evans (1988), Swales (1990), Dudley-Evans (1994), and Yang and Allison (2003). All these studies explicitly or implicitly point to the delimitation problems: conclusions often combine discussion, closing remarks, and teaching implications, all of which may have the status of separate text sections that follow the presentation of results.

Hopkins and Dudley-Evans (1988) identify eleven moves that tend to recur in sections following results in articles and dissertations in the field of biology and agricultural engineering, the most prominent being Statement of results, Reference to previous research, Explanation (of the observed differences or unexpected results), Exemplification (in support of the proposed explanation), Deduction and Hypothesis (both aiming at generalisation), and Recommendation (with suggestions for future work), a list later extended with Limitation (on the findings, methodology, or applicability of the results; Dudley-Evans 1994: 225). This model, with some modifications, was later successfully applied to other disciplines (Holmes 1997) and to the native/non-native speaker distinction (Peacock 2002).¹

¹ More recently, discussion / concluding sections of research articles have also been analysed in terms of promotional steps taken by authors writing in English and in Spanish (Moreno 2021) and in terms of linguistic variation across disciplines (Liu – Xiao 2022).

Commenting on the existing research on discussion sections, Swales (1990: 172) points to their well-attested cyclic nature: compared to introductions, sections that follow results are rarely “chunked”; rather, the moves tend to form recurring patterns, often depending on the predictability of the results or their compatibility with previous findings. Swales observes that, generally, the cycles proceed from tangible results to their broad context and real-world significance. Thus, the final organisation of concluding sections will to a large extent depend on how authors situate their findings among other research, how far they are prepared to move beyond the results and hypothesise on their possible implications, and how broadly they define their research interests. This suggests that the rhetorical structure of sections that follow results is likely to evolve with the academic development of the author, evolution which may be more noticeable here than in sections that do not rely so heavily on the contextualisation of research findings and the awareness of others’ needs and disciplinary challenges.

The delimitation problems involved in concluding sections are perhaps most explicitly addressed in Yang and Allison (2003). Focusing on research articles in applied linguistics, the authors remark that discussion and conclusions may form a single text section (variously headed), or appear as distinct units in the same text, sometimes followed by separate pedagogic implications. Table 1 presents the moves and steps they identify in sections that follow results. Moves 1-4 have not been attested in conclusions if these follow discussion as a separate text section.

Table 1. Moves and steps in sections that follow results according to Yang and Allison (2003)

Moves	Steps
Move 1 Background information	
Move 2 Reporting results	
Move 3 Summarizing results	
Move 4 Commenting on results	Interpreting results Comparing results with literature Accounting for results Evaluating results
Move 5 Summarizing the study	
Move 6 Evaluating the study	Indicating limitations Indicating significance/advantage Evaluating methodology

Move 7 Deductions from the research	Making suggestions
	Recommending further research
	Drawing pedagogic implications

Apart from providing insights into specific communicative goals of concluding sections in the field of applied linguistics, the framework proposed by Yang and Allison (2003) has proved useful in comparative analyses of texts by Anglophone and EAL (English as an Additional Language) scholars. For example, Rañosa Madrunio (2012) analysed research papers prepared by graduate students of the English programme at a Manila university and found out that their conclusions typically included Summarising results but omitted Evaluating the study, a move that occurred quite regularly in Yang and Allison’s data. The model was also used in Kashiha (2015), a comparative study of the distribution of lexical bundles across concluding moves in English-language applied linguistics articles by Anglophone and Iranian writers. The author shows that Iranian scholars tend to rely on a smaller set of bundles which they use repeatedly in various moves and steps. By contrast, lexical bundles used by Anglophone authors often occur only in one move or step; thus, they more closely correspond to a specific communicative goal. Kashiha notes that a possible reason for this less move-specific use of bundles by Iranian scholars may be a smaller range of vocabulary and convenience, as the writers rely on common and well-familiar phrases. It seems, though, that another explanation is also possible: in spite of following the prescribed move-and-step formula, non-Anglophone writers may be less aware of specific communicative goals accomplished through individual steps and hence use fewer structures that directly correspond to these goals (such as explicit references to current needs or conditional verb forms in Deductions from the research).

In contrast to previous studies on the rhetorical structure of concluding sections, this paper looks into texts produced by one scholar across thirty years of her publishing life. It aims to document the development of genre competence throughout her research and second-language writer career by analysing the move structure of concluding sections at various stages of her academic life. Additionally, it looks for possible changes in the degree of authorial presence and markers of epistemic commitment. As Kuo (1999) notes, references to the self reveal writers’ perception of their relationship with the discipline and set the findings in a broader context of their research interests and previously obtained results, so the degree to which authors mark their presence in texts is likely to change with experience. Epistemic

markers, in turn, reflect authors' commitment to the truth of what is being said, different in reporting results and in making generalisations from the findings, recommendations, and predictions; they also reflect the author's authority, the awareness of the state of the art in the discipline, and the understanding of established patterns of interaction (Hyland 2005; Biber 2006). Thus, their use is also likely to change with the academic development of the writer. Concluding sections, which situate the results within previous research, show their significance, point out implications, and indicate possible lines of study, seem a promising text segment to study authorial presence and epistemic commitment.

3. Methodology

Looking into a set of concluding sections produced by the same author through the period of 30 years, this paper seeks answers to the following questions: Do the rhetorical structure of conclusions and relative prominence of particular moves change with the academic development of the scholar? Is authorial presence marked in different ways at different stages of her academic career? Is epistemic modality used differently at early and advanced stages of her academic development?

Throughout the paper, the terms concluding sections and conclusions are used interchangeably and understood broadly as sections that follow the presentation of results. In practice, this means that the term covers segments that can be identified as discussion and/or conclusions, an approach that seems justified in view of the fact that the two segments often overlap, with discussion taking over the role of concluding remarks and concluding remarks providing extended commentary. The fact that discussions and conclusions tend to overlap was also noted by Yang and Allison (2003), whose framework is used as the basis for move analysis in this study.

The approach taken in this paper is case study: the analysis focuses on an individual and the evolution of selected characteristics of texts she produced over an extended period of time; the observed differences are then interpreted in the context of her developing expertise as a researcher and her building confidence as a second language writer (Flyvbjerg 2011; Casanave 2015). It aims to provide some insight into how selected text features evolve together with the academic development of the author and add to the understanding of how growing awareness of the context (including existing research, the readers, and disciplinary conventions), expanding research

interests, and a growing sense of independence as a researcher are reflected in the rhetoric of concluding sections. Thus, it falls into the category of instrumental case study (Stake 2003), since it aims to better understand the rhetorical change that takes place in the process of development of academic identity (a process the present case illustrates) rather than to describe and understand the case itself “in all its particularity and ordinariness” (Stake 2003: 136).

The research method is text analysis supported by an unstructured interview with the author, a prolific writer, widely-cited researcher, respected mentor, and academic journal editor, whose main field is applied psycholinguistics, the first language Polish, and the main language of publication English, her second language in terms of the order of acquisition. Her first academic publication dates back to 1990. She obtained her PhD in 1993, post-doctoral degree (habilitation) in 2005, and full professorship in 2013. The interview was a preliminary interview conducted at the stage of corpus compilation (Rose et al. 2020). Its main goals were: to select texts for the analysis and check their accessibility (especially in the case of early texts); to discuss the division of the material into time-frames (while the stages are centred on academic degrees, a decision had to be made on the points of division); and to gather additional information about the author’s academic development. In the last-mentioned case, the intention was to determine whether there had been any sharp turns in the author’s academic career, such as a sudden change of academic discipline or area, attributable to other reasons than a natural evolution of research interests. No such events were identified.

The material for this study is limited to concluding sections of selected monograph chapters and journal articles (to the exclusion of monographs and the unpublished PhD thesis). It comprises 28 texts published in the years 1990-2019, all of them in English and single-authored, beginning with early, pre-doctoral publications, through post-doctoral degree texts, to full professorship contributions. The overall size of the corpus is about 18,000 running words.

The selection of texts was made after consultation with the author, the intention being to reach a compromise between two elements: the need to give fair treatment to all stages of her publishing life and the need to choose texts that the author herself considered as significant for her academic development. All texts were saved as text files, with figures, tables, fragments of interviews and think-aloud protocols, and more extensive examples in other languages than English removed. The corpus was then divided into three parts corresponding

to various stages of the academic career of the author: the novice stage (NS), the mature stage (MS), and the expert stage (ES). These time-frames are built around major academic achievements, that is, the PhD degree in 1993, post-doctoral degree in 2005, and full professorship in 2013; the points of division were discussed with the author. The final decisions regarding the corpus and the time-frames are shown in Table 2.

Table 2. The corpus

Stage	Time-frame	No of texts included (CH+JA)	Running words	Mean length	Shortest	Longest
Novice stage (NS)	1990-1996	4 (2+2)	1,880	470	87	725
Mature stage (MS)	2000-2009	9 (7+2)	4,580	508	84	771
Expert stage (ES)	2010-2019	15 (15+0)	11,480	766	223	1965
Total:	1990-2019	28 (24+4)	17,940			

CH – book chapters; JA – journal articles

As can be seen in Table 2, the three time-frames differ in length – the novice stage being shorter than the other two – and, most importantly, in the number of texts. Also, concluding sections in the expert stage are considerably longer than those in the earlier stages. These differences will necessarily limit the validity of the findings, but at the same time, they reflect the author's development as a writing scholar and anticipate some differences in the rhetorical organisation of texts.

The analysis was carried out with the help of QDA Miner 6. In the first part (rhetorical moves), the framework proposed by Yang and Allison (2003) was applied, but the system was slightly modified to accommodate text segments that combined the functions of text sections traditionally referred to as discussion and conclusions. Moves and steps attested in the corpus are shown in Table 3.

In the second part of the analysis (authorial presence), the focus was on first-person pronouns, self-citations, and other ways of referring to the self (such as *the present author* or *the writer*). In the third part (epistemic modality), searches were run using a list of words compiled on the basis of literature for a different study (Warchał 2015). An abbreviated version of the list is presented in Table 4. In the second and third part of the analysis, the search results were checked and cleared manually to eliminate accidental records. The results of the analysis are presented in the next section.

Table 3. Moves attested in the corpus (based on Yang & Allison, 2003)

Moves		Steps
Move 1	Summarising the study	1.1 Main problem/ goals/ hypotheses 1.2 Main results
Move 2	Commenting on results	2.1 Interpreting results 2.2 Comparing with literature (also invoking similar studies) 2.3 Accounting for results
Move 3	Evaluating the study	3.1 Indicating significance/ advantage 3.2 Indicating limitations 3.3 Evaluating methodology
Move 4	Deductions from the research	4.1 Recommending further research (also promising further research) 4.2 Drawing pedagogic implications

Table 4. Epistemic modality markers used as search words (Warchał 2015)

Modal value/ category of marker	High-value epistemic modality markers	Middle-value epistemic modality markers	Low-value epistemic modality markers
(quasi)modal verbs	be bound to, be going to, can't, couldn't, have (got) to, must, need (to), will	be supposed to, ought to, should, would	could, may, might
modal modifiers	by no means, certainly, definitely, doubtless, for certain, for sure, incontestably, incontrovertibly, indisputably, indubitably, on no account, no doubt, surely, unarguably, undeniably, undoubtedly, unquestionably, without (a shadow of a) doubt	arguably, in all likelihood, (most) likely, presumably, probably, supposedly	allegedly, conceivably, maybe, perhaps, possibly, purportedly

adjectives with a clausal complement	certain, convinced, impossible, inconceivable, not possible, sure, undeniable	(most) likely, plausible, probable	conceivable, doubtful, not likely, possible, uncertain, unlikely
nouns with a clausal complement	little doubt, no doubt	likelihood, feeling, impression	possibility, doubts
verbs of mental states and processes with a clausal complement		assume, believe, expect, imagine, presume, suppose, think, appear, seem	doubt, guess, speculate, suspect

4. Results

The first part of the analysis looked into the move-and-step structure of conclusions at different stages of the author's academic career. As can be seen in Table 5, different sets of rhetorical steps are characteristic of the novice stage and of the other two stages. The formula for early-career texts seems to be: [Main problem + Comparing with literature + Indicating significance + Recommending further research + Drawing pedagogic implications], while for the more advanced stages, it tends to follow the sequence: [Main results + Interpreting results + Comparing with literature + Drawing pedagogic implications]. Restating the main problem appears to be less important at the later stages, while the presentation of main results gains in importance, as does interpretation of the results. Comparing with literature and Drawing pedagogic implications seem to be relatively stable throughout the studied period of time, a feature that may be characteristic of this field of study or specific research area. However, Comparing with literature may play a slightly different role in the NS than in the more advanced stages. In the NS, it is not used to highlight, corroborate, or contextualise the findings, which are rarely restated in this set of texts, but seems to take place of the summary of results. Since in the material analysed the results are in line with previous research, a possible explanation for this use of literature may be the need to invoke the authority of more experienced researchers and in this way to gain credibility for the analysis. In all time-frames, Accounting for results, Evaluating methodology, and Indicating limitations seem to be of lesser importance.

Somewhat unexpectedly, Indicating significance, a move that may be thought to follow from the interpretation of the results, appears with a greater regularity in NS texts, where Interpreting results is rarely attested, than in the MS or ES, where Interpreting results occurs quite regularly. A possible explanation may be that in the early stage, Indicating significance tends to appear in place of interpretation rather than as its natural consequence; in other words, in the NS, the significance of the results seems to be announced rather than demonstrated to the reader. It is also worth noting that, perhaps surprisingly, the rhetorical structure of MS and ES texts is similar in spite of the fact that ES conclusions are markedly longer than MS concluding sections (see Table 2).

Table 5. Move occurrence in conclusions at different career stages²

Move/Step	NS		MS		ES	
	no of texts	%	no of texts	%	no of texts	%
1.1 Main problem/ goals/ hypotheses	2	50	3	33	4	27
1.2 Main results	1	25	7	78	11	73
2.1 Interpreting results	1	25	5	56	10	67
2.2 Comparing with literature	3	75	6	67	13	87
2.3 Accounting for results	1	25	3	33	1	7
3.1 Indicating significance/ advantage	3	75	2	22	6	40
3.2 Indicating limitations	1	25	3	33	1	7
3.3 Evaluating methodology	1	25	1	11	2	13
4.1 Recommending further research	3	75	3	33	6	40
4.2 Drawing pedagogic implications	2	50	4	44	8	53

The prominence of a rhetorical move was operationalised as the relative length of the text segment where the move was realised. Table 6 presents the results (expressed in the percentage of words) for texts belonging to a particular stage. Naturally, if the move occurs in a larger number of texts in a given time-frame, the percentage is expected to be higher; this can be seen, for instance, in Main results and Interpreting results, which are rare in the NS but quite common in the MS and ES. Still, text segments associated with Interpreting results tend to be considerably longer in the two later stages, as

² Numbers refer to the number of texts where a move was attested, also expressed as a percentage of the total number of texts in a given time-frame.

can be seen in Examples 1-3, which illustrate this move in the NS (Ex. 1) and in the MS and ES (Ex. 2 and 3, respectively).

It is interesting to note, however, that moves that tend to occur regularly in texts regardless of the career stage, such as Comparing with literature or Drawing pedagogic implications, may become more or less prominent in time. In the present case, Comparing with literature takes less text space in the MS and ES than in the NS, while Drawing pedagogic implications appears to be slightly on the rise. To illustrate, a text segment associated with Comparing with literature drawn from a NS text (Ex. 4) is markedly longer than corresponding text segments drawn from MS and ES texts (Ex. 5 and 6, respectively).

Table 6. Relative length of moves (in percentage of words) at different career stages

Move/Step	NS	MS	ES
1.1 Main problem/ goals/ hypotheses	3.7	3.6	3.7
1.2 Main results	9.2	20.3	23.4
2.1 Interpreting results	1.4	25.6	27.4
2.2 Comparing with literature	55.7	16.6	23.3
2.3 Accounting for results	4.3	6.4	0.9
3.1 Indicating significance/ advantage	4.9	3.2	4.6
3.2 Indicating limitations	2.1	2.6	0.2
3.3 Evaluating methodology	2.9	3	1
4.1 Recommending further research	6.5	7.1	3.3
4.2 Drawing pedagogic implications	9.3	11.6	12.2
Total	100	100	100

- (1) So we can probably conclude that they came with the already mentioned, growing consciousness of how to learn as a result of individual learning experience.
(01-NS-1990-JA-C; Move 2 Commenting on results 2.1 Interpreting results; 25 words)
- (2) It can be assumed that the reversed order of instruction may develop and facilitate meaningful learning, and thus bring about more positive learning outcomes. The main assumption of the meaningful learning is that success in learning depends on the ability to discover the relationships between ideas. This general definition when applied to this specific context will mean that conceptualisation will act as an

initial stage of the discovery of the link between something seemingly distant from the verbal expression, i.e. body with its physicality, and language itself. The three modes of learning distinguished by J. Bruner: enactive (manipulation), iconic (visualisation) and symbolic (abstract) are the best illustration of how conceptualisation can facilitate language learning by discovery through analysis and comparison. Another variable emphasised in meaningful learning is prior knowledge, i.e. knowledge of the world, here the familiarity of the feeling and the ability to describe ANGER as a phenomenon, and knowledge of one's mother tongue.

(08-MS-2002-CH-C; Move 2 Commenting on results 2.1 Interpreting results; 158 words)

- (3) Comparing the questionnaire data and narrative data, it can be observed that the focus on learners' responsibility for classroom climate is expressed strongly only in the former, emerging from the more structured way of expressing one's views that most questionnaires demand. When asked to reflect in an open reflective narrative, the subjects seem to focus almost entirely on the teachers as the creators and facilitators of classroom climate. This is also visible in the comments on their past positive and negative experiences of classroom climate at different levels of education, in which teachers are either blamed or praised for how they coped with or ignored the issues concerning classroom climate. What comes as a surprise is that the trainees, future FL teachers, do not appear to be aware of the role a foreign language itself can play in establishing a positive classroom climate. As a vehicle for communication, for instance, off-task interaction between the teacher and learners can be a factor contributing to the development of more individually-oriented contacts, and thus an individualized approach to learners.
(26-ES-2016-CH-C; Move 2 Commenting on results 2.1 Interpreting results; 179 words)

- (4) Psycholinguistic investigations of the mental lexicon of bilinguals have produced contradictory results, some researchers advocating the existence of separate L1 and L2 lexicons, while others assume the existence of one lexical store for both L1 and L2. Albert and Obler, as quoted in Channell (1988), sum up the controversy: "It is clear that words in one language and their translation equivalents (when such exist) are related in the brain in a non-random way, much as a word

and its synonyms in the same language may be connected in an associational network". Numerous word-association tests in L1 have brought to light the links between lexical items, establishing various types such as synonymy, super-ordination and collocation. Since collocational links seem to be powerful and lasting, the research into L2 associations could be of value in the teaching of foreign-language collocations (see Author, forthcoming, for pilot results).

(02-NS-1992-CH-C; Move 2 Commenting on results 2.2 Comparing with literature; 148 words)

- (5) In the case of L1 gesticulation, full automatisation and lack of awareness of the use of gestures was reported, which would indicate that just like the verbal part of the message, it is fully automated, or perhaps, as it was suggested by Krauss et al. (2000), the two processes, i.e. non-verbal language production and speech are inseparable, however, the model accounts only for the L1 messages, and may be different for L2 production.

(10-MS-2004-CH-C; Move 2 Commenting on results 2.2 Comparing with literature; 75 words)

- (6) As Moser (1999) puts it: "Like many other key concepts in psychological research, the self is a 'classical' research topic for metaphor analysis, because people can only speak metaphorically about the complex and abstract matter of the 'self'. Analyzing metaphors thus not only gives access to the tacit knowledge and mental models which shape the individual understanding of the self, but also to the cultural models provided by language to express individuality, self-concept and the 'inner world'." (Moser 1999: 144)

(28-ES-2017-CH-C; Move 2 Commenting on results 2.2 Comparing with literature; 82 words)

The first part of the analysis can be summed up in the following points: i) In terms of the occurrence of rhetorical moves, MS and ES conclusions appear quite similar; they differ from texts produced in the NS in the centrality of Main results and Interpreting results, both rare in the NS; ii) Comparing with literature, a step commonly attested in all three time-frames, seems to play a different role in NS than in the more advanced stages, since it often replaces Main results rather than helps to interpret them; and iii) Compared to MS and ES texts, conclusions produced in the NS tend to give more prominence to Comparing with literature. By contrast, Main

results and Interpreting results are the most prominent moves in the more advanced stages.

The second part of the analysis focused on authorial presence as expressed by personal pronouns, lexical references to the self, and self-citations. The results are presented in Table 7. Phrases such as *this author*, *the present author*, or *the author/the researcher* – referring to the writer herself – were not attested in the corpus. As can be seen from Table 7, the overall frequency of self-references in texts representing the three time-frames stays on a similar level, between 59 to 66 references per 10,000 words. The preferred type of expression changes only slightly, as can be seen in Table 8, which presents the relative frequency of particular self-referring expressions at different stages of academic career. While inclusive *we* is the most frequent choice at all stages (Ex. 7-9), first-person singular appears somewhat more frequent in the MS than in the other stages (Ex. 10-12). Also, self-citations are on the rise, which is to be expected in view of the growing output and consistency of the author's academic interests.

With regard to first-person singular, the most frequently attested functions in all time-frames are announcing or explaining the author's intentions – e.g., *I would like to draw some conclusions* (NS); *I would like to emphasize the significance of* (MS); and *In this article, I have tried to demonstrate* (ES) – and modifying epistemic commitment, as in *It seems to me that a psycholinguistic approach* (NS); *I think that it may be fairly safely assumed that* (MS); and *I believe that although based on very limited data* (ES). Direct references to research decisions or to the research context are rare and tend to appear in the more advanced stages, e.g., *the extent to which my data is in concord with* (MS) and *the subject in my study who* (ES). In the ES, self-citations sometimes concur with the use of first-person singular, as in *In my previous comparative study*; *As I observed in my previous studies*; and *In my earlier studies on culture-grounded concepts*.

Table 7. Self-referring expressions in conclusions at different career stages

Self-referring expression	NS	per 10,000 words	MS	per 10,000 words	ES	per 10,000 words
1 singular	2	11	9	20	15	13
1 plural inclusive	8	42	16	35	44	38
1 plural exclusive	0	0	2	4	0	0
Self-citation	1	5	3	7	14	12
Total	11	59	30	66	73	64

Table 8. Relative frequency of self-referring expressions (in percent) for different career stages

Self-referring expression	NS	MS	ES
1 singular	18.2	30	20.5
1 plural inclusive	72.7	53.3	60.3
1 plural exclusive	0	6.7	0
Self-citation	9.1	10	19.2
Total	100	100	100

- (7) It is only one example of how this method can contribute to **our** knowledge of how learners process language (both their L1 and L2).
(04-NS0-1995-JA-C; 1 person plural inclusive)
- (8) If **we** relate both models directly to the issue of L2 grammar instruction, what clearly emerges is the need for both explicit and implicit teaching and learning (see Table 6)... The development of grammatical competence and its complexity derives from the complexity of grammar in **our** minds; thus learning becomes a complex process in need of support from various approaches which do not exclude but rather complement each other.
(13-MS2-2007-CH-C; 1 person plural inclusive)
- (9) If **we** first of all consider the teacher to be an agent of change, **we** have to focus on his/her role as a facilitator expressed in the different dimensions of facilitation (Table 5).
(26-ES2-2016-CH-C; 1 person plural inclusive)
- (10) Analysing the reports with respect to the strategies and techniques described by the learners **I** would like to draw some conclusions.
(01-NS0-1990-JA-C; 1 person singular)
- (11) Tomiyama reports on the findings of various studies in L2 attrition and draws up a list of concluding statements deriving from them. **I** would like to compare the extent to which **my** data is in concord with the reported results by Tomiyama.
(09-MS1-2003-JA-C; 1 person singular)
- (12) Generally, the perceptions of **my** subjects and those in Czekaj's study (2009) as highlighted above were fairly homogenous but not totally.
(16-ES1-2010-CH-C; 1 person singular)

With regard to epistemic modality, the focus of the third part of the analysis, there is a downward tendency in the use of the markers from 128 per 10,000 words in the NS (Ex. 13-14), through 111 in the MS (Ex. 15), to 57 in the ES (Ex. 16-18). The most rapid fall, however, occurs between the MS and ES and is primarily connected with low-value epistemic markers, whose frequency per 10,000 words fell from 69, through 55 to 18 in the ES (Table 9). When we look at the relative frequency of epistemic markers at different stages of academic career, shown in Table 10, it can be seen that low-value is the most frequently marked modality type in the first two stages while middle-value modality predominates in the ES. The relative importance of high-value modality seems to be increasing, but the differences are very small.

This downward trend in the numbers of epistemic markers is an unexpected result. Previous research has shown that epistemic modality is marked considerably more frequently in English research articles than in Polish-language articles, and that this difference holds for all three modal values, although it is more significant in low- and high-value markers than in middle values markers (Warchał 2015). It could be expected, then, that when writing in English, Polish writers would transfer their habits and intuitions from L1 and mark epistemic modality less often compared to Anglophone authors. Also, one would expect that – with the author's growing academic expertise as well as experience and confidence as an L2 writer – these preferences and intuitions would change in time, becoming more like those of Anglophone scholars (which would mean, in principle, more epistemic markers and a higher ratio of low- and high-value markers). In the present case, however, these expectations prove incorrect. To understand these findings, more research is needed into the use of epistemic markers by Polish scholars writing in English.

Table 9. Epistemic modality markers in conclusions at different career stages

Epistemic value	NS	per 10,000 words	MS	per 10,000 words	ES	per 10,000 words
High-value markers	2	11	7	15	10	9
Middle-value markers	9	48	19	41	35	30
Low-value markers	13	69	25	55	21	18
Total	24	128	51	111	66	57

Table 10. Relative frequency of epistemic modality markers (in percent) for different career stages

Epistemic value	NS	MS	ES
High-value epistemic markers	8.3	13.7	15.2
Middle-value epistemic markers	37.5	37.3	53
Low-value epistemic markers	54.2	49	31.8
Total	100	100	100

- (13) Though the reason **may** be found in the fact that little or no instruction is given to learners in how to learn, i.e. how to make conscious use of certain language regularities (for example, use of affixes, inferencing or guessing meanings from the context).
(01-NS0-1990-JA-C; low-value EM [1]; Accounting for results)
- (14) An understanding of the metaphorical nature of language **may** have implications for the acquisition of, for example, verb + noun collocations of the type to waste/spare time. Besides lexical phrases (collocations), metaphors **could probably** be used in teaching Pre^o positions and phrasal verbs, especially so-called orientational metaphors which refer to the relation of our bodies to our surroundings.
(02-NS0-1992-CH-C; low-value EM [2], middle value EM [1]; Drawing pedagogic implications)
- (15) In the context of LI, they clearly are inborn, intuitive, implicit and as if automatic. In the context of a FL it **will** be language exposure, for example the language instructor's input – his/her choice of modality, the type of didactic/non-didactic materials used...
(12-MS2-2006-JA-C; high-value EM [1]; Interpreting results)
- (16) Literature on the subject does not comment at any significant length on the above discussed issues, **perhaps** with one exception of the issue of a seating arrangement and its significance for the effectiveness of different student groupings. The theme of classroom space and its individual dimensions as presented here **would** be worth a more in-depth research.
(16-ES1-2010-CH-C; low-value EM [1], middle value EM [1]; Recommending further research)
- (17) It **can be assumed** that this relation shows the link between learning different languages in a sequence (and not simultaneously) and the

role the subjects attach to their prior experiences and their individual ways of approaching learning in the past.

(21-ES1-2013-CH-C; middle-value EM [1]; Interpreting results)

- (18) These comments **seem** to express a rather traditional view, in which learners appear to be passive recipients of what is more or less imposed upon them.

(26-ES2-2016-CH-C; middle-value EM [1]; Interpreting results)

5. Discussion and concluding remarks

This paper looked into concluding sections of two academic genres, journal articles and book chapters, in an attempt to see whether the rhetorical organisation of the final sections evolved with the growing academic experience of the writer, whether the way of marking authorial presence changed in time, and whether the use of epistemic modality markers was different at different stages of the author's academic career. Considering the amount and type of data used for analysis, one should perhaps start with the constraints they impose on the interpretation of the results. Apart from the usual limitations inseparable from case studies, which focus on what is unique rather than what can be generalised (Casanave 2015; Rose et al. 2020), an important limiting factor is the necessarily unequal number of texts that represent different time-frames. The novice stage of early-career researchers is naturally shorter and, in many cases, less prolific than more advanced stages of the academic career. This is exactly the case in the present study. Although it reflects the dynamics of academic development, the limited number of NS texts may be a factor that distorts the results. Next, since the results come from a single-case study, they may be more useful in formulating questions for further research into the development of academic writers than in providing a set of strong conclusions.

The first important observation that emerges from this study is that in the ES, conclusions tend to be much longer than in the earlier stages. The length of sections that follow the presentation of results reflects the increased need to provide an in-depth interpretation of the findings – by pointing to possible reasons for or consequences of the studied problems or relationships – and to contextualise the results more thoroughly in the existing body of research by identifying these aspects in which they complement, contradict, or tally with previous studies.

Second, in spite of the difference in length, conclusions in the MS and ES are rhetorically similar, the core formula comprising the following moves/steps: Main results, Interpreting results, Comparing with literature, and Drawing pedagogic implications. By contrast, NS conclusions tend to focus on restating the problem rather than summarising the main results. Also, Interpreting results is a rare step in the NS; instead, much more space is devoted to Comparing with literature. The rhetorical similarity between the MS and ES texts, on the one hand, and the differences between these texts and the NS texts, on the other, very well reflect the two-fold distinction between novice and expert scholars, often applied in English for Academic Purposes research (e.g., Li – Flowerdew 2007; Dressen-Hammouda 2008; Rowley-Jolivet – Carter-Thomas 2014; Katsampoxaki – Hodgetts 2022). At the same, the results for all three stages confirm the importance of references to previous studies in linguistics research articles. In his interdisciplinary analysis, Peacock (2002) notes that this move is more often attested in Language and Linguistics than in other disciplines, approaching 90% of non-native speakers' texts. Although in the present case the percentage is lower (67-87%, depending on the stage), Comparing with literature is one of the two most frequently occurring moves in all three time-frames.

Interestingly, references to literature seem to play a somewhat different role in the NS than in the more advanced stages. In the NS, they often report others' findings which, in a sense, give credibility to the author's results presented in the preceding section. By contrast, in the MS and ES, they complement, support, or provide additional context for the results, which are discussed and interpreted in the same section. The use of sources in the NS – in place of the interpretation of the author's own findings – may reflect the reluctance to rely on one's own voice and budding expertise, which interpretation of results demands. These findings are not incompatible with Mansourizadeh and Ahmad (2011), an analysis of citation practices among expert and novice writers in the field of chemical engineering. While the authors note that expert writers tend to include more citations than novices, they point out that they do that for different reasons. Novice scholars cite mainly for attribution (i.e., to acknowledge the source of information), while expert writers do so to establish links with and between previous studies and to support their own claims or justify the findings. The tendency to combine citations with stance-taking in non-native speaker expert writers' texts (including distancing oneself from the original author's claim, conceding the cited claim, and expressing commitment to it) is also pointed out by

Rowley-Jolivet and Carter-Thomas (2014). Our data, although subjected to a different type of analysis, seem to reveal a similar pattern.

With regard to authorial presence as expressed by personal pronouns and self-citations at different career stages, the differences are small, with inclusive *we* being the most frequently applied self-referring pronoun in all time-frames. The only significant but expected change is the rise in the number of self-citations in the ES, a result of the author's long-standing academic interests, growing output, and increasing academic impact. It is worth noting that the preference for the plural first-person pronoun was also noted by Kuo (1999) in a corpus of journal articles from computer sciences, electronic engineering, and physics. In this case, however, exclusive *we* was the most frequent type, a difference that could perhaps be explained by disciplinary variation.³

Perhaps least expected are the results concerning epistemic modality. The frequency of epistemic markers was found to fall rapidly at the ES, which may be somewhat counterintuitive for two reasons. First, one would expect epistemic modality markers to concur with Interpreting results, a step that involves moving beyond reporting facts and that in our data was more often attested in MS and ES conclusions than in the NS.⁴ Second, previous research has shown that Polish-language articles contain considerably fewer epistemic markers than do articles by Anglophone academic scholars (Warchał 2015), a difference that is a likely source of transfer to English as an L2. This may lead us to expect that at the later stages of the academic career, the author, by that time an experienced L2 researcher, would more often mark her epistemic stance than in the early texts. These expectations are not confirmed by the data. Interpreting these findings would require more research into the use of epistemic modality by novice and experienced academic writers.

Additionally, the analysis draws attention to the role of book chapters in the dissemination of results in the humanities and social sciences in the Polish context. Despite the current emphasis on journal articles, reflected in evaluation systems and various incentives for academic authors to publish in journals rather than contribute to multi-authored monographs, the latter have long been important venues for presenting research results in these

³ See also Wang et al. (2021), where exclusive *we* was found to dominate over the inclusive use of the pronoun in a corpus of articles from the field of electrical and electronic engineering.

⁴ See also Warchał (2023) for the occurrence of modality markers in specific rhetorical moves in conclusions to research articles.

fields of scholarship (see also Engels et al. 2018 for a discussion of the future of monographs and book chapters in the social sciences and humanities).

A single-case study would offer a very poor starting point for even cautious generalisations, yet generalisations are not the aim of this analysis. Rather, it is hoped that the findings may help identify some directions for future research on the process of developing academic identity and writing competence. A particularly interesting problem seems to be the evolution of epistemic modality marking in texts produced at various stages of academic career. Following Moreno (2021), it would also be interesting to look into promotional strategies employed by novice and experienced scholars. Still another promising direction would be the role of references to literature in interpreting one's own results, an aspect that seems to distinguish experts from novice writers. Also, the fact that conclusions produced in the novice stage are markedly different in their rhetorical structure from those produced at later stages offers additional support to the idea of writing accountability groups and other forms of writing facilitation as solutions that may help young academics build writer's confidence and develop as writers (Chu 2022).

Acknowledgements

I would like to thank the Author whose writings provided the material for this analysis for her consent to use the texts, for making some of the earlier publications available for the study, and for her time and patience with the questions I asked during the compilation of the corpus.

REFERENCES

- Arnaudet, M. L. – M. E. Barrett
 1984 *Approaches to Academic Reading and Writing*. New York: Prentice-Hall.
- Bhatia, V. K.
 1993 *Analysing Genre. Language Use in Professional Settings*. London: Longman. <https://doi.org/10.4324/9781315844992>.
- Biber, D.
 2006 "Stance in spoken and written university registers", *Journal of English for Academic Purposes* 5, 97-116.
<https://doi.org/10.1016/j.jeap.2006.05.001>.

- Bruce, I.
2008 "Cognitive genre structures in Methods sections of research articles: A corpus study". *Journal of English for Academic Purposes* 7, 38-54.
<https://doi.org/10.1016/j.jeap.2007.12.001>.
- Carter-Thomas, S. – E. Rowley-Jolivet
2003 "Analysing the scientific conference presentation (CP). A methodological overview of a multimodal genre", *ASp* 39-40, 59-72. <https://doi.org/10.4000/asp.1295>.
- Casanave, C. P.
2015 "Case studies". In: B. Paltridge – A. Phakiti (eds.) *Research Methods in Applied Linguistics*. London: Bloomsbury, 119-135.
- Chu, A. G.
2022 "Research Rituals – Finding the value of writing accountability groups", *LSE Academic Writing Blog*, 17th June, <https://blogs.lse.ac.uk/impactofsocialsciences/2022/06/17/research-rituals-finding-the-value-of-writing-accountability-groups/>, accessed November 2023.
- Connor, U. – A. Mauranen
1999 "Linguistic analysis of grant proposals: European Union research grants", *English for Specific Purposes* 18 (1), 47-62.
[https://doi.org/10.1016/s0889-4906\(97\)00026-4](https://doi.org/10.1016/s0889-4906(97)00026-4).
- Cross, C. – C. Oppenheim
2006 "A genre analysis of scientific abstracts", *Journal of Documentation* 62 (4), 428-446. <https://doi.org/10.1108/00220410610700953>.
- Dressen-Hammouda, D.
2008 "From novice to disciplinary expert: Disciplinary identity and genre mastery", *English for Specific Purposes* 27, 233-252.
<https://doi.org/10.1016/j.esp.2007.07.006>.
- Dudley-Evans, T.
1994 "Genre analysis: An approach to text analysis for ESP". In: M. Coulthard (ed.) *Advances in Written Text Analysis*. London/New York: Routledge, 219-228.
- Engels, T. C. E. et al.
2018 "Are book publications disappearing from scholarly communication in the social sciences and humanities?", *Aslib Journal of Information Management* 70 (6), 592-607.
<https://doi.org/10.1108/AJIM-05-2018-0127>.
- Feng, H. – L. Shi
2004 "Genre analysis of research grant proposals", *LSP & Professional Communication* 4 (1), 8-30.
- Flyvbjerg, B.
2011 "Case study". In: N. K. Denzin – Y. S. Lincoln (eds.) *The SAGE Handbook of Qualitative Research* (4th edn.). Thousand Oaks: SAGE, 301-316.

- Hewings, M. – C. Thaine
2012 *Cambridge Academic English: An Integrated Skills Course for EAP. Advanced*. Cambridge: Cambridge University Press.
- 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 (4), 321-337.
[https://doi.org/10.1016/s0889-4906\(96\)00038-5](https://doi.org/10.1016/s0889-4906(96)00038-5).
- Hopkins, A. – T. Dudley-Evans
1988 "A genre-based investigation of the discussion sections in articles and dissertations", *English for Specific Purposes* 7 (2), 113-121.
[https://doi.org/10.1016/0889-4906\(88\)90029-4](https://doi.org/10.1016/0889-4906(88)90029-4)
- Hyland, K.
2005 "Stance and engagement: A model of interaction in academic discourse", *Discourse Studies* 7 (2), 173-192.
<https://doi.org/10.1177/1461445605050365>.
- Kashiha, H.
2015 "Recurrent formulas and moves in writing research article conclusions among native and non-native writers", *3L, Language, Linguistics, Literature: The Southeast Asian Journal of English Language Studies* 21 (1), 47-59. <https://doi.org/10.17576/3L-2015-2101-05>.
- Katsampoxaki-Hodgetts, K.
2022 "The emergence of a new inclusive meta-genre; 'the Bigger Picture'", *Journal of English for Academic Purposes* 57, 101-114,
<https://doi.org/10.1016/j.jeap.2022.101114>.
- Kuo, C. H.
1999 "The use of personal pronouns: Role relationships in scientific journal articles", *English for Specific Purposes* 18 (2), 121-138.
[https://doi.org/10.1016/s0889-4906\(97\)00058-6](https://doi.org/10.1016/s0889-4906(97)00058-6).
- Li, Y. – J. Flowerdew
2007 "Shaping Chinese novice scientists' manuscripts for publication", *Journal of Second Language Writing* 16 (2), 100-117.
<https://doi.org/10.1016/j.jslw.2007.05.001>.
- Liu, J. – Xiao, L.
2022 "A multi-dimensional analysis of conclusions in research articles: Variation across disciplines", *English for Specific Purposes* 67, 46-61.
<https://doi.org/10.1016/j.esp.2022.04.003>.
- Lorés Sanz, R.
2004 "On RA abstracts: From rhetorical structure to thematic organisation", *English for Specific Purposes* 23 (3), 280-302.
<https://doi.org/10.1016/j.esp.2003.06.001>.
- Luzón Marco, M. J.
2002 "A genre analysis of corporate home pages", *LSP & Professional Communication* 2 (1), 41-56.

Mansourizadeh, K. – U. K. Ahmad

- 2011 "Citation practices among non-native expert and novice scientific writers", *Journal of English for Academic Purposes* 10 (3), 152-161.
<https://doi.org/10.1016/j.jeap.2011.03.004>.

Moreno, A. I.

- 2021 "Selling research in RA discussion sections through English and Spanish: An intercultural rhetoric approach", *English for Specific Purposes* 63, 1-17. <https://doi.org/10.1016/j.esp.2021.02.002>.

Peacock, M.

- 2002 "Communicative moves in the discussion section of research articles", *System* 30 (4), 479-497. [https://doi.org/10.1016/s0346-251x\(02\)00050-7](https://doi.org/10.1016/s0346-251x(02)00050-7).

Pinto dos Santos, V. B. M.

- 2002 "Genre analysis of business letters of negotiation". *English for Specific Purposes* 21 (2), 167-199. [https://doi.org/10.1016/s0889-4906\(00\)00028-4](https://doi.org/10.1016/s0889-4906(00)00028-4).

Rañosa Madrunio, M.

- 2012 "Move sequences in graduate research paper introductions and conclusions", *Journal on English Language Teaching* 2 (1), 1-14.
<https://doi.org/10.26634/jelt.2.1.1615>.

Rose, H. – J. McKinley – J. Briggs Baffoe-Djan

- 2020 *Data Collection Research Methods in Applied Linguistics*. London: Bloomsbury. <https://doi.org/10.5040/9781350025875>.

Rowley-Jolivet, E. – S. Carter-Thomas

- 2005 "The rhetoric of conference presentation introductions: Context, argument and interaction", *International Journal of Applied Linguistics* 15 (1), 45-70. <https://doi.org/10.1111/j.1473-4192.2005.00080.x>.
 2014 "Citation practices of expert French writers of English: Issues of attribution and stance". In: A. Lyda – K. Warchal (eds.) *Occupying Niches: Interculturality, Cross-culturality and Aculturality in Academic Research*. Cham: Springer, 17-34. https://doi.org/10.1007/978-3-319-02526-1_2.

Shamsudin, S. – S. J. Ebrahimi

- 2013 "Analysis of the moves of engineering lecture introductions", *Procedia Social and Behavioral Sciences* 70, 1303-1311.
<https://doi.org/10.1016/j.sbspro.2013.01.191>.

Stake, R. E.

- 2003 "Case studies". In: N. K. Denzin – Y. S. Lincoln (eds.) *Strategies of Qualitative Inquiry* (2nd edn.). Thousand Oaks, CA: Sage, 134-164.

Swales, J. M.

- 1981 *Aspects of Article Introductions*. Birmingham: The University of Aston.
 1990 *Genre Analysis: English in Academic and Research Settings*. Cambridge: Cambridge University Press.

Swales, J. M. – C. B. Feak

- 2012 *Academic Writing for Graduate Students: Essential Tasks and Skills*. Ann Arbor: University of Michigan Press (3rd edn.).
<https://doi.org/10.3998/mpub.2173936>.

Thompson, S. E.

- 1994 "Frameworks and contexts: A genre-based approach to analyzing lecture introductions", *English for Specific Purposes* 13 (2), 171-186.
[https://doi.org/10.1016/0889-4906\(94\)90014-0](https://doi.org/10.1016/0889-4906(94)90014-0).

Wallace, M. – A. Wray

- 2016 *Critical Reading and Writing for Postgraduates* (3rd edn.). Thousand Oaks, CA: Sage.

Wang, S. P. – W. T. Tseng – R. Johanson

- 2021 "To *we* or not to *we*: Corpus-based research on first-person pronoun use in abstracts and conclusions", *SAGE Open* 11 (2), 1-18.
<https://doi.org/10.1177/21582440211008893>.

Warchał, K.

- 2015 *Certainty and Doubt in Academic Discourse: Epistemic Modality Markers in English and Polish Linguistics Articles*. Silesia: University of Silesia Press. <https://opus.us.edu.pl/info/book/USL84c6adad609d4d188dcd8eec214aacb9/>.
- 2023 "Beyond mere facts: Epistemic profiles of conclusions to English- and Polish-language articles", *Studia Linguistica Universitatis Iagellonicae Cracoviensis* 140 (2), 165-189.
<https://doi.org/10.4467/20834624SL.23.008.17757>.

Yang, R. – D. Allison

- 2003 "Research articles in applied linguistics: Moving from results to conclusions", *English for Specific Purposes* 22 (4), 365-385.
[https://doi.org/10.1016/s0889-4906\(02\)00026-1](https://doi.org/10.1016/s0889-4906(02)00026-1).

Address: KRYSTYNA WARCHAŁ, Faculty of Humanities, University of Silesia, ul. Grot-Roweckiego 5, 41-200 Sosnowiec, Poland.

ORCID code: <http://orcid.org/0000-0002-8422-4911>.

English MA theses at a German university before and after the Bologna reform: Comparing global rhetorical structures and stance in Linguistics and Cultural Studies¹

Josef Schmied and Marina Ivanova
Chemnitz University of Technology

ABSTRACT

This contribution analyses changing practices of academic writing with special reference to MA theses. It considers theses as the central genre that introduces students to independent academic writing and thinking and the first step towards an academic writing career. A small empirical case study compares 20 MA theses written at Chemnitz University of Technology after the introduction of the Bologna reform (2012-20) with 20 Magister theses written before (2002-12). Two well-known metadiscourse variables are analysed: global rhetorical structures, i.e., IMRAD, from Introduction/issue, Methodology, Research (question) to Discussion or conclusion, and the expression of personal evaluation through metalanguage, i.e. stance. The focus is on a particularly interesting stance variable, evaluative *that* complement clauses (e.g., *suggest that*, *claim that*) and their functions indicating different strengths of authorial stance. A corpus-linguistic analysis reveals important differences between the English subdisciplines. The MA theses show similar teaching-induced trends, though to a different degree. Linguistics follows the patterns and perceived standards in international social science models more than Cultural Studies. This can be interpreted as functional adaptation to changing rhetorical situations in a wide social context.

Keywords: MA theses, global rhetorical structures, stance, *that* complements, explicitness, writer identity, Bologna reform.

¹ We would like to thank all students who sent us their “masterpieces” to be included in the corpus and to all student assistants who helped to compile and anonymise the texts over the years.

1. Introduction

This contribution focusses on the beginning of academic writing at German (and European) universities, the longest texts (60+ pages) in Humanities studies in Germany, called Magister thesis before and Master theses after the Bologna reform. These texts are the first research attempts by young graduates, who have to come to terms with their own personal topic, combine a digested literature review with their own analyses, and deliver a complex readable pattern of sections and paragraphs to their supervisors. Thus their thinking and writing is narrowed down from a seemingly infinite variation of structures to what they perceive as the most presentable text in a given institutional context. These theses usually anticipate their first research articles, which have been the focus of academic writing research over the last few decades. This is the formative period for academic novices, in which they experiment with developing their identity as academic writers and which (together with their PhD supervision afterwards) is the beginning of “shaping rhetorical subjects” (cf. Paré – Starke-Meyerring – McAlpine 2011).

One specific feature of the genre has to be mentioned here: in theory, MA students write for a wider academic audience; in practice they may attempt to please their supervisors and thus reproduce their models to fulfil their expectations probably more than when writing for a journal or conference. This is important for developments in academic genres, which have become more dialogic, both rhetorically and structurally, at the surface (with pronouns addressing the reader) and at a deeper level of organisation (with argumentative structures). As other qualification texts from BA to PhD theses, the dialogue widens, until other core genres like conference presentations, reviews and journal articles address theoretically the entire research community in the subdiscipline. English departments in Germany usually include at least the four sections Literature and Cultural Studies, Linguistics and Methodology and, although there may be considerable overlap (especially between the first two and the second two), they all have their specific research practices and functional conventions and thus are called (sub-)disciplines here (Schmied 2015: 11). In the overview by Pérez-Llantada (2021: 28), MA theses are not listed, because they are written not only before the tenure-track genres, but also even before the early career-related genres, which start with PhD theses; hence, MA theses are under-researched, although they are the decisive test whether a student should venture an academic career.

The Bologna reform (since its declaration in 1999) offered the opportunity to modernise degree programmes more radically than before, not only taking account of the teaching experience at home, but also of discussions with European colleagues:

The Bologna Declaration ... aims to create a European Higher Education Area (EHEA) that ... stipulates comprehensible and comparable degrees, a two-cycle system of study programmes, the introduction of a credit transfer system (ECTS), quality assurance mechanisms and enhanced mobility, and the development of an European dimension in HE [Higher Education]. (Kuhlee 2017: 301)

The second cycle was explicitly defined as scientific, thus, the MA theses analysed here represent the beginning of the personal development as young scientific writers. The reform included an inherent tension between creating a uniform EHEA and maintaining the diversity of national HE systems. The resulting simultaneous process of convergence with and divergence from Anglophone academic writing norms led to hybridization (Pérez-Llantada 2013: 264) or ‘glocalization’ (Swales 2004: 11; Sancho Guinda 2015: 29) of the discourse of non-Anglophone scholars. The diversity of culture-specific non-native variants of academic English was particularly aptly grasped by the term ‘alternative academic written Englishes’ (Mauranen – Pérez-Llantada – Swales 2010: 671).

However, apart from globalisation, digitalisation has also changed higher education in Germany: online access to different academic genres has made it much easier to find models for discipline-specific theses, self-study materials and complete lectures on academic writing, etc. All this may influence students’ rhetorical choices and cannot be controlled in a small corpus. The trend towards digitalisation is linked to the trend to Americanisation, for instance in text processing, where American spelling and style rules are default, and in style sheets, where APA conventions have become standard in Linguistics like MLA conventions in Literary Studies.

Finally, recent trends in scientific work have focused on replicability, e.g., in the context of the Common Language Resources and Technology Infrastructure (CLARIN) as part of the European Open Science Cloud. Here new digital affordances are linked with new political perspectives that see academic knowledge not as a commercial commodity, but as a general public good (Luzón – Pérez-Llantada 2019). This cumulative pressure of technological and societal challenges on all members of the academic

community to adapt to these trends has caused accelerated publication and citation cycles in a “publish in English or perish” world (Pérez-Llantada 2012), which is particularly challenging for young non-native writers. This has led to altered academic discourse trends and developments: especially young authors have to aim at enhancing their visibility and credibility as members of a research community by increasing their promotional writer identity, producing individual texts characterised by informativeness and surveyability. This can be achieved by adding features of explicitness (i.e., *that* clauses, stance markers, etc.) and reader-oriented metadiscourse (see below, elaborated in section 3.3), following modern professional (sub-) disciplinary specific conventions in dominant discourse communities and prestigious journals, especially when texts are presented and read only in digital form in direct competition with similar publications.

2. Pre- and post-Bologna systems at German universities:
Conventions

Although the length and preparation time of the final theses can vary slightly across German universities and disciplines, the general structure of the Magister, Bachelor and Master programmes were harmonised by documents compiled by selected university professors in the name of the German Rectors Conference (HRK). Table 1 is based on the study regulations of the English Studies programme at Chemnitz University of Technology, as the data analysed in this paper come from the same university as part of the ChemCorpus (Schmied – Dheskali 2015). This database comprises student academic works over 20 years and is especially useful for the internal analysis of metadiscourse (Schmied 2015).

Table 1. Magister, Bachelor and Master studies structure at Chemnitz University of Technology

Magister	Bachelor
<ul style="list-style-type: none">• Introductory studies (1st-4th semester)• Main studies (5th-8th semester)• Magister thesis (80-120 p., 24? weeks)	<ul style="list-style-type: none">• Main studies (1st-5th semester)• Bachelor thesis (6th semester, 40-60 p., 16 weeks)
	Master
	<ul style="list-style-type: none">• Main studies (1st-3rd semester)• Master thesis (4th semester, 60-80 p., 23 weeks)

Before the Bologna reform, the most common university qualification as a first degree in Humanities was Magister, which lasted at least four years and consisted of four semesters of introductory studies up to an “intermediate exam”, followed by at least four semesters of main studies. After a sufficient number of course certificates were collected from a wide range of options, the studies were completed by a Magister thesis as well as written and oral exams. After the Bologna reform, a two-tier structure was implemented: Bachelor with six semesters of main studies including a Bachelor thesis in the final semester, and Master with four semesters of main studies including a Master’s thesis in the final semester. The new system was modularised in a European Credit Transfer System (with 30 points every semester) to make it more coherent or strict and to adapt study times and content to international European standards to make international mobility simple and fully integrated. Apart from the obvious length differences (more years, but shorter theses), the teaching programme has undergone changes in structure and general academic skills, especially in academic writing instruction. The effects of these changes are explored in this paper by comparing 20 Magister with 20 Master’s theses, written by students with German as their major (academic) language and mostly as their first language (18/20) in non-native, academic *lingua franca* English in the subdisciplines (and MA specialisations) Linguistics and Cultural Studies. This choice is motivated by departmental conventions, but also by the focus of Cultural Studies on native English cultures from a literary or social science perspective and the focus of Linguistics on a more global empirical perspective.

3. Review of key concepts

3.1 Global rhetorical structures: Moves

Global rhetorical structures (Pérez-Llantada 2013) have been analysed and compared since Swales’ seminal work (e.g., 1990, 2004). His Creating a Research Space [C.A.R.S.] Model is based upon his analysis of journal articles (representing a variety of discipline-based writing practices) and attempts to explain and describe the organizational pattern of writing the introduction to scholarly research studies. It has been expanded to a complete sequence of rhetorical moves. It has also been used in teaching for decades (e.g., Swales – Feak 2012 [1994], cf. Pérez-Llantada 2021: 161-197). Table 2 presents the model outlined by Pho (2013: 31) and also the Discussion/Conclusion moves by Yang and Allison (2003: 378). Although the database and the categories

of research articles and their abstracts may not always be identical with that for MA theses, they are, slightly adapted, a good basis for our comparison.

Table 2. Research article moves (adapted based on Pho 2013: 31 and Yang & Allison 2003: 378)

Section	Moves	
Abstract	M1: Situating the research	
	M2: Presenting the research	
	M3: Describing the methodology	
	M4: Summarizing the findings	
	M5: Discussing the research	
Introduction	M6: Establishing a territory	
	M7: Establishing a niche	
	M8: Presenting the present work	
Methods	M9: Describing the data and data collection procedure	
	M10: Describing the data analysis procedure	
Results	M11: Preparing for the presentation of the Results section	
	M12: Reporting specific/individual results	
	M13: Commenting on specific results	
	M14: Summarizing results	
Discussion– Conclusions	Pho (2013: 31)	Yang & Allison (2003: 378)
	M15: Preparing for the presentation of the Discussion section	M15: Background information
	M16: Summarizing the study	M16: Reporting results
	M17: Highlighting overall research outcome	M17: Summarising results
	M18: Discussing the findings of the study	M18: Commenting on results
	M19: Drawing conclusions of the study/Stating research conclusions	M19: Summarizing the study
	M20: Evaluating the study	M20: Evaluating the study
		M21: Deductions from the research

Like in many other European universities, these principles of research writing were introduced at Chemnitz University of Technology with

the new 2-year MA programme in the English department in 2006. They were taught to all MA students in their first semester, but they were not emphasised by all supervisors in the practical research seminars at the end of the programme when students wrote their Master's theses. A teaching example is Table 3, which summarises the research article AIMAC structure, based on Schmied's revised IMRAD model (Schmied 2015: 6).

Table 3. Research article AIMAC structure – revised IMRAD model (based on Schmied 2014: 6)

Structure	Substructure	Key lexemes/phrases (as indicators)
A = abstract	Keywords in context	Focus, discuss, approach
I = issue	New	Not enough research yet
	Relevant	Important, practical application
	Focused	Concentrate, emphasise, purpose
M = methodology	Previous research, i.e., lit. review incl. evaluation	Concept developed, review, refer to, proceed to, claim
	Hypotheses possible?	Research question
	Data base	Corpus, data collection
	Tests/procedure	Calculate
A = analysis	Examples as evidence	Illustrate, show, prove
	Statistical tables as summaries	Table, figure, diagram, graph, bar
	Significance to generalise	Significant
C = conclusion	Summary	In conclusion, finally/at last, we have shown, discussed above
	Interpretation	This proves that
	Contextualisation	In a wider perspective, apply
	Limitations	More data, beyond the scope
	Outlook	Further research is necessary, predict, application of results

Finally, the new emerging pattern of rhetorical structures is partly reflected in the titles and subtitles of the linguistics theses. Data come to the fore, occasionally through an explanatory prototypical title quotation (1) and partly

through explicit mention of well-known international data collections (2) or parallel own compilations (3):

- (1) “Want to see more?” Null subjects in Facebook status updates (MAL12Ft_DB)
- (2) Modals in Kenyan English: A contrastive corpus analysis of modals in the ICE-K and the ICE-GB (MGL04Ft_KS)
- (3) Sentence Connectors in English Academic Writing – An Empirical Comparison of Research Articles by German and Native English Writers (MGL08Ft_EW)

Table 4 shows the clear differences between the MA theses in the two periods, which have to be seen parallel to the developments in related genres. Thus, it is not surprising that the Czech research articles (Dontcheva-Navratilova this issue) show a parallel development in the functional focus (Table 4) as well as the formal complexity of titles (Table 5). The increasing complexity of thesis titles can be seen in the choice of subtitles and complex phrases with compounds and other noun phrase complexities in general – all can be seen in a Linguistics title like (4), whereas simple titles can still be found in Cultural Studies (5). Even so, Cultural Studies titles have generally become longer and thereby more similar to those of Linguistics, as evident from Figure 1 and examples (6) and (7). This increases the specificity of a text and the individuality of young writers, features that seem desirable to those who wish to establish an academic identity in today’s competitive academic world.

- (4) Corpus-Linguistic and Cognitive Approaches to Determiner Usage in Chinese Student Writing. Testing the Fluctuation Hypothesis (MAL16Mt_SA)
- (5) Devolution and National Identity in Wales and Scotland (MAC11Mt_ES)
- (6) Britain and the transatlantic slave trade (MGC07Ft_ID)
- (7) Lobbyism during the Prohibition Era of the United States – The Impact of Organized Crime on Federal Politics between 1919 and 1933 (MAC14Mt_MM)

Table 4. Content-related features of MA theses titles

Degree	Topic	Topic + Dataset	Topic + Example	Topic + Method	Topic + Results	Topic + Conclusion
Cultural studies	20	8	2	3	8	1
Magister	10	3	1	1	4	1
Master	10	5	1	2	4	0
Linguistics	20	14	3	13	6	1
Magister	10	6	0	7	4	0
Master	10	8	3	6	2	1
Grand Total	40	22	5	16	14	2

Table 5. Formal features of MA theses titles

Degree	Nominal	Prep. phr.	V-ing phr.	Compound	Sub title	Length (M SD)	
Cultural studies	9	18	4	4	15	12.3	4.28
Magister	3	8	2	2	7	10.9	3.70
Master	6	10	2	2	8	13.7	4.55
Linguistics	14	18	4	13	18	14.6	5.07
Magister	7	9	1	5	8	14.5	3.84
Master	7	9	3	8	10	14.7	6.29
Grand Total	23	36	8	17	33	13.45	4.78

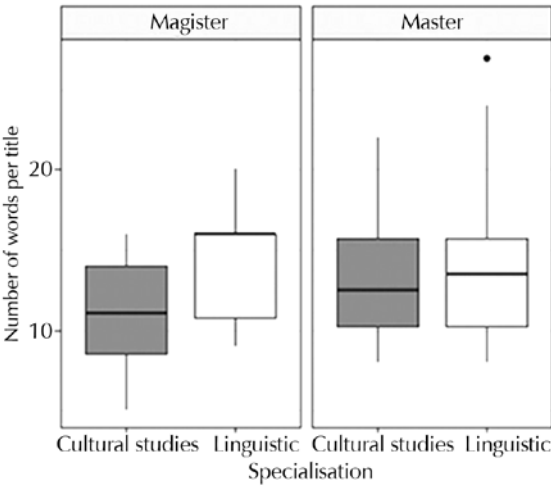


Figure 1. Word count of Magister and Master's thesis titles in Cultural studies and Linguistics

3.2 Stance

Stance indicates the emphasizing or softening of a speech act's illocutionary force (Holmes 1984: 346). It can be classified in terms of affect (positive or negative stance) and evidentiality (certainty or doubt) (Biber – Finegan 1989: 98). It channels the writer's textual 'voice' – how writers present themselves and convey their judgements, opinions, and commitments (Hyland 2005a: 176). It conveys "attitudes that a speaker has about information, how certain they are about its veracity, how they obtained access to the information, and what perspective they are taking" (Biber 2006: 87). Stance establishes dialogistic positioning (Martin – White 2005: 97), a negotiation of the discussed information and an implicit dialogue with the scientific community.

In the vast literature on stance, there are different views on the scope of stance. In appraisal theory, all utterances are seen to be in some way stanced or attitudinal (Martin – White 2005: 92). Other works identify emotive and nonemotive controlling words (Pho 2013: 37), which implies an existing neutrality. Expressing attitude or stance can be viewed as an act of evaluation (Thompson – Hunston 2000). Evaluation has been widely studied; for instance, attention has been paid to how peer reviews and book reviews express praise and criticism; here Diani (2004, 2017) showed that this is often achieved through adversative connectors. Evaluation in professional digital communication is another interesting case – while consumer reviews obviously differ generically from academic peer reviews, they share similarities in the negotiation of praise and criticism through concession (Ivanova 2020; Schmied 2021). Although Thompson and Hunston (2000) could not include recent developments in professional digital communication, the close connection between stance and evaluation was mentioned explicitly in their definition:

evaluation is the broad cover term for the expression of the speaker or writer's attitude or stance towards, viewpoint on, or feelings about the entities or propositions that he or she is talking about. That attitude may relate to certainty or obligation or desirability or any of a number of other sets of values. (Thompson – Hunston 2000: 5)

3.3 Stance in *that* complement clauses

In the long lists of metadiscourse features (e.g., Hyland 2005b: 218-224), *that* complement clauses are hardly analysed quantitatively, although or because

they are used very frequently. Even though authors may be less aware of this stance feature, it is a particularly interesting variable in metadiscourse corpus-linguistic analyses (Hyland – Tse [2005a] and Kim – Crosthwaite [2019]), since they are used more often in disciplines that place greater emphasis on authors generating their arguments than on authors presenting “facts”. From a functional perspective, *that* complement clauses are powerful structures commonly used for evaluation through extraposition. Hyland and Tse (2005a) refer to them as “evaluative *that*-structures” and define them as “a grammatical pattern in which a *that* complement clause is contained in a higher superordinate clause to complete its construction and which together project the writer’s attitudes or ideas about something” (Hyland – Tse 2005a: 40). The evaluated entity is postponed through extraposition (Quirk et al. 1985: 1391-1393). Example (8) from a Master thesis shows the evaluated entity in the complement clause (*that meronymy and hyponym are not used in their data...*) after the critical evaluation (*claim*) and its source (*the researchers*).

- (8) the researchers claim | that meronymy and hyponym are not used in their data... (MA16Ft_JZ)

super-ordinate (matrix) clause: source and evaluation | complement clause: evaluated entity

Biber (2006: 106) found that *that* complement clauses generally are more common in the spoken university registers than in the written registers, where, more specifically, certainty verbs (*know that*) and likelihood verbs (*guess that*) were most common. Related research on the use of reporting verbs to express stance and evaluation in academic writing can be found on expert and learner level. Bondi (2009) investigated monologic and dialogic evaluation through reporting verbs in book reviews and found cultural variation: the English corpus had more dialogic representations, while the Italian had more monologic expressions. In student writing, Dontcheva-Navratilova (2008) found increased use and variation of reporting verbs in Czech English graduate theses compared to undergraduate theses. These research results are relevant here when we want to observe whether these trends are confirmed in our theses.

In the rich literature, we find at least three major frameworks on the expression of stance through *that* clauses – the categorisations of Martin and White (2005: 97-98), Hyland and Tse (2005b: 134), and Pho (2013: 37-38). In our analysis, we take the most recent complex classification system by Pho (2013), although the earlier ones would probably have produced similar

results. The framework's structure is visualized in Figure 2. According to Pho (2013), emotivity (positive/negative/neutral) "refers to the writer's attitude or feeling towards the proposition or the original authors" (Pho 2013: 38). A word like *show* expresses a positive attitude as in example (9) while *claim* expresses a negative attitude (10). There are also nonemotive words like *state* (11) and ambiguous words which depend on the context, for instance *imply* can express a negative implication like a hidden suggestion (12) or a positive implication like a suggested research finding (13).

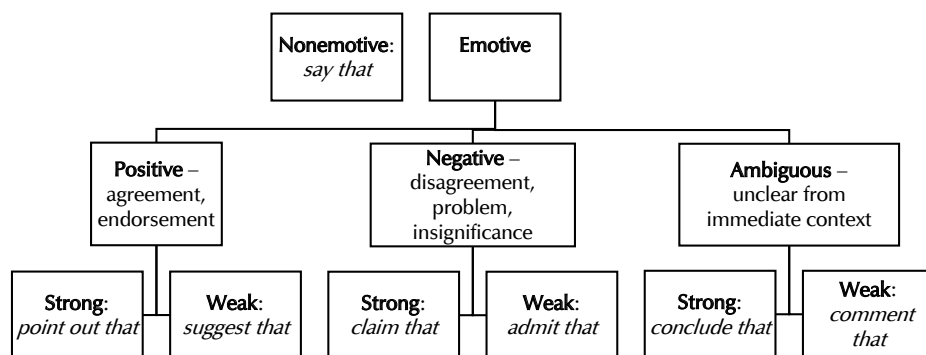


Figure 2. Categorisation of stance expressed through that clauses based on Pho (2013: 37-38)

- (9) The present findings *show that* object coordinates conjoining 2sg. you and 1sg. I/me are far less prone to non-normative case marking than subject coordinates and largely employ standard objective case. (MA12Mt_SS)
- (10) Connectionists *claim that* the mind contains network maps from the stems of all verbs to the past tense forms of their representatives (Patterson, McClelland; 2002: 464). (MA14Ft_MD)
- (11) Other sources *state that* the so-called "punishment attacks" have also been an issue in local papers from time to time. (MAC13Ft_CW)
- (12) Although their research tends to *imply that* writing quality can be measured by analysing grammatical and lexical cohesion, the linguists are aware that other factors have an influence on quality as well (genre, audience, etc.). (MAL16Ft_JZ)
- (13) Still, the results of the analysis *imply that* high quality texts use more lexical cohesive devices and more complex ones than low quality texts do. (MAL16Ft_JZ)

Controlling words are also classified in terms of the commitment to the proposition into strong and weak words (Pho 2013: 39) – e.g., *demonstrate* expresses strong commitment (14) whereas *suggest* expresses weak commitment (15).

- (14) With respect to each groups' feeling of closeness to their own group as presented in Table 2 and Table 3 with the background of rating the importance of "race" for the respondents, numbers *demonstrate that* Blacks are generally more likely to feel closer to their own group than Whites. (MAC14Ft_CH)
- (15) These findings *suggest that* race relations have the potential to improve, but that Whites are considerably weaker in their perception of differences among the races, which limits their support for race-based remedies. (MAC14Ft_CH)

As numerous examples of stance can be found in all theses, it was necessary to concentrate on structural elements where stance was expressed regularly and systematically. Thus, the focus of this small study is on the combination of the global structure, the function stance and the form *that* complements, since in modern academic writing the trend is towards explicit tentative hypotheses in certain sections of the thesis (and even the abstract). In such cases, the *that* clause becomes a central part of the entire thesis and indicates scholarly thinking in a nutshell. This is also the reason why *that* clauses also occur frequently in abstracts, especially when central claims or research hypotheses are formulated and research results reported (cf. Ivanova this issue). Global rhetorical structures and stance *that* complements are two central characteristics of academic literacy: they indicate a comprehensive discourse strategy and a researcher identity or authorial persona or voice. Both sociocognitive skills are extremely important for researchers intending to be accepted into a research community.

4. Corpus and method

4.1 The ChemCorpus as data

The ChemCorpus contains 497 (anonymised) academic texts (6.9 million words): Magister, Bachelor, and Master's theses as well as Bachelor term papers and project area reports. As described in Table 6, the sample used for

this study consists of 4 times 10 Magister and Master's theses in Linguistics and Cultural Studies (1.16 million words):

Table 6. ChemCorpus (Schmied & Dheskali 2015-2020) sample for this study

Field	Magister theses (2002-2013)		Master's theses (2012-2020)		Total words
	texts	words	texts	words	
Linguistics	10	321,967	10	265,088	587,055
Cultural Studies	10	324,137	10	254,106	578,243
Total	20	646,104	20	519,194	1,165,298

As Master's theses are considerably shorter than Magister theses, results are usually normalised to 100,000 words, which also allows comparisons to other studies. The advantage of the ChemCorpus is that it can be used as a comparative basis, e.g. to analyse cultural differences with other non-native (academic *lingua franca*) theses, such as African or Chinese in a similar formative training programme (Schmied 2015). The ChemCorpus has also been used as a reference corpus in a pilot study on Cameroon English dissertations (Cosmina 2020), which analyses the literature review in 30 theses (3×2×5) in Linguistics, Literature, and TESOL by Anglophone and Francophone writers (150,000 words). The findings on the differences in the subdisciplines will be compared to this study.

The current study compares texts from the same institution; this means that the teaching staff and approach is fairly similar, the major change was the new curriculum. The main variable in this study is the course content at two different periods of time, before and after the "Bologna" reform, which introduced the BA and MA programmes in Germany (i.e. the MA thesis is the second thesis written whereas the Magister thesis was the first, albeit after the same time of studies) – and of course, the parallel development of conventions as perceived by the teachers and supervisors responsible. The institutional background is the same, probably typical of a small modern English department in Germany. Thus, the corpus is controlled for the general teaching (and related personal consultation) in the 1st MA semester, but not for the 4th MA semester personal consultation with the final supervisor. It includes only exam texts with a good grade to control developmental stages in written C1 proficiency. Other factors influencing academic writing cannot be controlled, such as exposure to external academic instructional texts that autonomous learners were encouraged to read. The small time frame of less

than 20 years also ensured that global changes in academic writing conventions were kept to a minimum, but they cannot be excluded since the digitalisation of universities over the last 20 years has had vast influences on their work, in their communication, but also in their teaching and research practices. This comparison is thus a case study in the effect of formal instruction to young scholars, which has been advocated for a long time (as Paltridge 2020 reiterates).

4.2 Corpus linguistic methodologies

The methodologies used in this study must be combined: manual annotation, automated retrieval, and manual verification. The global structure elements of the texts were manually recorded according to the keywords signalling the detailed moves in Issue, Research Question/Hypothesis, Methodology, Analysis, and Conclusion outlined above.

The *that* clauses in the texts were analysed with AntConc (Anthony 2020). The keyword “that” was used for the query and the results were manually cleaned, excluding unrelated uses such as in relative clauses and as determiners/demonstrative pronouns. The remaining thousands of *that* complement clauses had to be categorised manually. The list of reporting verbs or *verba dicendi* (in a wide sense) in the literature (cf. Fig. 1 above) made this relatively easy.

5. Results and Discussion

5.1 Global rhetorical structures

The most important result of our analysis of the global structure of the theses is that strong disciplinary variation can be found in the Methodology and Analysis sections. Mainly, there are more theses in Cultural Studies without an explicit Methodology description. As can be seen in Figure 3, Cultural Studies writers continue structuring their theses with implicit Methodology sections even in the Masters’ studies, whereas in Linguistics there are barely any such theses, especially since the new MA programme was introduced. The comparison shows that structures of Linguistics theses have changed more than others in a direction that had started before, i.e., making headlines and structures explicit. Still, it should be noted that theses without explicit Methodology description do have a methodological approach. The method can be specified in the Introduction, as in the following example:

- (16) The employed method of analysis is based on the examination of different approaches towards understanding Americanization developed in the course of social debates. (MA15Ft_IL)

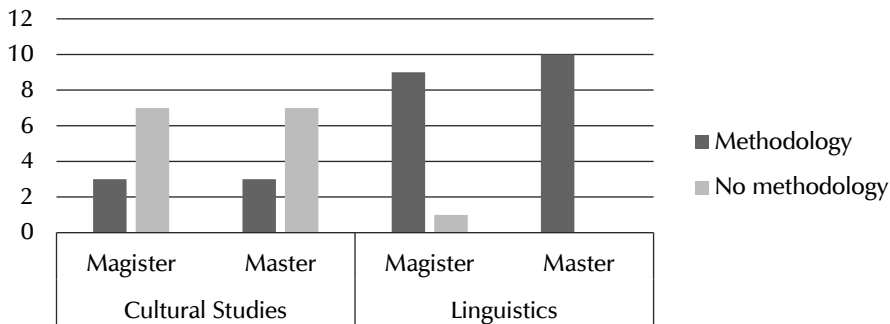


Figure 3. Number of theses with and without an explicit methodology section (N = 40)

Even if the Methodology structure remains similar in both specialisations, graduate theses display variation in the analysis section after the Bologna reform. Analysis sections become more explicit and uniform (Figure 4):

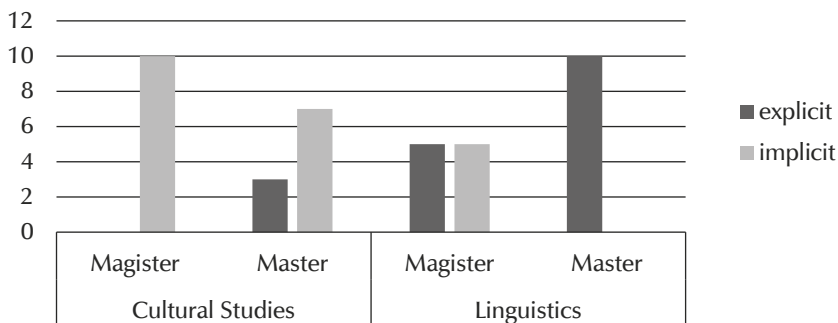


Figure 4. Number of theses with an explicit and an implicit Analysis section (N = 40) showing a growing explicitness and uniformity of thesis structure after the Bologna reform

Examples of explicit analysis sections include: “Results and Discussion”, “Research Results”, “Findings”, “Analysis and Discussion”, “Results”, “Results and Interpretation”. Examples of implicit analysis sections include: “Pre-War England: Women’s Position in Society and Its Reflection in Football”, “Italian Immigration to the United States”, and “Testing Explanatory Effectiveness”. These section headlines clearly show that explicit structures also depend on

the topic – and many topics can be dealt with from a more hermeneutic or a more social science perspective.

Similar trends can be identified for the use of research questions and hypotheses. As evident from Figure 5, the use of research questions (RQs) or the combination of RQs and hypotheses increases in both disciplines, the sole use of hypotheses decreases. RQs and hypotheses can be presented explicitly as a list and implicitly as part of the reading flow. When they are listed, they are used as signposts to structure the thesis as in this example:

(17) Consequently, I will try to answer the following research questions:

1. What are types and numbers of different lexical cohesive devices used in Cameroonian M.A. theses?
2. Are there any differences between Cameroonian male and female M.A. students in the usage of lexical cohesive devices in their M.A. theses?
3. Does the usage of lexical cohesive devices influence writing quality in Cameroonian M.A. theses?
4. Are there any differences between Cameroonian anglophone and francophone M.A. students in the usage of lexical cohesive devices in their M.A. theses? (MA16Ft_JZ)

When RQs and hypotheses are embedded in the reading flow, they can still function as reference:

(18) Contrary to my hypothesis that the learners would clearly prefer the overt relative pronouns who/which > that > zero, learners seem to handle who and which differently. (MA13Ft_SK)

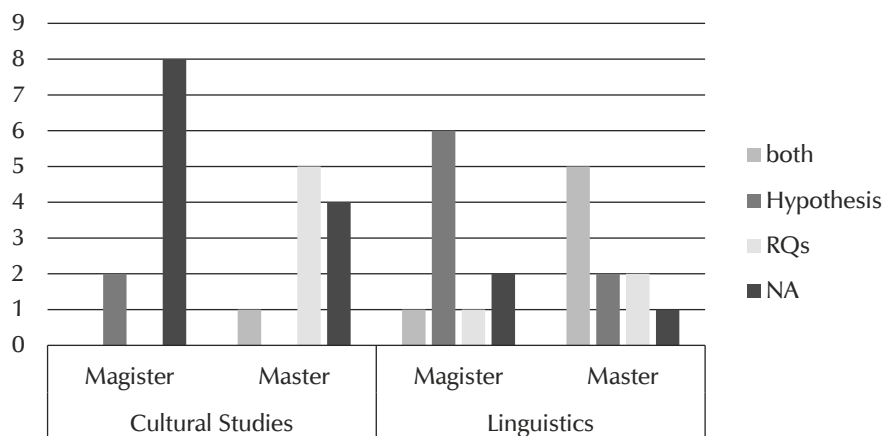


Figure 5. Occurrence of hypotheses and research questions in the theses (N = 40)

5.2 Stance in *that* clauses introduced by verbs

A comprehensive overview of *that* clause type per thesis type and discipline can be found in Table 10 in the Appendix. Table 7 is only a summary of the development in the use of *that* clauses before and after the Bologna reform:

Table 7. Relative frequency of *that* clause verb types in Magister and Master's theses

<i>that</i> clause type	Magister per 100,000	Master per 100,000
negative weak	46	48
negative strong	34	67
ambiguous weak	55	68
ambiguous strong	79	87
positive weak	68	142
positive strong	154	256
nonemotive	166	277
Total	603	947

There is a noticeable increase in the use of all controlling verbs from 603 to 947 occurrences per 100,000 words. This indicates a higher accordance with international academic writing conventions. It also shows that student academic writing becomes more dialogic – the students feel as part of a community. Parallel, there is an evident rise in negative strong and positive weak and positive strong verbs. This means that young scholars start to employ more expressive stance. Still, the most frequent choice in all text categories are neutral, nonemotive verbs, i.e. students continue to rely on them to avoid expressing evaluation. These results contradict the findings in Hyland and Jiang (2017), who noticed a decreasing of stance markers throughout the years. One reason for this difference is probably the different time scale of analysis – while Hyland and Jiang compare research articles from 1965, 1985 and 2015, the Magister and Master's theses in our study span between 2002 and 2020. Another possible reason may lie in our research focus on non-native student works. Non-native academic writing tends to be more explicit in terms of cohesive devices (Schmied 2011: 10), so it could be expected that students rely on *that* clauses as a convenient structuring device. A replication study comparing native and non-native writers from different academic levels would be an interesting endeavour for future research.

Table 8 shows the use of *that* complement clauses in the two specialisations, Linguistics and Cultural Studies, which generally includes theses in literary studies in this perspective (but only 3/10 Magister theses in this subcorpus):

Table 8. Relative frequency of *that* clause verb types in Magister and Master's theses from Cultural Studies and Linguistics

<i>that</i> clause type	Cultural Studies per 100,000	Linguistics per 100,000
negative weak	42	52
negative strong	61	40
ambiguous weak	62	61
positive weak	73	137
ambiguous strong	93	74
positive strong	181	229
nonemotive	201	243
Total	713	836

Students in Cultural Studies use more negative strong and ambiguous strong verbs, whereas those in Linguistics clearly use more positive (weak and strong) verbs. These results are similar to the findings of the pilot study by Cosmina (2020: 28), where linguistic theses by Cameroon writers also include more positive verbs. Positive verbs are seen to make conclusions more compelling and help promote the value of their findings or of the reported proposition. In this case, they are used to convey a clear and positive strong attitude to the other author's original proposition (Cosmina 2020: 49). However, in the current study the overall number of verbs in the disciplines is similar in contrast to Cosmina (2020: 24), who counted a higher number of reporting verbs in harder disciplines (Linguistics) than in softer disciplines (Literature, TESOL).

Table 9 presents an alphabetical summary of the top 20 verbs controlling *that* complement clauses per thesis type (left) and per discipline (right). The most general result of our study comparing stance in *that* complement clauses before and after the Bologna reform is – as expected – an increase in the subset totals: higher frequencies in Master than in Magister and higher frequencies in Linguistics than in Cultural Studies. This result correlates with developments in other genres like research articles (cf. Dontcheva-

Navratilova 2023). If we look at the most frequent lexemes, we notice the use of strong verbs, mainly of *show* (positive strong) in Master's theses, predominantly in Linguistics, as well as *claim* (negative strong). This indicates a higher awareness of author identity, which may be teaching induced and corresponds to international trends in the subdiscipline (cf. Dontcheva-Navratilova 2023). Few verbs are significantly more used in Cultural Studies, e.g., *believe* (ambiguous weak) and *demonstrate* (positive strong). The overall absolute and relative figures for the top 20 verbs can be found in Table 11 in the Appendix.

Table 9. Top 20 verbs controlling that complement clauses per thesis (left) and per discipline (right) in alphabetical order

Verb	Evaluation type	Magister per 100,000	Master per 100,000	Cultural Studies per 100,000	Linguistics per 100,000
argue	ambiguous strong	37	53	52	38
assume	negative weak	34	33	26	41
believe	ambiguous weak	11	17	22	7
claim	negative strong	29	59	50	38
conclude	ambiguous strong	18	16	16	18
demonstrate	positive strong	6	18	20	4
explain	nonemotive	8	10	12	6
imply	ambiguous weak	20	21	12	29
indicate	positive weak	19	41	17	43
know	positive strong	13	5	15	3
mean	nonemotive	41	61	55	46
note	nonemotive	26	37	23	40
point out	positive strong	33	22	24	31
reveal	positive weak	7	19	6	21
say	nonemotive	27	51	37	40
see	nonemotive	3	28	4	27
show	positive strong	49	128	61	116
state	nonemotive	36	65	51	49
suggest	positive weak	27	51	24	54
think	positive weak	8	20	21	7
Total		450	754	548	656

6. Conclusions

Since the Bologna reform, scientific writing conventions have been spreading from the natural and social sciences into English Studies and its subdisciplines (here Linguistics and Cultural Studies), i.e., from ‘hard’ to ‘soft’ fields (cf. Hyland – Bondi 2006).

The analysis of global rhetorical structures shows a trend towards more uniformity and explicit structuring, which confirms the changes identified in previous studies. There is a rise in the explicit formulation of research questions and their combination with hypotheses. The IMRAD / AIMAC structure is becoming more popular, often signalled through explicit section labels in headlines and references (such as *Methodology*, *Analysis*). These explicit labels can also be found in abstracts and titles, contributing to their complexity and specificity, which also allows authors to distinguish themselves more clearly from related writers and their texts. For young academics, whom such formulas can help to write more systematically, these smaller chunks may be an important scaffold, if the writing process gets difficult. The stereotypical form does look uniform and even unimaginative, but it can be individualised by more specific phrasing later in the career.²

The metadiscourse case study showed that after 2012, non-native thesis writers from English Studies rely significantly more on *that* complement clauses to express their stance. Their writing has become more evaluative through the increasing use of positive and negative strong verbs but students still rely extensively on nonemotive expressions. From the reviewed disciplines, theses in Linguistics tend to incorporate more positive verbs (e.g., *show*), both in the German and the Cameroon context (Cosmina 2020). This may mean that empirical linguistics fits better into the

² This balance between genre-specific standard patterns and individual creativity is particularly important in view of the new sophisticated “chatbots” like ChatGPT, which were hotly discussed in universities and beyond in early 2023. The large language models combining huge text databases and neural networks created texts that passed exams and thus obviously met the (minimum) expectations of academic teachers. Many were surprised about the “quality” of texts produced by “Artificial Intelligence”, which is, strictly speaking, neither “art” or “intelligent”, but rather “craft” and “statistics”, because algorithms are used to “predict” most likely patterns and word combinations, irrespective of truth value and semantic creativity. OpenAI warns in the limitations section of its ChatGPT website that “ChatGPT sometimes writes plausible-sounding but incorrect or nonsensical answers” (<https://openai.com/blog/chatgpt/>). Instead of excluding such complex text bots from university, they could be used to create rough stereotypical drafts and students could be trained in adding individuality and specificity to personalise their texts within expected genre conventions as analyzed in this contribution.

international science patterns or that young linguists feel a greater need to adapt to international conventions.

The developments described in this study imply that student writing will continue being influenced by professional international publication models. This will take place both on a global structural scale (IMRAD) as well as in local signposting of the sections (in titles, abstracts, and headlines). For the discussion of genre development, these very early career texts allow us to observe a researcher and a learner perspective at the same time: on the one hand, changes are likely to be based on the perception of changes in genre conventions by teachers, who then incorporate the detected conventions not only into their own writings, but also into their teaching; on the other hand, changes must be evaluated positively by young writers, either because they are taken over from their teachers or from other models in the literature. This process of academic knowledge dissemination leads to convergence in a highly competitive market of young graduates, whose MA theses serve as evidence of qualification for doctoral and further careers.

This case study used a small but consistent dataset, one which had only 40 cases for the variable global rhetorical structures, but over 4000 for the variable *that* complement clause. Thus, it needs to be supplemented by comparable studies from other universities and countries³. In many academic contexts, it is still difficult to find the appropriate data, but the digitalisation of university exams makes open repositories possible and allows them to be exploited (as demonstrated by Schmied 2013 for South Africa). However, it is safe to say that the academic writing principles discussed here as well as the changes traced in the variables analysed (more IMRAD, more research questions/hypotheses, more stance in *that* complements) show basically the same patterns as in other genres, especially in research articles (cf. Bondi – Nocella, Ivanova, Dontcheva-Navratilova, all this issue). In contrast to published research articles, MA theses leave the final decision on IMRAD structures, for instance, to the writers; while editors can finally decide on the publication of articles, supervisors cannot control the submission of theses, only the students decide. This suggests that the decisions to change academic writing (as analysed here) are laid in modern MA degree programmes already; they are made by academic novices who write to develop their authorial identities in their academic discourse communities. This case study

³ From a cultural perspective, it is interesting that the well-discussed contrast between German and English discourse structure (Clyne 1987, 1993) is not prominent in the texts written in English by German students.

is only a small stone in a large mosaic that illustrates much wider changes, as has already been argued by Shaw – Vassileva (2009: 300):

Some of these linked sets of developments are shared by all the journals – that is, the same sorts of change seem to happen over the period in all the journals. One such set has the effect of making the articles more explicit. There is less reliance on the reader's background knowledge of recent publications and issues: the reader gets more guidance as to how to read the article, and how to place it in the disciplinary debate. Thus references have become conventionalised and explicit in form, the initial sporadic and non-standardised division into sections has become obligatory and uniform, and introductions have become more explicit in stating aims, article structure, and connections to current issues in research and/or policy.

Finally, it remains to be studied whether the changing discourse features indicating explicitness and writer identity analysed in this case study can be found on a similar timescale in other academic languages, like German or French. This could be used as an empirical basis to discuss whether they are essential features of a particular dominant national or language culture or functional adaptations to changing rhetorical situations, new sociocognitive responses to technological affordances and community needs.

REFERENCES

Primary sources

Schmied, J. – D. Dheskali
2015 *ChemCorpus*. <https://tinyurl.com/3cj7kbzp>, accessed February 2023.

Secondary sources

- Anthony, L.
2020 *AntConc* (Version 3.5.9). Tokyo: Waseda University. Available from <https://www.laurenceanthony.net/software>, accessed February 2023.
- Biber, D.
2006 *University Language*. Amsterdam: John Benjamins.
- Biber, D. – E. Finegan
1989 "Styles of stance in English: Lexical and grammatical marking of evidentiality and affect", *Text-Interdisciplinary Journal for The Study of Discourse* 9 (1), 93-124.

- Bondi, M.
2009 "Historians at work: Reporting frameworks in English and Italian book review articles". In: K. Hyland – G. Diani (eds.) *Academic Evaluation. Review Genres in University Settings*. Basingstoke: Palgrave Macmillan, 179-196.
- Clyne, M.
1987 "Cultural differences in the organization of academic texts: English and German", *Journal of Pragmatics* 11 (2), 211-241.
1993 "The German language after Unification: Adapting assumptions and methodologies to the 'New World Order'", *International Journal of the Sociology of Language* 100/101, 11-27.
- Cosmina, G.
2020 *Stance Through Verb that- Complement Clauses: A Three-Level Analysis*. Dissertation (Master's). Chemnitz: TU Chemnitz.
- Diani, G.
2004 "A genre-based approach to analysing academic review articles." In: M. Bondi – L. Gavioli – M. S. Silver (eds.) *Academic Discourse, Genre and Small Corpora*. Rome: Officina Edizioni, 105-126.
2017 "Criticism and politeness strategies in academic review discourse: A contrastive (English-Italian) corpus-based analysis", *Kalbotyra* 70, 60-78.
- Dontcheva-Navratilova, O.
2008 "Reporting verbs as indicators of stance in academic discourse", *Porta Lingua Debrecen*, 97-103.
2023 "Self-mention in L2 (Czech) learner academic discourse: Realisations, functions and distribution across master's theses", *Journal of English for Academic Purposes* 64, 101272.
<https://doi.org/10.1016/j.jeap.2023.101272>.
- Holmes, J.
1984 "Modifying illocutionary force", *Journal of Pragmatics* 8 (3), 345-365.
- Hyland, K.
2005a *Metadiscourse: Exploring Interaction in Writing*. London: Continuum.
2005b "Stance and engagement: A model of interaction in academic discourse", *Discourse Studies* 7 (2), 173-192.
<https://doi.org/10.1177/1461445605050365>.
- Hyland, K. – M. Bondi (eds.)
2006 *Academic Discourse Across Disciplines*. Bern: Peter Lang.
<https://doi.org/10.3726/978-3-0351-0446-2>.
- Hyland, K. – F. (K.) Jiang
2017 "'We believe that...': Changes in an academic stance marker", *Australian Journal of Linguistics* 38 (2), 139-161.
<https://doi.org/10.1080/07268602.2018.1400498>.
- Hyland, K. – P. Tse
2005a "Hooking the reader: A corpus study of evaluative *that* in abstracts", *English for Specific Purposes* 24 (2), 123-139.

- 2005b "Evaluative *that* constructions: Signalling stance in research abstracts", *Functions of Language* 12 (1), 39-63.
- Ivanova, M.
2020 "Semantic and pragmatic truth implications of concession in single- and double- blind open peer reviews". In: J. Schmied – J. Dheskali (eds.) *Conflicting Truths in Academic and Journalistic Writing*. Göttingen: Cuvillier, 141-156.
- Kim, C. – P. Crosthwaite
2019 "Disciplinary differences in the use of evaluative *that*: Expression of stance via *that*-clauses in business and medicine", *Journal of English for Academic Purposes* 41, 1-14. <https://doi.org/10.1016/j.jeap.2019.100775>.
- Kuhlee, D.
2017 "The impact of the Bologna Reform on teacher education in Germany: An empirical case study on policy borrowing in education", *Research in Comparative and International Education*, 12 (3), 299-317.
- Luzón, M.-J. – C. Pérez-Llantada
2019 "Connecting traditional and new genres trends and emerging themes". In: M.-J. Luzón – C. Pérez-Llantada (eds.) *Science Communication on the Internet, Old Genres Meet New Genres*. Amsterdam: John Benjamins, 1-18. <https://doi.org/10.1075/pbns.308>.
- Martin, J. R. – P. R. R. White
2005 *The Language of Evaluation. Appraisal in English*. Basingstoke/New York: Palgrave Macmillan.
- Mauranen, A. – C. Pérez-Llantada – J. M. Swales
2010 "Academic Englishes: A standardised knowledge?". In: A. Kirkpatrick (ed.) *The Routledge Handbook of World Englishes*. London/New York: Routledge, 659-676.
- Paltridge, B.
2020 "Writing for academic journals in the digital era", *RELC Journal* 51 (1), 147-157. <https://doi.org/10.1177/0033688219890359>.
- Paré, A. – D. Starke-Meyerring – L. McAlpine
2011 "Knowledge and identity work in the supervision of doctoral student writing: Shaping rhetorical subjects". In: D. Starke-Meyerring et al. (eds.) *Writing in Knowledge Societies*. Fort Collins, CO: The WAC Clearinghouse, 215-236. <https://doi.org/10.37514/PER-B.2011.2379.2.11>.
- Pérez-Llantada, C.
2012 *Scientific Discourse and the Rhetoric of Globalization: The Impact of Culture and Language*. London: Continuum.
2013 "'Glocal' rhetorical practices in academic writing", *European Journal of English Studies* 17(3), 251-268. <https://doi.org/10.1080/13825577.2013.867180>.
2021 *Research Genres Across Languages*. Cambridge: Cambridge University Press.

- Pho, P. D.
 2013 *Authorial Stance in Research Articles: Examples from Applied Linguistics and Educational Technology*. Basingstoke: Palgrave Macmillan.
- Quirk, R. et al.
 1985 *A Comprehensive Grammar of the English Language*. London: Longman.
- Sancho Guinda, C.
 2015 "Genres on the move: Currency and erosion of the genre moves construct", *Journal of English for Academic Purposes* 19, 73-87.
<https://doi.org/10.1016/j.jeap.2015.07.001>.
- Schmied, J.
 2011 "Academic writing in Europe: A survey of approaches and problems". In: J. Schmied (ed.) *Academic Writing in Europe: Empirical Perspectives*. Göttingen: Cuvillier Verlag, 1-22.
 2013 "Academic knowledge presentation in MA theses: From corpus compilation to case studies of disciplinary conventions", *Brno Studies in English* 38, 149-164.
 2015 "Graduate academic writing in Europe in comparison: A research-based approach to metalanguage and genre". In: J. Schmied (ed.) *Academic Writing for South Eastern Europe: Practical and Theoretical Perspectives*. Göttingen: Cuvillier Verlag, 1-24.
 2021 "Applying Appraisal Theory to conference reviews: Systemic functional perspectives to corpus-based genre-specific discourse analyses". In: J. Schmied et al. (eds.) *Critical Discourse and Corpus Approaches to Systemic Functional Grammar*. Göttingen: Cuvillier, 1-22.
- Shaw, P. – I. Vassileva
 2009 "Co-evolving academic rhetoric across culture; Britain, Bulgaria, Denmark, Germany in the 20th century", *Journal of Pragmatics* 41 (2), 290-305. <https://doi.org/10.1016/j.pragma.2008.07.009>.
- Swales, J. M.
 1990 *Genre Analysis: English in Academic and Research Settings*. Cambridge: Cambridge University Press.
 2004 *Research Genres: Explorations and Applications*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139524827>.
- Swales, J. M. – C. B. Feak
 2012 *Academic Writing for Graduate Students: Essential Tasks and Skills* (3rd edn.). Ann Arbor, Michigan: University Press.
<https://doi.org/10.3998/mpub.2173936>.
- Thompson, G. – S. Hunston
 2000 "Evaluation: An introduction". In: S. Hunston – G. Thompson (eds.) *Evaluation in text*. Oxford: Oxford University Press, 1-27.
- Yang, R. – D. Allison
 2003 "Research articles in applied linguistics: Moving from results to conclusions", *English for Specific Purposes* 22 (4), 365-385.
[https://doi.org/10.1016/S0889-4906\(02\)00026-1](https://doi.org/10.1016/S0889-4906(02)00026-1).

APPENDIX

Table 10. Stance of *that* complement clauses controlled by verb in Magister and Master's theses in the two subdisciplines. *Note:* Table ordered by relative frequency of the clause total (last column)

<i>that</i> clauses controlled by verb	Magister Cultural Studies		Magister Linguistics		Master Cultural Studies		Master Linguistics		Total	
	Σ	per 100,000	Σ	per 100,000	Σ	per 100,000	Σ	per 100,000	Σ	per 100,000
nonemotive	264	81	273	85	303	119	419	158	1259	108
positive strong	238	73	261	81	274	108	393	148	1166	100
positive weak	77	24	142	44	125	49	247	93	591	51
ambiguous strong	162	50	95	30	110	43	117	44	484	42
ambiguous weak	77	24	99	31	96	38	80	30	352	30
negative strong	45	14	64	20	120	47	53	20	282	24
negative weak	71	22	79	25	52	20	74	28	276	24
Total	934	288	1013	315	1080	425	1383	522	4410	378

Table 11. Top 20 verbs controlling that complement clauses in the corpus sample ordered by relative total frequency

Verb	Evaluation Type	Magister Cultural		Magister Linguistics		Master Cultural		Master Linguistics		Total	
		Σ	per 100,000	Σ	per 100,000	Σ	per 100,000	Σ	per 100,000	Σ	per 100,000
show	positive strong	70	22	87	27	99	39	237	89	493	42
mean	nonemotive	77	24	54	17	80	31	78	29	289	25
state	nonemotive	55	17	60	19	87	34	81	31	283	24
argue	ambiguous strong	79	24	40	12	71	28	67	25	257	22
claim	negative strong	33	10	62	19	102	40	49	18	246	21

suggest	positive weak	33	10	55	17	36	14	97	37	221	19
say	nonemotive	45	14	41	13	59	23	73	28	218	19
assume	negative weak	46	14	64	20	29	11	57	22	196	17
note	nonemotive	31	10	53	16	34	13	63	24	181	16
indicate	positive weak	8	2	53	16	38	15	70	26	169	15
point out	positive strong	43	13	62	19	27	11	30	11	162	14
imply	ambiguous weak	11	3	53	16	23	9	32	12	119	10
conclude	ambiguous strong	27	8	31	10	19	7	23	9	100	9
see	nonemotive	4	1	5	2	6	2	67	25	82	7
believe	ambiguous weak	33	10	4	1	30	12	14	5	81	7
think	positive weak	21	6	5	2	36	14	15	6	77	7
reveal	positive weak	4	1	19	6	11	4	39	15	73	6
demonstrate	positive strong	15	5	4	1	39	15	7	3	65	6
know	positive strong	37	11	4	1	8	3	4	2	53	5

Address: JOSEF SCHMIED, English and American Studies, Chemnitz University of Technology, 09107 Germany. josef.schmied@phil.tu-chemnitz.de.

ORCID code: <https://orcid.org/0000-0001-8499-3158>.

Address: MARINA IVANOVA, English and American Studies, Chemnitz University of Technology, 09107 Germany. marina.ivanova@phil.tu-chemnitz.de.

ORCID code: <https://orcid.org/0000-0001-7414-6668>.

“The aim of this paper is...”: Frame markers in English as a lingua franca of academic writing

Tereza Guziurová
University of Ostrava

ABSTRACT

This study explores variation in metadiscourse patterns in English as a lingua franca academic writing. The paper aims to investigate discourse reflexivity in English-medium research articles written by non-native speakers from ten different L1 backgrounds included in the SciELF corpus. Specifically, the paper focuses on one reflexive category, frame markers, which signal text boundaries, announce discourse goals, and label text stages (Hyland 2005), thus making the discourse organisation more explicit. The corpus comprises 72 articles from the field of social sciences and humanities, totalling over 432,000 words. The findings are compared with a specialized corpus of 72 published research articles written by Anglophone authors (approximately 621,000 words), which has been designed as a corpus comparable to the SciELF. The results indicate differences in the forms and functions of certain frame markers in the two corpora, suggesting that this type of discourse reflexivity shows language and culture-specific diversity.

Keywords: metadiscourse, frame markers, English as a lingua franca, research articles, SciELF corpus.

1. Introduction

English as a lingua franca is now a relatively established research field. Most of ELF research has focused on spoken interactions (Cogo – Dewey 2012), which were analysed using spoken ELF corpora, such as the Vienna-Oxford International Corpus of English (VOICE) launched in 2001, or the corpus of English as a Lingua Franca in Academic Settings (ELFA) completed in 2008 (ELFA 2008). Studies on the ELFA corpus yielded interesting results

on the nature of spoken ELF. Mauranen (2017), for example, has identified several processes typical of ELF speech: on the one hand it is structural simplification, which can be manifested in the regularization of irregular forms, or lexical simplification, i.e. a tendency to higher representation of the most frequent lexical items in ELFA (in comparison with L1 English). On the other hand, there are complexifying tendencies such as approximation, when speakers use approximate equivalents of target expressions whose meaning is easily recognizable (e.g. *on the other side*). This process is not seen as an individual idiosyncrasy (such as omitting articles by a given speaker), but as a more general tendency of speakers across different speech events and lingua-cultural backgrounds, that is a property of ELF (Mauranen 2017).

One of the observed tendencies of ELF speech at discourse level is enhanced explicitness (Mauranen 2017). Since speakers from different linguistic and/or cultural backgrounds cannot often rely on the shared cultural context and their interactions might be associated with uncertainty or unpredictability, they tend to use clearer and more explicit expressions. Moreover, clarifying and explicating strategies are cooperative (Mauranen 2012: 167) as they increase the chances of getting the message across. Enhanced explicitness takes many forms (e.g. rephrasing in speech), but a common manifestation of explicitation is discourse reflexivity (Mauranen 2017: 246), which makes discourse organisation more visible and enhances clarity by guiding readers through the discourse. Mauranen et al. (2016: 4) argue that “discourse reflexivity is central to academic discourse, and particularly relevant for academic ELF, where it can help increase clarity and explicitness among speakers from different linguistic and cultural backgrounds” (Mauranen et al. 2016: 46). Moreover, discourse reflexivity in ELF was found to have similar functions, forms and distributions across different L1 speakers (*ibid.*).

What has received less attention so far is written academic ELF. Among the research topics that have already been addressed are, for example, phraseological approximations in spoken and written academic ELF (Carey 2013), hedging modal verbs in ELF research articles (Mur Dueñas 2016), the rhetorical structure of research article abstracts (Lorés-Sanz 2016), science blogs (Mauranen 2013), and multi-word units of meaning in L2 Master’s theses (Vetchinnikova 2014). It should be mentioned that there exists a substantial body of research on academic discourse in English written by non-native speakers. Drawing on the tradition of contrastive rhetoric (Connor 1996), researchers examined grammatical, lexical and discursive

features of academic texts written by L2 English speakers and contrasted them with L1 English texts. The issue with this kind of research was that the writing produced by native speakers was perceived as an ideal, a norm that the others should aim to meet (Mauranen – Hynninen – Ranta 2016). Nevertheless, the majority of writers and readers of academic texts nowadays are likely to be academics with other than L1 English background because English has gradually become the lingua franca of academia (ibid.).

In 2015, the first large database of written academic ELF was compiled – the WrELFA corpus (Written English as a Lingua Franca in Academic Settings). It consists of 1.5 million words and has three parts: PhD examiners' reports, research blogs, and unedited research articles (the SciELF corpus). The last part contains 150 research articles written by L2 users of English from 10 different L1 backgrounds, which have not undergone professional proofreading or checking by English native speakers. The SciELF corpus has already been examined from several linguistic viewpoints (see e.g. Rowley-Jolivet 2017; Mur Dueñas 2018; Murillo 2018; Wu et al. 2020; Shchemeleva 2022). For example, Mur Dueñas (2018) has compared the use of a lexicogrammatical structure, the anticipatory *it* pattern, in ELF research articles and ENL research articles from similar disciplines. She has found that the *it* pattern most commonly expresses attitudinal meanings in both corpora, but generally it is used more frequently by ELF authors. Moreover, ELF scholars use some chunks which are absent in ENL articles, for example a wider range of discourse verbs in the pattern *It V-link ADJ to*, which suggests possible innovations or creativity when expressing interpersonal meanings in ELF (Mur Dueñas 2018).

Another study by Bondi and Borelli (2018) is particularly relevant for this paper because it investigates how non-native English speakers use metadiscursive resources in the economics component of the SciELF. Analysing positive (overused) and negative (underused) keywords in ELF in comparison with the reference corpus of published research articles, the study shows that the SciELF corpus is characterized by some prototypical metadiscursive elements, such as neutral evidentials (*according to*), general labelling nouns identifying elements of cognition (e.g. *characteristics*) and text (e.g. *paper*) rather than event (e.g. *change*) and discourse (e.g. *claim*), and also prototypical frame markers pointing to topic and focus (e.g. *consider*) rather than arguments (e.g. *show*). The ELF articles also show a marked underuse of textual and personal deixis (*I, this, these*) compared to published articles which are characterized by a greater authorial presence in the text (Bondi – Borelli 2018).

Generally, discourse reflexivity or metadiscourse is an interesting research area which can help reveal some distinctive features of written academic ELF. Compared with other writing cultures, Anglo-American academic English tends to be described as “reader-oriented” as it is associated with a higher level of interactivity and it puts the responsibility for clarity and understanding on the writer rather than the reader (Hyland 2005; Čmejrková 1996). In terms of discourse organization, it has been characterized as “more explicit about its structure and purposes”, containing a noticeable amount of metadiscourse (Hyland 2005: 117; Swales – Feak 2012). On the other hand, as mentioned above discourse reflexivity proved to be particularly relevant for academic ELF, where mutual understanding and explicitness are important (and maybe more important than correctness) (Bondi – Borelli 2018). The aim of the present study is to investigate discourse reflexivity in written academic ELF, represented by the SciELF corpus, and to find out whether a higher level of explicitness typical of spoken ELF also characterizes ELF academic writing. The study focuses on one reflexive category, frame markers, which signal text boundaries or elements of text structure (Hyland 2005). They can sequence parts of the text, announce goals or label text stages and are signalled by such expressions as *first*, *next*, *my purpose is*, *to conclude* etc. Frame markers refer to discourse acts or stages, making discourse organization more explicit and accessible to readers.

Specifically, the present study addresses the following research questions:

1. What are the forms and functions of frame markers in ELF research articles?
2. Are there any differences in the use of frame markers between ELF research articles and published research articles written by Anglophone authors?

2. Frame markers in academic writing

Frame markers are part of Hyland’s (2005) metadiscourse model that defines metadiscourse as “self-reflective expressions used to negotiate interactional meanings in a text, assisting the writer (or speaker) to express a viewpoint and engage with readers as members of a particular community” (Hyland 2005: 37). Understanding metadiscourse primarily as writer-reader interaction, Hyland draws a distinction between two types of metadiscourse – interactive and interactional. Interactive resources are used to organize

a text taking account of the reader's knowledge and processing needs. They comprise transition markers (conjunctions and adverbial phrases signalling relations in the text); endophoric markers which refer to other parts of the text (e.g. *noted above*); evidentials, i.e. citation practices; code glosses which help readers understand the text by rephrasing or explaining what has been said (*in other words*); and frame markers which signal text boundaries (*in conclusion*).

Interactional resources concern ways writers comment on and evaluate the content of propositions, expressing epistemic and affective stance. They also express writer-reader interaction as they aim to involve readers in the arguments. While interactional resources, including hedges, boosters and engagement markers, have been extensively researched in academic writing (e.g. McGrath – Kuteeva 2012; Hyland 2004; Hu – Cao 2015; Hyland – Jiang 2016; Dontcheva-Navratilova 2021, etc.), interactive resources have attracted less attention (probably with the exception of transition markers and reformulation markers).

Linguistic expressions referred to as frame markers are present in most conceptions of metadiscourse. For instance, one of the early taxonomies developed by Crismore et al. (1993) contains several categories that overlap with frame markers: sequencers which indicate ordering of material (*first, next*), topicalizers indicating a shift in topic (*now; in regard to*) and partly illocution markers which name the act performed (*to sum up*). Hyland (2005) reorganized Crismore et al.'s (1993) categories and introduced the notion of frame markers to denote rhetorical units which mark elements of text structure and help readers follow the development of the discourse. In his metadiscourse model, Hyland characterizes frame markers as follows:

Frame markers signal text boundaries or elements of schematic text structure. [...] Items included here function to sequence, label, predict and shift arguments, making the discourse clear to readers or listeners. (2005: 51)

Hyland does not elaborate the definition much further in his original framework, but he provides a more detailed account of individual categories in his later study (Hyland – Zou 2020). Specifically, frame markers can be divided into:

- **sequencers** used to sequence parts of the text or to internally order an argument (*first, next*);

- **labellers** which explicitly label text stages (*to summarize, in conclusion*); they often indicate the speech act that will be performed;
- **goal announcers** which state the author's purpose in the text (*my purpose is, the paper proposes*);
- **topic shifters** which indicate a shift in the direction of the text (*well, now, let us return to*) (Hyland – Zou 2020).

Frame markers thus seem to be salient in the overall organization of texts, especially of research articles, which need to be carefully structured to convey the author's intended message and also to conform to conventions of academic writing and genre expectations.

A number of studies have investigated frame markers, usually as part of the whole metadiscourse framework, across different genres, languages and disciplines. A detailed analysis of frame markers is provided by Hyland and Zou (2020) who explored how academics recontextualise their scientific findings from journal articles to academic blogs. Analysing 50 blogs and 50 articles with the same authors and topics, they have found that bloggers generally use more frame markers to present complex research material to lay audience. They have also shown that labellers and topic shifters occur more frequently in research articles to assist specialists follow longer and more complex arguments, while sequencers (especially listing sequencers) are more numerous in blogs to help general readers understand connections in the text, facilitating processing of the material (Hyland – Zou 2020).

Cross-linguistic studies of metadiscourse have confirmed that its use in academic writing is influenced by writers' linguistic and cultural backgrounds (Mauranen 1993; Peterlin 2005; Mur Dueñas 2011; Mu et al. 2015). For example, Mauranen (1993) in her seminal study of text reflexivity finds that Anglo-American writers use more reflexive text than Finnish writers, which indicates that Anglo-American writers are more concerned with guiding the reader in the discourse, showing more explicit presence in the text. A comparison of the use of metadiscourse in English and Chinese research articles from applied linguistics has revealed that English articles contain more interactive metadiscourse features, including frame markers, which is attributed to different rhetorical strategies between the languages and the fact that Chinese writers share more background knowledge as they address the local discourse community (Mu et al. 2015). Dahl (2004) investigated the relation between language and discipline as two important variables influencing the use of metadiscourse. She examined textual metadiscourse in research articles across three languages (English, French

and Norwegian) and three disciplines (economics, linguistics and medicine). Her findings suggest that the language variable is key within economics and linguistics, where English and Norwegian display similar patterns, using frame markers much more than French. The situation is different in medical texts, which show almost identical frequencies of frame markers in all three languages, pointing to stable disciplinary practices in medicine (Dahl 2004: 1822).

To summarise, previous studies have shown that frame markers as part of interactive metadiscourse are important tools of organising academic texts and they are sensitive to their social and rhetorical context (Hyland – Zou 2020). Since this study focuses on unedited research articles which have not been shaped by language professionals, reviewers, editors and other "literacy brokers" (Lillis – Curry 2010), the results might shed light on how L2 English scholars structure their scientific texts and what their rhetorical preferences are.

3. Data and methodology

In order to investigate how L2 writers use frame markers in English as a lingua franca academic writing, two corpora have been compared. The first one is composed of articles from the SciELF corpus, which comprises 150 unedited research articles written by L2 users of English, totalling 759,300 words. The papers have not undergone professional proofreading or checking by English native speakers, and most of them are final drafts of unpublished manuscripts. The corpus is divided into two broad disciplinary domains – sciences (labelled 'Sci') and social sciences and humanities (labelled 'SSH'). The Sci part contains 78 articles (326,463 words), which are drawn from natural sciences (79%), medicine (18%) and agriculture and forestry (3%). The SSH part contains 72 articles (432,837 words) and it includes texts from social sciences (45%), humanities (34%) and behavioural sciences (21%). Since disciplinary variation in the use of metadiscourse is high (Hyland 2005), I have limited my focus to social sciences and humanities articles (the SciELF-SSH subcorpus) in this study.

The authors of the papers come from ten different L1 backgrounds (see Table 1). The number of articles in each language group varies, but the goal of ELF research is not to make L1-based comparisons, but to examine how people from different language backgrounds use English as a lingua franca. Specific disciplines represented in the corpus are psychology, sociology,

educational sciences, economics, classical philology, linguistics, philosophy, law, theatre studies, anthropology, history, urban design, literary studies, information sciences, social policy, and art history (SciELF 2015).

Table 1. Distribution of the L1 language groups in the SciELF-SSH corpus

First author's L1	No. of articles	No. of words	% of total words
Chinese	10	44,196	10%
Czech	10	59,569	14%
Finnish	10	59,118	14%
French	8	48,373	11%
Italian	5	31,249	7%
Portuguese (Brazil)	6	39,223	9%
Romanian	4	25,197	6%
Russian	6	38,834	9%
Spanish	7	51,383	12%
Swedish	6	35,695	8%
Total	72	432,837	100%

The SciELF-SSH corpus is contrasted with a reference corpus, which was compiled in 2019 (labelled 'CSSH'). The reference corpus is composed of 72 English-medium articles, which were published in prestigious academic journals. The CSSH corpus is comparable in terms of disciplines so the number of articles in each discipline corresponds to those in the SciELF-SSH corpus. For the purposes of corpus compilation, some specific disciplines were subsumed into a more general category; for instance, an article from corpus linguistics was included in the 'linguistics' category. The articles were drawn from the Web of Science and Scopus databases. I selected only well-established journals based on their impact factor (Web of Science data 2018) or SCImago rankings (Scopus 2018); only journals in Quartile 1 were included. If one discipline was represented by numerous articles in the SciELF (e.g. economics), papers from several journals were selected; therefore, the CSSH corpus is comprised of research articles from 41 different journals altogether.

With regard to the language background of the authors, the intention was to select articles written by Anglophone writers. The SciELF corpus includes 45 single-authored articles and 27 multi-authored ones, so the reference corpus has been compiled accordingly. In the case of single-authored articles, it was possible to assume that the author's L1 was English

based on their names, affiliations and professional CVs (if available). The situation was more complicated with the multi-authored ones, since it was sometimes difficult to find texts where all the authors were native speakers. Therefore, in the case of multi-authored articles, the first author is always a native speaker of English, but in 10% of papers not all the other authors are L1 English speakers. However, since all the papers have been published in highly ranked journals, it is safe to assume that the texts have undergone editorial changes and/or have been proofread.

The CSSH corpus comprises the same number of articles as the SciELF-SSH, i.e. 72, totalling 621,267 words, with 8,629 average word count per article. The articles are on average longer than those in the SciELF, which is mainly caused by the length of economic papers published in high-impact journals; the average word count in economic papers is 13,112. All the papers were published between 2016-2018. Table 2 shows the composition of both corpora.

Table 2. Composition of the SciELF-SSH corpus and CSSH corpus

	SciELF-SSH corpus	CSSH corpus
No. of articles	72	72
No. of words	432,837	621,267
Average words /article	6,012	8,629
Texts	final drafts of unpublished RAs	published RAs
Disciplines (no. of articles)	educational sciences (15), economics (15), linguistics (13), sociology (6), psychology (3), history (3), social policy (3), classical philology (2), philosophy (2), law (2), theatre studies (2), anthropology (2), urban design (1), literary studies (1), information sciences (1), art history (1)	educational sciences (15), economics (15), linguistics (13), sociology (6), psychology (3), history (3), social policy (3), classical philology (2), philosophy (2), law (2), theatre studies (2), anthropology (2), urban design (1), literary studies (1), information sciences (1), art history (1)

Prior to the analysis, the texts in the reference corpus were processed. I followed the *SciELF Corpus Manual* (Carey 2015) to make sure that the corpora were comparable, so bibliographic references, block quotes and long stretches of foreign text were omitted in the plain text files (used for concordances), while abstracts were kept.

First, the corpora were searched for specific features which could potentially act as frame markers using *AntConc* (Anthony 2019). The selection of frame markers was based on the list from Hyland and Zou's study (2020) and my previous research on metadiscourse (Guziurová 2018), including the subcategories of sequencers, labellers, goal announcers and topic shifters. Then all retrieved items were examined in context to ensure they functioned as frame markers. Since one of Hyland's (2005) principles of metadiscourse is that it is distinct from propositional aspects of discourse (see also Mauranen 1993), all cases had to be manually checked to ensure they were reflexive. For example, when the word *now* refers to the extralinguistic context functioning as a typical time adverbial, it is not regarded as reflexive (1). However, when it refers to the current text, organizing writer's arguments, it is classified as a frame marker (2).

- (1) World Travel & Tourism Council estimates that from direct and indirect activities, the tourism sector *now* provides a remarkable 9.2% of world GDP. (SSH51)
- (2) Let us *now* proceed to a systematic overview, and a further characterisation, of the four varieties. (SSH11)

Since the two corpora are not equal in size, all the results have been normalized to 100,000 words. Then log-likelihood tests were performed to determine differences of statistical significance. All statistical tests were performed on raw data using the online calculator developed by Paul Rayson (Rayson 2008). If the p-value was <0.05 (the threshold level usually set in linguistics), the results were regarded as statistically significant. The results will be discussed in the following section.

4. Results and discussion

4.1 An overview of frame markers in the two corpora

Overall, the results of the quantitative analysis show that there are no significant differences in the frequency of frame markers between the two corpora. As Table 3 demonstrates, frame markers are more numerous in published articles written by Anglophone authors, but the difference is not statistically significant ($G^2 = 0.13$, $p > 0.05$). However, the distribution of individual types of frame markers varies in the two corpora. In the ELF

corpus, the most prominent are goal announcers, accounting for almost 40% of all frame markers. In the CSSH corpus, sequencers are the most frequent category (40.9% of frame markers), with goal announcers being the second most frequent type, accounting for 32% of the total. In both corpora, topic shifters are the least numerous; however, ELF writers use almost twice as many shifters as Anglophone writers and the difference is statistically significant ($G^2 = 10.83$, $p < 0.01$).

Table 3. Overall incidence of frame markers in the two corpora

Frame markers	SciELF-SSH corpus			CSSH corpus		
	Total no.	Freq. per 100,000 w	%	Total no.	Freq. per 100,000 w	%
Sequencers	195	45.0	27.4	425	68.4	40.9
Goal announcers	281	64.9	39.5	335	53.9	32.2
Labellers	163	37.7	22.9	221	35.6	21.2
Topic shifters	73	16.9	10.2	59	9.5	5.7
Total	712	164.5	100	1040	167.4	100

The distribution of individual subcategories in the CSSH corpus is in line with Hyland and Jiang's (2020) study of interactive metadiscourse, which shows that sequencing devices are by far the most frequent frame markers in soft fields, followed by goal announcers. Similarly, sequencers were most prominent in Cao and Hu's (2014) study of published research articles in three soft disciplines. In the following sections, each subcategory of frame markers is discussed in detail.

4.2 Sequencing text

Sequencers in Hyland's (2005) model correspond to "linking adverbials" expressing enumeration (Biber et al. 1999) or a category of "listing conjuncts" in traditional Quirk et al.'s (1985) terms. More specifically, they form an open class of "enumerative conjuncts" which give a particular structure to a list; the enumerative function "connotes relative priority and endows the list with an integral structure, having a beginning and an end" (Quirk et al. 1985: 636). Since they explicitly signal connections between units of discourse, they are important cohesive devices.

As mentioned above, only sequencers expressing internal relations in discourse are considered metadiscursive. This is in line with internal and external types of conjunctive relations distinguished by Halliday and Hasan (1976). Therefore, example (3) is regarded as reflexive, since the frame marker organizes writer's arguments, but example (4) is considered propositional, as it expresses the time sequence of events in the world (in this particular case, the process of data collection).

- (3) *Finally*, I will present evidence of change in students' understanding of the CCC and DCI as represented in conceptual models [...]. (CSSH17)
- (4) *Finally*, I collected the instructional artifacts Jennifer used while teaching this unit, including presentation slides, handouts, grading rubrics, and assigned readings. (CSSH06)

Table 4 shows the frequency of sequencers in the SciELF-SSH corpus and the reference corpus. The results demonstrate that published RAs contain significantly more sequencers than ELF articles ($G^2 = 24.41$, $p < 0.0001$). The difference is particularly striking with enumerating markers *First/Firstly*, *Second/Secondly* and *finally*. This is in line with Hyland and Jiang (2020), who found that although sequencers have declined in soft knowledge fields (applied linguistics and sociology) over the last 50 years and increased in hard sciences, the forms showing the largest rises overall in their corpus are *first*, *second* and *finally*.

Table 4. Frequency of sequencers in the SciELF-SSH and CSSH corpora

Marker	SEQUENCERS			
	SciELF-SSH corpus		CSSH corpus	
	Total no.	Frequency per 100,000 words	Total no.	Frequency per 100,000 words
finally	42	9.7	97	15.6
First / Firstly	56	12.9	132	21.2
Last / lastly	13	3.1	11	1.8
Next	6	1.4	9	1.5
Second / Secondly	43	9.9	109	17.5
subsequently	0	0	1	0.2
Then	11	2.6	3	0.5
Third / Thirdly	8	1.8	43	6.9

begin	8	1.8	16	2.6
start	8	1.8	4	0.6
Total	195	45.0	425	68.4

It is also interesting to note that CSSH authors overwhelmingly prefer the forms without -ly, i.e. *First* and *Second*, as linking adverbials (95% of cases), while there is more variability among ELF authors. *First* is preferred over *firstly* in 71%, whereas *Secondly* is favoured over the shorter form in 65% of cases in ELF articles. Since -ly is a suffix commonly associated with adverbs, L2 English writers might be more inclined to use this form in the function of a frame marker.

Sequencing devices fulfil a range of functions in research articles. In the introduction, they can signal the overall organization of the paper, as in example (5). They can also present contributions of the study (6) when "occupying the niche" that exists in previous research (Swales – Peak 2012). In the conclusion, sequencers can help summarize findings, present implications for practice or list limitations of the research (7). In general, sequencing devices structure writer's arguments, explanations, reasons, findings, methodological difficulties, aims etc. Numerical sequencers are especially favoured in the genre of research article, where they contribute to the clearness and legibility of the text, facilitating the reader's understanding of the message (Hyland – Zou 2020).

- (5) The paper unfolds as follows: *First*, I review the literature on age, access, and motivation and then utilize this literature to propose a series of relationships that comprise the model. *Next*, I present the methods and results for a study that provides support for the hypothesized relationships. (CSSH41)
- (6) By more accurately measuring agency costs imposed by criminal prosecutions, the present research contributes to understanding both corporate crime and agency theory in three ways. *First*, confirming predictions made by applying agency theory to a legal context defends agency theory against growing skepticism about its predictive validity and calls to reconceptualize it (e.g., Lan & Heracleous, 2010; Pepper & Gore, 2015). (CSSH26)
- (7) It also appeared that in some respects the results were biased due to the corpus used. *Firstly*, the size of the corpus was not large enough to produce a sufficient number of valid instances. (SSH22)

4.3 Announcing goals

The expression of purpose is a powerful rhetorical tool in research articles, and writers are aware of its value in signalling the direction of their arguments (Hyland – Zou 2020). The data show the importance of goal announcers in both corpora as they account for the largest subcategory in ELF articles and the second largest in the CSSH corpus (see Table 5). While ELF writers use more goal announcers than CSSH writers, and the difference is statistically significant ($G^2 = 5.23$, $p < 0.05$), the effect size measured by Phi coefficient is very low ($\Phi = 0.0022$), which suggests a very small correlation.

Table 5. Frequency of goal announcers in the SciELF-SSH and CSSH corpora

Marker	GOAL ANNOUNCERS			
	SciELF-SSH corpus		CSSH corpus	
	Total no.	Frequency per 100,000 words	Total no.	Frequency per 100,000 words
aim*	68	15.7	32	5.2
desire* to	0	0	0	0
focus*	44	10.2	78	12.6
goal*	14	3.2	5	0.8
intend* to	6	1.4	4	0.6
intention	4	0.9	3	0.5
objective*	7	1.6	4	0.6
outline*	7	1.6	16	2.6
propose*	29	6.7	48	7.7
purpose*	32	7.4	28	4.5
seek*	5	1.2	16	2.6
want* to	8	1.8	8	1.3
wish* to	3	0.7	4	0.6
would like to	14	3.2	1	0.2
(in this) section	40	9.2	88	14.2
Total	281	64.9	335	53.9

The importance of explicitly stating author’s purposes has been corroborated by a diachronic study of Hyland and Jiang (2020) that followed the changes of interactive metadiscourse in research articles over the past 50 years. The

study has shown that the only subcategory of frame markers which increased across all disciplines was announcing goals and purposes, ensuring that objectives can be understood by readers (Hyland – Jiang 2020).

Goal announcers can indicate the overall purpose of the text (8) or a local goal for a specific part of the text (9). Outlining the overall purpose of the text is a key part of journal articles; it is one of the obligatory moves of RA introductions in the "create-a-research-space" (CARS) model (Swales – Feak 2012: 331).

- (8) *The aim of this article is to analyse family businesses and approaches to them from the perspective of two extremely different cultures.* (SSH14)
- (9) *In the next section, we succinctly present the theoretical background that supports the development of this study.* (SSH48)

Stating purposes can take a wide range of forms and the two corpora differ in this respect. The most common marker among ELF writers is *aim*, accounting for 24% of all goal announcers. As Table 5 indicates, the marker is almost three times more common in the Sci-ELF corpus than in the CSSH corpus. This is in line with Bondi and Borelli (2018), who found that *aim* was one of the positive (overused) keywords in the economics segment of SciELF, especially in its nominal form. Example (8) is thus a typical statement of purpose in ELF articles. In the CSSH corpus, the range of means expressing writer's purposes was more varied. The most frequent marker in the table, (*in this*) *section*, was followed by a range of verbs, such as:

- (10) *This / The next section discusses / presents / describes / examines* etc. (CSSH)
In this section, we investigate / explore / argue etc. (CSSH)

It also indicates that CSSH writers take greater effort in stating what will be presented in each part of the article, indicating local goals.

Another difference between the corpora can be found in the use of *goal*. This marker was not favoured by either group: the SciELF corpus includes 14 occurrences, out of which 10 are followed by a textual product (11). In published RAs there are only 5 occurrences of *goal* in the function of a frame marker, none of which collocate with a textual product. Rather, the authors talked about goals of *research*, *exploration* or *study* (in the sense of research) (12).

- (11) *The main goal of this paper* is to assess if socio-demographic determinants play a different role in European countries characterised by different welfare systems [...]. (SSH63)
- (12) *One goal of this exploration* is to identify this phenomenon and provide a systematic review of how it operates. (CSSH36)

Finally, the corpora differ in the use of the structure *would like to*, which occurs 14 times in the SciELF as a frame marker, but only once in the CSSH. Overall, ELF writers were more tentative in framing the discourse, using hedges (e.g. *would like to*, *try to*, *endeavour*, *attempt to*). This is not to say that they used more hedges in general, but that they were more tentative in announcing their goals (13) or labelling discourse stages (14) than Anglophone writers. This expression of modesty and politeness might help them gain acceptance of their claims by the discourse community or shield them from potential criticism.

- (13) *In my paper I would like to show* how Constantine's participation in the Council of Nicaea (325) has changed the political thought of Christian theologians. (SSH41)
- (14) *We will now try to summarise* the main problems that lexicographers have to face when considering to write/publish a dictionary of collocations, with regards to the definition of collocation. (SSH42)

4.4 Labelling text stages

Another subcategory of frame markers labels discourse stages. These expressions "signal the current discourse activity and offer an explicit means for writers to mark upcoming text stages or rhetorical functions" (Hyland – Zou 2020: 38). According to Hyland and Jiang (2020), labellers in academic writing are most commonly used to summarise or draw conclusions from an argument at certain points in the text. The writers thus explicitly point out how readers should interpret the preceding message (15). However, labellers can also pause the discourse and signal what will follow (16).

- (15) *To summarise*, better environmental performance is not primarily influenced by SL of any of the three frames. (SSH71)
- (16) However, there are two problems, especially pertinent to my current effort, that I would like to mention *at this point*. (SSH11)

The frequency of markers labelling stages is similar in both corpora (see Table 6). It seems that recapitulating key points in the text, or announcing what will come next, are "conventional rhetorical strategies" in research articles (Hyland – Zou 2020).

Table 6. Frequency of labellers in the SciELF-SSH and CSSH corpora

Marker	LABELLERS			
	SciELF-SSH corpus		CSSH corpus	
	Total no.	Frequency per 100,000 words	Total no.	Frequency per 100,000 words
all in all	2	0.5	0	0
at this point	6	1.4	5	0.8
at this stage	0	0	0	0
briefly	3	0.7	3	0.5
conclu*	67	15.5	52	8.4
in a word	1	0.2	0	0
in brief	1	0.2	0	0
in short	4	0.9	13	2.1
in sum	7	1.6	15	2.4
in summary	0	0	16	2.6
now	29	6.7	11	1.8
on the whole	1	0.2	1	0.2
overall	8	1.8	62	10.0
repeat*	2	0.5	6	0.9
so far	6	1.4	2	0.3
summaris*	21	4.9	29	4.7
thus far	0	0	5	0.8
to sum up	5	1.2	1	0.2
TOTAL	163	37.7	221	35.6

Although the overall frequency of labellers across the corpora is very similar, some of the markers are used differently. The greatest difference can be seen in the use of *overall* as a sentence adverb. While this is the most common labeller in the CSSH corpus (10 occurrences per 100,000 words), it is rare in the ELF corpus (1.8 instances per 100,000 words). This expression seems to be a useful tool for writers to summarise their findings (17), summarise results of previous studies or provide a synthesis of arguments.

- (17) *Overall*, these results revealed that there are different types of participants who play CRYSTAL ISLAND, and by being able to differentiate and identify these types of participants, we can move toward developing adaptive GBLEs that scaffold participants based on their gameplay behaviors. (CSSH63)

It should be noted that the marker *now* is multifunctional, as is often the case with metadiscourse devices. The respective functions are not connected with the adverb alone but result from the structure in which it occurs, i.e. the semantics of the whole phrase, and the context. If *now* occurred with illocutionary verbs or similar expressions indicating the function of the part of text (such as *mention*, *discuss*, *distinguish* or *introduce*), it was classified as a discourse label (examples 18 and 19). If *now* followed or preceded verbs which clearly indicate the change of the topic (*turn to*, *switch to*, *move on*), it was regarded as a topic shifter (20). Although there were some ambiguous cases, the context usually helped determine which category it belonged to.

- (18) As I began to write this article, I invited Anna to read and respond to the analysis of her narrative accounts. [...] My sense was that she may still have found the incident uncomfortable to some degree and did not wish to return to it again. I *now* introduce Anna and the setting in which the reported conflict took place. (CSSH20)
- (19) However, the three dictionaries mentioned by *now* [fino ad ora] do not include idioms within their entries. (SSH42)
- (20) Let me turn *now* to the significance of (rare) exceptions to Conceptual Role Determinism. (CSSH71)

4.5 Shifting topics

The last type of frame markers are topic shifters. They indicate a change in the direction of the discourse, moving from one issue the writer wishes to address to another (Hyland – Zou 2020). Functionally, they correspond to a category of topicalizers introduced by Mur Dueñas (2011) which are used to change the topic, introduce related topics or resume a topic introduced earlier. Topic shifters account for the lowest proportion of frame markers in both corpora, so it is not possible to draw any general conclusions. However, the results in Table 7 indicate that they were used more frequently by ELF writers, and the difference is statistically significant ($G^2 = 10.83$, $p < 0.01$). This

finding is consistent with Mur Dueñas (2011), who reveals that topicalizers (together with endophoric markers) are the only metadiscourse devices which have been found to be significantly more frequent in the Spanish RAs than in English RAs written by American scholars. She concludes that it makes argumentation in Spanish RAs less linear and speculates that it might be one of the traits of texts written in languages favouring "reader-responsible style" (Hinds 1987), such as Spanish.

Table 7. Frequency of topic shifters in the SciELF-SSH and CSSH corpora

Marker	TOPIC SHIFTERS			
	SciELF-SSH corpus		CSSH corpus	
	Total no.	Frequency per 100,000 words	Total no.	Frequency per 100,000 words
back to	5	1.2	4	0.6
digress	1	0.2	0	0
in regard to	1	0.2	0	0
mov* on	2	0.5	1	0.2
now	11	2.5	9	1.4
regarding	31	7.2	8	1.3
resum*	0	0	0	0
return* to	4	0.9	12	1.9
revisit*	2	0.5	1	0.2
shift* to	1	0.2	2	0.3
to look more closely	0	0	1	0.2
turn* to	9	2.1	16	2.6
well	0	0	0	0
with regard to	6	1.4	5	0.8
Total	73	16.9	59	9.5

As can be seen in Table 7, the preposition *Regarding* which introduces a new or related topic in the discourse (21) is characteristic of ELF papers, accounting for 42% of all topic shifters. The expression is simple in terms of lexico-grammatical choices and seems to be a useful signal of a change of focus for the reader. In published RAs, the expressions are more varied, including more personal structures such as (22).

- (21) *Regarding* gender differences, men are more prone to volunteer in professional and political activities as well as in education and leisure activities. (SSH63)
- (22) *We now turn to* discussing the theoretical and managerial implications of the research findings. (CSSH44)

This confirms Bondi and Borelli's (2018) observation that ELF articles show an underuse of personal self-reference in comparison with published RAs.

5. Conclusions

The present study has investigated one aspect of discourse reflexivity, frame markers, in English as a lingua franca academic writing. The findings have shown that ELF writers recognise the importance of this kind of discourse organisation since there is no significant difference between the frequency of frame markers in their articles and published RAs written by Anglophone authors. The ELF writers in the corpus are mostly junior or senior academics (not e.g. undergraduate students) who are clearly aware of rhetorical practices in their fields. However, there is considerable variation in the individual types of frame markers across the two corpora, with published RAs containing more sequencers that explicitly organise arguments and fewer topic shifters which indicate digressions and make the discourse possibly less linear.

The results show that ELF is in many respects similar to ENL as writers must recognize rhetorical expectations of their readers to be successful in their argumentation. Moreover, the research article is a key academic genre which is largely standardized, and writers need to follow disciplinary conventions. Nevertheless, ELF writers tend to use different means of discourse organisation, as the ELF corpus seems to be characterized by a prototypical use of certain devices announcing goals (e.g. *aim*), more tentative presentation of discourse goals and labelling stages, and also simplification as the ELF writers rely on simpler structures (*regarding*) and a limited number of devices. This may support the claim that even though writers from different cultural backgrounds do not have to differ in the overall amount of metadiscourse used, they may diverge in specific realisations of metadiscursive categories (Dontcheva-Navratilova 2021).

The results of this small-scale study should be interpreted with caution since there are several methodological limitations. First, this type of corpus-based metadiscourse research relies on a limited set of metadiscursive items which cannot capture all types of frame markers in the texts. It would also be desirable to investigate other types of interactive metadiscourse, such as endophoric markers or transitions, to get an overall picture of how ELF writers organise their texts. Furthermore, we would need more qualitative studies focusing on specific textual histories (see Lillis – Curry 2010), which would help us understand the role of "literacy brokers" in the process of publishing scientific texts and how metadiscourse is concerned. For example, the case study of Flowerdew and Wang (2016) suggests that revision changes of articles written by Chinese doctoral students concerned among other things textual cohesion and additions of some metadiscursive elements (e.g. endophoric markers). If we knew the final published versions of the articles in the SciELF corpus, we would be able to identify these changes. In spite of these limitations, the current study has identified language variation in metadiscourse patterns on the example of English as a lingua franca academic writing.

REFERENCES

Primary sources

ELFA. *The Corpus of English as a Lingua Franca in Academic Settings*.

<http://www.helsinki.fi/elfa>, accessed June 2022.

SciELF. The SciELF Corpus.

<https://www2.helsinki.fi/en/researchgroups/english-as-a-lingua-franca-in-academic-settings/research/wreifa-corpus/scielf-corpus>, accessed June 2022.

Secondary sources

Anthony, L.

2019 *AntConc* (Version 3.5.8). Tokyo: Waseda University. Available from <https://www.laurenceanthony.net/software>, accessed June 2022.

Biber, D. et al.

1999 *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education Limited.

- Bondi, M. – C. Borelli
 2018 "Publishing in English: ELF writers, textual voices and metadiscourse". In: P. Mur-Dueñas – J. Šinkūnienė (eds.) *Intercultural Perspectives on Research Writing*. Amsterdam/Philadelphia: John Benjamins, 217-235.
<https://doi.org/10.1075/aals.18.10bon>.
- Cao, F. – G. Hu
 2014 "Interactive metadiscourse in research articles: A comparative study of paradigmatic and disciplinary influences", *Journal of Pragmatics* 66, 15-31. <https://doi.org/10.1016/j.pragma.2014.02.007>.
- Carey, R.
 2013 "On the other side: Formulaic organizing chunks in spoken and written academic ELF", *Journal of English as a Lingua Franca* 2 (2), 207-228.
<https://doi.org/10.1515/jelf-2013-0013>.
 2015 *SciELF Corpus Manual. ELFA Project*. University of Helsinki.
- Čmejrková, S.
 1996 "Academic writing in Czech and English". In: E. Ventola – A. Mauranen (eds.) *Academic Writing*. Amsterdam/Philadelphia: John Benjamins, 137-152. <https://doi.org/10.1075/pbns.41.11cme>.
- Cogo, A. – M. Dewey
 2012 *Analysing English as a Lingua Franca: A Corpus-driven Investigation*. London: Continuum.
- Connor, U.
 1996 *Contrastive Rhetoric: Cross-Cultural Aspects of Second Language Writing*. Cambridge: Cambridge University Press.
- Crismore, A. – R. Markkanen – M. Steffensen
 1993 "Metadiscourse in persuasive writing", *Written Communication* 10 (1), 39-71. <https://doi.org/10.1177/0741088393010001002>.
- Dahl, T.
 2004 "Textual metadiscourse in research articles: A marker of national culture or of academic discipline?", *Journal of Pragmatics* 36 (10), 1807-1825.
<https://doi.org/10.1016/j.pragma.2004.05.004>.
- Dontcheva-Navratilova, O.
 2021 "Engaging with the reader in research articles in English: Variation across disciplines and linguacultural backgrounds", *English for Specific Purposes* 63, 18-32. <https://doi.org/10.1016/j.esp.2021.02.003>.
- Flowerdew, J. – S. H. Wang
 2016 "Author's editor revisions to manuscripts published in international journals", *Journal of Second Language Writing* 32, 39-52.
<https://doi.org/10.1016/j.jslw.2016.03.004>.
- Guziurová, T.
 2018 *Metadiscourse in Undergraduate Textbooks and Research Articles in Linguistics*. Ostrava: University of Ostrava.

- Halliday, M. A. K. – R. Hasan
1976 *Cohesion in English*. Harlow: Longman.
<https://doi.org/10.4324/9781315836010>.
- Hinds, J.
1987 "Reader versus writer responsibility: A new typology". In: U. Connor – R.B. Kaplan (eds.) *Writing across Languages: Analyses of L2 Texts*. Reading: Addison-Wesley, 141-152.
- Hu, G. – F. Cao
2015 "Disciplinary and paradigmatic influences on interactional metadiscourse in research articles", *English for Specific Purposes* 39, 12-25.
<https://doi.org/10.1016/j.esp.2015.03.002>.
- Hyland, K.
2004 "Engagement and disciplinarily: The other side of evaluation". In: G. Del Lungo Camiciotti – E. Tognini Bonelli (eds.) *Academic Discourse – New Insights into Evaluation*. Bern: Peter Lang, 13-30.
<https://doi.org/10.3726/978-3-0351-0784-5>.
2005 *Metadiscourse: Exploring interaction in writing*. London: Continuum.
- Hyland, K. – F. (K.) Jiang
2016 "'We must conclude that...': A diachronic study of academic engagement", *Journal of English for Academic Purposes* 24, 29-42.
<https://doi.org/10.1016/j.jeap.2016.09.003>.
2020 "Text-organizing metadiscourse: Tracking changes in rhetorical persuasion", *Journal of Historical Pragmatics* 21 (1), 137-164.
<https://doi.org/10.1075/jhp.00039.hyl>.
- Hyland, K. – H. (J.) Zou
2020 "In the frame: Signalling structure in academic articles and blogs", *Journal of Pragmatics* 165, 31-44.
<https://doi.org/10.1016/j.pragma.2020.05.002>.
- Lillis, T. – M. J. Curry
2010 *Academic Writing in a Global Context. The Politics and Practices of Publishing in English*. London/New York: Routledge.
- Lorés Sanz, R.
2016 "ELF in the making? Simplification and hybridity in abstract writing", *Journal of English as a Lingua Franca* 5 (1), 53-81.
<https://doi.org/10.1515/jelf-2016-0003>.
- Mackenzie, I.
2014 *English as a Lingua Franca. Theorizing and Teaching English*. London/ New York: Routledge. <https://doi.org/10.4324/9781315890081>.
- Mauranen, A.
1993 *Cultural Differences in Academic Rhetoric: A Textlinguistic Study*. Frankfurt: Peter Lang.
2012 *Exploring ELF. Academic English Shaped by Non-native Speakers*. Cambridge: Cambridge University Press.

- 2013 "Hybridism, edutainment, and doubt: Science blogging finding its feet", *Nordic Journal of English Studies* 12 (1), 7-36.
<https://doi.org/10.35360/njes.274>.
- 2017 "A glimpse of ELF". In: M. Filppula et al. (eds.) *Changing English: Global and Local Perspectives*. Berlin/Boston: De Gruyter, 223-254.
<https://doi.org/10.1515/9783110429657-013>.
- Mauranen, A. – N. Hynninen – E. Ranta
- 2016 "English as the academic lingua franca". In: K. Hyland – P. Shaw (eds.) *The Routledge Handbook of English for Academic Purposes*. Abingdon: Routledge, 44-55. <https://doi.org/10.4324/9781315657455>.
- McGrath, L. – M. Kuteeva
- 2012 "Stance and engagement in pure mathematics research articles: Linking discourse features to disciplinary practices", *English for Specific Purposes* 31, 161-173. <https://doi.org/10.1016/j.esp.2011.11.002>.
- Mu, C. et al.
- 2015 "The use of metadiscourse for knowledge construction in Chinese and English research articles", *Journal of English for Academic Purposes* 20, 135-148. <https://doi.org/10.1016/j.jeap.2015.09.003>.
- Mur-Dueñas, P.
- 2011 "An intercultural analysis of metadiscourse features in research articles written in English and in Spanish", *Journal of Pragmatics* 43 (13), 3068-3079. <https://doi.org/10.1016/j.pragma.2011.05.002>.
- 2016 "Modal hedging verbs in English as a Lingua Franca (ELF) Business Management research articles", *Kalbotyra* 69, 153-178.
<https://doi.org/10.15388/Klbt.2016.10371>.
- 2018 "Exploring ELF manuscripts: An analysis of the anticipatory *it* pattern with an interpersonal function". In: P. Mur-Dueñas – J. Šinkūnienė (eds.) *Intercultural Perspectives on Research Writing*. Amsterdam/Philadelphia: John Benjamins, 277-297.
<https://doi.org/10.1075/aals.18.13mur>.
- Murillo, S.
- 2018 "Not the same, but how different? Comparing the use of reformulation markers in ELF and in ENL research articles". In: P. Mur-Dueñas – J. Šinkūnienė (eds.) *Intercultural Perspectives on Research Writing*. Amsterdam/Philadelphia: John Benjamins, 237-253.
<https://doi.org/10.1075/aals.18.11mur>.
- Peterlin, A. P.
- 2005 "Text-organising metatext in research articles: An English-Slovene contrastive analysis", *English for Specific Purposes* 24 (3), 307-319.
<https://doi.org/10.1016/j.esp.2004.11.001>.
- Quirk, R. et al.
- 1985 *A Comprehensive Grammar of the English Language*. London: Longman.

Rayson, P.

- 2008 "Log-likelihood and effect size calculator."
<http://ucrel.lancs.ac.uk/llwizard.html>, accessed June 2022.

Rowley-Jolivet, E.

- 2017 "English as a Lingua Franca in research articles: The SciELF corpus",
ASp 71, 145-158. <https://doi.org/10.4000/asp.4987>.

Shchemeleva, I.

- 2022 "When the unconventional becomes convention: Epistemic stance
in English as a lingua franca research articles", *Journal of English as
a Lingua Franca* 11 (1), 265-279. <https://doi.org/10.1515/jelf-2022-2075>.

Swales, J. M. – C. B. Feak

- 2012 *Academic Writing for Graduate Students: Essential Tasks and Skills*
(3rd edn.). Ann Arbor, Michigan: University Press.
<https://doi.org/10.3998/mpub.2173936>.

Vetchinnikova, S.

- 2014 *Second Language Lexis and the Idiom Principle*. Dissertation. Helsinki:
University of Helsinki.

Wu, X. – A. Mauranen – L. Lei

- 2020 "Syntactic complexity in English as a lingua franca academic writing",
Journal of English for Academic Purposes 43, 100798.
<https://doi.org/10.1016/j.jeap.2019.100798>.

Address: TEREZA GUZIUROVÁ, Department of English Language Teaching, Faculty
of Education, University of Ostrava, F. Šrámka 3, 709 00 Ostrava, Czech Republic.
ORCID code: <http://orcid.org/0000-0002-7739-8980>.

European research project websites and corporate websites: Patterns of evaluation and genre evolution

Enrique Lafuente Millán

University of Zaragoza

ABSTRACT

Websites are fundamental tools for knowledge communication and for strategic identity construction. In academic as well as professional websites, visibility and promotion are constructed via evaluative strategies instantiated through multimodal resources made possible by new affordances provided by this medium. This study aims to investigate the strategies used for promotion and identity construction in academic websites. It also aims to shed light on the way these genres evolve due to technological and socioeconomic factors. To do this, I carry out a comparative analysis of 12 European research project websites comparing them with a reference corpus of 12 corporate websites focusing on their showcasing genres. Then, I complement this analysis with qualitative data from interviews with specialist informants. The results of the analysis show that specific contextual factors largely determine the rhetorical purposes of these websites and their use of evaluative resources. However, despite their contextual differences, the websites in the two subcorpora both seek social validation and construct strikingly similar identities to fulfil that function.

Keywords: digital genres, websites, showcasing genres, evaluation, multimodality.

1. Introduction

In recent years, the way scientific work is shared among members of the scientific community and communicated to different audiences has substantially changed due to the new affordances offered by the Web and

other ICT tools. Besides publishing their research in specialised journals, researchers now engage in various writing practices to conduct, discuss and share research, promote research products and make their work visible for a global community. As a result, increasing attention has been paid in the literature to the impact of technology on knowledge dissemination and the emergence and evolution of technology-mediated genres in science communication (Campagna et al. 2012; Kuteeva – Mauranen 2018; Luzón – Pérez Llantada 2019).

Research on knowledge dissemination has also explored how writers use digital genres for rhetorical and interpersonal purposes. For example, scholars use academic homepages to construct a credible online identity, enhance their reputation and gain visibility (Hyland 2011). Similarly, studies have found that research groups turn to online media such as blogs (Luzón 2018) and Twitter (Kuteeva 2016; Pascual – Mur-Dueñas 2021) to create reputable academic identities and to increase their visibility. Comparable processes of strategic identity construction have been identified in other areas of knowledge communication, such as institutional knowledge dissemination (Engberg 2020) and corporate websites (Domenec 2014).

The need to use technology mediated communication to disseminate and promote their work is perhaps more evident for members of publicly-funded research projects since they are required by funding entities to elaborate plans to communicate their results and activities so as to account for the public funding they have received (Flecha et al. 2018; Gertrudix et al. 2021). Project websites are generally the main channel to achieve these objectives (Marín-González et al. 2017).

In the literature on professional and academic websites the importance of the homepage and other related showcasing genres (*about us, the project, objectives*, etc.) has often been underscored, insofar as they act as the official gateway to the website and they help to promote the institution represented in the website (Askehave – Nielsen 2005). The homepages of academic websites have been described as an ideal tool for strategic self-representation and identity construction (Hyland 2011). In a similar vein, corporate webpages (CWs) have been identified as a means for projecting a company's identity so as to gain reputational or social legitimacy (Domenec 2014).

Research has shown that, in order to perform these functions, the central mode of representation in websites is no longer just text. Identity and promotion are constructed by means of evaluative strategies instantiated through multimodal resources that are made possible by

new affordances provided by this medium (Shepherd – Watters 1998). As a result, it becomes necessary to adopt a multimodal perspective in order to study how different modes aggregate in webpages to create meanings and project certain values.

The main objective of this study is to investigate the (meta)textual and visual strategies used for promotion and strategic identity construction. More particularly, my study will examine the incidence of these rhetorical strategies and the preferred values across the two corpora. Second, despite their contextual differences, I aim to establish potential similarities (and differences) in the rhetorical functions of research project and corporate websites, as well as in the identities projected to help fulfil those functions. Finally, I will explore social and technological factors affecting the production and modification of websites so as to identify patterns of variation and change in the production of these digital genres motivated by these technological and social factors. The present study will be limited to the showcasing genres of research project websites and of corporate websites, since they are regarded as crucial in identity promotion (Askehave – Nielsen 2005) and are consistently found in both types of websites.

2. A theoretical framework for the study of evaluation in websites: Multimodal genres and identity construction

2.1 Genre and genre evolution

Genres are “dynamic rhetorical forms” (Berkenkotter – Huckin 1993: 479) and, as such, they are constantly evolving. Several studies in the field of genre analysis have established that genres and generic practices are shaped not only by social groups and organizations but also by the medium (Askehave – Nielsen 2005; Giltrow – Stein 2009; Bawarshi – Reiff 2010; Rowley-Jolivet – Campagna 2021). Studies have also explored how established print genres are imported into a new medium or how genre variants or even new genres develop and emerge in electronic environments. When digital genres are perceived as emanating from existing printed genres a process of genre *repurposing* takes place, while genre variants or new genres may appear as a result of a process of genre *re-mediation* (Heyd 2015).

To account for these processes, Crowston and Williams (2000) established three types of genre: *reproduced*, *reconfigured* (adapted) and *emergent* genres. Even if traditional written genres (e.g. the research article)

are sometimes reproduced in the digital medium with few or no major changes, they may acquire added value (Luzón 2007) when adapted for online publication, by including for example hyperlinked citations (Crownston), or may be enhanced with add-on genres such as author videos or graphical abstracts (Luzón – Pérez-Llantada 2019).

In contrast, it has been found that certain digital genres display a more hybrid nature (Crownston). For example, Askehave and Nielsen (2005) posited that homepages are hybrid in purpose, as they combine promotional and informational features. In addition, homepages exploit the affordances of the digital medium while displaying characteristics of newspaper discourse (Askehave – Nielsen 2005: 124). Similarly, Alejo González (2005) claimed that pages of commercial websites such as *Home*, *About us*, *Contact us*, or *Products and services* may have originated and evolved from printed genres such as sales promotion letters, since they share the same move structure.

Due to their dynamic nature, the analysis of the impact of the digital medium on the processes of evolution and change in these genres continues to be relevant. Digital genres emerge and change to accommodate new rhetorical needs and other contextual factors. As a result, exploration of such processes needs to consider “the various social, economic, and technological factors that occasion the production, reproduction, or modification of different genres in different sociohistorical contexts” (Yates – Orlikowski 1992: 320). To account for the influence of these factors on genre evolution, the present study uses multi-perspective methods of critical genre analysis (Bhatia 2008) which take into account textual and contextual aspects of the genre.

2.2 Identity and authorial evaluation

Identities (and authorial evaluation) are social positions that writers take when they interact, and both are constructed at the level of discourse (Flowerdew – Wang 2015). Important aspects of identity are indexed through linguistic resources such as evaluative markers and other metadiscourse strategies. Webpage writers must be able to strategically use evaluation to construct credible and reputable identities and to promote the institutions described in these websites by assembling attributes and values that are consistent with social expectations and which are culturally approved and esteemed in their field or context (Hyland 2011).

In the last decades there has been extensive research on the interpersonal potential of evaluation in written academic genres including research articles (Giannoni 2005; Hyland 2005), research article abstracts (Martín Martín – Burgess 2004; Stotesbury 2003), referee reports (Fortanet 2008) and book reviews (Alcaraz-Ariza 2011). More recently, research has also studied evaluation in academic and professional websites (Caiazzo 2009; Lorés Sanz 2020; Suau-Jiménez 2019). Numerous different perspectives including concepts such as *attitude*, *evaluation*, *stance*, *appraisal* or *affect* have been used in the literature to define roughly similar and sometimes overlapping aspects such as the writers' personal response, attitude or value judgement of the entities they are referring to, the people they are interacting with or the material they discuss.

Following Hunston (1993, 2011) evaluation expresses an attitude towards a person, situation or entity and is both subjective and located within a societal value system. Hunston (1993) and Thompson and Hunston (2001) explain that evaluation can be performed along different parameters like value, relevance and status. Yet, in this study, evaluation of status or certainty (also known as epistemic evaluation) will be left out of the analysis since it does not directly contribute to promotion and identity construction.

As Thompson and Hunston (2001) contend, evaluation allows writers to persuade their readers to see things in a certain way. When writers use certain values to represent themselves, evaluative language functions as a promotional strategy that enables writers to construct credible and reputable identities. The consideration of what counts as positive value, however, depends on the attributes and values that are culturally approved and appreciated by the social group.

2.3 Multimodality

The study of digital genres requires discourse analysts to examine modes, multimodal configurations, and their semiotic functions in discourse. Writers construct meaning through the selection and configuration of modes and therefore text or meaning cannot be studied in isolation (Jewitt 2015). Multimodal analyses need to investigate how choices in the modes of communication affect meaning and generate social effects.

Previous research contends that the use of the visual mode has increased in professional writing. This extensive use of images is due to the limitations of the verbal mode in fulfilling the cognitive and functional goals

writers set for their texts (Rowley-Jolivet 2002). The use of images in texts has great potential for meaning making and for strategically influencing the readers' attitudes and beliefs (Wekesa 2012). Not surprisingly, visual elements play an important role in constructing identities in academic websites (Hyland 2011), as well as in promoting corporate values in annual reports (David 2001) and corporate websites (Domenec 2014).

Research on multimodality has often emphasized the importance of the social context for meaning making (Kress 1997, 2010). Meaning is viewed as socially situated since modes of communication have been shaped by the social functions they have historically been used to accomplish (van Leeuwen 2005; Kress 2010). As Hyland (2011) explains, writers combine linguistic, content and design features to construct self-representations that will be recognized and valued by their intended readers and, in doing so, they strategically manage the impression they make on them.

The model of "visual metadiscourse" developed by De Groot et al. (2016) is suitable for the analysis of multimodal genres, as it allows us to identify and account for the way visual and verbal elements are used by writers to convey meaning and perform specific rhetorical functions such as evaluation and identity construction. This model draws on Hyland and Tse's (2004) textual metadiscourse framework and on Kress and Van Leeuwen's (2006) work on multimodal discourse. It focuses on writer-reader interaction and also explores the socio-cultural aspects present in pictures. As De Groot et al. (2016) explain, visual metadiscourse allows writers to create and communicate ideas about reality or to convey (subjective) attitudes and evaluative meanings to guide the readers interpretation of the text. Images derive their meaning both from their intrinsic features and the textual and cultural context they are inserted in. Using this approach to visual metadiscourse, I will study the way writers exploit the possibilities of multimodal texts to strategically construct their identities.

3. Methodology

To undertake this study, I selected a corpus of 12 websites of research projects funded by the EU H2020 programme for research and innovation. These websites were part of a database of 100 websites (EUROPROW4), collected as part of the InterGEDI research project on digital scientific discourse analysis. One criterion for inclusion in my corpus was that at

least one of the participants in the project was linked to the Universidad de Zaragoza.

For the sake of comparability and coherence, all the websites dealt with the topic of sustainable energy, which is key in the EU research agenda, and were subject to the same requirements for dissemination of project results from the EU H2020 programme. A parallel corpus of 12 corporate websites from top international corporations in the field renewable energies was compiled and used as a reference corpus (Table 1).

To ensure the comparability of the two corpora, all the texts included in the corpus were examples of showcasing genres (*home, about, the project, objectives*, etc). When the name of the web page was not provided or was different from the ones commonly identified as showcasing genres, its communicative purpose was established and used as the criterion for selection.

Table 1. H2020 and corporate websites in the corpus

H2020 projects in the field of renewable energies (UZ participants)	Corporations in the field of renewable energies
ADREM	Acciona wind power
AGROinLOG	Avangrid Renewables
BuildHEAT	Berkshire Hathaway Energy (BHE)
FLEXICIENCY	EDF Energy
GreenGain	GE Energy
Indus3es	Iberdrola
Medeas	Jinko
MIGRATE	NextEra Energy, Inc.
SCOOPE	Orsted (FKA DONG Energy)
SteamBio	Siemens
uP_Running	Suzlon
WASTE2FUELS	Vestas
TOTAL WORDS IN CORPUS: 13160	TOTAL WORDS IN CORPUS: 13911

The corpus was read manually several times. Instances of evaluation in verbal and visual resources (see for example Figures 1 and 2) were identified and compiled in a data base. Successive readings allowed for data reduction and a set of values was generated from the data. Results were sorted and labelled using Microsoft Excel database affordances.



Figure 1. Extracted from <https://www.aspire2050.eu/adrem> (latest access May 2020)

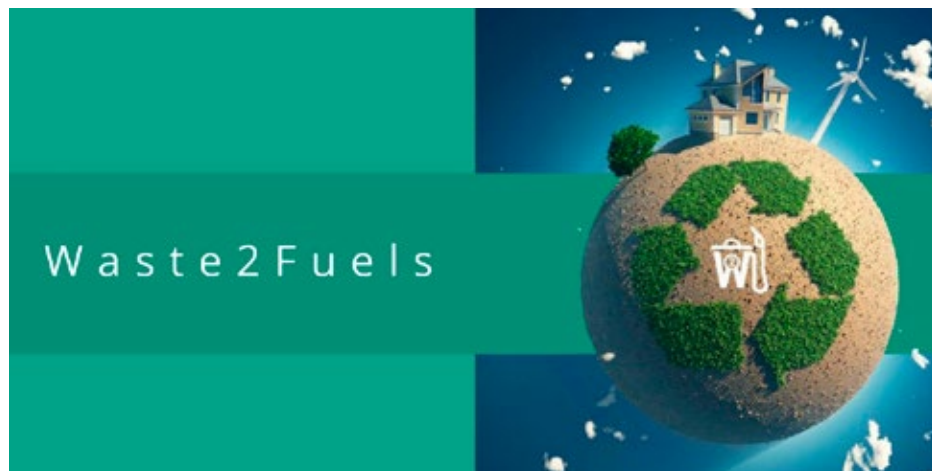


Figure 2. Extracted from www.waste2fuels.eu (latest access June 2019)

Moving from text to context, I also enquired into the professional and discursive practices of the target communities using qualitative data from interviews with three informants involved in supervising, designing, and drafting these websites:

1. A chief technical director and consultant with extensive experience in digital project management for private and public companies
2. A senior manager for a research institution participating in over 60 H2020 projects
3. A junior communications manager for the abovementioned research institution

Prior to the interviews, a number of questions addressing key issues related to the functions, objectives and process of design of a website were generated by the author. These questions were then edited and revised with the help of two other scholars belonging to the InterGedi research group. The final interview schedule (see appendix 1) was used to carry out a semi-structured interview with each informant. Two of interviews were performed face to

face. Due to restrictions during the Covid-19 pandemic, the third interview was held online. All interviews were recorded and transcribed. Interview transcriptions were then analysed and findings were used to triangulate the results obtained from the corpus analysis.

4. Results

Important aspects of identity can be strategically constructed by means of evaluation in texts. The manual analysis of the webpages in the H2020 and in the CWs subcorpora revealed the preferred values website writers projected in their texts through the use of evaluative markers and images.

4.1 Evaluation at the verbal level

As we can see in Table 2, writers in H2020 project websites prioritised presenting the project as “innovative” and “technically useful”, which together represent 28% (N = 57) of the evaluative markers in the texts in this subcorpus.

Table 2. Evaluative markers (raw counts) in the H2020 and the CWs corpora

H2020 website corpus			Corporate websites corpus		
Relevant / Important	13	6%	Relevant/ important	87	27.8%
Accurate / Reliable / Solid	10	5%	Accurate / Reliable / Solid	5	1.6%
Problematic / Complex	6	3%	Problematic/ Complex	4	1.3%
Profitable	10	5%	Profitable	7	2.2%
Cost effective	26	13%	Cost effective	20	6.4%
Beneficial for society	29	14%	Beneficial for society	27	8.6%
Green	15	7%	Energy efficient	24	7.7%
Sustainable	8	4%	Sustainable	45	14.4%
Energy efficient	28	14%	Innovative / Modern	21	6.7%
Innovative / Modern	29	14%	Pioneer	20	6.4%
Technically useful	28	14%	Technically useful	53	16.9%
Total	202	100%	Total	313	100%

As might be expected given the common field shared by these websites, values related to energy efficiency (“green”, “sustainable”, “energy efficient”) were also frequently found in H2020 texts, accounting for one fourth (25%) of all the evaluative markers. In a similar vein, writers often described

the project and its products as “cost-effective” and “profitable” (13% and 5% respectively). Aside from discussing its monetary perks, writers of H2020 webs also stressed the benefits these research projects entailed for the well-being of society in general (14%). In brief, evaluation in H2020 websites mainly focuses on promoting its technical and innovative edge, as well as the benefits they can provide for society and the environment.

It is worth noting that the sets of values appearing in the texts in the CW subcorpus are notably akin to the values in the H2020 websites. Despite this apparent similarity between the two subcorpora, there is considerable variation in the total number of evaluative markers and the relative incidence of some of these values, i.e. the preferred values conveyed in the two subcorpora. Writers in the CW subcorpus, for example, conveyed the notion that the companies and their products were “relevant or important” some 87 times (27.8%), compared to only 13 times in the H2020 corpus. These writers also accentuated their sustainability (N = 45) more often and portrayed themselves as a “pioneer” in the field (N = 20) and “technically useful” (N = 53). In contrast, the incidence of other values such as “beneficial for society”, “cost-efficient”, “profitable”, “innovative” or “energy efficient” in the CW subcorpus was comparable to that in the H2020 corpus. Verbal texts in corporate websites are very brief yet highly promotional. To successfully do this, evaluative markers stressing the positive values of the company are compactly packed in these texts, as we can see in examples 1 and 2.

- (1) [The company’s] technical expertise, comprehensive portfolio and long-standing experience are helping to pioneer a sustainable future across the globe.
- (2) [The company] is leading this transformation and co-creating the future of energy with our customers, providing safe, efficient, reliable, and affordable power to drive economic growth and raise living standards around the world.

We can conclude from these data that evaluative markers in CWs prioritise positioning companies as important, technically advanced and pioneering, thus helping to carve a space for them in the competitive corporate field.

4.2 Evaluation through visuals

Table 3 shows the evaluative meanings website authors conveyed through the use of visual metadiscourse in the webpages under analysis. Once again,

the values projected in the text to construct their identities were surprisingly similar in both subcorpora. The most obvious difference across the two sets of texts lies in the frequency and quality of these visual elements. The choice of values projected by these visuals may also be described as rather strategic, as visuals tended to be used most frequently to stress socially sanctioned values, i.e. *people-oriented* and *socially minded*; *green* and *sustainable* (N = 58), as well as to advertise their dominant status in the field (*innovative/technologically advanced* and *global*).

Table 3. Evaluative meanings conveyed through visuals in the H2020 and the CWs corpora

H2020 website corpus			Corporate websites corpus		
European	3	6%	Relevant/ important	5	3%
Neutral	8	17%	Neutral	9	5.4%
Green / Nature oriented	6	13%	Green/ Nature oriented	12	7.2%
Sustainable	10	21%	Sustainable	31	18.7
Team player/ collaborative	5	11%	Team player/ People Oriented	37	22.3%
Socially minded	3	6%	Socially Minded	21	12.7%
Innovative / technological	8	17%	Innovative/ Technological	29	17.5%
Energy efficient	4	9%	Efficient	5	3%
Total	47	100%	Pioneer/ oriented to future	6	3.6%
			Global	11	6.6%
			Total	166	100%

Table 4. Types of visuals in the H2020 and the CWs corpora

H2020 website corpus			Corporate websites corpus		
XL Images (Full screen)	4	8.2%	XL Images (Full screen)	42	20%
L / M Images	25	51%	L / M Images	83	39.7%
S / XS Images	1	2%	S / XS Images	44	21.1%
Logos / Icons	12	24.5%	Logos / Icons	34	16.3%
Videos	4	8.1%	Videos	6	2.9%
Others (Diagram, Flag, Graph)	3	6.2%	Total	209	100%
Total	49	100%			

Corporate webpages display a much more intensive use of high-quality visuals indexing evaluation (see Figure 3). Most of the pictures are large or even full screen pictures (Table 4) which often constitute the central element in multimodal ensembles (Figure 1 above). In these webpages the screen is sometimes organized by the logic of image, instead of the text.

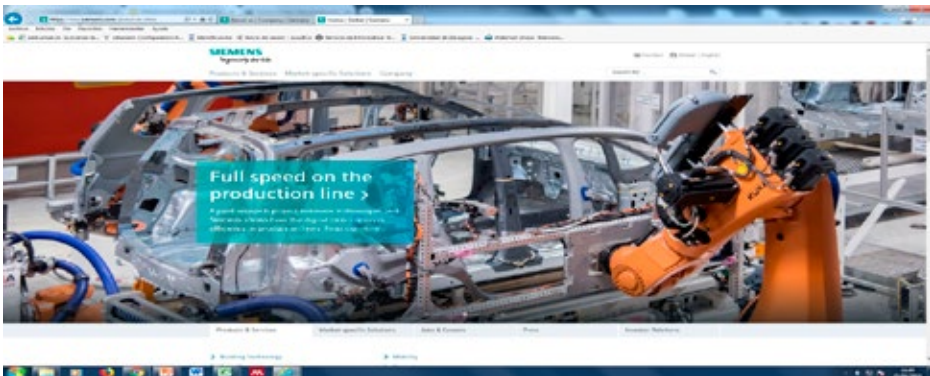


Figure 3. Extracted from www.siemens.com (latest access June 2019)

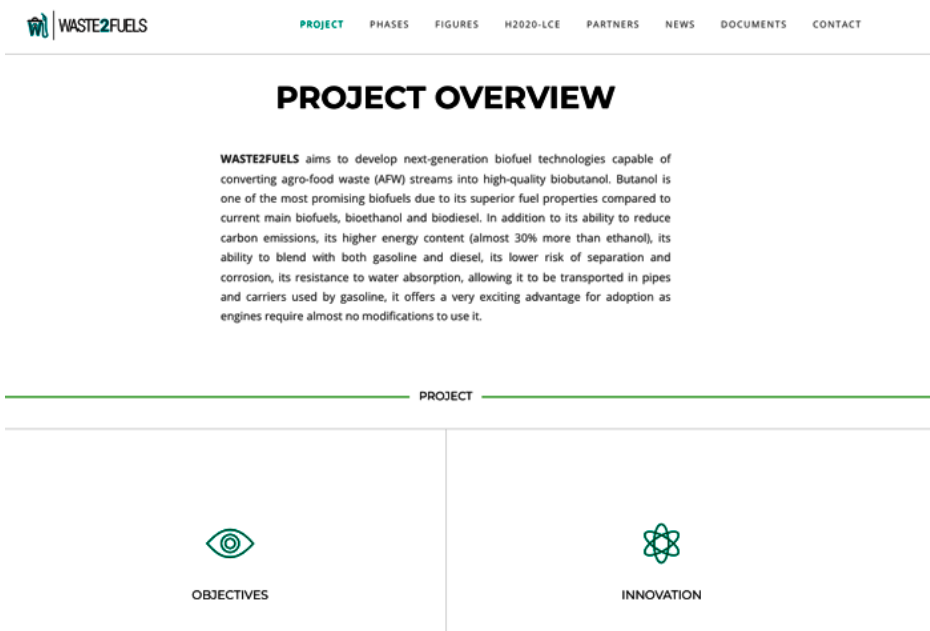


Figure 4. Extracted from www.waste2fuels.eu (latest access June 2019)

In contrast, the use of visuals in H2020 websites is much less significant in terms of frequency, size and quality. The values which are more recurrently

underlined in their visuals are more field dependent (*sustainable, innovative, green*) and are mainly aimed at constructing the identity of experts in the field of renewable energies. Moreover, in these websites authors create multimodal ensembles combining text and small images or logos where simplicity is prioritised (see Figure 4).

The results presented reveal significant differences in the overall incidence of both evaluative markers and visuals across the subcorpora as well as in the type of evaluative meanings conveyed through the two modes analysed. The intensive use of evaluation allows writers to strategically create preferred identities and to steer readers to intended interpretations. Nevertheless, in order to understand the goals and motivations guiding these rhetorical uses and the processes underlying text creation we need to shift from text to context and adopt a more qualitative perspective of analysis.

4.3 Analysis of contextual factors

As discussed in section 2, genres evolve to adapt to the authors' rhetorical needs, which in turn are determined by contextual factors. To account for the evaluative strategies and the identities projected in these texts, as well as for the construction and evolution of these genres, we need to consider the social, economic and technological factors present in the context in which they are used (Yates – Orlikowski 1992). In order to do this, I will re-examine the results derived from the analysis of visual and verbal metadiscourse using the qualitative data obtained from interviews with three specialist informants.

Webpages are dynamic in nature; therefore, the processes of evolution and change are largely influenced by the digital medium and by other technological developments. One of the most obvious features of H2020 websites is that their designers prioritize simplicity and the use of static sections that tend to follow a similar structure. Simplicity is also a feature of corporate websites and it can be linked to specific economic and technological factors, as well as to the priorities and reasons motivating the design of the website. As informant 1 reported (example 3), simplicity of content and design allows a website to upload quickly, which is a characteristic valued by users. In addition, when web designers operate on a limited budget, they tend to adopt open access programs that provide templates, making web design simple and accessible. As a result of this, these websites often display a somewhat similar structure.

- (3) Today there is a tendency towards simplicity of content and of web design. The objective is for the website to upload as quickly as possible. [...] The most popular programs used for designing websites (e.g. Word press) are open access and incorporate templates (Informant 1)

Web designers also try to exploit the digital affordances of websites in order to capture the readers' attention and to make it easier for them to process the information in the text. Informants 2 and 3 stress that, in H2020 websites in particular, technical affordances of webs are used not only to increase the number of visits but also its duration, so that visitors interact with the contents and these contexts have the desired impact in the target audience (examples 4 and 5).

- (4) We try to take advantage of the possibilities the web has to offer. We try to make the web dynamic using pop-ups, hyperlinks, shapes that upload as you hover over them... This way the reader, who is not an expert, is not overwhelmed and gets to see things little by little, which is more appealing. (Informant 3)
- (5) Reaching the target audience is not trivial. We analyse the data we can obtain from the web itself to find the number of visits and the time they spend on the website, because if people visit our websites and leave within seconds this means they are not interested in what they see. Therefore, there are different indicators we use, including social media and other channels, to measure the impact of the website and to find out if communication is being effective. (Informant 2)

Traffic to the websites is determined by technological, as well as contextual and economic factors. According to the data obtained from the informants (example 6), it is far more difficult for an H2020 website to attract traffic than for the website of a well-known corporation, as users will not look for the project name or will visit its homepage on their own initiative. Traffic to project websites is promoted through google searches of specific technical terms or through other social media (e.g. twitter) where specific events, news or products are communicated and may become viral.

As a result of these new practices and strategies for finding and accessing websites, H2020 project homepages and other showcasing genres are losing importance in favor of more dynamic sections which will be linked to social media and which contain key terms that may come up in Google searches (example 7).

- (6) The most complicated thing is to create a website that people can reach directly, especially if the website is very technical and is not frequently updated. It is easier to get to people through the news because we send them to national online media and the project is always named. This generates visits. (Informant 3)
- (7) It is difficult to get visits through the homepage. Google usually takes users directly to the news or demos. Perhaps, if they are interested, they stay on the website and they look around. (Informant 2)

A similar process seems to be taking place in corporate websites. According to informant 1, the homepage has lost importance in recent years and has given up its role as the key internet genre because internet users no longer access websites through the homepage to find and access content. Google searches, in fact, do not direct users to the homepage of a website (example 8).

- (8) Homepages are no longer the entrance gate to the different sections and content. The algorithms used by search engines to determine traffic are designed not to measure traffic through the homepage but direct traffic towards the pages with specific content within the website. The reason for this is that internet users use google, not homepages to find content. Another example of this is that the open source framework for web development designed by Google, which is known as AMP (Accelerated Mobile Pages), does not even include homepages among its options. (Informant 1)

The main goal in H2020 project websites is to publicize the project so as to reach and inform a non-specialized audience that may be interested in the topic. Unlike in CWs, self-promotion or establishing reputation are not priority objectives since these projects have a short life span and they assume they can reach only a relatively small number of internet users. These competitive research projects are required by their funding entities to design and implement communication plans to disseminate their research activities to the general public (Gertrudix et al. 2021) so as to inform citizens of how public research funding is being used, thus increasing accountability and transparency (examples 9 and 10).

- (9) The European Union requires us to transfer and communicate our results as the project develops. The main aims of the website are to make the results of the project known [...] One key function is

knowledge dissemination; so that anyone can access information that they can understand even if they are not experts. (Informant 2)

- (10) Every call for projects includes a task involving dissemination, communication or something similar. Within that task, there are several subtasks and one of them is the design of a website and, more recently, another subtask is setting up accounts in social media. (Informant 3)

The description of scientific research processes and results through a medium which can reach multiple audiences can potentially cause confusion among readers and even context collapse. A communications manager needs to be aware of the importance of considering the audiences these texts are drafted for and how this determines the complexity of the language used, the structure of the text, the amount of explicit information and the use of digital resources to facilitate text processing (example 11). Aspects related to text structure and specific language in these web-generated genres, however, fall beyond the scope of this study and will not be examined here.

- (11) Our target is a non-specialized audience who wants to find out what we are doing. Specialists have other means to access more specific information on the topic. [...] As a communications manager I need to ensure the information is simple enough and easy to understand. (Informant 3)

Crownston claims that some digital genres such as the digital version of RAs display a hybrid and transitional nature, as the limits of the genre (e.g. the use of citations, whether the digital text will be considered as a merit for academic advancement, etc.) are negotiated in the context of knowledge production and dissemination. According to informant 3, however, these are new emergent genres which have little or no similarities with other previously existing written genres (example 12).

- (12) I do not find any similarities between these webpages and other written genres. At least since I have been working here [in the design of websites for research projects] they are nothing like that. (Informant 3)

While the main aim of an H2020 website is two-fold – to disseminate and account for the research being done and to stress the benefits the project will bring about for citizens – corporate websites are designed to promote

a company and its products, establish a niche for the company against the competition and create a reputable and socially sanctioned identity. This could explain why the use of promotional strategies through evaluative markers and visuals conveying positive values is less frequent and less strategically relevant in H2020 websites than in CWs (example 13).

- (13) We do not pay a lot of attention to images. We do not use them as much as corporate websites, where they are extremely important and they consider them carefully. You will find them mostly in the homepage. (informant 3)

Corporate websites use evaluation to strategically create specific identities in order to manage their reputation in the eyes of a more general audience. Besides using self-promotional strategies which will allow them to construct identities as *leading*, *technologically advanced* or *global* companies, they focus intensively on values such as being *green*, *sustainable* or *socially minded*, which allow them to construct socially approved identities and position the audience in favor of the company, as well as its products and activities. To achieve these promotional objectives, CWs devote many more resources to take advantage of multimodal ensembles combining different modes (verbal, visual, aural) to produce the intended unified meaning, as evidenced by the quantitative data presented in the previous section. This united orchestration of different modes allows the designer to obtain what Lemke calls a “multiplying effect” (1998).

5. Conclusions

This study set out to investigate academic writing in online environments and explore in a comparative manner the possible similarities and differences in the use of verbal and visual evaluation for promotion and identity construction in websites related to sustainable energies: by this means aiming to better understand web-based communication practices in the scientific world vis-à-vis the corporate world. The quantitative results showed notable differences across the subcorpora in the overall incidence of verbal evaluative markers and visual metadiscourse (De Groot et al. 2016), as well as in the attributes and values most frequently assembled in each subcorpus. Nevertheless, there were also notable similarities in the types of values conveyed in the two corpora.

A second objective of this analysis was to establish potential similarities and differences in the rhetorical functions these web-mediated texts fulfilled, as well as in the identities projected to perform those functions. The informants' responses confirmed that the main functions of websites in the H2020 and in the CW corpora were largely determined by their particular contexts. The main function of H2020 websites is to publicize the project so as to reach a non-specialist general audience and inform that audience about their research activity, thus accounting in the eyes of society for the funding they received. The success of these websites depends on their overall impact, as measured by the number of visits and the duration of those visits. In contrast, corporate websites are designed to help companies establish a reputation in the field and to help them create a niche.

The preference for certain values and identities projected in the websites in the corpus may also be traced back to the main functions intended for the websites. H2020 websites aim at dissemination of the projects' findings and activities, as well as transparency and accountability. Hence, when drafting their websites, writers are less aware of the need to adopt promotional strategies and to establish their value against competitors. In contrast, strategic projection of identity through evaluation is far more prominent in corporate websites, as it enables them to manage their reputation and to gain social legitimacy. To project these identities, corporate websites exploited the use of visual metadiscourse extensively. As suggested by Rowley-Jolivet (2002), the increasing use of images in professional discourse may be explained because linguistic modes do not allow writers to fulfil their functional or rhetorical goals, in this case strategic manipulation of the readers' attitude, as efficiently as the visual mode.

Despite their contextual differences, the websites in the two corpora also share a similar rhetorical function, i.e. gaining social acknowledgement and acceptance. As described by Engberg (2020), public institutions often need to publicize their work to reduce feelings of strangeness and animosity from society, whose members may otherwise question their usefulness and may fail to see the justification for their funding. Corporations in the field of renewable energies, for their part, also use websites to position the audience and public opinion in their favor. To do this, corporate websites project values (*beneficial for society, socially minded/ people-oriented, energy efficient or sustainable*) which are socially approved attributes and which allow these corporations to construct a socially acceptable and valued identity. In contrast, they simultaneously use evaluation and promotion to construct powerful competitive identities to establish their primacy, relevance and technological superiority. Websites

therefore constitute potent tools in order to manage and shape their reputation and, together with other marketing actions and tools, may be used to project values and identities which are in line with the values and demands of society as part of the company's purpose. Following Selzer's terminology (2004, in Afros – Schryer), promotion in websites therefore relies on field specific attributes (*logos*) such as innovation and efficiency, as well as on values shared and approved by the community (*pathos*).

Together with these socio-economic factors, technological factors play an important part in shaping the way the digital affordances of research project websites are used and on the evolution of the genre. To achieve their aims, these websites need to be made appealing, simple and easy to upload. Furthermore, technical standards for web design have also led to webpages becoming more and more standard in structure.

Social and technological factors have also affected the way traffic is generated and have changed the role homepages and other showcasing genres play in internet communication. Since H2020 research projects have a short shelf-life and are not known to the general public, in order to attract users to their websites project members have to publicize their activities through their news sections or the use of their own social media. In addition, internet users in general no longer access websites through the homepage but rather do so through search engines and social media. As a result, the homepage and other showcasing genres are becoming less relevant, and even disappearing. It could be concluded from this that, just like the affordances of websites determine the contents and metadiscourse strategies used in websites, the way users find and access contents (largely through Google) influence the way these affordances change over time.

The present study has used a quantitative and qualitative analysis to obtain a better understanding of evaluative strategies in websites by examining the context and professional practices. Nevertheless, this study has some limitations that future research should take into consideration. An important objective of this study was to throw light on the use of linguistic features for expressing evaluation in the same form of communication but in two professional contexts (academic vis-à-vis corporate). One of the limitations when trying to obtain qualitative data to do this was that the specialist informants I interviewed were largely unaware of the evaluative strategies they used when writing these texts and of what makes these genres effective. In addition, the qualitative data obtained from informants revealed that static parts of a website, such as showcasing genres, are decreasing in importance, while dynamic sections of websites such as news and demos are now fundamental to direct traffic to the websites. Future

studies on websites will need to investigate these dynamic sections as well the use of other social media (Twitter, LinkedIn, etc.) as a means to interact with the target audience.

REFERENCES

Primary sources

EUROPROW4 (*European projects websites 4*)
<http://www.intergedi.unizar.es/>, accessed November 2023.

Secondary sources

- Afros, E. – C. F. Schryer
 2009 “Promotional (meta)discourse in research articles in language and literary studies”, *English for Specific Purposes* 28 (1), 58-68.
<https://doi.org/10.1016/j.esp.2008.09.001>.
- Alcaraz Ariza, M. A.
 2011 “Evaluation in English-medium medical book reviews”, *International Journal of English Studies* 11 (1), 137-153.
- Alejo González, R.
 2005 “Textual metadiscourse in commercial websites”, *Iberica* 9, 33-53.
- Askehave, I. – A. E. Nielsen
 2005 “Digital genres: A challenge to traditional genre theory”, *Information Technology & People*, 18 (2), 120-141.
<https://doi.org/10.1108/09593840510601504>.
- Bawarshi, A. S. – M. J. Reiff
 2010 *Genre: An Introduction to History, Theory, Research and Pedagogy*. Lafayette, LA: Parlor Press.
- Berkenkotter, C. – T. N. Huckin
 1993 “Rethinking genre from a sociocognitive perspective”, *Written Communication* 10 (4), 475-509.
<https://doi.org/10.1177/0741088393010004001>.
- Bhatia, V. K.
 2008 “Towards critical genre analysis”. In: V. K. Bhatia – J. Flowerdew – R. H. Jones (eds.) *Advances in Discourse Studies*. London: Routledge, 166-177.
- Caiazzo, L.
 2010 “The ‘promotional’ English(es) of university websites”. In: R. Cagliero – J. Jenkins (eds.) *Discourses, Communities and Global Englishes*. Bern: Peter Lang, 43-60.

- Campagna, S. et al. (eds.)
2012 *Evolving Genres in Web-Mediated Communication*. Bern: Peter Lang.
<https://doi.org/10.3726/978-3-0351-0436-3>.
- Crowston, K.
2010 "Internet genres". In: M. Bates – M. Niles Maack (eds.) *Encyclopedia of Library and Information Science*. Boca Raton, FL: CRC Press.
<https://doi.org/10.1201/9780203757635>.
- Crowston, K. – M. Williams
2000 "Reproduced and emergent genres of communication on the World-Wide Web", *The Information Society* 16 (3), 201-216.
<https://doi.org/10.1080/01972240050133652>.
- David, C.
2001 "Mythmaking in annual reports", *Journal of Business and Technical Communication* 15 (2), 195-222.
<https://doi.org/10.1177/105065190101500203>.
- De Groot, E. et al.
2016 "Picture this: Developing a model for the analysis of visual metadiscourse", *Journal of Business and Technical Communication* 30 (2), 165-201. <https://doi.org/10.1177/1050651915620235>.
- Domenec, F.
2014 "Monsanto's and Chevron's home pages in the US and the UK: Corporate discourse as a reflection of social trends?", *Iberica* 27, 51-76.
- Engberg, J.
2020 "Multimodal institutional knowledge dissemination and popularization". In: G. Tessuto et al. (eds.) *The Context and Media of Legal Discourse*. Newcastle upon Tyne: Cambridge Scholars Publishing, 55-76.
- Flecha, R. – A. Radauer – P. Besselaar
2018 *Monitoring the Impact of EU Framework Programmes: Expert Report*.
<https://data.europa.eu/doi/10.2777/518781>, accessed June 2021.
- Flowerdew, J. – S. H. Wang
2015 "Identity in academic discourse", *Annual Review of Applied Linguistics* 35, 81-99. <https://doi.org/10.1017/S026719051400021X>.
- Fortanet, I.
2008 "Evaluative language in peer review referee reports", *Journal of English for Academic Purposes* 7 (1), 27-37.
<https://doi.org/10.1016/j.jeap.2008.02.004>.
- Gertrudix, M. et al.
2021 "Scientific communication in the digital space: actions for the dissemination of research projects under the H2020 program", *El Profesional de la Información* 30 (1), 1-13.
<https://doi.org/10.3145/epi.2021.ene.04>.

- Giannoni, D. S.
 2005 "Negative evaluation in academic discourse. A comparison of English and Italian research articles", *Linguistica e Filologia*, 71-99.
https://doi.org/10.6092/LeF_20_p71.
- Giltrow, J. – D. Stein (eds.)
 2009 *Genres in the Internet: Issues in the Theory of Genre*. Amsterdam: John Benjamins. <https://doi.org/10.1075/pbns.188>.
- Heyd, T.
 2015 "Digital genres and processes of remediation". In: A. Georgakopoulou – T. Spilioti (eds.) *The Routledge Handbook of Language and Digital Communication*. London: Routledge, 87-102.
<https://doi.org/10.4324/9781315694344>.
- Hunston, S.
 1993 "Evaluation and ideology in scientific writing". In: M. Ghadessy (ed.) *Register Analysis: Theory and Practice*. London: Pinter Publishing, 57-73.
 2011 *Corpus Approaches to Evaluation: Phraseology and Evaluative Language*. New York: Routledge. <https://doi.org/10.4324/9780203841686>.
- Hyland, K.
 2005 "Stance and engagement: A model of interaction in academic discourse", *Discourse Studies* 7 (2), 173-192.
<https://doi.org/10.1177/1461445605050365>.
 2011 "The presentation of self in scholarly life: Identity and marginalization in academic homepages", *English for Specific Purposes* 30 (4), 286-297. <https://doi.org/10.1016/j.esp.2011.04.004>.
- Hyland, K. – P. Tse
 2004 "Metadiscourse in academic writing: A reappraisal", *Applied Linguistics* 25, 156-177. <https://doi.org/10.1093/applin/25.2.156>.
- Jewitt, C.
 2015 "Multimodal analysis". In: A. Georgakopoulou – T. Spilioti (eds.) *The Routledge Handbook of Language and Digital Communication*. London: Routledge. <https://doi.org/10.4324/9781315694344>.
- Kress, G.
 1997 *Before Writing: Rethinking Paths to Literacy*. London: Routledge.
 2010 *Multimodality: A Social Semiotic Approach to Contemporary Communication*. London: Routledge.
<https://doi.org/10.4324/9780203970034>.
- Kress, G. – T. van Leeuwen
 2006 *Reading Images: The Grammar of Visual Design* (2nd edn.). London: Routledge. <https://doi.org/10.4324/9781003099857>.
- Kuteeva, M.
 2016 "Research blogs, wikis and tweets". In: K. Hyland – P. Shaw (eds.) *The Routledge Handbook of English for Academic Purposes*. London: Routledge, 433-445. <https://doi.org/10.4324/9781315657455>.

- Kuteeva, M. – A. Mauranen
 2018 “Digital academic discourse: Texts and contexts. Introduction”, *Discourse, Context and Media* 24, 1-7.
<https://doi.org/10.1016/j.dcm.2018.06.001>.
- Lemke, J. L.
 1998 “Multiplying meaning: Visual and verbal semiotics in scientific text”. In: J. R. Martin – R. Veel (eds.) *Reading & Science*. London: Routledge, 87-113.
- Lorés, R.
 2020 “Science on the web: The exploration of European research websites of energy-related projects as digital genres for the promotion of values”, *Discourse, Context & Media* 35, 1-10.
<https://doi.org/10.1016/j.dcm.2020.100389>.
- Luzón, M.-J.
 2007 “The added value features of online scholarly journals”, *Journal of Technical Writing and Communication* 37 (1), 59-73.
<https://doi.org/10.2190/H702-6473-8569-2R3Q>.
 2018 “Constructing academic identities online: Identity performance in research group blogs written by multilingual scholars”, *Journal of English for Academic Purposes* 33, 24-39.
<https://doi.org/10.1016/j.jeap.2018.01.004>.
- Luzón, M.-J. – C. Pérez-Llantada
 2019 “Connecting traditional and new genres. Trends and emerging themes”. In: M.-J. Luzón – C. Pérez-Llantada (eds.) *Science Communication on the Internet. Old Genres Meet New Genres*. Amsterdam: John Benjamins, 1-18.
<https://doi.org/10.1075/pbns.308.01luz>.
- Martín-Martín, P. – S. Burgess
 2004 “The rhetorical management of academic criticism in research article abstracts”, *Text* 24, 171-195.
<https://doi.org/10.1515/text.2004.007>.
- Pascual, D. – P. Mur-Dueñas
 2022 “Dialogic interaction with diversified audiences in Twitter for Research Dissemination Purposes”, *Círculo de Lingüística Aplicada a la Comunicación* 90, 61-79. <https://doi.org/10.5209/clac.81307>.
- Rowley-Jolivet, E. – S. Campagna
 2021 “From print to web 2.0: The changing face of discourse for special purposes”, *LSP Journal* 2 (2), 44-51.
- Selzer, J.
 2004 “Rhetorical analysis: Understanding how texts persuade readers”. In: C. Bazerman – P. Prior (eds.) *What writing does and how it does it. An Introduction to Analyzing Texts and Textual Practices*. London: Lawrence Erlbaum Associates, 279-308.

- Shepherd, M. – C. Watters
 1998 “The evolution of cybergenres”, *Proceedings of the Thirty-First Hawaii International Conference on System Sciences* 2, 97-109.
<https://doi.org/10.1109/HICSS.1998.651688>.
- Stotesbury, H.
 2003 “Evaluation in research article abstracts in the narrative and hard sciences”, *Journal of English for Academic Purposes* 2, 327-341.
[https://doi.org/10.1016/S1475-1585\(03\)00049-3](https://doi.org/10.1016/S1475-1585(03)00049-3).
- Suau Jiménez, F.
 2019 “How can hotel websites discursively adjust to customer preferences using online criticism”, *Ibérica* 38, 203-226.
- Thompson, G. – S. Hunston
 2001 “Evaluation: An introduction”. In: S. Hunston – G. Thomson (eds.) *Evaluation in Text: Authorial Stance and the Construction of Discourse*. Oxford: Oxford University Press, 1-27.
- van Leeuwen, T.
 2005 *An Introduction to Multimodality*. London: Routledge.
<https://doi.org/10.4324/9781315638027>.
- Wekesa, N. B.
 2012 “Cartoons can talk? Visual analysis of cartoons on the 2007/2008 post-election violence in Kenya: A visual argumentation approach”, *Discourse & Communication* 6 (2), 223-238.
<https://doi.org/10.1177/1750481312439818>.
- Yates, J. – W. J. Orlikowski
 1992 “Genres of organizational communication: A structurational approach to studying communication and media”, *The Academy of Management Review* 17 (2), 299-326. <https://doi.org/10.2307/258774>.

APPENDIX

INTERVIEW PROTOCOL

I. Functions, characteristics, structure

1. What are the main functions of the website of a European research project? What are the main reasons for making the website?
2. Do you also use it to sell your products or contact potential buyers?
3. Do you think it successfully reaches the target audience? How do you assess this?
4. What are the characteristics of a good website for a European research project? What aspects make it effective?
5. Which sections of your website are the most relevant? Why?

II. Decision-making

6. How are the websites of European research projects in which you take part designed? What decisions need to be made?
7. What considerations are taken into account when designing the website, for example facing a certain sector or a certain company? When you are making decisions, what aspects do you take into account or want to prioritize?

III. Writing processes

8. Who is in charge of writing the texts of the web page? Is the website written in the local language and then translated to other languages?
9. You have mentioned that the news pieces are very important. What about the more static pages which present/showcase the website (Home, Project, about us). Who writes them? What are their functions?
10. How often/When is the content of the website updated? Which sections are most often updated/modified?

IV. Audience

11. Who do you think are the potential readers of the website?
12. Do you also use the website to inform other colleagues and project members?
13. What do you do to reach non-specialized audiences? How do you share technical information?

V. Visibility and digital identity

14. How do you increase traffic to the website and reach potential recipients?
15. Is the website an effective mechanism to make the project visible? Is it an efficient investment (of time/ money)?
16. What is the image or identity that you try to project? Do you emphasize any specific values?
17. How important is it to project specific identities/values?

VI. Multimodal Resources

18. What criteria/processes do you follow to choose the images, videos, ...?
19. Why are images or videos included on certain occasions?
20. In what sections do you think videos make the most sense?

Address: ENRIQUE LAFUENTE MILLÁN, Facultad de Educación, Universidad de Zaragoza, C/ S. Juan Bosco 7, 50009 Zaragoza, Spain.

ORCID code: <https://orcid.org/0000-0002-4783-3953>.

VARIA

Language, narrative and structure of storytelling in museum communication: A diachronic approach

Cecilia Lazzeretti

University of Bozen-Bolzano

ABSTRACT

Storytelling is at the core of museum activities (Bedford 2001), but little attention has been paid to its role in museum communication (MC) from a discourse-analytical perspective. This paper aims to show how narrative has developed as part of MC since the 1970s, appearing in a wide array of museum texts, from press releases to digital genres. Typically, museum storytelling goes hand in hand with evaluative language, lexical markers of AFFECT and forms of non-standard language (emoji). The methodological toolkit for the study is qualitative in focus and draws on frameworks developed for the analysis of museum discourse (Ravelli 2007). The dataset consists of a corpus of museum press releases, dating from 1950 to 2016, and a sample of digital museum texts, dating from 2015 to 2021. The study confirms that MC represents an interesting field for discourse studies, providing a repertoire of innovative practices not only in terms of storytelling techniques, but communication strategies in general.

Keywords: storytelling, narrative, museum communication, diachronic studies, discourse analysis, emotive language, emoji.

1. Introduction

The pervasive role played by storytelling in museum settings is widely recognised by practitioners, who have based most of their activities, from exhibitions to educational programs, on the art of telling stories. Bedford (2001) argues that the *raison d'être* of the museum has to be traced back to a story worth telling and passing on from generation to generation, thus claiming that stories are the “real thing” of museums (Bedford 2001: 33). Similarly, Nielsen (2017: 6) sees museums as storytellers “with a firm grasp of

what can be communicated and understood”, constantly creating meaning and eliciting individual interpretation from the public through the use of narrative resources.

In a global context of communication dominated by storytelling, the forms of narration implemented by museums seem to be qualified by an added value. As highlighted by Hughes (2021), museums are generally trusted by the public and perceived as a reliable source of information: they are still considered as a safe zone, unaffected by misinformation and fake news, where critical thinking can still be exercised and authentic knowledge dissemination take place. In the current times, when opinions are strongly polarised and society is exacerbated by divisions, museums are therefore called upon to assume a great responsibility: they have the opportunity to share thought-provoking stories, dealing with salient issues, such as globalisation, inequality, climate change, and migration, that are credible in the eye of the public.

Today, it is not only exhibitions that are organised following the framework of stories (Hughes 2021). Professionals in the education departments of museums also rely heavily on them, and have developed principles for effective storytelling according to different segments of the public (Di Blasio – Di Blasio 1983), involving verbal and non-verbal communicative practices (Burdelski et al. 2014). These forms of museum storytelling require a wide array of skills: Bourlakovs et al. (2017) argue that behind the narrative of a guided tour is the joint work of an educator, a performer, a psychologist, and a contractor, the latter being able to manage strict time constraints.

Storytelling is very much part of the picture even in the communication departments of museums, when it comes to prompt visitor interaction and participation, i.e. at the stage of reaching out to the audiences (see Krämper 2017 on the benefits of using storytelling in museum public relations). Yet, comparatively little attention has been paid to the role stories play in museum promotional and digitally-mediated communication (Nielsen 2017). Museum communication (MC), often considered as the domain of practitioners (media and public relations experts), actually represents a burgeoning research area for discourse studies, pioneered in the 1990s by Louise Ravelli (1996, 2006, 2007) and sharing common ground with the domains of art, media, and tourism discourse. Inherently discursive aspects of MC, such as lexis, structure and narrative, deserve attention on the part of linguists, as their analyses can add to the experience of practitioners and offer further insights into professional practice.

This paper aims to provide a contribution in this respect, by analysing the use of narrative in MC across diverse genres, such as press releases,

web pages and social media posts. Drawing on frameworks applied in the qualitative analysis of museum discourse (Ravelli 2007), the study explores the language features and structure underlying narrative in MC from a diachronic perspective. Two main research questions are addressed: 1) when did museums start using narrative in communication? and 2) what are the typical language features of storytelling used in MC?

The next section provides a brief presentation of the materials used for the study, as well as of the methodology adopted. The analysis will first deal with the identification of prototypical forms of narrative in museum press releases, considered as a founding genre for MC, on which all subsequent genres have been built, and then move on to an overview of recontextualised forms of narrative and micro-narrative in newly emerged museum genres, such as websites and social media. The discussion of the data will focus on the degree of manipulation and transformation of basic narrative structure across genres, taking into account the typical combination of narrative with emotive language and the gradual introduction of non-standard language features (emoji).

2. Materials and methods

The analysis carried out in this study is based on two main resources: a corpus of press releases issued by US and UK museums between 1950 and 2016 (*EPA Diacorp*, see Lazzeretti 2016), and a sample of web pages and social media posts collected from US and UK museums' digital platforms between 2015 and 2021. The non-homogeneous nature of the dataset's two components must be recognised in advance. Whereas in the case of press releases – which are 'finished documents' and, once published, cannot be changed – texts were collected across seven decades, web pages and social media were observed within a more limited time span, covering less than a decade. This is due to the ephemeral nature of Internet texts, which "were not designed to be considered 'finished products', but to be continuously built on and modified" (Tagg 2015: 34), or even deleted by their authors. These considerations affected the way digital museum texts were collected and led to focus on a period which could guarantee actual data availability. This can be roughly dated between the start of the author's research in the field of digital MC (September 2015) and January 2021.

While in the first part of analysis, aimed at tracing the origins and evolution of narrative in MC, an ad-hoc structured corpus of 430 press releases

(378,315 words) has been utilised in order to mirror language change over seven decades, in the second part, a selective approach has been adopted so as to illustrate how the use of narrative has been transferred and recontextualised in museum digital media: therefore, some notable examples have been chosen across the digitally-mediated communication produced by museums and dealt with in the analysis. The selection was made bearing in mind Glaser – Strauss’ (2006) principles of theoretical sampling, whereby “the process of data collection is controlled by the emerging theory” (2006: 45). Accordingly, the examples presented in this paper were chosen for their value in illustrating particular theoretical phenomena noted by the researcher.

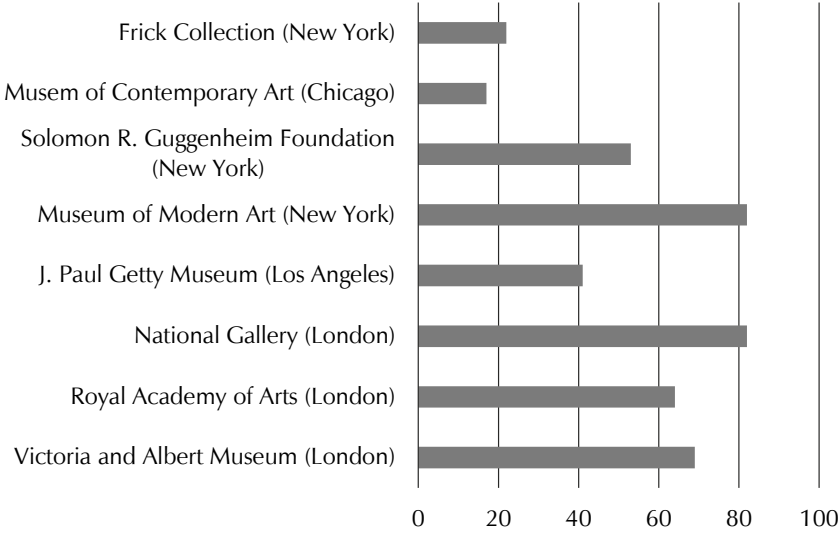


Figure 1. Museum sources of press releases

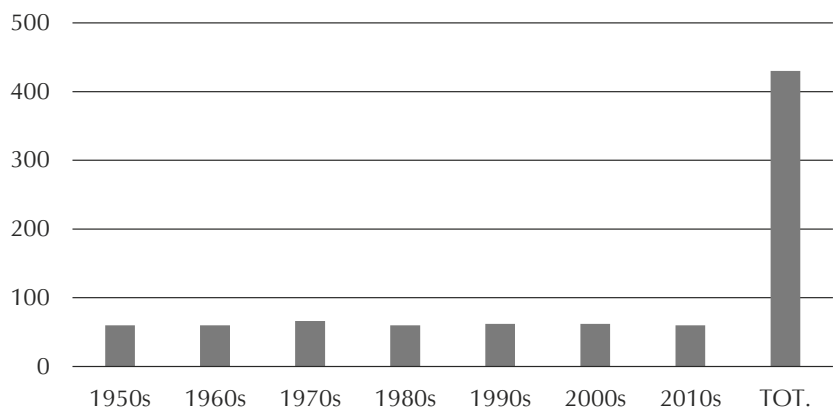


Figure 2. Number of press releases collected per decade

The graphs reproduced above (Figure 1 and Figure 2) show sources and number of press releases in use for the diachronic analysis, while Table 1 summarises the main features of the sample of museum digital media. References for web pages and social media posts taken into consideration in the second part of the study will be contextually provided throughout the analysis.

Table 1. Main features of the sample of museum digital media

Observation Period	September 2015 – January 2021
Number of observed museum websites	40
Number of observed museum social network sites	30
Number of collected web pages	60
Number of collected social media posts	120

The methodology adopted in the study is discourse-analytical and draws on frameworks applied in the qualitative analysis of texts in museums (Ravelli 2007). To a lesser degree, the study entails a quantitative approach, insofar as it aims to determine when MC began to rely on narrative techniques and to quantify their use over time. Consequently, the calculation of word frequency, made possible by Wordsmith Tools 8 (Scott 2020), and the manual analysis of texts were combined.

Since the main focus of the study is on storytelling in MC, some clarifications of the theoretical approach taken here need to be mentioned, starting from a terminological one. As pointed out by Pireddu (2018: 3), when speaking about museum storytelling, “we are dealing with technique rather than content” and, more specifically, with “a selection of techniques that can fit into the museum environment”. Emphasis, therefore, is placed on the skill of museum professionals as tellers and on their ability of telling a story which is compelling and engaging. Conversely, narrative is a theoretical notion and deals with more formal aspects of a story, as comprised of structural components. In the context of MC, storytelling and narrative appear closely intertwined, as museums, in consequence of being multimodal spaces (Hofinger – Ventola 2004), need to think both in terms of performance – the art of telling a story – and message – the content and articulation of the story itself. Bearing in mind this distinction, the present study would ideally take into account both aspects, but will necessarily focus on the material signs of narrative identified across diverse museum genres, i.e. on the discourse which

conveys the story. To that end, the canonical structure of narrative and its discursive features, as described by Toolan (2001), will serve as a framework for the present study and guide the analysis of museum texts. Narrative, in Toolan's sense, is "a recounting of things spatiotemporally distant" (Toolan 2001: 1), consisting of "events, characters and settings" (2001: 12). Furthermore, the discourse of narrative is defined by a series of typical discursive signs: a certain degree of constructedness and prefabrication; a 'trajectory', i.e. some sort of development, coming to a resolution, or conclusion; a teller (and, consequently, an addressee); linguistic markers of displacement, referring to things or events that are removed, in space and time, from both the speaker and the addressee (Toolan 2001: 4-5).

Evaluation, stance and appraisal are all terms which describe the ways the subjective views of speaker or writer are conveyed in language (Hunston – Thompson 2000; Martin – White 2005; Biber et al. 1999). In the theoretical approach of Appraisal Theory (Martin 2001), in particular, the function of AFFECT is understood as a semantic resource for constructing emotions and establishing an interpersonal relation with the reader. Language used for expressing feelings and opinions plays a key role in narrative, as authors often deploy evaluation to highlight the news story potential and emphasise aspects of narrative complication (van Dijk 1988). The use of narrative may also be combined with emotive language, relying on an intense vocabulary which triggers an emotional response in the readers (Ungerer 1997).

3. Analysis and discussion

3.1 Narrative in museum press releases

Museums have embraced public relations in the form of a structured practice, by hiring professionals in the field, dedicating a museum department to media relations and investing resources in communication projects, only after the end of the Second World War. In this respect, US museums have led the field and provided a model for other cultural institutions worldwide (McLean 2012). Hence, it is worth exploring when the fascination with stories began and storytelling emerged as a powerful communication tool for museums.

As suggested by Baker (2011), language change over time can tell us much about changes within culture, society, and everyday life. The appearance of new words, as well as the decline or re-conceptualisation of

old ones, across corpora containing texts from different periods, can give us a hint at phenomena on a larger level. Baker's suggestion is applied here to an historical corpus of museum press releases covering seven decades, from 1950 onwards, to corroborate the results of manual inspection of texts. More specifically, frequency of the words "story" and "stories" has been calculated over decades. These words have been selected for their metadiscursive function: they can be considered, in Hyland's (2004, 2005) sense, as metadiscoursal features, able to explicitly organise the propositional content and signal the writer's attitudes towards both their writing and their audience. As such, these words may reflect awareness of writers regarding the use of narrative at different points in time.

The graph below (Figure 3) shows the frequency of "story" and "stories" over decades – from 1950 to 2010 – in the corpus under scrutiny. Data were obtained by using the 'detailed consistency analysis' function in WordSmith Tools 8 (Scott 2020), enabling the analyst to compare multiple word lists.

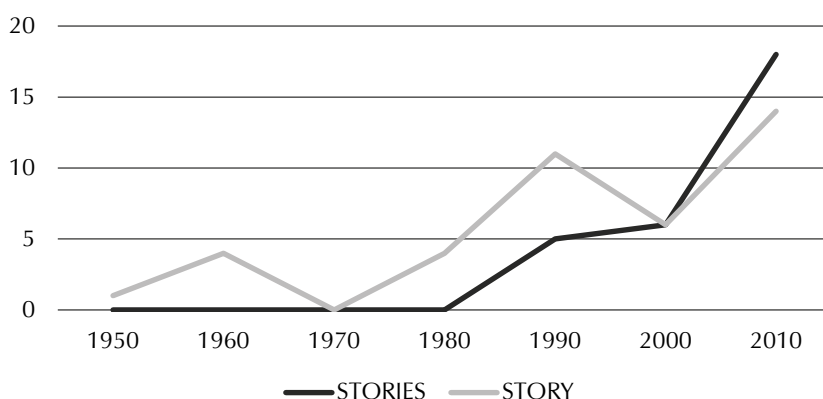


Figure 3. Absolute frequency of the nouns story and stories (1950-2010)

The frequency pattern shown by both nouns is that of an irregular increase over decades, going from none or very few occurrences in 1950 up to 15 or more in 2010. In particular, the period between 1970 and 1980 appears crucial, as from these years on the frequency of "story" and "stories" starts to increase sharply. These findings are in line with those provided by close reading of press releases and manual identification of prototypical narrative forms within their structure. It is in fact from the 1970s on that museum press releases start relying more consistently on a new structural element: narrative sections. These are textual sequences, varying in length and

illustrating facts or cases of particular interest for the journalistic readership, structurally characterised by the presence of the basic components of stories, i.e. characters, events and settings (Toolan 2001). The following excerpt (1) provides an early example of these sections.

- (1) “During these same years Rossetti had moved to a house in Cheyne Walk in Chelsea. A large and brilliant group of friends met there – Ruskin, Swinburne, Morris, Whistler and Burne-Jones among them, and along with the amazing menagerie of wombats, kangaroos and armadilloes that were kept there, they proved an irresistibly fascinating target for Rossetti’s contemporaries. [...] However from 1870 onwards it slowly began to break up under the pressure of Rossetti’s worsening physical and mental health which turned him into a recluse. [...] It seems that Rossetti’s growing feelings of remorse over Elizabeth Siddal’s death became identified in his mind with that hopeless grief which Dante felt after the death of his Beatrice”.

(Press Release of the Royal Academy of Arts, London – 1 January 1972)

The example above is taken from a press release on Pre-Raphaelite painters, whose works were exhibited at the Royal Academy of Arts, London, in 1972. In particular, this narrative section was retrieved from a specific rhetorical step within the organisation, or macrostructure, of museum press releases, which is the biography of the featured artists. In fact, it is in this step, appearing in the body of the press release, that narrative is more likely to appear, as part of a strategy aimed at arousing interest and curiosity around an artist. In this case, the biography of Dante Gabriel Rossetti is told in a way that activates imaginative processes and connects with our experiences and emotions. Emphasis is placed on the unusual and exceptional company the artist used to keep, on his wife’s tragic death and his subsequent illness.

The second example (2) was also retrieved within the biography step of a press release announcing a major retrospective dedicated to German artist Joseph Beuys by the Guggenheim Museum in 1979. The narrative focusses on a serious accident Beuys had during the war, which, it turns out, was crucial for his career and determined his production as an artist, including the choice of using felt and fat for his sculptures. Readers are led to empathise with the leading character of the story told and to re-live his experiences. The story ignites a process of self-reflexivity (Giddens 1991), disclosing emotions and feelings readers can relate to (Georgakopoulou 2015).

- (2) “In World War II, as a combat pilot in the German AirForce, Beuys was seriously injured several times. After a near-fatal plane crash during a snowstorm on the Russian Front, he was found by Crimean Tartars who saved his life by wrapping him in a thick insulation of fat and felt. [...] He had never fully recovered from his shattering war experience and withdrew into a state of depression that lasted throughout the 50s. Beuys emerged from this period of crisis with new convictions which led him to testify against Nazism”.

(Press Release of the Solomon R. Guggenheim Museum, New York – 1 November 1979)

In both examples 1 and 2, a narrative prototypical structure can be recognised on the basis of the presence of basic components, as identified by Toolan (2001): characters (Dante Gabriel Rossetti/Joseph Beuys, both flanked by side characters), spatio-temporally distant settings (XIX century, Pre-Raphaelite Circle / World War II, Russian front), and a development of events culminating in a conclusion or resolution (popularity and success – death of wife – grief and illness / plane crash – rescue – depression – creativity and artistic awareness). The aim of emotional engagement is, however, obtained through the combination of a narrative structure with evaluative language (Hunston – Thompson 2000). In both cases, evaluation is key to the success of the narrative. Based on Martin and White’s (2005) APPRAISAL system, the evaluative markers used in the narrative sections presented above, such as *a large and brilliant group of friends, amazing menagerie, irresistibly fascinating target, worsening health, growing feelings of remorse, hopeless grief, near-fatal plane crash, seriously injured, never fully recovered, shattering war experience*, can be regarded as linguistic realisations of AFFECT, addressing positive or negative feelings, moods, dispositions and attitudes (Martin – White 2005). This category includes “a range of grammatical metaphors, including nominalised realisations of qualities (joy, sadness, sorrow) and processes (grief, sobs, constriction in his throat)” (Martin – White 2005: 46). Markers of AFFECT allow the attitude of the narrators toward the narrative emerge: authors project their subjective views on situations, unveiling their emotional state and establishing relations with readers.

The use of narrative sections in museum press releases has increased over the decades. As shown by Table 2, these were extremely rare in early examples of the genre, while today at least 1 out of 3 press releases – 35% – include narrative sections. This result may suggest an evolution in the practice of writers, who have gradually become more aware of the potential of narrative in terms of reader involvement and newsworthiness.

Table 2. Number of press releases including narrative sections across decades

	N. of PRs	%
1950	2	3%
1960	2	3%
1970	9	14%
1980	15	25%
1990	16	25%
2000	15	23%
2010	21	35%
TOT	80/430	18%

To explain the increasingly important role played by narrative in museum press releases, one has to bear in mind that stories are particularly valued by journalists, the “professional storytellers of our age” (Bell 1991: 147). Press releases with a narrative potential have news value in the eye of reporters and therefore greater chances of being used and retold in subsequent news stories. As van Dijk (1988: 87) puts it, newsmakers tend to “make the news report livelier... conveying both the human and the dramatic dimension of the news event”. Narrative sections provide

therefore that *quid* of ‘newsworthiness’ which seems to be an indispensable element for the implementation of what has been defined by Jacobs (1999) as the metapragmatics of the press release, i.e. its function of reuse by journalists and its pre-formulation aimed at such reuse.

3.2 Narrative in museum websites

In the mid-to-late nineties, with the development of the first websites, museums started to slowly venture into digital media. At an early stage, museum websites mostly replicated the printed material online and offered basic information, while later on many institutions began to digitise their collections and offer virtual visits (Rizzo – Mignosa 2013).

The storytelling tradition, so firmly embedded in museums, was automatically applied to the new media. As observed by Geismar (2018), museum websites reproduce older representational frameworks, showing the same issues of “classification, narration, value, and perspective that are on display in the galleries” (2018: 78). Furthermore, web narrative is used across different areas of MC, as in the case of *ad hoc* storytelling for educational web content (Glover Frykman 2009), or specific narratives conveying architectural identity in museum websites (Pierroux – Skjulstad 2011).

Narrative sections, in the form already identified for press releases, typically feature in specific pages of the museum website, such as those on the history of the institution, or on the statement of museum identity and mission (‘who we are’ and ‘about’ pages). However, forms of narratives

in nuce, or micro-narratives, which promise or imply a story, even though not fully developed in structural terms, can be disseminated also elsewhere in museum websites. This is the case, for instance, of the Gardner Museum website, where storytelling is used as the predominant way to engage with visitors.¹

The institutional image of the Gardner Museum is built entirely around its founder, Isabella Stewart Gardner, an eccentric and wealthy Bostonian heiress who collected more than 16,000 works of art throughout her life. The website of the Gardner Museum aims to reflect her unique personality by sharing a selection of meaningful stories, set in a colourful and lively context. Isabella is the main character of a narrative pervading the website both at the verbal and visual level, depicting an ambitious, unconventional woman, unafraid of pursuing dreams and desires, even if it implied risks. Beside the main narrative devoted to Isabella Gardner, a series of further, secondary narrative layers are disseminated throughout the website, such as those on the Venetian palace built in Boston to house her collection and the infamous theft happened at the museum in 1990.^{2,3}

Right from the homepage (see Figure 4), visitors are invited to 'meet Isabella', called by her name all across the website to establish a sense of familiarity, and 'learn about the daring visionary and the museum she created'.



Figure 4. Homepage of the website gardnermuseum.org

¹ <https://www.gardnermuseum.org/>, accessed October 2021.

² <https://www.gardnermuseum.org/about/building-isabellas-museum>, accessed October 2021.

³ <https://www.gardnermuseum.org/about/theft>, accessed October 2021.

Bearing in mind the attitudinal provocation (Martin – White 2005) inherent in the evaluative coupling ‘daring visionary’, this invitation sounds as a prelude to a story and can serve as a narrative cue. As noted by Baumbach (2015), these forms of micro-narratives, or narratives *in nuce*, are particularly involving for readers, who need to activate their cognitive schemata to interpret them and complement the information provided with their previous “knowledge of images, figures and narratives” (Baumbach 2015: 29). The storytelling approach characterising the website is then made fully explicit by means of metadiscourse elements, as the final link guides users to discover ‘stories and more’: the upcoming narrative content is therefore clearly anticipated by the audience, who knows exactly what to expect.

Following the link on the homepage, users willing to ‘meet Isabella’ are sent to a richly illustrated ‘about’ page, which is structured as a major, multimodal extension of the narrative sections previously identified in museum press releases (see Figure 5).⁴ The page, divided into subsections, provides a detailed account of Isabella Gardner’s ‘unconventional life’, from her early life, to her marriage, travels, friendships, and lastly her ‘visionary’ legacy. Isabella is often defined as ‘a visionary’ throughout the website: she is called ‘a daring visionary’ in the homepage, ‘a visionary founder’ in the mission statement. The noun ‘vision’ (like its synonym ‘dream’) is therefore key to the characterisation of her personality. Lexical choices also rest on the theme of sensuality: Isabella created an immersive environment and she used to say that her collection was her pleasure. The museum is therefore described as a source of ‘pleasure’ – as it was for Isabella, it can also be for visitors – and as a place to ‘awaken your senses’.

In addition to evaluative language, quotations are strategically used throughout the text to amplify the main character of Isabella, appearing as a sort of literary heroine that would fit well in a novel by Henry James or Edith Wharton (who, by the way, were both friends of Gardner in real life). As shown by the examples below (1-6), the voices quoted in the web page are those of side characters in Isabella’s story: friends, acquaintances, reporters.

- (3) *“Mrs. Jack Gardner is one of the seven wonders of Boston. There is nobody like her in any city in this country. She is a millionaire Bohemienne. She is the leader of the smart set, but she often leads where none dare follow... She imitates nobody; everything she does is novel and original.”* – a Boston reporter

⁴ <https://www.gardnermuseum.org/about/isabella-stewart-gardner>, accessed October 2021.

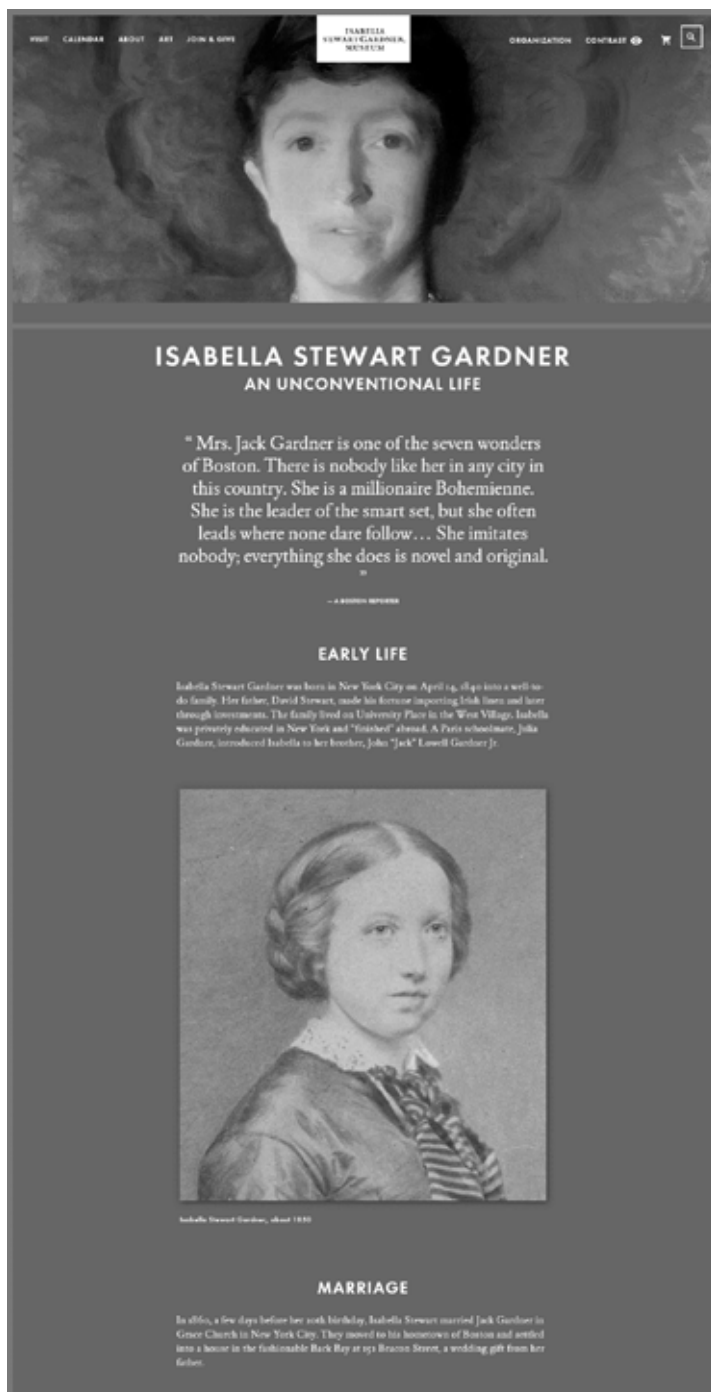


Figure 5. Upper part of the 'about' page on Isabella Stewart Gardner

- (4) *"She lives at a rate and intensity, with a reality that makes other lives seem pale, thin and shadowy."* – Bernard Berenson
- (5) *"You said to me... that if ever you inherited any money that it was yours to dispose of, you would have a house... filled with beautiful pictures and objects of art, for people to come and enjoy. And you have carried out the dream of your youth."* – Isabella's friend, Ida Agassiz Higginson, 1923
- (6) When Renzo Piano did his first walk-through of the Gardner Museum, he [...] said, *"This lady was mad. I have to quit this job. No one can do it."*

The above characterisation of Isabella is achieved by means of other-presentation, typically occurring "when a character or person provides explicit information about someone else" (Culpeper 2001: 167). When a number of different characters characterise the target character in the same way the validity of the characterisation itself is reinforced and the aim of high consensus is achieved.

The last quote (6), by Italian architect Renzo Piano, who was hired in 2005 to design a new wing for the museum, makes an exception, as it comes from a contemporary character who never met Isabella in person. However, Piano seems to have perfectly grasped the temper of Mrs. Gardner, by defining her as 'mad' due to her unrealistic demands: according to Gardner's will, no one would ever have been allowed to alter the museum's basic premise after her death. Even though Piano accepted the job in the end and the new wing opened in 2012, he had realised from the beginning that Mrs. Gardner, even after her death, would be a very difficult client to please. The characterisation of Isabella is therefore well-rounded and does not exclude negative aspects, such as, for instance, that she could be very stubborn and spoiled.

The insertion of Renzo Piano as a secondary character in the main narrative level devoted to Isabella also allows the introduction of a story within a story, dealt with in another page of the website: that of the 'personal museum' Mrs. Gardner built in Boston to house her collection, which was modelled after the 15th century Palazzo Barbaro in Venice. Similarly, a secondary level of narrative has been inserted in the website, regarding the theft which occurred in 1990, when thirteen works of art, including some paintings by Vermeer and Rembrandt, were stolen from the museum. Guards admitted two men posing as police officers responding to a disturbance call; the thieves then tied the guards up and looted the museum over the next

hour. The case is still unsolved, no arrests have been made and no works have been recovered. The empty frames of the stolen paintings still hang on the wall of the museum as a memento. This infamous event in the history of the museum, which would best fit in crisis communication manuals, is defined as “the single largest property theft in the world” in the website and told in a captivating way, as if it was crime fiction. It is not surprising that Netflix has developed a docu-fiction on this story.^{5, 6}

3.3 Narrative in museum social media

As pointed out by De Fina (2016), the transposition of narrative practices to social media implies that stories become subject to “recontextualisations, contestations, and all sorts of manipulations that would not have been possible in offline environments” (2016: 493). Thus, when we move on to the observation of museum social media, despite the limitations in terms of textual length imposed by the genre (Carter et al. 2013), we do not have to expect the demise of narrative, but rather its reconfiguration and reinterpretation according to the characteristic features of these highly interactive and multimodal platforms. Indeed, stories conveyed in museum social media can still comply with the criteria of a canonical narrative, yet, attention is not placed only on their content, but also on their production and circulation: on the way they are “shared, recontextualised, commented upon” (De Fina 2016: 477). This reinforces the hypothesis that museum communicators have become increasingly aware of the potential of narrative as a discursive strategy. Having experienced the effectiveness of narrative in other museum genres (e.g. in press releases and websites), they were highly motivated to extend its use to different contexts.

Typically, in the case of museum social media, we come across forms of minimal narrative, containing at least two clauses told in the same order as that of the original events (Labov 1972: 360; Labov 2006: 38). An example is provided by Figure 6, reproducing a Facebook post of the Victoria & Albert Museum, where an object in the museum’s collection – a medieval beaker – is not presented through a description, but by telling ‘a romantic story’ condensed into two propositions.⁷ As seen in previous examples,

⁵ <https://www.gardnermuseum.org/about/theft>, accessed November 2021.

⁶ <https://www.netflix.com/title/81032570>, accessed November 2021.

⁷ <https://www.facebook.com/victoriaandalbertmuseum/photos/the-romantic-story-of-this-medieval-beaker-ca-1350-says-that-the-fairies-left-it/10153431981188880/>, accessed November 2021.

the narrative dimension is also here announced *a priori* and made explicit by means of metadiscourse (“story”), while the evaluative lexical choice confirms a preference for markers of AFFECT (“romantic”).

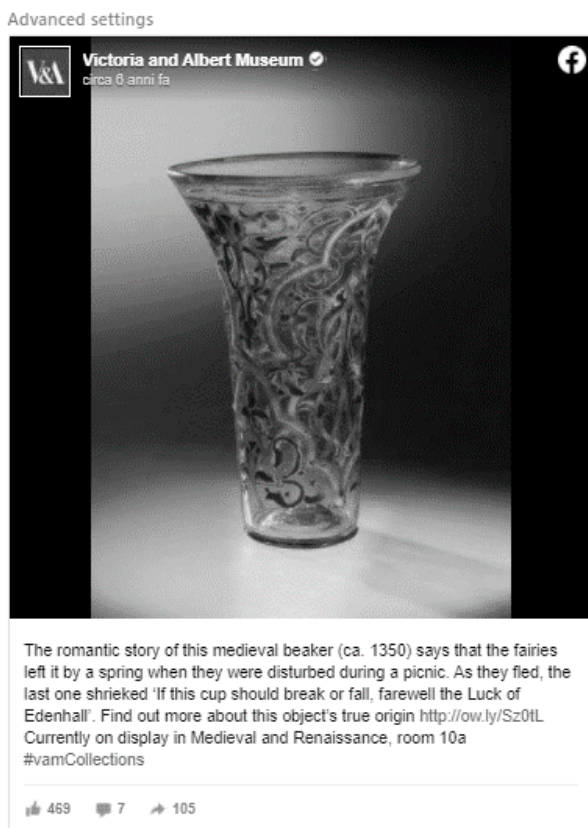


Figure 6. Micro-narrative in museum social media

Even within the limited space of a Facebook post, the basic components of narrative can be identified in the text: settings – an imaginary Eden –, characters – the fairies –, and an essential plot (the fairies were having a picnic, then were disturbed by something, so they had to leave the cup by a spring and flee, but they cast a spell on the object). The final call to action (“Find out more about this object’s true origin”) invites users to discover the story in its entirety, as told in the linked webpage, and to fill in the details left unsaid in the post.

Another aspect of the recontextualisation of narrative in museum social media relies on the use of emoji, a form of non-standard language based on pictograms (Danesi 2017: 2) and increasingly used in digitally

mediated communication “to add visual annotations to the conceptual content of a message” (Danesi 2016: 10). While at the scholarly level debate on the semiotics of emoji, alternatively interpreted as an independent language (Ge – Herring 2018), or as a paralinguistic modality (Gawne – McCulloch 2019; Logi – Zappavigna 2021, Zappavigna – Logi 2021), is in full bloom, at the level of professional communication, museums have been among the first to acknowledge the ground breaking contribution of emoji to digital communication. Notably, in 2016, the Museum of Modern Art (MoMA) in New York acquired the original set of 176 emoji characters designed by Shigetaka Kurita and contextually organised an exhibition celebrating the new acquisition (Lee 2018). In justifying their choice, the MoMA stressed that emoji were “reasserting the human within the deeply impersonal, abstract space of electronic communication”.⁸

In the context of museum social media, emoji are often used in combination with language to create innovative pictorial representations and generate new forms of storytelling, which range from “improvisational approaches to narrative” (Sergeant 2019: 167), drawing on the conventions of comics, e.g. the use of bubbles to express the interior concerns of characters, to straightforward emoji stories, whose interpretation is co-constructed and shared among members of a close community (Kelly – Watts 2015).

The following example (Figure 7) can be ascribed to the first category of minimal narratives, where emoji provide users with a visual hint for interpreting the story. Introducing an artwork by Henri Matisse in the Tate collection, the Facebook post tells a story about the author, who, in his elderly age, was confined to bed and couldn’t paint anymore.⁹ Matisse had therefore to switch to a less tiring art technique, that of collage.

As it is often the case with abstract art, the subject of the collage shown in the post would be hard to recognise at first sight, if it weren’t for the caption and a small emoji representing a snail placed at the end of the post. Emoji are therefore used here as a fundamental complement to the narrative: it allows users to identify a relevant character in the story, the snail, while, at the same time, it suggests the emotional state of the artist, who perhaps, during his illness, likened his condition to that of a snail. The evaluative dimension of AFFECT, revealing the author’s stance and leading

⁸ <https://stories.moma.org/the-original-emoji-set-has-been-added-to-the-museum-of-modern-arts-collection-c6060e141f61>, accessed November 2021.

⁹ <https://www.facebook.com/tategallery/photos/a.117432718992/10159481646913993>, accessed November 2021.

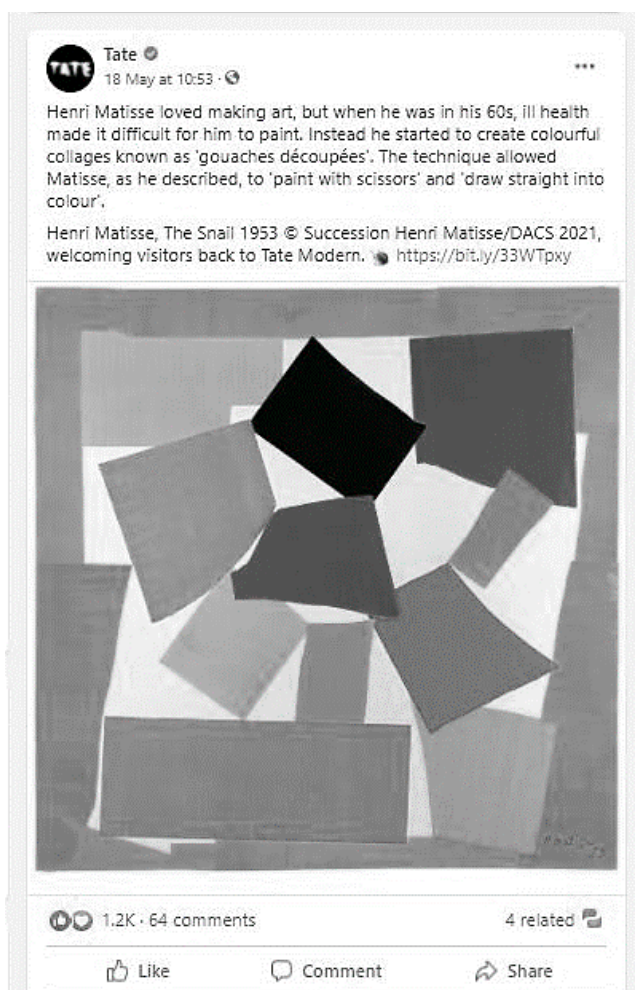


Figure 7. Micro-narrative combined with emoji

users toward empathy, is therefore still present in the narrative, yet, it is not fulfilled by lexical items, but by emoji, effectively replacing the use of emotive standard language.

A second, less frequent category of narrative which can be identified across museum social media is that of micro-stories told in emoji, where pictograms are used to replace single lexical items or even entire strings of meaning. This leads to a most peculiar communicative strategy, that of “emojification” (Seargeant 2019: 163), generally signalled in the Web by the hashtag #emojireads and especially applied in MC to involve users in games or puzzles, as shown in the example below (Figure 8).



Figure 8. Micro-narrative and 'rebus use' of emoji

In this Tweet published by the J. Paul Getty museum during the 2020's lockdown, users were invited to guess eight different ancient Greek myths composed directly in emoji.¹⁰ As suggested by Seargeant (2019), the interpretation of 'emojification' like these heavily relies on our pre-knowledge and skills: "we try to decode the story based on our understanding of how narratives work" (2019: 163), expecting emoji to represent the fundamental elements of a story (characters, events, settings) and guessing their sequence. However, given the complexity of this example, full knowledge of the relevant matter, and, in this specific case, of the kind of art, themes and stories on display at the Getty, seems to be an indispensable prerequisite for understanding. One has to be familiar with the museum environment, in the first place, and be part of its community of art lovers, sharing similar interests and knowledge, to decipher 'emojification' as the one shown above. This observation seems to be in line with what Kelly – Watts (2015) suggested when highlighting the narrative potential of emoji in creating "shared and secret uniqueness" in personal interactions.

¹⁰ <https://twitter.com/gettymuseum/status/1295776721740865537>, accessed November 2021.

4. Conclusions

The aim of this study was to illustrate the fundamental role played by storytelling in MC, an area that has received little attention from narrative scholars. While the use of storytelling techniques is fully recognised in ordinary museum activities, such as exhibition organisation, guided tours, education and public outreach programmes, its applications in MC deserve further investigation. Particular attention has therefore been paid to the discursive and communicative features characterising storytelling across textual genres on which MC typically relies (press releases, websites, social media networks), taking into consideration the language, structure and narrative strategies in use. A diachronic approach has been adopted, taking into consideration a period covering seventy years, from 1950 to the present.

The analysis has shown how narrative forms have been part of MC almost from the beginning of this professional practice, which has been embraced by museums in a structured way only after the Second World War. In particular, it was not until the 1970s that stories began to be used more consciously by communication professionals, with the increasing insertion of narrative sections within press releases. From the 1990s onwards, storytelling has been extended to digital media, undergoing a natural process of gradual transformation that has neither threatened its survival nor altered its essential features. The diachronic analysis suggests that museum communicators have become gradually aware of the effectiveness of narrative as a communication strategy. Its success seems to be rooted in the textual genre of press releases, which increasingly made use of stories for the purpose of newsworthiness (van Dijk 1988, Bell 1991). Once writers had experienced the positive effects of narrative on a restricted professional community – journalists – they extended it to other genres, aimed at a wider audience. Since then, storytelling has been used in MC as a public engagement tool and has been adapted to an increasingly innovative and complex multimodal environment.

The affordances that typically characterise new media provided an opportunity to develop narrative strategies well beyond the space of a single textual section, as in the case of press releases. Websites, in particular, offer the possibility of disseminating hypertexts with narrative cues, and inserting stories into the story, as in the case of the Gardner Museum website, where secondary stories are intertwined with the main level of the narrative centred on the museum's founder. By contrast, the space constraints imposed by social media platforms have led to a contraction of the narrative structure,

reduced to micro-narratives. Yet, museum communicators have not been constrained by the challenges that new media entail, performing a re-contextualisation of former narrative practices (De Fina 2016) and relying on new media affordances, such as multimodality, hyper-textuality, and non-standard language (emoji). This evolution over time and across different communicative genres bears witness to a creative and dynamic appropriation of narrative practices by museums, anticipating further developments, some of which are probably already underway.

A salient aspect of this diachronic study is that the typical features of narrative, i.e. i) its canonical structure, consisting of characters, events and settings (Toolan 2001), ii) its evaluative language (Hunston – Thompson 2000), and, in particular, iii) its emotive language (Ungerer 1997) appear consistently across traditional and digital genres. Already from the first experiments with narrative sections shown in press releases, the use of evaluative language has been strategic. In particular, the semantic dimension of AFFECT (Martin – White 2005), revealing the subjective stance of the author and activating empathy processes in the reader, seems to play a key role in engagement. Interestingly, in the case of social media posts, where the available textual space is necessarily reduced, the emotive function, previously performed by standard language, can be assumed by alternative semiotic resources, i.e. emoji, which allow for a visual representation of feelings and moods. The study suggests, therefore, that emoji might gradually replace evaluative language in museum storytelling, at least as far as social media networks are concerned. It is not excluded, moreover, that emoji will come out of the spaces to which they have been relegated up to now to reach other, less obvious, museum genres.

In consideration of the developments outlined above, the field of MC appears as a rich research area for discourse and communication studies. Museums play a valuable role as incubators of innovation and creativity in communication practices and, as such, can inspire professional communicators and researchers from different areas and disciplines.

REFERENCES

Primary sources

- Corpus of museum digital media (2015-2021)*. Collected by C. Lazzeretti.
EPA Diacorpus (Exhibition Press Announcements Diacorpus, 1950-2016). Collected by C. Lazzeretti.

Secondary sources

- Baker, P.
 2011 "Times may change, but we will always have money: Diachronic variation in recent British English", *Journal of English Linguistics* 39 (1), 65-88.
- Baumbach, S.
 2015 *Literature and Fascination*. London: Palgrave Macmillan
- Bedford, L.
 2001 "Storytelling: The real work of museums", *Curator: The Museum Journal* 44 (1), 27-34.
- Bell, A.
 1991 *The Language of News Media*. Oxford: Blackwell. Biber, D. et al.
 1999 *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education Limited.
- Bourlakos, I. et al.
 2017 "Formalisation and visualisation of the narrative for museum guides". In: J. Szymański – Y. Velegrakis (eds.) *Semantic Keyword-Based Search on Structured Data Sources*. Cham: Springer, 3-13.
https://doi.org/10.1007/978-3-319-74497-1_1.
- Burdelski, M. et al.
 2014 "Storytelling in guided tours: Practices, engagement, and identity at a Japanese American museum", *Narrative Inquiry* 24 (2), 328-346.
- Carter, S. et al.
 2013 "Microblog language identification: Overcoming the limitations of short, unedited and idiomatic text", *Language Resources and Evaluation*, 47 (1), 195-215.
- Culpeper, J.
 2001 *Language and Characterisation: People in Plays and Other Texts*. New York: Routledge.
- Danesi, M.
 2016 *The Semiotics of Emoji: The Rise of Visual Language in the Age of the Internet*. London: Bloomsbury Publishing.
 2017 "Emoji in advertising", *International Journal of Semiotics and Visual Rhetoric* 1 (2), 1-12.
- De Fina, A.
 2016 "Storytelling and audience reactions in social media", *Language in Society* 45 (4), 473-498.
- Di Blasio, M. – R. Di Blasio
 1983 "Constructing a cultural context through museum storytelling", *Roundtable Reports* 8 (3), 7-9.
- Dijk, T. A. van
 1988 *News as Discourse*. Hillsdale, NJ: Lawrence Erlbaum Associates.

- Gawne, L. – G. McCulloch
2019 “Emoji as digital gestures”, *Language@Internet* 17, article 2.
- Ge, J. – S. C. Herring
2018 “Communicative functions of emoji sequences on Sina Weibo”, *First Monday* 23 (11). <https://doi.org/10.5210/fm.v23i11.9413>.
- Geismar, H.
2018 *Museum Object Lessons for the Digital Age*. London: UCL Press.
- Georgakopoulou, A.
2015 “Small stories research”. In: A. De Fina – A. Georgakopoulou (eds.) *The Handbook of Narrative Analysis*. Chichester: John Wiley & Sons, 255-272.
- Giddens, A.
1991 *Modernity and Self-Identity: Self and Society in the Late Modern Age*. Stanford: Stanford University Press.
- Glaser, B. G. – A. L. Strauss
2006 *The Discovery of Grounded Theory: Strategies for Qualitative Research*. New York: Aldine.
- Glover Frykman, S.
2009 “Stories to tell? Narrative tools in museum education texts”, *Educational Research* 51 (3), 299-319.
- Hofinger, A. – E. Ventola
2004 “Multimodality in operation: Language and picture in a museum”. In: E. Ventola et al. (eds.) *Perspectives on Multimodality*. Amsterdam / Philadelphia: John Benjamins, 193-209.
- Hughes, P.
2021 *Storytelling Exhibitions: Identity, Truth and Wonder*. London: Bloomsbury.
- Hunston, S. – G. Thompson (eds)
2000 *Evaluation in Text: Authorial Stance and the Construction of Discourse*. Oxford: Oxford University Press.
- Hyland, K.
2004 *Disciplinary Discourses: Social Interactions in Academic Writing*. London: Longman.
2005 *Metadiscourse. Exploring Interaction in Writing*. London: Continuum.
- Jacobs, G.
1999 *Preformulating the News: An Analysis of the Metapragmatics of Press Releases*. Amsterdam / Philadelphia: John Benjamins.
- Kelly, R. – L. Watts
2015 “Characterising the inventive appropriation of emoji as relationally meaningful in mediated close personal relationships”. Paper presented at *Experiences of Technology Appropriation: Unanticipated Users, Usage, Circumstances, and Design*, Oslo, Norway, 20th September 2015. https://purehost.bath.ac.uk/ws/portalfiles/portal/130966701/emoji_relational_value.pdf, accessed December 2023.

- Kramper, A.
2017 *Storytelling für Museen. Herausforderungen und Chancen*. Bielefeld: transcript Verlag.
- Labov, W.
1972 "The transformation of experience in narrative syntax", *Language in the Inner City*, 354-396.
2006 "Narrative pre-construction", *Narrative Inquiry* 16 (1), 37-45.
- Lazzeretti, C.
2016 *The Language of Museum Communication: A Diachronic Perspective*. London: Palgrave Macmillan.
- Lee, S.
2018 "Emoji at MoMA: Considering the 'original emoji' as art", *First Monday* 23 (9). <https://doi.org/10.5210/fm.v23i9.9401>.
- Logi, L. – M. Zappavigna
2021 "A social semiotic perspective on emoji: How emoji and language interact to make meaning in digital messages", *New Media & Society* 25 (12), 3222-3246. <https://doi.org/10.1177/14614448211032965>.
- Martin, J. R.
2001 "Beyond exchange: Appraisal Systems in English". In: S. Hunston – G. Thompson (eds.) *Evaluation in Text*. Oxford: Oxford University Press, 141-175.
- Martin, J. R. – P. R. R. White
2005 *The Language of Evaluation: Appraisal in English*. Basingstoke/New York: Palgrave Macmillan.
- McLean, F.
2012 *Marketing the Museum*. New York: Routledge.
- Nielsen, J. K.
2017 "Museum communication and storytelling: Articulating understandings within the museum structure", *Museum Management and Curatorship* 32 (5), 440-455.
- Pierroux, P. – S. Skjulstad
2011 "Composing a public image online: Art museums and narratives of architecture in web mediation", *Computers and Composition* 28 (3), 205-214.
- Pireddu, S.
2018 "Storytelling in museums: Construing language, heritage and places in the aftermath of history". Paper presented at *Aftermath: the Fall and Rise after the Event*, Krakow, Poland, 25-26 October 2018. <https://zenodo.org/records/3774567>, accessed December 2023.
- Ravelli, L. J.
1996 "Making language accessible: Successful text writing for museum visitors", *Linguistics and Education* 8 (4), 367-387.
2006 "Genre and the museum exhibition", *Linguistics & the Human Sciences* 2 (2), 299-317.

- 2007 *Museum Texts: Communication Frameworks*. New York: Routledge.
- Rizzo, I. – A. Mignosa (eds.)
- 2013 *Handbook on the Economics of Cultural Heritage*. Cheltenham, UK: Edward Elgar Publishing.
- Seargeant, P.
- 2019 *The Emoji Revolution: How Technology is Shaping the Future of Communication*. Cambridge: Cambridge University Press
- Scott, M.
- 2020 *WordSmith Tools version 8*. Stroud: Lexical Analysis Software.
- Tagg, C.
- 2015 *Exploring Digital Communication: Language in Action*. New York: Routledge.
- Toolan, M.
- 2001 *Narrative: A Critical Linguistic Introduction*. New York: Routledge.
- Ungerer, F.
- 1997 "Emotions and emotional language in English and German news stories". In: S. Niemeier – R. Dirven (eds.) *The Language of Emotions: Conceptualization, Expression, and Theoretical Foundation*. Amsterdam: John Benjamins, 307-328.
- Zappavigna, M. – L. Logi
- 2021 "Emoji in social media discourse about working from home", *Discourse, Context & Media* 44, 100543.
<https://doi.org/10.1016/j.dcm.2021.100543>.

Address: CECILIA LAZZERETTI, Faculty of Education, Free University of Bozen-Bolzano, Regensburger Allee 16 – viale Ratisbona 16, 39042 Brixen-Bressanone, Italy.
 ORCID code: <http://orcid.org/0000-0002-0415-9129>.

“It is a long road from sorrow to joy”: Metaphors of *happiness* and *sadness* in Late Modern English private correspondence

Gloria Mambelli

University of Verona

ABSTRACT

This paper investigates the metaphorical conceptualisation of the two basic emotions of *happiness* and *sadness* in the private correspondence of eighteenth- and nineteenth-century men and women of letters. Grounded in the framework of Conceptual Metaphor Theory, it exploits metaphorical pattern analysis (MPA) and the metaphor identification procedure (MIP) in a corpus-based analysis of Late Modern English letters. The aim is to identify metaphorical expressions in the corpus and illustrate the conceptual mappings involved. First, a set of lexical items belonging to the two target domains under investigation was retrieved and subsequently used for concordance analysis. Secondly, a sample of metaphorical expressions was examined and mapped according to the metaphorical mappings described in the literature. Additional mappings are proposed for the conceptualisations identified in the analysed sample. Findings confirm the effectiveness of MPA in the identification of metaphors in corpus-based investigations and offer insights into the use of metaphor in Late Modern English educated, non-literary discourse.

Keywords: metaphor, conceptual metaphor theory, metaphor mapping, emotion, corpus-based approach, Late Modern English, private correspondence.

1. Introduction

Metaphor has attracted considerable attention in cognitive linguistic scholarship since the publication of Lakoff and Johnson's *Metaphors We Live By* (1980), a ground-breaking contribution to the study of metaphor.

The renowned monograph illustrates how, contrary to most people's beliefs, metaphor is not a device confined to poetry and rhetoric but a phenomenon pervading both our everyday language and thoughts. The authors dismiss the classical view of metaphor as a mere rhetorical device and state that "[o]ur ordinary conceptual system, in terms of which we both think and act, is fundamentally metaphorical in nature" (Lakoff – Johnson 1980: 3). By treating metaphor as a matter of concepts and language rather than language only, the aforementioned volume paved the way for Conceptual Metaphor Theory (CMT), which played a significant role in the development of cognitive linguistics and is now a well-established framework in metaphor studies.

The theory of metaphor presented by Lakoff and Johnson is based on the assumption that "the essence of metaphor is understanding and experiencing one kind of thing in terms of another" (Lakoff – Johnson 1980: 5); in other words, metaphor involves the establishment of correspondences between two conceptual domains. These types of correspondence are usually called "mappings" in the literature: metaphors are the result of conceptual mappings from a source domain onto a target domain. For instance, the linguistic expression "your claims are *indefensible*" conceptualises ARGUMENT as WAR in that an argument is expressed by means of words belonging to the domain of war, implying a correspondence between the two domains. Other examples of mappings are TIME IS MONEY, IDEAS (OR MEANINGS) ARE OBJECTS, HAPPY IS UP, SAD IS DOWN.

New approaches to the study of metaphor have arisen over the years to complement and supersede the traditional introspective method of metaphor identification, "based on more-or-less explicit commonsensical intuitions of the part of the researcher" (Stefanowitsch 2006a: 10): procedures for the identification of metaphors based on empirical methods such as MIP and MIPVU have been developed (see Pragglejazz Group 2007; Steen et al. 2010) and corpus-based studies of metaphor are flourishing at the present time (cf. Deignan 2005; Semino 2017; Stefanowitsch – Gries 2006). The majority of metaphor investigations are based on Present-Day English (PDE) data; however, the increasing availability of electronic corpora and digitised editions of historical texts and dictionaries has allowed historical linguists to contribute to the discussion by means of a diachronic perspective (Allan 2009, Gevaert 2005, Hintikka 2007, Patterson 2018, Tissari 2006, among others).

This paper employs corpus methods to examine emotion metaphors in Late Modern English correspondence written by eighteenth- and nineteenth-century men and women of letters. Specifically, it presents a case study focusing on the conceptualisations of *happiness* and *sadness*.

Conceptual metaphors of emotion are a subclass of metaphors involving a correspondence between an emotion (the target domain) and a physical source domain. The pioneer of research on this type of metaphor is undoubtedly Kövecses, who carried out extensive studies of emotion concepts in PDE from a cognitive linguistic perspective (1986, 1990, 2000). His most comprehensive work on the subject is *Metaphor and Emotion*, in which he effectively illustrates "the role of figurative language in the conceptualization of emotion" (2000: 1). The study of emotion metaphor focuses on target domains related to the sphere of emotions, as illustrated by examples such as *boiling with anger* or *to be governed by anger*, in which ANGER is the target domain and the source domains are A HOT FLUID IN A CONTAINER and A SOCIAL SUPERIOR respectively. MetaNet,¹ a publicly accessible repository of conceptual metaphors relating to a wide range of domains and topics, offers a list of mappings including those that are most commonly involved in the metaphorical description of emotions, e.g. EMOTION IS A FORCE, EMOTION IS PRESSURE IN A CONTAINER, EMOTIONAL INTENSITY IS TEMPERATURE, EMOTIONS ARE OBJECTS, EMOTIONS ARE SUBSTANCES. Among the various types of emotion concepts, ANGER is one of the most extensively investigated from a cognitive perspective (Glynn 2002; Kövecses 1986, 1995a, 1995b; Lakoff – Kövecses 1987), alongside LOVE (Glynn 2002; Kövecses 1986, 1988; Tissari 2001) and HAPPINESS (Kövecses 1991; Stefanowitsch 2004), all of which represent basic emotions. It has been observed that "[s]peakers of a given language appear to feel that some of the emotion words are more basic than others" (Kövecses 2000: 3) in that they perceive them as more 'prototypical' (in the sense proposed by Rosch [1975, 1978]) within a taxonomical framework (Fehr – Russell 1984; Shaver et al. 1987). Ekman (1999) identified a set of basic emotions found in all cultures, which includes the two opposite concepts of *happiness* and *sadness*.

1.1 Metaphors of *happiness* and *sadness*

Like other basic emotions, *happiness* and *sadness* have been the object of cognitive linguistic studies investigating metaphorical expressions in English and other languages. As a result, a considerable number of metaphorical mappings for these concepts are available in the literature.

¹ MetaNet is a repository of conceptual metaphors provided by the MetaNet Project. The list of conceptual mappings is available here: <https://metaphor.icsi.berkeley.edu/pub/en/index.php/Category:Metaphor>, accessed June 2022.

As far as metaphors involving the target domain *HAPPINESS* are concerned, Kövecses (2000: 24-25) provides the following list of mappings based on expressions occurring in the English language, drawn from Kövecses (1991) and Lakoff and Johnson (1980):

HAPPINESS/BEING HAPPY IS

UP
 BEING OFF THE GROUND
 BEING IN HEAVEN
 LIGHT
 VITALITY
 WARM
 HEALTH
 A PLEASURABLE PHYSICAL SENSATION
 A FLUID IN A CONTAINER
 A CAPTIVE ANIMAL
 AN OPPONENT IN A STRUGGLE
 A RAPTURE/HIGH
 INSANITY
 A NATURAL FORCE

A HAPPY PERSON IS

AN ANIMAL THAT LIVES WELL

Additional patterns were identified by Stefanowitsch (2006b: 84), whose corpus-based investigations of metaphor proved effective in the retrieval of metaphorical expressions and the formulation of less frequent mappings:

HAPPINESS/BEING HAPPY IS

HEAT/FIRE
 A LIQUID
 A SUBSTANCE IN A CONTAINER (UNDER PRESSURE)
 A MIXED/PURE SUBSTANCE
 A DESTROYABLE OBJECT
 DISEASE
 AGGRESSIVE ANIMAL BEHAVIOUR
 AN ORGANISM
 BLOOD
 A SHARP OBJECT

INTENSITY OF HAPPINESS IS

DEPTH

According to Kövecses (2000: 26), metaphors for *sadness* involve some general source domains which are shared among emotion metaphors alongside some more specific ones. The latter "mostly have to do with negative evaluation of the concept of sadness and, as such, form the opposites of several of the source domains for happiness" (2000: 25-26). The following list of metaphorical mappings, provided by Kövecses (2000: 25-26), was drawn from Barcelona (1986) and partially modified by the author himself:

SADNESS/BEING SAD IS
 DOWN
 DARK
 A LACK OF HEAT
 A LACK OF VITALITY
 A FLUID IN A CONTAINER
 A PHYSICAL FORCE
 A NATURAL FORCE
 AN ILLNESS
 INSANITY
 A BURDEN
 A LIVING ORGANISM
 A CAPTIVE ANIMAL
 AN OPPONENT
 A SOCIAL SUPERIOR

Additionally, Stefanowitsch (2006b: 88) identified the following metaphorical patterns:

SADNESS/BEING SAD IS
 A MIXED/PURE SUBSTANCE
 DEPTH
 A SUBSTANCE IN A CONTAINER (UNDER PRESSURE)
 A LIQUID
 AN AURA
 A SOUND
 A WEATHER PHENOMENON
 TASTE
 HEAT

I chose the mappings proposed by Kövecses (2000) and Stefanowitsch (2006b) as representative of the results obtained in the investigation of

these emotion metaphors due to the different methods involved in their identification. On the one hand, Kövecses's studies are representative of the traditional introspective method; on the other hand, Stefanowitsch's innovative studies involving corpus-based methods can be considered as representative of corpus-based approaches to metaphor.

The metaphorical patterns presented above are all based on expressions occurring in PDE. By investigating the ways in which eighteenth- and nineteenth-century poets and writers conceptualised the two basic emotions of *happiness* and *sadness*, I intend to ascertain whether the metaphorical patterns described in the literature are mirrored in Late Modern English private writing and provide new insights into the conceptualisation of emotions that may possibly prove beneficial for diachronic comparison.

1.2 Late Modern English writing

The language used in Britain during the eighteenth and nineteenth centuries is generally referred to as Late Modern English. Unlike the preceding stages of the development of the English language, whose start is conventionally associated with significant external factors such as the Norman Conquest in 1066 (marking the beginning of the Middle English period) and the introduction of the printing press in 1476 (marking the beginning of the Early Modern English period), the boundaries of Late Modern English are undoubtedly more blurred. Nevertheless, it is conventionally said to have spanned between 1700 and 1900 (Mugglestone 2006, Tieken-Boon van Ostade 2009).

This period of the history of English was chosen as the object of this study in view of "the increasing numbers of private letters that were produced" at the time (Tieken-Boon van Ostade 2009: 1), making it a suitable historical framework for a corpus-based investigation of private correspondence. Like the preceding ones, it can be considered from both a diachronic and synchronic perspective: the latter is adopted in this study, since the focus is on the use of metaphor on the part of a specific community of practice rather than on its development in time. Moreover, as observed by Patterson (2018: 60), texts dating from before 1900 are not subject to copyright and can be freely accessed, a matter that is worth considering when creating a corpus.

The focus on Late Modern English educated writing rather than on the 'purely non-literary language' of ordinary people aims to find a meeting point between metaphor viewed as "a device of the poetic imagination and the rhetorical flourish" and metaphor as "a matter of [...] ordinary language"

(Lakoff – Johnson 1980: 3), i.e. the metaphorical conceptualisations observable in the language of everyday life. The purpose of this study is to emphasise the continuity between the two dimensions by providing insights into the ways in which men and women of letters, who were used to exploiting metaphor in poetic discourse, resorted to conceptual metaphors to communicate basic emotions in ordinary situations. Starting from the assumption that "details about the distribution, function, and effect of metaphor in literature versus outside literature need to be collected and examined by means of corpus-linguistic [...] studies" (Semino – Steen 2008: 233), a corpus of letters was used to retrieve a set of metaphorical expressions involving the target domains HAPPINESS and SADNESS and to map them according to the patterns described in the literature. A target-domain oriented approach was chosen in view of the advantages it offers in the identification of metaphors on the basis of corpus data. As observed by Stefanowitsch (2006b: 66), "we can retrieve a large number of instances of a target domain item [...] from a corpus and exhaustively identify the metaphorical patterns that it occurs with."

2. Data and methodology

This research is based on a non-annotated corpus consisting of a selection of letters drawn from the volume entitled *Selected English Letters (XV-XIX Centuries)*, available via *Project Gutenberg*. The anthology, edited by M. Duckitt and H. W. Elgee, was published in 1913 and contains a selection of letters written by British intellectuals between the fifteenth and nineteenth centuries, grouped per author in chronological order. In line with the purpose of this study, only the authors whose letters were written during the eighteenth and nineteenth centuries were selected: as a result, the corpus consists of a total of 169 letters dating from between 1710 and 1852. These were written in varying numbers by 34 authors, namely writers, including Jonathan Swift and Walter Scott, and poets, including Thomas Gray and John Keats (see Appendix 1). The assembled corpus has a total of 109,600 words and 128,263 tokens.

By combining semi-automated corpus methods with qualitative analysis, examples of metaphorical expressions relating to the concepts of HAPPINESS and SADNESS were examined.

The procedure involved the following steps:

1. Retrieval of lexical items associated with the two emotion concepts;
2. Manual identification of metaphorical expressions;

3. Mapping of a sample of metaphorical expressions by recurring to mappings described in the literature or, if not available, design of new mappings.

The method used for the identification of metaphors in the corpus is metaphorical pattern analysis (MPA), a procedure developed by Stefanowitsch (2006b) to overcome the limits of the traditional introspective method in terms of data coverage. A metaphorical pattern is “a multi-word expression from a given source domain (SD) into which one or more specific lexical item from a given target domain (TD) have been inserted” and represents a fitting object of study for this research in view of the fact that MPA provides an ideal “basis for target-domain oriented studies on the basis of corpus data” (2006b: 66). Since emotion concepts manifest through groups of lexical items, the investigation of the words belonging to a target domain appears to be a productive procedure in corpus-based studies. A similar approach was adopted, for instance, by Tissari (2006), who made use of the term *shame* as a keyword for searches in diachronic corpora of English in order to understand how shame was conceptualised through time.

MPA allows us to identify potential metaphors by retrieving the occurrences of one or more words representing a target domain from a corpus. It consists in the following procedure:

we choose a lexical item referring to the target domain under investigation and extract (a sample of) its occurrences in the corpus. In this sample, we then identify all metaphorical expressions that the search word is a part of and group them into coherent groups representing general mappings. (Stefanowitsch 2006b: 64)

Although it can only be used to identify metaphorical expressions containing target domain items, MPA proves effective to map out metaphors for a specific target domain. Moreover, the advantage of dealing with metaphorical patterns lies in the fact that the target domain is made explicit through the target domain lexis, whereas uncertainties may arise when tackling metaphorical expressions that do not contain lexical items associated with the target domain in question. For instance, it could be argued that the metaphorical expression “He fled from her advances”, for which the target domain of LOVE was identified by Lakoff and Johnson (1980: 49), may involve different target domains such as LUST or DESIRE (Stefanowitsch 2006b: 66-67).

The adoption of MPA entails the identification of lexical items representing the target domain under investigation. In the case of HAPPINESS

and SADNESS, limiting the search to the words *happiness* and *sadness* in the corpus yielded a limited number of results. Therefore, since "each of these emotions has a set of semantically similar lexical items associated with it" (Stefanowitsch 2006b: 71), I decided to use a set of near synonyms as target domain items to verify if they participate in the same metaphorical mappings. In this respect, Stefanowitsch's findings suggest that near synonyms are generally associated with the same metaphors, since his investigation of the words *joy* and *happiness* produced similar results.

In order to search for near synonyms of *happiness* and *sadness* in the corpus, it was necessary to retrieve a reliable list of representative keywords. To this end, Wmatrix, a software tool for corpus analysis and comparison developed at Lancaster University (Rayson 2009), was used for the extraction of semantic domains and subsequent identification of keywords. By means of the UCREL² Semantic Analysis System (USAS) and Constituent Likelihood Automatic Word-tagging System (CLAWS) corpus annotation tools, the software offers the possibility to retrieve key semantic domains, as well as frequency lists and concordances. As explained by Koller et al., "semantic domain tagging can be used to identify instances of metaphorical expressions in a lengthy text or corpus with a greater recall than other methods that have previously been used", thus making it "possible to expand on the results of earlier analyses of metaphor in discourse" (2008: 141-142). Although Wmatrix is mostly designed to carry out quantitative analyses, its function of allocating semantic domain tags proved crucial in the identification of representative target domains items in the present investigation. Thanks to the USAS semantic annotations, Wmatrix is able to compare the frequency of these tags to subsets of the British National Corpus (BNC), the latter functioning as an external reference corpus to determine keyness based on statistical significance. Even though the BNC covers the late 20th century and is therefore not comparable to the corpus under investigation, the subset selected for the comparison, namely *BNC Sampler Written*,³ proved beneficial in the identification of keywords for target domains. The software retrieved a series of USAS semantic domains included in the key concept EMOTION, among which the following ones were selected and considered as corresponding to the target domains under investigation:

² UCREL is the acronym for the University Centre for Computer Corpus Research on Language based at Lancaster University. <https://ucrel.lancs.ac.uk/>, accessed June 2022.

³ See <https://ucrel.lancs.ac.uk/bnc2sampler/sampler.htm>, accessed December 2023.

- i. E4.1 Happy
- ii. E4.1 Sad
- iii. E4.2 Content
- iv. E4.2 Discontent

The selected semantic domains provided a set of lexical items semantically associated with *happiness* and *sadness*. Given the high number of concordance lines to be analysed (over 200 in each of the two domains), I carried out a selection procedure in order to reduce the list of keywords provided by Wmatrix. First, I decided to focus on nouns and therefore discarded adjectives, adverbs, and verbs. Secondly, I removed all lexical items associated with concrete rather than abstract meanings (e.g. *smile, tears, howl, joke*) and those not in a relation of (near) synonymy with the emotions under investigation (e.g. *remorse, jealousy, frolic, frivolity, drollery, mourning, comedy, tragedy*). Finally, I manually extracted the words *anguish, enjoyment, dismay, distress*, which had not been identified by Wmatrix, and added them to the dataset. The selected keywords are listed in Table 1, which reports the lexical items in alphabetical order below the respective semantic domain.

Table 1. Selected keywords representing the target domains HAPPINESS and SADNESS in the corpus

HAPPINESS	SADNESS
Bliss	Anguish
Cheerfulness	Dejection
Delight(s)	Depression
Enjoyment	Despair
Exultation	Discontent
Fun	Dismay
Gaiety	Distress(es)
Happiness	Grief(s)
Jollity	Melancholy
Joy(s)	Misery(es)
Merriment	Pathos
Mirth	Regret(s)
Relief	Sadness
	Sorrow(s)
	Suffering(s)
	Unhappiness
	Woe

The selected words were used as keywords to perform searches in the corpus with the aid of the Sketch Engine (Kilgarrieff et al. 2004, 2014), a tool designed for both corpus building and analysis. Specifically, concordance lines were manually examined by using the Key Word In Context (KWIC) concordance format in order to identify metaphorical expressions. For this purpose, the metaphor identification procedure (MIP) developed by the Pragglejaz Group (2007) was applied. According to this procedure, a lexical unit is considered to be used metaphorically if “it has a more basic contemporary meaning in other contexts than the one in the given context”

and "the contextual meaning contrasts with the basic meaning but can be understood in comparison with it" (2007: 3). In order to determine what the most basic meaning of a lexical unit is, a dictionary of contemporary language should be used. In our case, since the language under investigation is Late Modern English, dictionaries of current English focusing on present-day meanings prove inadequate. Hence, the *Oxford English Dictionary* (OED), which tracks the semantic history of words since the earliest stages of the language, was consulted alongside two dictionaries that were published during the Late Modern period, namely Johnson's *Dictionary of the English Language* (1755) and Webster's *Complete Dictionary of the English Language* (1886).

The results obtained through the application of MPA and MIP led to the identification of metaphorical mappings specific to the emotion concepts under investigation. As illustrated in the following section, these were subsequently compared to those described in the literature.

3. Findings and discussion

Combination of MPA and MIP yielded a higher number of metaphorical occurrences associated to *sadness*. As illustrated in Table 2, out of the 103 metaphorical expressions identified, 63 (accounting for 61% of the total) are related to *sadness* and 40 (39%) to *happiness*.

Table 2. Number of metaphorical expressions of *happiness* and *sadness* identified in the corpus

	N. of occurrences	%
Happiness	40	39%
Sadness	63	61%
Tot	103	100%

The analysis carried out on this sample unveiled patterns discussed in the literature, i.e. by Kövecses (2000) and Stefanowitsch (2006b) in particular (see section 1.1), and led to the identification of metaphors requiring the formulation of appropriate mappings.

The patterns emerging more frequently in the sample are conceptualisations of the emotions under investigation in terms of objects or substances that can be quantified or measured. These types of metaphor are based on a process by means of which we conceptualise abstract concepts as concrete objects or substances and can therefore be applied to a broad range of emotions: for a more comprehensive classification of these frequent

patterns, the MetaNet repository suggests the mappings EMOTIONS ARE OBJECTS and EMOTIONS ARE SUBSTANCES.

The examples below (1-6 representing HAPPINESS, 7-12 representing SADNESS) show that both *happiness* and *sadness* are often described by means of a plural form or adjectives, adverbs, and verbs expressing quantity (e.g. *great*, *some*, *little*, *to increase*, *to amount*), which imply the source domain A QUANTIFIABLE ENTITY. In particular, the use of the plural form emerged more frequently in the subset of metaphorical expressions manifesting *sadness*, where quantity is expressed by means of plural nouns such as *regrets* and *sorrows* (examples 9, 10).

- (1) ... she is young, and her conversation would be a **great** *relief* to me...
- (2) ... it has been a **great** *delight* to me to read Mr. Thackeray's work...
- (3) Dr. Johnson has **more** *fun* [...] than almost anybody I ever saw...
- (4) I should think that in due time a memorial might get **some** *relief* in this part of the appointment...
- (5) I had a **great deal of** *fun* in the accomplishment of this task...
- (6) ... and has as **little** *enjoyment* or pleasure in life at present as anybody in the world...
- (7) I had no pain, and so **little** *dejection* in this dreadful state...
- (8) ... and, to **increase** my *misery*, the knaves are sure to find me at home...
- (9) So I will cease my *regrets*...
- (10) I must not bother you too much with my *sorrows*...
- (11) ... and **small** *grief* at our parting...
- (12) Mine, you are to know is a white *Melancholy*, or rather Leucocholy for the most part; which, though it seldom laughs or dances, nor ever **amounts** to what one called *Joy* or *Pleasure*...

Example 12 deserves particular attention in that it represents an instance in which both concepts of *happiness* and *sadness* appear in the same metaphorical expression and it involves more than one metaphorical pattern.

Firstly, *melancholy*, an emotion associated with *sadness*, is conceptualised as a quantifiable entity: the author compares the amount of this emotion to that of *joy*, referred to by means of the same metaphorical mapping. Secondly, *melancholy* is referred to as a human being, capable of laughing and dancing, thus involving a personification metaphor expressed through both the capitalisation of the word and the choice of the verbs *to laugh* and *to dance*. Finally, the author conceptualises *melancholy* in terms of colour by specifying that the kind of emotion that he is experiencing is *white*, as opposed to the *black* one from which it differs considerably.⁴ This distinction has its root in the humoral theory underpinning medicine from ancient times up to the beginning of the nineteenth century, on the basis of which it is possible to link this metaphorical pattern to a specific cultural phenomenon. According to the doctrine founded by Hippocrates, four humoral fluids regulate the human body, i.e. black bile, yellow bile, phlegm, and blood. These are responsible for health and disease, as well as for the four temperaments: melancholic, choleric, phlegmatic, and sanguine respectively. As a result, "the melancholic, gloomy and fearful, suffers from a constitutional excess of black bile" (Geeraerts – Grondelaers 2010: 156). Due to the influence of the traditional concept of the four humours, conceptualisation of melancholy in terms of colour was not unusual at the time: the corpus analysis of Burton's *Anatomy of Melancholy* carried out by Sandford from a cognitive linguistic perspective revealed considerable use of terms related to colour and vision, besides confirming that "[b]lackness is the key representative of melancholy, [...] of sadness" (217).

The sample includes another metaphorical expression in which the two emotions appear together: example 13 (below) illustrates how the opposite emotions are conceptualised in terms of two distant locations connected by means of a road. More specifically, *sorrow* represents the point of departure of the journey undertaken along the road and *joy* the destination. The author

⁴ The reported example is taken from Thomas Gray's letter addressed to Richard West, written on 27 May 1742, in which the distinction between white and black melancholy is presented: "Mine, you are to know is a white Melancholy, or rather Leucocholy for the most part; which, though it seldom laughs or dances, nor ever amounts to what one called Joy or Pleasure, yet is a good easy sort of a state, and *ça ne laisse que de s'amuser*. The only fault is its insipidity; which is apt now and then to give a sort of Ennui, which makes one form certain little wishes that signify nothing. But there is another sort, black indeed, which I have now and then felt, that has somewhat in it like Tertullian's rule of faith, *Credo quia impossibile est*; for it believes, nay, is sure of everything that is unlikely, so it be but frightful; and on the other hand excludes and shuts its eyes to the most possible hopes, and everything that is pleasurable; from this the Lord deliver us! for none but he and sunshiny weather can do it."

describes a change of emotional state by referring to a movement from one place to another, hence instantiating the mapping CHANGE OF STATE IS CHANGE OF LOCATION, recorded in the MetaNet repository. This is a metaphorical pattern that can be applied to a broad range of target domains including those related to emotions. Taking this example into account, it is possible to formulate a more specific mapping, HAPPINESS/SADNESS IS A LOCATION. This pattern can also be observed in example 14, where *despair* is referred to in terms of a location or point of departure of a path or trajectory leading to another location.

- (13) ... it is a long **road** from *sorrow* to *joy*...
- (14) ... there is **no middle way** between *despair* and the yielding up of the whole spirit unto the guidance of faith.

Another source domain extensively discussed in the literature that was identified in the sample is A SUBSTANCE IN A CONTAINER. The verbs *to contain* and *to fill* in examples 15 and 16 and the word *fullness* in example 17 suggest that the emotions in question are treated as substances with which the author (or their heart, as illustrated in example 15) is filled, thus implying a reference to the metaphorical pattern BODY IS A CONTAINER FOR EMOTIONS (as formulated in the MetaNet repository). Examples 15 and 16 show that the substances referred to through this type of metaphor can be measured or quantified. In addition, considering the sense “to carry as a burden” (Johnson’s Dictionary) as the basic meaning of the verb *to bear*, we can observe how example 15 instantiates the mapping SADNESS IS A BURDEN.

- (15) It surprises me that the human heart is capable of **containing** and **bearing** so **much** *misery*.
- (16) ... **has filled me with a degree of** *grief* and *dismay* which I cannot find words to express.
- (17) ... I often shed tears in the motley Strand from **fullness of** *joy* at so much life.

I will now focus on the conceptualisation of *happiness* by considering the metaphors associated with this emotion. On the basis of the collected data, *happiness* seems to be predominantly conceptualised by means of patterns that are described in the literature, even though a few cases require the formulation of additional mappings.

- (18) ... **enjoy** the *delights* of both city and country...
- (19) ... giving myself **the highest bliss** I know on earth.
- (20) ... We cannot arrive at **any portion of heavenly bliss** without in some measure imitating Christ...
- (21) ... I had many, many hours of **pure happiness**.
- (22) ... I know not what other **terrestrial happiness** would deserve **pursuit**.
- (23) They teach the youthful mind to sigh after beauty and *happiness* that never **existed**...
- (24) ... describe *happiness* that man never **tastes**.

Being *happiness* perceived as a positive emotion, it is not surprising to find the metaphor HAPPINESS IS A PLEASURABLE PHYSICAL SENSATION manifesting through the verb *to enjoy* (example 18). Other metaphors specific to this emotion concept are HAPPINESS IS UP (or VERTICALITY, as formulated in the MetaNet repository), exemplified by the adjective *high* in example 19, and HAPPINESS IS BEING IN HEAVEN, observable in example 20, where the presence of the adjective *heavenly* points to a reference to God's abode. In addition, this example conceptualises *bliss* as a quantifiable entity that can be divisible in portions, instantiating a metaphorical pattern that is commonly applied to a wide variety of emotions. Another common pattern is that instantiated by example 21, in which *happiness* is described in terms of a substance; however, the adjective *pure* suggests the involvement of the more specific source domain A PURE SUBSTANCE (cf. Stefanowitsch's mapping HAPPINESS/BEING HAPPY IS A MIXED/PURE SUBSTANCE in section 1.1).

Personification metaphors constitute another recurring pattern in the sample. The noun *pursuit* found in example 22 possibly implies a personification of *happiness*, if we consider the basic meaning of the verb *to pursue* as exemplified by the sense "to chase; to follow in hostility" (Johnson's *Dictionary*). In addition, this example is particularly interesting in view of the presence of the adjective *terrestrial*, which suggests the involvement of a source domain that is opposite to that referred to in Kövecses's mapping HAPPINESS IS BEING IN HEAVEN (see section 1.1), observed in example 20. If we take into account the opposition between an earthly dimension and a celestial one, as suggested by the sense "earthly; not celestial" found below the headword *terrestrial* in Johnson's *Dictionary*, HAPPINESS IS BEING ON EARTH seems an appropriate mapping for this metaphorical expression.

Example 23 can be understood by means of the mapping HAPPINESS IS A REAL ENTITY due to the presence of the verb *to exist*, representing an action that can be attributed to both living beings and inanimate objects. Another metaphor requiring the formulation of a specific mapping is example 24, in which the verb *to taste* suggests the mapping HAPPINESS IS FOOD.

The metaphorical mappings of HAPPINESS identified in the sample are summarised in Table 3.

Table 3. Metaphorical mappings of HAPPINESS identified in the sample

HAPPINESS/BEING HAPPY IS
A LOCATION
A PERSON
A PLEASURABLE PHYSICAL SENSATION
A PURE SUBSTANCE
A QUANTIFIABLE ENTITY
A REAL ENTITY
A SUBSTANCE IN A CONTAINER
AN OBJECT
BEING IN HEAVEN
BEING ON EARTH
FOOD
UP/VERTICALITY

I turn now to the conceptualisation of *sadness*. Besides being expressed through a higher number of metaphors in the corpus, it was found to be described metaphorically through a wider variety of mappings.

- (25) So I will cease my *regrets*, or **lay them by to be taken up and used** as arguments of comfort...
- (26) But in your absence the **tide** of *melancholy* **rushed in** again...
- (27) To save me from that horrid situation of at any time **going down** [...] to *misery*...
- (28) ... in **gloomy** *discontent* or **importunate** *distress*.
- (29) ... which premature hour arises, I suppose, from *sorrow* being **hungry** as well as **thirsty**.
- (30) I will endeavour to **bear** my *miseries* patiently.
- (31) ... but *despair* **is forced upon me** as a habit.
- (32) ... after an interval of several months, during which **my flesh wasted from me with** sickness and *melancholy*.
- (33) ... or the cold and brief compliments, with the **warm** *regrets*, of the Quarterly?

- (34) No *misery* is **to be seen** here...
- (35) ... the comfort and blessing that literature can prove in **seasons of** sickness and *sorrow*...
- (36) I **opened** to him all my *distresses*...
- (37) ... her *anguish* is as **sharp** as yours...

Example 25 is possibly the one which best represents the pattern SADNESS IS AN OBJECT: *regrets* are described by the author as objects or tools to be laid by or "taken up and used". On the other hand, example 26 appears as a representative instance of the mapping SADNESS IS A NATURAL FORCE, a pattern commonly applied to emotion concepts: by referring to the sea when the tide is rising, the author effectively describes the recurrence of *melancholy*. Furthermore, in view of the presence of the verb *to rush*, which usually refers to an action carried out by human beings, the same example includes a personification metaphor.

Examples 27 and 28 show metaphors that are typically used to refer to *sadness*: SADNESS IS DOWN, which represents the opposite of HAPPINESS IS UP, and SADNESS IS DARK. The mapping SADNESS IS A LIVING ORGANISM is exemplified by the metaphor identified in example 29, where *sorrow* is described as a hungry and thirsty living being. Other patterns that are common in *sadness* metaphors are illustrated in example 30, 31, 32, which instantiate the mappings SADNESS IS A BURDEN, SADNESS IS A PHYSICAL FORCE, and SADNESS IS AN ILLNESS respectively. Example 33 is particularly interesting in that it instantiates the mapping SADNESS IS HEAT, which involves a source domain that is more commonly found in happiness metaphors. Considering Kövecses's mapping SADNESS IS A LACK OF HEAT (see section 1.1), we would expect to find lexical items associated with cold in the *sadness* subset. However, this is not the case. In addition, due to the presence of a plural form (i.e. *regrets*), this example involves a conceptualisation of *sadness* in terms of a quantifiable entity.

Infrequent mappings, found only once in the sample, are illustrated in examples 34, 35, 36, 37. Examples 34 and 35 suggest involvement of the source domains A VISIBLE ENTITY and A PERIOD OF THE YEAR respectively. Interestingly, example 36 involves the source domain A CONTAINER, usually found in the mapping BODY IS A CONTAINER FOR EMOTIONS (MetaNet repository). The source domain A SHARP OBJECT, illustrated in example 37, was found to be associated with *happiness* by Stefanowitsch (2006b) but is here used to describe *sadness*, which suggest its association with both emotion concepts.

The metaphorical mappings of SADNESS identified in the sample are summarised in Table 4.

The mappings identified in this study confirm Stefanowitsch’s findings suggesting that “the emotions referred to by *happiness* and *sadness* are not primarily understood as opposites, but [...] each of them is conceptualised (and presumably experienced) on its own terms” (2006b: 102). Besides the prototypical opposition between HAPPINESS IS UP and SADNESS IS DOWN, the metaphorical expressions found in the corpus show how, notwithstanding the semantic opposition between the two, these emotions manifest the general patterns that are common to a wide variety of emotions (A SUBSTANCE IN A CONTAINER, A QUANTIFIABLE ENTITY, AN OBJECT, etc.) and some specific mappings (e.g. HAPPINESS IS BEING IN HEAVEN, SADNESS IS A BURDEN).

The analysis presented above revealed that eighteenth- and nineteenth-century poets and writers recurred to metaphor to refer to the concepts of *happiness* and *sadness* not only when producing literary works but also in their private correspondence. In order to do so, they exploited common mappings still observable in PDE non-literary discourse (i.e. HAPPINESS IS UP, SADNESS IS DOWN), as well as infrequent mappings that are not recorded in the literature, possibly more likely to be found in literary discourse (i.e. HAPPINESS/ SADNESS IS A LOCATION, HAPPINESS IS FOOD, SADNESS IS A PERIOD OF THE YEAR).

4. Conclusion

This investigation of metaphorical expressions involving the target domains HAPPINESS and SADNESS in a corpus of Late Modern English private correspondence has firstly confirmed the effectiveness of MPA as a method for metaphor identification in corpus-based studies, especially in the study of emotion concepts. Such target-domain oriented approach, which allows us

Table 4. Metaphorical mappings of SADNESS identified in the sample

SADNESS/BEING SAD IS
A BURDEN
A CONTAINER
A LIVING ORGANISM
A LOCATION
A NATURAL FORCE
A PERIOD OF THE YEAR
A PERSON
A PHYSICAL FORCE
A QUANTIFIABLE ENTITY
A SHARP OBJECT
A SUBSTANCE IN A CONTAINER
A VISIBLE ENTITY
AN ILLNESS
AN OBJECT
DARK
DOWN
HEAT
WHITE/BLACK

to identify metaphors via a set of lexical items belonging to the target domain in question, proved effective in terms of results and less time-consuming if compared to the traditional introspective method. In this regard, the adoption of Wmatrix to allocate semantic domain tags and subsequently to select a set of representative lexical items has proved essential for the identification of potentially metaphorical expressions associated with the emotions under investigation. The sample of metaphors found in the corpus as a result of the use of the aforementioned lexical items as keywords to perform concordance searches has confirmed Stefanowitsch's claim that "near synonyms will broadly be associated with the same metaphors (and thus, that *it is possible to investigate emotion concepts via individual lexical items*)" (2006b: 101-102, italics mine).

Secondly, the analysis carried out on the selected sample of metaphorical expressions identified in the corpus by means of MIP has led to results that confirm Kövecses's (2000) and Stefanowitsch's (2006b) findings concerning the conceptualisation of these basic emotions and to the formulation of additional mappings that are specific to the metaphorical expressions found in this corpus. Further investigation and a larger dataset may reveal additional mappings and possibly shed light on the frequency of occurrence of such metaphorical patterns in Late Modern English educated discourse related to private contexts. Moreover, the application of quantitative analyses may provide results concerning the most significant patterns and the (basic and non-basic) emotion concepts that are most represented in the corpus.

Even though a larger amount of data is required to obtain better results, the analysis of both target domains, usually conceptualised by means of antonyms, has proved beneficial in finding commonalities and divergences in the metaphorical expressions used to describe these opposite emotions. It was illustrated how Late Modern English authors resorted to source domains that are common to a wide range of emotion metaphors (i.e. A QUANTIFIABLE ENTITY, A (VALUABLE) OBJECT, A SUBSTANCE IN A CONTAINER) as well as to a variety of domains that are specific to the emotions under investigation, in some cases unveiling mappings that are not described in the literature. These findings confirm that *happiness* and *sadness* "do not fall into pairs of opposing metaphors" (Stefanowitsch 2006b: 102).

The study shows how eighteenth- and nineteenth-century men and women of letters made extensive use of metaphor not only in their literary work but also in their private correspondence, confirming both the pervasive nature of metaphor pointed out by Lakoff and Johnson (1980) and "the *continuity* between metaphor in literature and metaphor in non-literary

language" (Semino & Steen 2008: 233). As suggested by the metaphorical expressions identified in the corpus, these two dimensions converge in the private correspondence of educated speakers. Moreover, the adoption of MIP disclosed a higher tendency to conceptualise *sadness* by resorting to metaphor as opposed to *happiness*.

Undoubtedly, the historical, cultural, and literary context in which these letters were produced played a role in shaping the ways in which poets and writers interacted with their emotions and imagination, calling for further discussion on the relationship between Romanticism and the conceptualisation of emotions in Late Modern English.

REFERENCES

Primary sources

- Duckitt, M. – Elgee, H. W.
 2004 *Selected English Letters (XV-XIX centuries)*. Project Gutenberg.
<https://www.gutenberg.org/ebooks/12515>, accessed June 2022.
Johnson's Dictionary Online
<https://johnsonsdictionaryonline.com/>, accessed February 2023.
Oxford English Dictionary Online (OED online)
<http://www.oed.com/>, accessed February 2023.
 Webster, N.
 1886 *Complete Dictionary of the English Language*. London: George Bell & Sons. <https://archive.org/details/websterscomplete00webs/mode/1up>, accessed February 2023.

Secondary sources

- Allan, K.
 2009 "An inquest into metaphor death: Exploring the loss of literal senses of conceptual metaphors", *Cognitive Semiotics* 5 (1-2), 291-311.
 Barcelona, A.
 1986 "On the concept of depression in American English: A cognitive approach", *Revista Canaria de Estudios Ingleses* 12, 7-33.
 Daignan, A.
 2005 *Metaphor and Corpus Linguistics*. Amsterdam: John Benjamins.
 Ekman, P.
 1999 "Basic emotions". In T. Dalgleish – M. J. Power (eds.) *Handbook of Cognition and Emotion*. Chichester: John Wiley & Sons, 45-60.

- Fehr, B. – J. A. Russell
 1984 "Concept of emotion viewed from a prototype perspective", *Journal of Experimental Psychology: General* 113 (3), 464-486.
- Geeraerts, D. – S. Grondelaers
 2010 "Looking back at anger: Cultural traditions and metaphorical patterns". In: J. R. Taylor – R. E. MacLaury (eds.) *Language and the Cognitive Construal of the World*. Berlin: De Gruyter, 153-180.
- Gevaert, C.
 2005 "The anger is heat question: Detecting cultural influence on the conceptualization of anger through diachronic corpus analysis". In: N. Delbecq et al. (eds.) *Perspectives on Variation: Sociolinguistic, Historical, Comparative*. Berlin: De Gruyter, 195-208.
- Glynn, D.
 2002 "Love and anger: The grammatical structure of conceptual metaphors", *Style* 36 (3), 541-559.
- Hintikka, M.
 2007 "Sickness as metaphor in Early Modern and Present-Day English". In: M. Rissanen et al. (eds.) *Change in Meaning and the Meaning of Change: Studies in Semantics and Grammar from Old to Present-Day English*. Helsinki: Société Néophilologique, 91-112.
- Kilgarrriff, A. et al.
 2004 "Itri-04-08 the Sketch Engine", *Information Technology* 1, 105-116.
 2014 "The Sketch Engine: Ten years on", *Lexicography* 1, 7-36.
- Koller, V. et al.
 2008 "Using a semantic annotation tool for the analysis of metaphor in discourse", *Metaphorik.de* 15 (1), 141-160.
- Kövecses, Z.
 1986 *Metaphors of Anger, Pride and Love: A Lexical Approach to the Structure of Concepts*. Amsterdam: John Benjamins.
 1988 *The Language of Love: The Semantics of Passion in Conversational English*. Lewisburg: Bucknell University Press.
 1990 *Emotion Concepts*. New York: Springer.
 1991 "Happiness: A definitional effort", *Metaphor and Symbolic Activity* 6 (1), 29-46.
 1995a "Metaphor and the folk understanding of anger". In: J. A. Russell et al. (eds.) *Everyday Conceptions of Emotion: An Introduction to the Psychology, Anthropology and Linguistics of Emotion*. Dordrecht: Springer, 49-71.
 1995b "The 'container' metaphor of anger in English, Chinese, Japanese and Hungarian". In: Z. Radman (ed.) *From a Metaphorical Point of View: A Multidisciplinary Approach to the Cognitive Content of Metaphor*. Berlin: De Gruyter, 117-145.
 2000 *Metaphor and Emotion*. Cambridge: Cambridge University Press.
- Lakoff, G. – M. Johnson
 1980 *Metaphors We Live By*. Chicago: The University of Chicago Press.

- Lakoff, G. – Z. Kövecses
 1987 “The cognitive model of anger inherent in American English”.
 In: D. Holland – N. Quinn (eds.) *Cultural Models in Language and Thought*. Cambridge: Cambridge University Press, 195-221.
- Mugglestone, L. (ed.)
 2006 *The Oxford History of English*. Oxford: Oxford University Press.
- Patterson, K. J.
 2018 *Understanding Metaphor through Corpora: A Case Study of Metaphors in Nineteenth Century Writing*. London/New York: Routledge.
- Pragglejaz Group
 2007 “MIP: A method for identifying metaphorically used words in discourse”, *Metaphor and Symbol* 22 (1), 1-39.
- Rayson, P.
 2009 *Wmatrix: A web-based corpus processing environment*. Computing Department, Lancaster University. Available:
<http://ucrel.lancs.ac.uk/wmatrix/>, accessed June 2022.
- Rosch, E.
 1975 “The nature of mental codes for color categories”, *Journal of Experimental Psychology: Human Perception and Performance* 1, 303-322.
 1978 “Principles of categorization”. In E. Rosch – B. B. Lloyd (eds.) *Cognition and Categorization*. Hillsdale, N.J: Lawrence Erlbaum, 27-48.
- Sandford, J. L.
 2017 “‘Active in sight, the eye sees the colour’ (1.157): A cognitive semantic analysis of seeing in Robert Burton’s *Anatomy of Melancholy*.”
 In: A. Volpone (ed.) *La felicità di essere tristi. Saggi sulla melanconia*. Perugia: Aguapiano, 195-218.
- Semino, E.
 2017 “Corpus linguistics and metaphor”. In: B. Dancygier (ed.) *The Cambridge Handbook of Cognitive Linguistics*. Cambridge: Cambridge University Press, 463-476.
- Semino, E. – G. Steen
 2008 “Metaphor in literature”. In: R. W. Gibbs (ed.) *The Cambridge Handbook of Metaphor and Thought*. Cambridge: Cambridge University Press, 232-246.
- Shaver, P. et al.
 1987 “Emotion knowledge: Further exploration of a prototype approach”, *Journal of Personality and Social Psychology* 52 (6), 1061-1086.
- Steen, G. J. et al.
 2010 *A Method for Linguistic Metaphor Identification: From MIP to MIPVU*. Amsterdam: John Benjamins.
- Stefanowitsch, A.
 2004 “Happiness in English and German: A metaphorical-pattern analysis”. In: M. Achard – S. Kemmer (eds.) *Language, Culture,*

- and Mind*. Stanford: Stanford University, Center for the Study of Language and Information, 137-149.
- 2006a "Corpus-based approaches to metaphor and metonymy".
In: A. Stefanowitsch – S. T. Griers (eds.) *Corpus-based Approaches to Metaphor and Metonymy*. Berlin: De Gruyter, 1-16.
- 2006b "Words and their metaphors: A corpus-based approach".
In: A. Stefanowitsch – S. T. Griers (eds.) *Corpus-based Approaches to Metaphor and Metonymy*. Berlin: De Gruyter, 63-105.
- Stefanowitsch, A. – S. T. Griers
2006 *Corpus-based Approaches to Metaphor and Metonymy*. Berlin: De Gruyter.
- Tieken-Boon van Ostade, I.
2009 *An Introduction to Late Modern English*. Edinburgh: Edinburgh University Press.
- Tissari, H.
2001 "Metaphors we love by: On the cognitive metaphors of LOVE from the 15th century to the present", *Studia Anglica Posnaniensia* 36, 217-242.
- 2006 "Conceptualizing shame: Investigating uses of the English word *shame*, 1418-1991". In: R. W. McConchie et al. (eds) *Selected Proceedings of the 2005 Symposium on New Approaches in English Historical Lexis (HEL-LEX)*. Somerville, MA: Cascadilla Proceedings Project, 143-154.

APPENDIX

Author	N. of letters
JONATHAN SWIFT	5
JOSEPH ADDISON	2
SIR RICHARD STEELE	7
JOHN GAY	1
ALEXANDER POPE	5
SAMUEL RICHARDSON	1
LADY MARY WORTLEY MONTAGU	6
PHILIP DORMER STANHOPE, EARL OF CHESTERFIELD	6
SAMUEL JOHNSON	7
LAURENCE STERNE	3
THOMAS GRAY	7
HORACE WALPOLE	10
OLIVER GOLDSMITH	4

WILLIAM COWPER	11
EDMUND BURKE	4
EDWARD GIBBON	2
FRANCES D'ARBLAY	3
GEORGE CRABBE	2
WILLIAM BLAKE	4
MARY LEADBEATER	2
ROBERT BURNS	3
WILLIAM WORDSWORTH	4
SIR WALTER SCOTT	10
SAMUEL TAYLOR COLERIDGE	6
ROBERT SOUTHEY	4
CHARLES LAMB	10
WILLIAM HAZLITT	3
LEIGH HUNT	5
GEORGE GORDON NOEL, LORD BYRON	8
PERCY BYSSHE SHELLEY	6
JOHN KEATS	4
THOMAS HOOD	4
ROBERT BROWNING and ELIZABETH BARRETT BROWNING	1
CHARLOTTE BRONTË	9

Address: GLORIA MAMBELLI, Dipartimento di Lingue e Letterature Straniere, Università degli Studi di Verona. Lungadige Porta Vittoria, 41, 37129, Verona, Italy.

ORCID code: <https://orcid.org/0000-0003-0944-6386>.

Towards a historical corpus of Canadian English letters and diaries

Giulia Rovelli

University of Bergamo

ABSTRACT

The paper reports on the compilation and illustrates the main features of a corpus of manuscript ego documents written in English by both adults and children in the nineteenth and early-twentieth centuries in Canada – a crucial time in the history of the country and in the development of the Canadian variety of English. The corpus is part of a larger project that intends to contribute to Canadian studies and English historical sociolinguistics by a) tracing the evolution of the specificities of Canadian English in a diachronic perspective; b) investigating the contribution of different languages or varieties of the same language to the development of Canadian English; and c) analyzing how demographic and social differences are encoded in and have influenced language use in Canada. Some preliminary findings are also presented.

Keywords: Canadian English, corpus linguistics, historical sociopragmatics, ego documents, personal letters, diaries.

1. A short history of Canadian English

The term “Canadian English” was first used in 1857 by Reverend A. Constable Geikie with regard to the “lawless and vulgar innovations” that, in his opinion, characterized the English language as spoken in Canada (DCHP-1, Avis et al. 1967; Chambers 1993). Non-derogatory uses of the expression, which may provide evidence of increasing language awareness and nationalism and, consequently, of the language having entered its endonormative stabilization phase (Schneider 2007; Dollinger 2014: 104), emerged only much later (Dollinger 2019a: 20). This is also suggested by

the relatively late publication of such metalinguistic materials as dictionaries (e.g., Avis et al.'s 1967 *Dictionary of Canadian English: The Senior Dictionary*, Avis et al.'s 1967 *Dictionary of Canadianisms on Historical Principles* and Barber's 2004 *Canadian Oxford Dictionary*) and usage guides (e.g., Fee and McAlpine's 1997 *Guide to Canadian English Usage*). However, while "Canadian English" gained currency only in the second half of the twentieth century, the English language has had a much longer history in Canada.

Although English was introduced to Newfoundland in 1497, the territory remained a separate political entity until 1949. For this reason, the language developed fairly independently from what is now referred to as Standard Canadian English (Schneider 2007: 238; Boberg 2010: 26). The origins of Canadian English (CanE) are therefore generally traced back to the late eighteenth century and to the first two of the five waves of immigration that shaped Canada's population and, ultimately, the main features of this variety (Dollinger 2015: 26). Indeed, while the Treaty of Paris (1763), which concluded the Seven Years' War (1756-1763), sanctioned the passage of France's North American colonies (i.e., Nouvelle-France) to Britain, British rule did not initially modify the demographic dominance of French (Edwards 1998: 19-22).

The first wave of immigration took place in the wake of the American Revolution, between 1776 and 1812. It consisted in two main groups of American settlers. The first, and smaller group, generally referred to as the United Empire Loyalists, was made up of anti-revolutionary Americans from Connecticut, Massachusetts, Maine and Rhode Island who chose to maintain their allegiance to the British Crown and settled mostly in Nova Scotia. The second, and more prominent group, consisted of the so-called "late Loyalists", Americans from Pennsylvania, New Jersey, New York and Vermont who, on the other hand, were more interested in new land for agricultural purposes and first settled modern-day New Brunswick and Ontario (Chambers 1998: 258-261; Boberg 2012: 163). Since the latter group of first-wave immigrants effectively represented the "founder population" (Schneider 2007: 240) of the territory that was to become the financial and industrial hub of the country (i.e., Ontario, Chambers 2004: 225), it has been described as the most significant from both a linguistic and cultural point of view (Bloomfield 1948; Mackey 1998: 22).

The second wave of immigration took place between 1812 and 1865 and consisted of mostly government-sponsored immigration from the British Isles, and specifically from Ireland, Northern England and Scotland,

which was supposed to have two main aims: on the one hand, it would “dilute” the American base in Canada, which was (wrongly) suspected of being disloyal to the British Crown, and, on the other, it could relieve Britain of paupers and demobilized soldiers (Boberg 2012: 164; Dollinger 2019a: 14). While this second wave of immigration increased the prestige of all things British, including language forms, on Canadian soil, its impact has been described as quite limited from a linguistic point of view as, with the exception of a few isolated regions (e.g., Cape Breton Island, the Ottawa valley and Peterborough County, see Chambers 2004: 228), it did not alter the basically North American character of the language (Chambers 1998: 262; Dollinger 2019a: 14).

The third wave of immigration has been traced back to the period between 1890 and 1914 and consisted of immigration from not only Scotland and Ireland, but also Continental Europe, especially from Germany, Scandinavia and Ukraine (Chambers 1998: 264). The fourth wave took place in the wake of WWII, between 1946 and 1970, and consisted of a highly diverse immigrant population from Europe, Asia and Latin America (Dollinger 2019a: 12). Finally, the fifth wave began in the 1990s and has been described as characterized by an even more varied immigrant population (Dollinger 2019a: 12). While certainly fundamental for the development of the country and its identity, the latter three waves of immigration had a more limited linguistic impact, as later immigrant populations tend to assimilate very quickly and conform to dominant patterns.

Because of this, for a long time Canadian English has been described simplistically as a conservative and homogenous language, and as a variety of American English with a British orientation – a rather restrictive definition, which fails to recognize the independence and autonomy of Canadian English (Dollinger 2008: 134, 2012b, 2015), effectively relegating it to a secondary position. Many studies have recently been carried out in an attempt to rectify the situation and highlight the specificities of this variety of English, especially from the point of view of lexis and phonology (e.g., Clarke et al.’s 1995 discussion of the Canadian Shift, Berger’s 2005 study of lexical variables, Chambers’s 2006 investigation of Canadian Raising, Dollinger’s 2008 analysis of the use of modal auxiliaries in early Ontario English, and Boberg’s 2010 monograph on Canadian English, among others). Even so, Canadian English continues to be a relatively under-researched variety, especially from a historical perspective (Dollinger 2008, 2012a, 2019b; Boberg 2010, 2017).

2. Aims

The present paper reports on the compilation of the new *Corpus of Canadian English Letters and Diaries* (CCanDL), a historical corpus of manuscript ego documents written in Canada between the late-eighteenth and early-twentieth centuries, which is being developed at the University of Bergamo, Italy, in collaboration with the *Rural Diary Archive* and the *McLaughlin Library* of the University of Guelph, Ontario (CA).¹ The project intends to contribute to Canadian English studies, which, as lamented by some scholars, still lack a diachronic dimension in full detail (Dollinger 2008, 2012a, 2019b; Boberg 2010, 2017). In addition, it aims to advance studies in English historical sociolinguistics by a) tracing the origins and diachronic evolution of Canadian English and its specificities; b) investigating the contribution of different languages (most importantly French and some of Canada's indigenous languages) and of different varieties of English to the development of Canadian English; and c) analyzing how demographic and social differences are encoded in and have influenced language use in Canada.

3. Corpus design

As the main goal of the project is to investigate the origins and diachronic evolution of the English language as used in Canada, CCanDL has been designed to include private letters and diaries written in the several provinces of Canada between the late-eighteenth and early-twentieth centuries. Personal letters and diaries have been chosen precisely because of their private nature: since they were written for the authors themselves (in the case of diaries) or for their immediate family (in the case of letters, but sometimes applicable to diaries as well), these genres generally provide evidence of less monitored linguistic choices, which might more closely reflect vernacular language – that is, how language is used in everyday situations (Tieken-Boon van Ostade 2005; Dossena 2012; Allen 2014; Dollinger 2014). Indeed, letters and diaries represent prototypical texts of immediacy, which are often characterized by orality, and for this reason they have been described as the closest approximation to spoken language (Elspeß 2012; Van der Wal – Rutten 2013). Although focusing on written

¹ Gratitude is expressed to both institutions for their generous cooperation. Once completed, the corpus will be made freely available in order to facilitate research into the historical development of Canadian English.

texts has its limitations, in that it restricts the investigation to the language use of literate informants who had the means of putting their thoughts on paper, this approach represents the only viable option – a necessary evil if one wants to analyze how language was used before the invention and widespread availability of the means to record and reproduce actual speech. Indeed, since the earliest recordings date back to the late nineteenth century,² much of the period under examination here would be excluded. Moreover, as stated by Hickey (2017), early recordings are normally very short, thus limiting the range of data; they generally consist of readings of set pieces, making free speech rarely available; and they tend not to present social stratification (Hickey 2017: 1-4).

The timeframe for the collection of such documents has been established as between 1776 and 1950. Following Dollinger (2008), the start date for the corpus has been set to 1776, the year which gave birth to the United States and, as a consequence, to British North America, namely, Canada (Chambers 2004: 225). The end date, on the other hand, has been set to 1950, as it was in the wake of WWI and WWII that linguistic nationalism started to peak in Canada, thus resulting in linguistic awareness and autonomy, which have been described as the direct consequences of the “coming of age” of Canada, and in the production of Canadian metalinguistic materials (Dollinger 2012b: 452).

The texts to be included in the corpus are being retrieved from the *Rural Diary Archive*, an online archive that collects rural diaries written in Ontario (which are being transcribed by volunteers all over the world),³ and from the *McLaughlin Library* manuscript archives. As their being based in Ontario would most likely bias the corpus towards that particular region and most certainly provide further evidence for the Loyalist Base Theory (cf. Bloomfield 1948), other archives will also be searched so as to provide as

² While the first attempts to record and reproduce sound waves date back to the phonautograph, a mechanical sound-recording device invented in 1857 by Édouard-Léon Scott de Martinville, the first device that could actually record and play back sounds was Thomas Alva Edison’s phonograph, invented in 1877 (cf. <https://www.britannica.com/science/acoustics/Amplifying-recording-and-reproducing#ref527605>).

³ Although using manuscript materials that have being transcribed by volunteers is not entirely unproblematic, as it might result in unsystematic biases and errors, this approach allows the compilation of a much bigger corpus than would be possible if the author alone were to transcribe the material. Moreover, the *Rural Diary Archive* offers prospective volunteer transcribers instructions on how to transcribe texts and deal with most issues. Finally, since the Archive collects both the digitized and the transcribed version of each file, the original text can be referred to and checked in case of doubtful transcriptions.

balanced an account as possible. In order to be included in the corpus, the authors of the manuscript materials in question must have either been born in Canada or emigrated there permanently. On the other hand, the corpus does not include the following text types:

- a) official or semi-official letters, whose authorship is more difficult to trace, as the sender might not have been the encoder of the letter, and, precisely because of their official or formal purpose, tend to be characterized by controlled, rather than spontaneous, linguistic choices;⁴
- b) published material, since it is impossible to differentiate the author's language from the editor's or publisher's on account of editorial interventions;
- c) memoirs and other types of autobiographical narratives that entail a certain degree of identity construction and reconstruction, which, because of their partly fictional nature, would be further removed from actual vernacular language use.

Each document included in the corpus will be accompanied by a detailed fact file (see Fig. 1) that reports the author's biographical and demographic information – data which are essential for sociopragmatic studies. These include the writers' name and gender; their place of birth and ethnicity, which might offer an indication of the contribution that different groups of speakers and their original dialects had on the development of Canadian English; their date of birth and age at the time of writing, which may help shed light on the diachronic evolution of the language; and location, which could provide some clues to its geographical differentiation. The fact files will also include information on the writers' religion and occupation, which could give an indication of their level of education; and close relatives, which, on the other hand, may be used to trace communities of practice. Such information will be retrieved directly from the archive and, when available, enriched by census data. The most important pieces of information will also be reported in the file name, which will include the diarist's name, gender, year of writing, any outstanding detail (e.g., ethnicity, religious affiliation, age, etc.), the standardized date of the entry, and the date as it appears in the diary itself (e.g. F. 1887 (German) – Middagh, V. – 1887-01-01 (1 Saturday); M.c. 1866 – Clarke, C.K. – 1866-02-21 (Wednesday Feb 21st 1866)). This will

⁴ Although more insecure writers might have relied on their more experienced relatives and friends, or even scribes, to help them write personal letters as well, this is less likely to be the case than for official or semi-official correspondence.

allow the creation of *ad hoc* sub-corpora that could be set up to investigate the contribution or idiosyncrasies of a specific writer or socio-demographic group.



Name: George Hill

Surname: Detlor

Gender: Male

Year of Birth: 1794

Year of Death: 1883

Age at Time of Writing: 28-55

Place of Birth: Ontario

Ethnicity: Irish

- Father's place of birth: US
- Mother's place of birth: US

Residence: Ontario

- County: Lennox & Addington
- Township: Fredericksburg

Religion: Methodist

Occupation: Farmer, Merchant, Mill-builder, Politician

Relationships:

- Father: Col. John Detlor (United Empire Loyalist)
- Mother: Jerusha (United Empire Loyalist)
- Wife: Maria Moore Roblin
- Sons: John St George, William Valentine, Titus Simon, Samuel McLean, Egerton Ryerson
- Daughters: Amelia Ann, Mary Matilda, Elizabeth Sophia, Elizabeth Lockwood, Helen, Jane Jerusha

Materials in corpus:

3 diaries (1822-1849)

Figure 1. Example of a fact file. Photo courtesy of *Rural Diary Archive*

As such, the corpus allows investigations in historical sociopragmatics at any level – phonological, morphological, syntactic, lexical, and pragmatic. Since CCanDL will include letters and diaries written at different times in

different parts of Canada and by different types of writers, diachronic (i.e., time-related), diatopic (i.e., regional) and diastratic (i.e., related to social class) studies may also be envisaged. The following section presents a brief case study which was carried out by retrieving data from the CCanDL corpus sampler that is available at this stage.

4. Preliminary findings

4.1 The CCanDL sampler

At the time of writing (July 2023), the CCanDL sampler includes a selection of 78 diaries written in present-day Ontario which were collected from transcribed material already available in the *Rural Diary Archive*, for a total of 1,221,244 tokens. Ontario was chosen as the starting point, firstly, for chronological reasons; secondly, for its importance in the history of Canada from a political, economic and linguistic point of view; and, thirdly, for the availability of material (Chambers 2004: 225; Baskerville 2005; Whitcomb 2007). The composition of the CanDL Sampler is shown in Table 1 below.

Table 1. Composition of the CCanDL Sampler

Gender	Males	26
	Females	18
Age	Adults (26-54yo)	42
	Young adults (20-25yo)	15
	Older people (60-84yo)	11
	Adolescents (17-19yo)	8
	Children (8-9yo)	2
Ethnicity	Canadians	28
	3rd-generation immigrants	13
	English ethnicity	4
	German ethnicity	2
	Irish ethnicity	5
	Scottish ethnicity	2
	2nd-generation immigrants	6
	English ethnicity	4
	Irish ethnicity	2

Ethnicity	no information	9
	African ethnicity	1
	English ethnicity	2
	German ethnicity	1
	Irish ethnicity	2
	Scottish ethnicity	3
	Immigrants	16
	American ethnicity	1
	Dutch ethnicity ?	1
	English ethnicity	4
	English ethnicity ?	2
	Irish ethnicity	1
	Scottish ethnicity	4
	Scottish ethnicity?	2
	Welsh	1
Religion	Presbyterians	11
	Methodists	10
	Anglicans	9
	Baptists	5
	Quakers	3
	Disciples	2
	Universalists	1
	Congregationalists	1
	Roman Catholics	1
Occupation	Farmers	18
	& newspaper men	1
	& minister	1
	& merchants	1
	& politician	1
	& students	1
	& millers and distillers	1
	& newspaper men	1
	Farm women	15
	& students	1
	& teachers	1

Occupation	Family of merchants	2
	Merchants	2
	Apprentice surveyor	1
	Carpenter	1
	Family of farmers	1
	Marble cutters	1
	Teachers	1

The diaries were penned by a total of 44 diarists, 26 of which were male and 18 female, who were born between 1794 and 1894 and died between 1870 and 1984. While the greatest share of diaries (42, corresponding to 56%) was written by adults (a category that includes people aged between 26 and 54), the corpus sampler also contains 15 diaries (19%) written by young adults (namely people aged between 20 and 25), 11 diaries (14%) written by an older person (which includes people aged between 60 and 84), 8 diaries (10%) written by adolescents (in this case, 17-19-year olds), and 2 diaries (3%) written by a child (an 8-9-year old, cf. Table 1).

The diaries so far included in the corpus were written between 1822 and 1919, thus offering a glimpse into Canadian English over much of the nineteenth century. That being said, however, the sub-sections are not entirely comparable to each other; the 1830s and 1840s are not represented in the corpus yet; while the 1820s are only scarcely present – see Fig. 2, which shows the proportion of the chronological sub-sections in the corpus sampler.

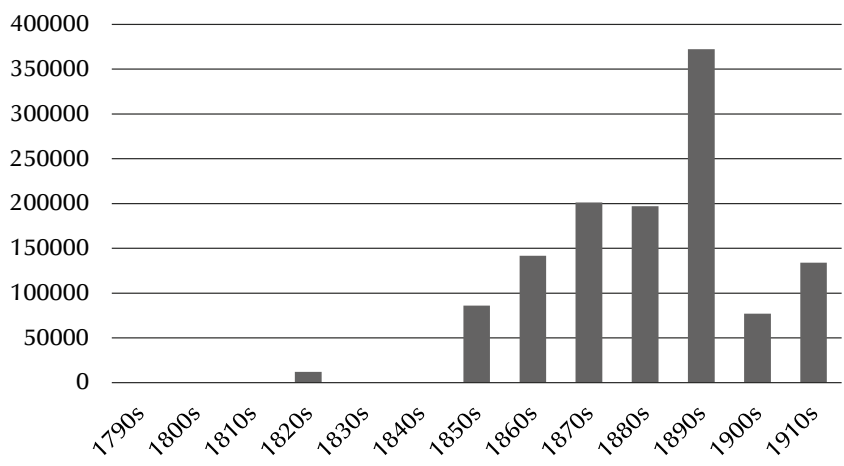


Figure 2. Proportion of the chronological sub-sections of the CCanDL corpus sampler

Although the corpus is not as balanced as seems possible given the scope of the material, this will soon be amended as new material is being made available and prepared.

Census information revealed that most diarists included in the corpus sampler (28) were born in Ontario. Of these, 13 may be described as at least third-generation immigrants, since their parents were born in Ontario as well; 6 may, on the other hand, be identified as second-generation immigrants (i.e., the sons and daughters of immigrants); while no information is available for the rest. Although they were born in Canada, 10 were originally from England, 9 from Ireland, 5 from Scotland, 3 from Germany and 1 was of African ethnicity – most likely a former American refugee.⁵ The corpus sampler, however, also includes a number of first-generation immigrants (11) who had emigrated to Canada from either Europe (4 each from England and Scotland, and 1 each from Ireland and Wales) or the US (1 diarist). No census information is available for the remaining 5 diarists, although the archive describes them as of English (2), Scottish (2) and Dutch (1) ethnicity (cf. Table 1). While the contribution of United Empire Loyalists and late-Loyalists (see above) is more difficult to trace as there was no such thing as an American ethnicity, emigrants who settled in Ontario later on might provide a clearer indication of the import of the second (and third) wave of immigration on Canadian English.

As shown in Fig. 3, which plots the distribution of the diaries in space, most documents cover much of present-day rural Ontario. However, there are some notable peaks: 12 diaries were written in Wellington County, a predominantly rural county originally formed in 1837 and which is part of the Greater Golden Horseshoe, while 4 diaries each were written in Oxford County, which was established in 1798, and Leeds and Grenville County, originally two separate counties which were created by Governor John Graves Simcoe in 1792 in an area that had already been settled by United Empire Loyalists.⁶

Religion-wise, the overwhelming majority of the diarists may, quite unsurprisingly, be generally described as Protestants, an umbrella term that covers a very rich array of denominations, the most significant of which are Presbyterians (11 diarists), Methodists (10 diarists), Anglicans (9 diarists) and

⁵ Canada was one of the “promised lands” of the Underground Railroad, a complex, clandestine network of people and safe houses that helped people enslaved in Southern plantations to reach freedom in North America, cf. <https://www.thecanadianencyclopedia.ca/en/article/underground-railroad>.

⁶ Cf. <https://sites.rootsweb.com/~onleedsg/>.

Baptists (5 diarists), but which also include Quakers (3 diarists), Disciples (2), Universalists (1) and Congregationalists (1). The corpus sampler also includes 1 Roman Catholic of Scottish origin (cf. Table 1).

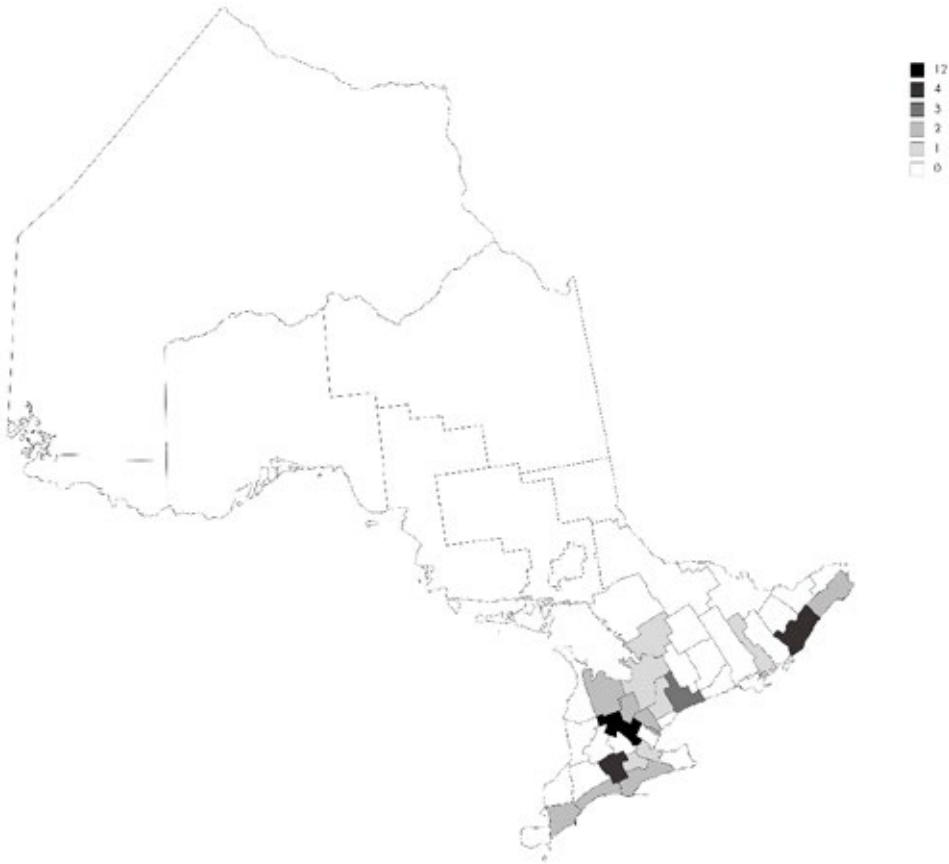


Figure 3. Distribution of the diaries in the different counties of Ontario

Although there are no explicit data on the level of education of the diarists, who were all at least literate enough to put pen to paper, some indications may be retrieved from their occupations. Quite predictably, the majority may be described as farmers (18) and farm women (15). However, while the farm was their primary occupation, some of them also covered other positions. The diarists thus also include 3 merchants, 2 newspaper men, 2 teachers, 1 apprentice surveyor, 1 carpenter, 1 marble cutter, 1 miller and distiller, 1 minister, and 1 politician, in addition to some members of their families (cf. Table 1). The corpus sampler therefore includes, at least in

theory, the diaries of both educated and, most importantly for the purposes of the project, minimally-schooled writers, whose language might more clearly reflect vernacular usage, thus allowing for diastatic investigations as well.

Finally, as shown in Table 2, the corpus includes both what may be described as personal diaries, which contain long entries focusing on the writer's personal thoughts and experiences, and other types of diaries, which may take the form of family books, account books and private chronicles, which generally record events of family and village life, weather reports and news, and which were also read and used by other members of the family (cf. Elspaß 2012: 163).

Table 2. Examples of diary entries included in the corpus

Examples of personal diary	Examples of family books / account books / private chronicles
<p>Mary came home from Nass. {Nassagawago} with her new teeth, they look just splendid better that ever expected, I picked ten qts of strawberries this afternoon, and put them down, I had no idea the little patch would do so well, sat up last night with Mrs McGregors Babies, twins, I think one of the little things will hardly live, John Hannah is real sick I am afraid Mrs. Hannah be worn out waiting on him, I am afraid the neighbours are not doing their duty. Mary had letter from Brittain. I think By the tone of it that he thinks I persuaded Anne not to go with him, I feel real hurt about it for I am sure I never said one word against him, but perhaps it will all come right sometime and if it never does I need not mind for God knows all things even the very inner most thoughts of our hearts – I went for a drive in the evening it was quite chilly. I don't think I was very good company, I felt so dull and my thoughts seemed so scattered [F. 1884-1885 – Hill, M. – 1885-07-10]</p>	<p>Cold East rain all day, don't think the rain ceased once Strong wind with it, finished the dress, also helped look over & can some gooseberries [F. 1888-1891 – Mott, P. – 1888-06-28]</p>
	<p>To swamp cutting wood. Went to Etta Ellerton's funeral. Bought a buggy pole from J. Hammond. [M. 1893-1896 (German) – Main, S. – 1894-01-22]</p>
	<p>Warm to day sowed our barley this afternoon William Morros stable caught fire but was put out [M. 1866-1867 – Beattie, W. – 1866-05-10]</p>
	<p>A lovely day, Sarah, Mary Ann, Carroll, and myself went out to Marcia's and spent the day. Bill was here for a few minutes, Andrew hoed the mangolds, Harold took Fannie and went out in the boat. [F. 1878 – Jones, F.A. – 1878-06-20]</p>

4.2 Corpus analysis

As spelling idiosyncrasies can reveal a lot about a language and its speakers, in what follows, a brief case study on orthography will be presented, in order to illustrate the different types of information that can be retrieved from the corpus, which in this case concern phonology, dialectology, pragmatics, and sociolinguistics.

Since Canadian English is generally described as a variety of American English, in order to retrieve idiosyncratic or marked spellings as compared to present-day norms, the corpus keywords were calculated using AmE06, a 1-million-word corpus of general American English from the Brown Family, as reference corpus. The keywords were then filtered according to their frequency in ascending order in the reference corpus, thus retrieving spellings that are only present in the target corpus and which, for this reason, may be considered typical of early vernacular Ontario English, the variety of English represented in the corpus sampler. The keywords list thus arranged is shown in Table 3.

As shown in Table 3, the majority of these (in normal type) may be described as simply erratic spellings (e.g. *nise*, *recieved*, *takeing*, *cloudy*, *kiled*), which are probably more indicative of the writers' level of education and literacy, rather than of their accent or orientation. Indeed, while certainly irregular by today's standards, these orthographies do not seem to reflect a different pronunciation of a specific word, but rather an attempt on the part of less educated and more insecure diarists to write on the basis of what we could describe as phonetic spelling (see Fairman 2000 and 2007).

On the other hand, some of the idiosyncratic orthographies included in Table 2 (highlighted in bold) appear in this list simply because they follow the British, rather than American, standard. While these do not necessarily give an indication of how the words were pronounced, such spellings are nonetheless extremely interesting, since they may be considered as indicative of the writers' background, schooling (cf. Gold 2004) and, possibly, of their orientation. Although connections between orthographical choices and ideology have been proven difficult to make (cf. Grue 2013), orientation represents a particularly significant concept with regard to Canadian English, which has frequently been described as alternatively attracted by either of the two poles: the US, where most Canadians actually came from and an influential neighbor in many respects; and Britain, the colonial administering power (Heffernan et al. 2010: 3; Dollinger 2012b: 451; Boberg 2016: 112; Dollinger 2019b: 56). Indeed, while most scholars agree

Table 3. Idiosyncratic or marked spellings as compared to present-day norms in CCanDL

WORD	FREQ (CCanDL)	FREQ (AmE06)	WORD	FREQ (CCanDL)	FREQ (AmE06)
ploug*	974	0	buisy	38	0
staid	595	0	makeing	37	0
nise	416	0	favour	37	0
<i>geting</i>	351	0	saviour	34	0
<i>cuting</i>	231	0	showry	33	0
waggon	204	0	cemetry	32	0
<i>metin</i>	174	0	<i>meatin</i>	31	0
recieved	93	0	<i>thrashin</i>	31	0
thrash	91	0	neighbour	30	0
threatning	89	0	<i>seting</i>	29	0
blustry	88	0	peices	27	0
choars	85	0	labour	26	0
harowed	77	0	parlour	26	0
<i>puting</i>	77	0	afternon	25	0
beaitful	76	0	calld	24	0
centre	72	0	dollers	24	0
takeing	71	0	choped	24	0
freinds	66	0	clening	22	0
grey	66	0	choping	22	0
<i>spliting</i>	65	0	borrowed	22	0
tok	64	0	finishd	22	0
stoped	62	0	burried	22	0
favourable	53	0	kiled	22	0
dissagreeable	49	0	togather	22	0
stopt	48	0	cheque	21	0
comeing	47	0	pullin	20	0
comenced	45	0	reference	19	0
comming	42	0	logged	19	0
profitable	42	0	privelege	19	0
peice	40	0	especialy	18	0
clowdy	39	0	sowd	18	0

on the American base of Canadian English, mostly as a result of the first wave of immigration into Canada (Dollinger 2008: 134, 2019b: 55), until the 1970s it was the British variety that carried prestige in official life (Boberg 2010: 34; Tagliamonte 2014: 201; Dollinger 2019b: 60), a phenomenon which is generally referred to as “Canadian Dainty” (cf. Chambers 2004). This seems to be confirmed by an analysis of the diarists’ origins. Indeed, of the 39 diarists who displayed a British orientation, 24 were born in Canada and only 4 in England (the place of birth of 3 diarists is unknown, 5 were born in Scotland, and 1 each in Ireland, the US and Wales). Moreover, 31 of them also made use of forms that by today’s standards are indexed as American, as shown in Examples 1 to 4, and, while showing preference for a particular spelling, at times they also spelled the same word following the other standard, as shown in Examples 5 to 8:

- (1) “I drove to Oshawa for a load of bran with the **wagon**” [M. 1899 – Geddes, M.D. – 1899-01-03 (Jan 3rd Tuesday)].
- (2) “Mr. Hudson rode down from Toronto yesterday with a **grey** mare he bought for us” [M. 1899 – Geddes, M.D. – 1899-03-21 (March 21)].
- (3) “Mr Ludy and Johnny worked at **parlor** and sitting room on Dec 8th” [F. 1901-1902 – Watson, J.].
- (4) “[...] placing them 16 in apart from **centre** to **centre**” [F. 1901-1902 – Watson, J.].
- (5) “A day or two after Mr Ingram, who has been our near **neighbor** ever since we came here, sent a lovely cup and saucer and plate” [F. 1877-1907 – Simpson, E. – 1893-01-30 (30)].
- (6) “Berrie and I went to call on our new **neighbour** Mr Carroll and had a very pleasand [sic.] time” [F. 1877-1907 – Simpson, E. – 1881-04-04 (4)].
- (7) “I finished **plowing** the south front field this forenoon and started to plow in the pea field” [M. 1896 – Sunter, W. – 1896-09-09 (SEPTEMBER WEDNESDAY 9 1896)].
- (8) “I have been **ploughing** today in the summer Fallow and got very well along” [M. 1857 – Sunter, W. – 1857-06-18 (THURDAY, June 18th, 1857)].

While American spellings (odd-numbered examples) generally prevail, most diarists alternate between the two conventions, not only in spelling different words, which may simply be the result of custom, rather than orientation, but also in spelling the same word, which, on the other hand, might be read as evidence of minimal schooling and of linguistic insecurity.

In addition to these first two categories, there are also a number of phonetic spellings in Table 2 (e.g. *blustry*, *clening*, *meatin*, *seting*) which are of particular interest from a dialectological point of view, since they seem to reproduce in writing the diarists' own pronunciation of these words. Of these, 9 involve vowel sounds (underlined in the table), while 8 involve consonants (in italics in the table). Although some of these phonetic spellings are found only in the diaries of one specific writer (e.g. "tok" in the diary of William Rea and "geting" in the diaries of Courtland Olds) and may thus be described as belonging to their idiolect, others are found in the works of several diarists, suggesting possible dialect and accent features that characterized this particular variety of English.

Among the most frequent phonetic spellings that involve vowels, we find the <thrash> spelling for "thresh" (and its inflections),⁷ which occurs 690 times in the corpus (98 instances of "thrash", 373 of "thrashing" and 219 of "thrashed"), against 482 occurrences of the <thresh> spelling (128 instances of "thresh", 294 of "threshing" and 60 of "threshed"). This might, again, be treated as a symptom of minimal schooling and literacy, as also evidenced by the fact that some diarists seem to be undecided about the two forms. However, this form appears in the diaries of 25 different writers (17 males and 8 females), most of whom had been born in Canada, even though they were of different ethnicities (7 each of English and Scottish, 5 of Irish, 2 of German, 1 each of Welsh and African origin, besides 2 whose ethnicity is unknown). Moreover, as shown in Table 4, it is found, albeit in different proportions, in all chronological subsections of the corpus, except for the pre-1850s one, which is also particularly unpopulated (only 12,066 tokens) and contains the diaries of only one writer. For these reasons, the <thrash> spelling might indicate that the word was pronounced with a lower /æ/ vowel, instead of the standard mid /ɛ/ vowel. While there are a number of possible alternative explanations, and further research is needed to either confirm or disprove this hypothesis, it may be read as evidence for

⁷ The <thrash> spelling is recorded in the *OED* as a now chiefly regional variant of "thresh" (first attested in 1364, https://www.oed.com/dictionary/thrash_v?tab=meaning_and_use#1344545110).

the beginning of the phenomenon commonly referred to as Canadian Shift (Dollinger 2008; Labov et al. 2006; Boberg 2019).

Table 4. Normalized frequency (10,000 words) of the two variants of “thresh” and its inflections in the corpus

	pre-1850s	1850s	1860s	1870s	1880s	1890s	1900s	1910s	WHOLE CORPUS
thrash*	0	32.4	6.5	2.1	3.8	3.6	7.1	1.2	5.7
thresh*	47.2	0.5	3.5	3.3	3.4	5.4	1.4	2.1	4

Another quite frequent and widespread pronunciation feature emerging from the corpus, evidenced by the <-ry> instead of the regular <-ery> spelling in such words as “showry” (showery), “blustry” (blustery) and “cemetry” (cemetery), appears to be schwa syncope, that is, the deletion of post-tonic /ə/, a phenomenon that is quite common in informal fast speech (Polgárdi 2015). Since these particular spellings emerge from the writings of 17 different diarists (11 males and 6 females) and, as shown in Table 5, are found in all decades covered by the corpus, with the exception of the very limited pre-1850s section, the phenomenon may be regarded not as part of the idiolect of one specific person, but as generally more widespread.

Table 5. Normalized frequency (10,000 words) of schwa syncope and retention in the corpus

	pre-1850s	1850s	1860s	1870s	1880s	1890s	1900s	1910s	WHOLE CORPUS
*ry	0	0.3	0.5	0.6	3.3	0.2	3.5	3	1.3
*ery	0	4.9	1.3	2	1.4	2.3	2.7	6.2	2.6

Although schwa syncope might in some cases also be indicative of Canadian Dainty (Chambers 2004) and may thus be described as characteristic of Canadian English, it is more probably related to the genre and register under investigation here, which, though written, display many features that are typical of texts of immediacy (Elspaß 2012: 157). As such, this finding may be more significant as evidence of the spontaneity and of the underlying oral nature of these specific texts, which seem to be particularly well-suited for the purposes of future investigations.

Phonetic spellings involving consonants, on the other hand, cluster in two groups. The first, evidenced by such spellings as “metin”, “meatin”

and “thrashin”, may be traced back to the very widespread phenomenon commonly (though improperly) referred to as g-dropping – that is, the pronunciation of final /ŋ/ as /n/, which in Late Modern times came to be associated with the lower social classes and lack of education (Beal 2004: 160–161). These spellings, however, are only found in the diaries of David Rea and his brother William, and may thus be considered characteristic of their own idiolects, rather than of Canadian English as a whole.

The second group consists of such spellings as “geting”, “cuting”, “spliting” and “seting”, which are used quite consistently by 4 different diarists. These may, once again, be considered as symptomatic of the limited level of education of the diarists – all of whom were farmers born in the first half of the nineteenth century. However, such spellings may also point to tapping or flapping – that is, the pronunciation of intervocalic /t/ as /ɾ/, thus suggesting an American orientation and the attempt, at least on the part of some diarists, to reproduce in writing what they probably perceived as a weaker sound.

5. Concluding remarks

Although confined to just one element – orthography – and carried out on a corpus sampler, the analysis has highlighted some features which may provide a better insight into the characteristics of early Canadian English, including further evidence of the so-called “dual-standard”, which has long been described as typical of this variety, and of possible early instances of the Canadian Shift. Moreover, as shown by the preliminary investigation, since the corpus focuses on prototypical texts of immediacy (Elspeß 2012: 157), where spontaneous and less controlled linguistic choices may have been more frequent, it may also help to shed light on the history of vernacular usage more in general. Furthermore, the corpus was also designed to complement and integrate itself with other similar corpora, including the *Corpus of Early Ontario English* (CONTE, Dollinger 2006), which includes letters, diaries and newspaper articles written in Ontario between 1776 and 1899, and the *Petworth Emigration to Canada Corpus* (PECC, Dollinger 2019c), which includes letters written by Southern English emigrants to Ontario between 1832 and 1842. Once completed, the *Corpus of Canadian English Letters and Diaries* may therefore represent a promising tool, as it is expected to make a significant contribution to our understanding of the characteristic features and diachronic (i.e., time-related), diatopic (i.e., regional) and diastratic (i.e., related to social class) developments of one of the major, albeit slightly under-researched, regional varieties of English.

REFERENCES

Primary sources

- Corpus of Canadian English Letters and Diaries (CCanDL)*. Compiled by G. Rovelli in collaboration with the *Rural Diary Archive* and the McLaughlin Library, University of Guelph, Ontario (CA).
- Oxford English Dictionary Online* (OED online)
<http://www.oed.com/>, accessed July 2023.

Secondary sources

- Allen, A.
 2014 "A non-standard standard? Exploring the evidence from 19th century vernacular letters and diaries". In: A. Auer – D. Schreier – R. J. Watts (eds.) *Letter Writing and Language Change*. Cambridge: Cambridge University Press, 202-222.
- Avis, W. S. – P. D. Drysdale – R. J. Gregg – M. H. Scargill (eds.)
 1967 *Dictionary of Canadian English: The Senior Dictionary*. Toronto: Gage.
- Avis, W. S. – C. Crate – P. Drysdale – D. Leechman – M. H. Scargill – C. L. Lovell (eds.)
 1967 *A Dictionary of Canadianisms on Historical Principles (DCHP-1)*. Toronto: Gage.
- Barber, K. (ed.)
 2004 *Canadian Oxford Dictionary* (2nd edn.). Toronto: Oxford University Press.
- Baskerville, P. A.
 2005 *Sites of Power. A Concise History of Ontario*. Oxford: Oxford University Press.
- Beal, J.
 2004 *English in Modern Times*. London: Hodder Education.
- Berger, C. M.
 2005 *The Dialect Topography of Canada: Method, Coverage, Interface and Analyses*. Dissertation (Master's). Vienna: University of Vienna.
- Bloomfield, M. W.
 1948 "Canadian English and its relation to eighteenth century American speech", *The Journal of English and Germanic Philology* 47 (1), 459-67.
- Boberg, C.
 2010 *The English Language in Canada. Status, History and Comparative Analysis*. Cambridge: Cambridge University Press.
 2012 "Standard Canadian English". In: R. Hickey (ed.) *Standards of English. Codified Varieties around the World*. Cambridge: Cambridge University Press, 159-178.

- 2016 "Newspaper dialectology: Harnessing the power of the mass media to study Canadian English", *American Speech* 91 (2), 109-138.
- 2017 "Archival data on earlier Canadian English". In: R. Hickey (ed.) *Listening to the Past: Audio Records of Accents of English*. Cambridge: Cambridge University Press, 375-394.
- 2019 "A closer look at the short front vowel shift in Canada", *Journal of English Linguistics* 47 (2), 91-119.
- Chambers, J. K.
- 1993 "'Lawless and vulgar innovations': Victorian views of Canadian English". In: S. Clarke (ed.) *Focus on Canada*. Amsterdam: John Benjamins, 1-26.
- 1998 "English: Canadian varieties". In: J. Edwards (ed.) *Language in Canada*. Cambridge: Cambridge University Press, 252-272.
- 2004 "'Canadian dainty': The rise and decline of Britishisms in Canada". In: R. Hickey (ed.) *Legacies of Colonial English. Studies in Transported Dialects*. Cambridge: Cambridge University Press, 224-241.
- 2006 "Canadian raising. Retrospect and prospect", *Canadian Journal of Linguistics* 51 (2-3), 105-118.
- Clarke, S. – F. Elms – A. Youssef
- 1995 "The third dialect of English: Some Canadian evidence", *Language Variation and Change* 7, 209-228.
- Dollinger, S.
- 2006 "Oh Canada! Towards the corpus of early Ontario English". In: A. Renouf – A. Kehoe (eds.) *The Changing Face of Corpus Linguistics*. Amsterdam and New York: Rodopi, 7-25.
- 2008 *New-Dialect Formation in Canada. Evidence from the English Modal Auxiliaries*. Amsterdam: John Benjamins.
- 2012a "Canadian English in a real-time perspective". In: A. Bergs – L. Brinton (eds.) *English Historical Linguistics. An International Handbook*. Berlin: De Gruyter Mouton, 1858-1879.
- 2012b "On the autonomy and homogeneity of Canadian English", *World Englishes* 31 (4), 449-466.
- 2014 "Emerging standards in the colonies: Variation and the Canadian letter writer". In: A. Auer – D. Schreier – R. J. Watts (eds.) *Letter Writing and Language Change*. Cambridge: Cambridge University Press, 101-113.
- 2015 "Canadian English: A conservative variety?", *Zeitschrift für Kanada-Studien* 35, 25-44.
- 2019a *Creating Canadian English. The Professor, the Mountaineer, and a National Variety of English*. Cambridge: Cambridge University Press.
- 2019b "English in Canada". In: C. L. Nelson – Z. G. Proshina – D. R. Davies (eds.) *The Handbook of World Englishes* (2nd edn.). Hoboken: Wiley Blackwell, 52-69.

- 2019c "‘I hope you will excuse my bad writing’. Shall vs. will in the 1830s Petworth Emigration to Canada Corpus (PECC)". In: R. Hickey (ed.) *Keeping in Touch. Emigrant Letters Across the English-Speaking World*. Amsterdam: John Benjamins, 43-66.
- Dossena, M.
2012 "The Study of correspondence: Theoretical and methodological issues". In: M. Dossena – G. Del Lungo Camiciotti (eds.) *Letter Writing in Late Modern Europe*. Amsterdam and Philadelphia: John Benjamins, 13-30.
- Edwards, J.
1998. *Language in Canada*. Cambridge: Cambridge University Press.
- Elspaß, S.
2012 "The use of private letters and diaries in linguistic investigation". In: J. M. Nandez-Campoy – J. C. Conde-Silvestre (eds.) *The Handbook of Historical Sociolinguistics*. Oxford: Wiley Blackwell, 156-169.
- Fairman, T.
2000 "English pauper letters 1800-34, and the English language". In: D. Barton – N. Hall (eds.) *Letter Writing as a Social Practice*. Amsterdam: John Benjamins, 63-82.
2007 "‘Lower-order’ letters, schooling and the English language, 1795 to 1834". In: S. Elspaß et al. (eds.) *Germanic Language Histories ‘from Below’ (1700-2000)*. Berlin: De Gruyter, 31-43.
- Fee, M. – J. McAlpine
1997 *Guide to Canadian English Usage*. Toronto: Oxford University Press.
- Gold, E.
2004 "Teachers, texts and early Canadian English 1791-1841". In: S. Burelle – S. Somesfalean (eds.) *Proceedings of the 2003 Annual Conference of the Canadian Linguistic Association*. <https://cla-acl.ca/actes/actes-2003-proceedings.html>, accessed December 2023.
- Grue, D.
2013 "Testing Canada’s ‘honour’: Does orthography index ideology?", *Strathy Student Working Papers on Canadian English*, 1-18. <https://qspace.library.queensu.ca/handle/1974/8649>, accessed December 2023.
- Heffernan, K. et al.
2010 "Preserving Canada’s ‘honour’. Ideology and diachronic change in Canadian spelling variants", *Written Language and Literacy* 13 (1), 1-23.
- Hickey, R.
2017 "Analysing early audio recordings". In: R. Hickey (ed.) *Listening to the Past: Audio Records of Accents of English*. Cambridge: Cambridge University Press, 1-12.
- Labov, W. – S. Ash – C. Boberg
2006 *The Atlas of North American English. Phonetics, Phonology and Sound Change*. Berlin: De Gruyter.

- Mackey, W. F.
1998 "The foundations". In: J. Edwards (ed.) *Language in Canada*. Cambridge: Cambridge University Press, 13-35.
- Polgárdi, K.
2015 "Syncope, syllabic consonant formation, and the distribution of stressed vowels in English", *Journal of Linguistics* 51 (2), 383-423.
- Schneider, E. W.
2007 *Postcolonial Englishes. Varieties Around the World*. Cambridge: Cambridge University Press.
- Tagliamonte, S. A.
2014 "System and society in the evolution of change: The view from Canada". In: E. Green – C. F. Meyer (eds.) *The Variability of Current World Englishes*. Berlin: De Gruyter, 199-233.
- Tieken-Boon van Ostade, I.
2005 "Eighteenth-century English letters: In search of the vernacular", *Linguistica e Filologia* 21, 113-146.
- Van der Wal, M. J. – G. Rutten
2013 *Touching the Past. Studies in the Historical Sociolinguistics of Ego-Documents*. Amsterdam and Philadelphia: John Benjamins.
- Whitcomb, E.
2007 *A Short History of Ontario*. Ottawa: From Sea to Sea Enterprises.

Address: GIULIA ROVELLI, Dipartimento di Lingue, Letterature e Culture Straniere, Università degli Studi di Bergamo. Piazza Rosate 2 – 24129 Bergamo, Italy.
ORCID code: <https://orcid.org/0000-0001-6282-4097>.

Acquiring epistolary literacy in nineteenth-century New England*

Ingrid Tieken-Boon van Ostade

University of Leiden

ABSTRACT

The concept of “epistolary literacy”, or the skill of letter writing, was first applied by Susan Whyman in *The Pen and the People* (2009) when analysing the letters of English families from the period 1660-1800. The present study applies the concept to nineteenth-century New England, and presents five case studies of families from different socio-economic backgrounds in order to study the question of how they acquired epistolary literacy. Since letter writing was not taught in schools, this paper investigates other means that were available to people eager to communicate with distant relatives and friends. Three possibilities are explored: the use of letter-writing manuals, the example of letters received, and practicing the skill through letter writing itself. The various skills identified in the letter collections, linguistic as well as epistolary, collocate with the

* This paper is based on a PhD project carried out by Bas van Elburg under my supervision at the Leiden University Centre for Linguistics between 2008 and 2021. The completed manuscript of the thesis did gain formal approval by the official examination committee instituted by the University of Leiden, but was never submitted for the purpose of obtaining the PhD degree. If it had been, it would have become available to the academic world as a published book. To ensure that the work done within the project should not be entirely wasted, I decided to present the main findings from van Elburg’s study in this article. In doing so I based myself on his selection, transcription and analysis of the epistolary material which he studied in the context of socio-economic and educational developments in nineteenth-century New England, but placed and interpreted his findings within the context of my own extensive research on Late Modern English letters and letter writing. At times this led to a slight rearrangement of his material into a perspective that was more fitting to a sociohistorical linguistic analysis than the approach originally taken by van Elburg. Van Elburg has granted me permission to make use of his work for publication purposes (personal communication by email in February 2022), and he read and commented on the paper before it was submitted for publication.

families' different socio-economic backgrounds. While the most highly placed family could draw on a teacher-caretaker's efforts to assist their children in communicating with their parents, the family at the lowest end of the social scale struggled with the need to acquire sufficient epistolary literacy to be able to stay in touch. One of the families, whose main proponent emigrated to Peru and lost his native language in the process, shows how their Spanish-speaking descendants used letter writing with American relatives both to acquire this skill, and to regain a command of English.

Keywords: epistolary literacy, letter writing, nineteenth-century New England, socio-economic stratification, letter-writing manuals, emigration, language loss.

1. Introduction

During the period covered by his book *Literary Publishing in America 1790-1850*, William Charvat notes, New England had "the most extraordinarily concentrated book-buying and reading public in the whole country" (1959: 30). At the time, according to Charvat, the area boasted many district schools and academies (1959: 32), which not only brought literacy to a great many people but also illustrates the importance New Englanders attached to education, their own and that of their children. This possibly accounts for the popularity of Noah Webster's *Grammatical Institute of the English Language* (Part II, 1784) around that time; the grammar was published in many editions and reprints throughout New England and adjacent states down to the early 1840s (Alston 1965; Tieken-Boon van Ostade 1996: 15).

While teaching literacy, or the ability to read and write, was the responsibility of elementary schools, in whatever form they took at the time (see e.g. Butts – Cremin 1953), acquiring epistolary literacy was a different matter. "Epistolary literacy" is a term coined by Susan Whyman in her book *The Pen and the People* (2009), which studies the development of the letter writing skills among members of lower and middle-class English families from the 1660s down to 1800 on the basis of unexplored letter collections from a variety of archives across the UK. Her book demonstrates that letter writing during this period was not a skill that was limited to the upper regions of society. Even parents from the lower social classes stimulated the acquisition of epistolary literacy among their own children as an important means to get on in society. Eventually, letter-writing manuals became available, but Whyman (2009: 30) shows that letter writing as a skill was largely passed on from parents to their children rather than acquired directly from books. This was also one of the conclusions drawn by Frances Austin (1973) on

the basis of her analysis of the Clift family correspondence, which covers the period 1792-1846 (Austin 1991), and my own study of the language of Jane Austen's letters similarly suggested that Jane Austen never had a letter-writing manual at her disposal: an inventory of her father's library did not include a single item of this popular text type (Tieken-Boon van Ostade 2014: 9). Like the working-class Clifts, Jane Austen acquired the skill of letter writing through extensive epistolary communication within her family, as well as from practice. Analysing the language of her letters I discovered for instance that her letter writing style changed over the years in subtle but linguistically important ways. She thus gradually learnt, for instance, that usage of flat adverbs, as in example (1), was less appropriate to an epistolary style, no matter how informal, than as a means to characterise colloquial language in her novels (Tieken-Boon van Ostade 2013):

- (1) we had an **exceeding** good ball last night (Jane Austen to her sister Cassandra, 9-10 January 1796; ed. Le Faye 2011; bold used for emphasis throughout this paper).

The present study addresses the question of how epistolary literacy evolved at a time and place different from those analysed and described by Whyman (2009). It does so by focusing on nineteenth-century New England, where acquiring a solid education was considered of great importance, as noted above. I will do so on the basis of an analysis of the letters of five families from the region. Letter writing, or indeed the teaching of composition writing in general, was not part of the school curriculum at the time, and only grew in importance once the pedagogical ideas of the Swiss educationist Johann Heinrich Pestalozzi (1746-1826) began to gain recognition in American schools; the subsequent increase in attention to writing instruction in the schools, according to Schultz (1999: 6), was a "tremendous pedagogical innovation". Pestalozzi's ideas spread from the 1830s onwards and eventually led to the implementation of common schooling – common, that is, in primarily addressing white children (Kaestle 1983: 176). Before the 1830s, the acquisition of epistolary literacy thus largely took place outside the schools, and the analysis presented here will illustrate several means by which this happened. While letters by ordinary people from the eighteenth century are not easily available according to Dierks (2009: xvi), the situation for the century following that period proved to be more favourable. Hewitt (2016) argues that the eighteenth century was "the golden age of the letter for Europe", and this was true for England as well (Tieken-Boon van Ostade 2009: 1), but in North America "letter writing first truly flourished" only a century later

(Hewitt 2016: 1). This is largely due to the fact that the Industrial Revolution started later in America than in England, so that significant infrastructural improvements which eventually facilitated better transportation by road, waterways and rail happened only later. These developments caused an increase in geographical mobility (Henkin 2006: 23), calling for a need to stay in touch by correspondence. During the same period, moreover, and profiting from better means of transport, the postal system greatly improved (Henkin 2006: 22), which consequently facilitated as well as encouraged letter writing as a social activity.

Among other major factors that encouraged letter writing during the nineteenth century by people from all walks of life there was the Civil War (1860-1865), when military service took many men away from their families. Many of their letters were preserved, as in the case of one member of one of the five families studied here, who served as a lieutenant in the Union Army, and for whom as many as thirty-three letters from his relatives have come down to us. The American whaling industry, which developed separately after a period of intense collaboration with the British (1817-1842) came to an end (Davis et al. 1997: Chapter 12), also it took men away from home, while trade between America and the former mother country sometimes caused entire families to spend large parts of their lives at sea; this indeed was the case with another family studied here. Communication with the home front was complicated because letters could only be dispatched when a ship was in port. And while nineteenth-century America witnessed the arrival of large numbers of immigrants from various European countries, such as the many families who left Ireland after the Great Famine (1845-1849), there were also Americans who emigrated themselves, either as adventure seekers or for economic reasons. Over the years, as we will see, such people kept in touch with their relatives at home, and in doing so, as Raymond Hickey puts it, they “took their native variety of English with them” (2019: 5). Once settled, however, and after growing families of their own, their native variety dwindled and was sometimes lost altogether. We will see an example of the desire among descendants to revive their parents’ former skills in English to be able to communicate with distant relatives in the US and even to explore the possibility of returning to their ancestor’s home country. This inevitably involved acquiring epistolary literacy *in English* as well.

The need to maintain contact with family and friends was thus a major motivation for people to engage in letter writing during the nineteenth century, and many correspondence collections have come down to us for analysis. The importance of studying emigrant letters has recently drawn

the attention of scholars of Late Modern English (see various papers in Hickey 2019), as such letters provide important insights into how the English language developed worldwide, and how other Englishes across the world emerged. All these letter writers needed to possess epistolary skills, which differed depending on when and where they had acquired such skills. The present paper deals with the question of how epistolary skills developed at a time and in a place – nineteenth-century New England – where such skills were not formally taught. Presenting five case studies, it focuses on a variety of American families that are all in one way or another representative of the socio-economic developments outlined in the previous paragraph. That these letters have survived is no coincidence, since they symbolised the tangible results of continuing relationships with distant relatives. But that these letters are available for analysis is something we owe to efforts like those of the Mystic Seaport Museum, situated in Mystic Seaport, a small village on the coast of Connecticut with a history of shipbuilding that goes back to the nineteenth century. Affiliated with the museum is the G. W. Blunt White Library (BWL), which houses a wealth of documents that are all related to the shipping industry.¹ All letters are reproduced on the BWL website in the form of digitised images, and are freely available for research.

The past twenty-five years has seen a surge of scholarly interest in letters and letter-writing from the Late Modern English period, particularly from a sociolinguistic perspective. Examples are the collection of papers by Marina Dossena and myself, called *Studies in Late Modern English Correspondence: Methodology and Data* (2008), Raymond Hickey's *Keeping in Touch: Emigrant Letters across the English-Speaking World* (2019) already mentioned, and various articles in the recently published *Intra-Writer Variation in Historical Sociolinguistics* (ed. by Markus Schiegg and Judith Huber, 2023). In my own work, I have particularly focused on individual letter writers like, most recently, Jane Austen (Tieken-Boon van Ostade 2014), while other scholars, like Anni Sairio (2009), have studied letter writers in the context of their social networks. In this light, it is of interest to mention the recently launched Mary Hamilton project,² which is already producing valuable results. All this type of research is being done within the context of what is known about letter writing practice from the period,

¹ See <https://www.mysticseaport.org/> and <https://librarytechnology.org/library/1297> for access to BWL.

² Made available online in 2022, it is called "Unlocking the Mary Hamilton Papers: A Window onto Eighteenth-Century Life, Literature and Language" (<https://www.maryhamiltonpapers.alc.manchester.ac.uk/>).

as analysed by Whyman (2009), but also before that, as discussed in the various contributions to the collection of studies published by David Barton and Nigel Hall, *Letter Writing as a Social Practice* (2000). Previously, there was the pioneering work by Frances Austin on the language and letter-writing practices of a single English working-class family from the late eighteenth and first half of the nineteenth century, the Clift family referred to above. The present paper aims to help widen the interest in the language of the period by demonstrating the sociolinguistic potential of hitherto unexplored resources held by such institutions as the Mystic Seaport Museum with its phenomenal G. W. Blunt White Library. The five case studies presented here will hopefully stimulate more research into the enormous array of private letters available there.

2. Selecting the letters

With BWL thus being a treasure trove for Late Modern English scholars, historians and historical sociolinguists alike, a selection was made of family letters that would be suitable for analysis. One New England family was suggested through Grover (2001), which deals with a number of Quaker families involved in the whaling business as well as the abolition movement; BWL contains a considerable number of private letters of one family, called Morgan. The Morgan collection includes in-letters, i.e. letters addressed to Charles W. Morgan (1796-1861) and his wife Sarah Rodman (1793-1888), as well as out-letters, letters addressed by the Morgans to a variety of correspondents.³ In the course of his life, Charles Morgan, according to the BWL website, became a wealthy businessman, and his life has been documented in studies of New England merchant families like Grover (2001).

To supplement accounts from relatively highly placed members of New England society, like the Morgan family, with those from letter writers further down the social scale, other letter collections were selected from the BWL website. The need to focus on a wider variety of letter writers is stressed by Marina Dossena and Gabriella Del Lungo Camiciotti in their book *Letter Writing in Late Modern Europe* (2012) for studies that aim to offer a more socially

³ For the terms “in-letters” and “out-letters”, as well as for an excellent introduction to the materialistic aspects of eighteenth-century letter-writing practice, see Baker (1980: 123).

inclusive view of developing epistolary traditions. Thus letter collections from two middle and two somewhat more lower-class families were selected from BWL: letters from Aaron H. Wood and his wife, the Fish Family Letters, the James W. Egleston Papers, and letters from William Douglas Goldsmith and relatives. The Wood family, whose history has been described extensively in Dooling (2014), has its origins in farming and the building trade; their letters are of interest here because of Aaron H. Wood's (1836-1895) ambition to have a career at sea, something he achieved largely through private study. The letters analysed are out-letters from Aaron and his wife Isabel Pearse (1842-1903) written during their seafaring trips to relatives at home. Members of the Fish family, like Nathan G. Fish (1804-1870) and his son Simeon (1837-1906), were connected with Mystic's shipbuilding industry (BWL website);⁴ only the letters addressed to Simeon were selected for analysis, most of them written by relatives while he was away from home. James W. Egleston (1816-1875), whose surname is occasionally spelled "Eggleston", likewise had his origins in the building trade (his father was a stone cutter). Leaving home at the age of sixteen, James spent a short time at sea, upon which he emigrated to Peru, where he settled and had a family. From Peru he kept in touch with his parents and siblings in Connecticut by letter, while his Spanish-speaking children and grandchildren later corresponded – in English – with their American relatives. Both sets of letters were analysed for this study. The Goldsmith letter collection in BWL is of interest because it includes letters written by William D. Goldsmith (b. 1824) to his sister Mary (1830-1914), whose education was of great concern to him, particularly her developing epistolary literacy so they would be able to communicate with each other directly. Though William and Mary, both of them orphans, were not from New England but from Louisiana, the letters, as their envelopes show, were sent to an address in Mystic, where Mary was living with her adoptive family. This explains why the letters eventually ended up in BWL. Like Aaron Wood, William Goldsmith was largely self-educated, though to a much lower degree.

Table 1 presents an overview of the selected letters, and attempts a general social classification of the five families concerned. This classification into differing socio-economic categories reflects the assumption that more affluent people would have greater means of educating their children before the days of compulsory education (cf. Carr 2003: 51), which might at

⁴ Efforts were made to ascertain the life dates of all letter writers and their families; occasionally, however, such biographical information could not be retrieved.

the same time have encouraged their letter writing activities. Before the rise of equal educational opportunities for all, according to Carr et al. (2005: 5), “[literacy often developed outside school, at work and at play, and at a range of other sites, including the family home, community settings, lyceums, and Sunday schools”]; this would have included acquiring epistolary literacy as well. As Hall (2000: 88) argues with reference to the situation in England, “the development of mass education”, alongside the effects of the Industrial Revolution, the growth of the railway system and cheaper postage, “contributed to letter writing becoming more widespread”. Much of this applies to New England as well. Evidence of the socio-economic status of some of the families is, however, largely circumstantial, and not always easy to interpret. Because the letters analysed are private, and more often than not communicate information *in medias res* rather than giving the type of full context that would benefit interpretation by outsiders, they typically lack information that would serve a more precise classification; this is particularly true for those families that have not been the subject of previous historical study, i.e. the Fish, the Egleston and the Goldsmith families. The attempted categorisation should therefore be seen as relative rather than absolute in the sense of the rather more rigid class division that characterised Late Modern England (cf. Garrard 2002).

Table 1. Letters analysed for the purpose of studying the acquisition of epistolary literacy

Letter collection	BWL source	Letters analysed	Dates	Type of letters	Social status
Charles W. Morgan	MS Coll. 27	55	1810-1866	in-letters	Higher
Aaron H. Wood	VFM 1496	26	1873-1895	out-letters	Mid
Fish family	MS Coll. 178	97	1849-1886	in-letters	Mid
James W. Egleston	MS Coll. 276	47 25	1832-1862 1862-1889	out-letters	Lower
William D. Gold-smith	VFM 1691	20	1838-1849	out-letters	Lower
<i>total</i>		270	1810-1895		

The classification in Table 1 was informed by the following considerations. Towards the end of his life, Charles W. Morgan bequeathed a thousand dollars to a Quaker Academy of which he was one of the founders (letter by Edmund Rodman and Thomas A. Green to Sarah Morgan, 15 September 1866). By modern standards, this was a considerable amount of

money,⁵ and by donating it to an educational institution he demonstrated the importance of acquiring a good education. What is more, he and his wife were able to afford employing a schoolteacher to look after their children during a prolonged period of absence from home (see Section 3.2). Both facts, moreover, demonstrate their affluent status in life. Aaron H. Wood and his wife and the Fish family were classified as occupying a middle position in society compared to that of the Morgan family on the one hand and those of James W. Egleston and William D. Goldsmith on the other. Aaron H. Wood, despite his family background (see above), married a schoolteacher and eventually became a ship-owner, which according to Dooling (2014: 125) earned him a relatively respected position in society. Nathan Fish, Simeon's father, had been a ship's captain, and later became a shipbuilder, while Simeon went to sea at the age of 19 or 20 and later enlisted in the Union Army during the Civil War. Evidence of these various career paths in Simeon's life transpires from the correspondence with his parents, as in the following passages from a letter by his mother Emeline (1805-1871) and one by his father written five years later:

- (2) If we had known what would be **the decision about the Ship** we should long ere had a box of goods making its passage towards you we have shirts and stockings all ready Your Pa says he can so direct them that you will receive them. (Emeline Fish to Simeon Fish, 29 January 1857).
- (3) I hope you ^{will} be able ^{to} fill your place with fidelity and be useful in whatever circumstances or condition you may be placed. at any rate look after **the men of your company** consider it your business to care for them conduct yourself with dignity & kindness towards them they ^{have} placed you in a respectable and responsible position & have a claim on your attention. (Nathan Fish to Simeon Fish, 18 November 1862).

The forms of address on the accompanying envelopes show that Simeon had been given the rank of lieutenant, even though he had no previous military experience. That he was not enlisted as a common soldier very likely reflects his middle-class background.

⁵ According to the CPI Inflation Calculator, the equivalent of this sum today would be \$32,853.30 (<https://www.in2013dollars.com/>, date of access 31 May 2022).

By comparison, James Egleston's early career as a sailor (see above) puts him into a social category different from that of Aaron H. Wood. His failure to make a fortune in Peru would not have earned him a higher position in American society either. His disappointment about his failed ambitions is evident from a letter to his father from 1845:

- (4) I am **disgusted with this Country** (although it is the place to make money) for Political Revolutions succeed one another the same as one year succeeds another (James Egleston to his father, 18 January 1845).

As for the Goldsmith family, much about their history is unfortunately unclear, which makes it difficult to classify them socially. William and his sister Mary never appear to have featured in any sociohistorical studies (unlike the Morgan and Fish families), so all we know about them and their position in society derives from their letters. After being orphaned, they appear to have been raised by different families, and while by 1854 Mary was married to a certain Nelson Lamb according to the last letter in the collection, her brother seems to have become a gold digger in California – not a social position that is comparable to those of any of the men described above:

- (5) You need not write to me as I shall leave Galveston in a few days for California a **"Gold Digging"** and I hope with the Blessing of God to return in a year or two if successful (William Goldsmith to his sister Mary, 29 January 1849).

Though with only twenty letters the Goldsmith collection is the smallest of the five case studies presented here, it is nevertheless of considerable interest with respect to the question of how epistolary literacy was acquired. Despite his lowly position in life, William strongly believed that his sister's developing epistolary literacy would be an important asset for her in life, and very likely it was.

As example (3) illustrates, all letters were transcribed as closely to their originals as possible, including superscripts and other self-corrections. Self-corrections are "the result of afterthoughts upon reading over the letter or of revisions during the writing process" (Tieken-Boon van Ostade 2014: 84), and their occurrence suggests care at producing a neat version of the letter. Another example is from a letter by Sarah Morgan to her husband Charles:

- (6) Benjⁿ [Benjamin] & Hannah ^{with their only son} had just got settled at housekeeping and were looking forward to a winter of comfort and

greater happiness than they had enjoyed for many years (Sarah Morgan to Charles Morgan, 23 November 1855).

Example (3) also shows that as late as the early 1860s, long <s> was still in use, something in which New England epistolary practice does not differ greatly from that of English letter writers (Fens-de Zeeuw – Straaijer 2012: 333). Emeline Fish, for instance, like her husband, still used long <s> during the early 1860s (*businefs*, *usefulnefs*, *kindnefs*). In 1855 Sarah Morgan, however, no longer used it (*happiness*) as example (6) shows; this may reflect on the different ways in which she and Nathan Fish and his wife had learnt to spell, even though they only differed some ten years in age. But Frederick Egleston, James's son, also still wrote *businefs* in a letter from 12 April 1866 to his aunt Ellen in Connecticut. Possibly, he copied the usage from the letters of his grandparents, with whom he corresponded. This shows that tracing the development of particular linguistic features is complicated when different generations of letter writers follow the example of older members of the family – willy-nilly in Frederick's case, due to his father's emigration and the family's resulting language loss (see below).

3. The letters analysed

As shown in Table 1, the letter collections selected comprise either in-letters (Morgan and Fish) or out-letters (Wood, Egleston and Goldsmith). The senders and recipients of the letters were either relatives or friends: selecting relatively close correspondents ensures access to the most informal letters available, which is important from a linguistic perspective (Tieken-Boon van Ostade 1987: 23-24). But such letters also provide the best possible material when studying how people learnt to write letters, especially when we are dealing with communication between parents and children, as we will see below. More formal letters will not yield a great deal of such private information. Accordingly, the letters analysed reflect the following relationships:

Morgan family – 55	From husband to wife (15) and wife to husband (16)
<i>In-letters</i>	From siblings or cousins (4)
	From their children (3)
	From a niece (1)
	From a nephew (2)
	From friends (14)

Wood collection – 26 <i>Out letters</i>	Husband (8) and wife (18) to husband's brother and his wife
Fish family – 97 <i>In-letters</i>	From Simeon's father (18) and mother (9) From Simeon's sisters (19) and brothers (46) From Simeon's cousin and a friend (1 each) From Simeon's daughters (3)
Egleston letters – 47 + 25 <i>Out-letters</i>	James to his parents (23) and sister Ellen (21) James to another sister and two brothers (1 each) To Ellen: James's children (13) and grandchildren (9) To Ellen: James's sister-in-law (1) and another relative (2)
Goldsmith letters – 20 <i>Out-letters</i>	William to his sister (17) Letters between relatives/friends (3)

Given the scope of this paper, only those letters that are relevant to the topic concerned will be cited.

All five letter collections include references to the postal system of the period as well as to other major socio-economic and political developments of the time, such as the Civil War, the educational system, the growth of the whaling industry, the effects of the ongoing Industrial Revolution and the emergence of a market economy. For this reason, those letters that have not yet been the object of sociohistorical study, the Fish, Egleston and Goldsmith letters, are worth further analysis, particularly since these families reflect the experiences of less highly placed people in society. Here, the interest lies in their contents only from the perspective of how the letter writers acquired the kind of epistolary literacy that enabled them to set up and maintain contact with distant relatives and friends. This might have come about as a result of explicit teaching in the schools, through access to letter-writing manuals, by being exposed to letters to and from family members, or, simply, through practice, something for which at times considerable encouragement was needed, as will be seen below. But as already mentioned, letter writing was not as a rule taught in school, and since we do not find any references to such instruction in the letters analysed, I will focus only on the role of letter-writing manuals in the acquisition of epistolary literacy (Section 3.1), and on the actual practice of letter writing within the families and in which certain family members played an important role (Section 3.2). The example of letters received within a family played a significant role in this process as well.

3.1 Letter writing and letter-writing manuals

Letter-writing manuals largely taught by example: they included practical information on how to address different kinds of correspondents, though many of them traditionally also included sections on English grammar (Fens-de Zeeuw 2008; Yáñez-Bouza – Rodríguez-Gil 2013: 145). Primarily though, they presented sample letters to show how, for instance, to write “A Letter from a Son to his Father”, “A Letter from a Youth at School to his Parents” or a letter from “A Brother to a Sister”. These examples are from *The Instructor; or Young Man’s Best Companion* (Fisher 1735?), a popular letter-writing manual that was originally published in England but that soon became available on the American market as well, where it appeared with the adjective *American* added to the title (Monaghan 2005: 391).⁶ Such sample letters could be copied with minimal changes by someone not yet experienced enough to compose letters of their own. The letter collections studied here do not include any references to letter-writing manuals, but there is evidence that some writers did not make use of them, while others do show a certain amount of familiarity with the material the manuals contained.

To begin with, the letters in the five collections are so personal that the possibility of borrowing entire letters from a letter-writing manual can be ruled out. What is more, quite a few letters are rather haphazardly punctuated, while several even lack punctuation altogether, so this reflects the practice of the writers themselves as well. Examples (2) and (3) from the Fish family collection illustrate this. Other examples are given below:

- (7) I wrote you soon after Mr Goldsmith Died he had nothing to leave me and I did not expect any thing for he done a greate deal for me (James Egleston to his sister Mary, 9 May 1846).
- (8) my love to David tell him to be a good boy and keep his nose clean tell him to write me a letter as I want to see if he knows how to write (James Egleston to his brother Joseph, 9 November 1849).
- (9) I have the pleasure to inform you of my good health allso pa and ma (William Goldsmith to his sister Mary, 13 December 1839).

⁶ The database Eighteenth Century Collections Online (ECCO) includes an American edition, already published in 1748 in Philadelphia. Fens-de Zeeuw (2008) provides a detailed comparative analysis of the contents of selected popular English and American letter-writing manuals from the Late Modern English period.

Capitalisation was likewise often irregular, as in (7) but also in (10). Verbs would not normally be capitalised at the time when capitalisation practice was at its height (see Osselton 1998).

- (10) Capt Morgan has gone to Jamaica. I **Expect** he will soon be back as he has been gone a long time (William Goldsmith to his sister Mary, 18 July 1840).

Exceptions are the Morgan letters and those of Aaron and Isabel Wood. The difference in attention paid to punctuation and capitalisation thus correlates with the letter writers' social background, since the Morgan family was more highly placed than the other families studied here. Isabel Wood, moreover, had been a teacher before her marriage, so she must have been familiar with the rules of punctuation professionally.

The spelling errors we find in James Egleston's letters, as in (11) and (12), may be explained similarly: going to sea at the age of sixteen, as mentioned in Section 2, would have caused a lack of interest in academic matters like correct spelling.

- (11) I have now by great casualty met a young man **beloning** to East Harfford Marten L. Roberts who says that he will see this forwarded which I trust he will as perhaps you have not recieved the others that I have wrote. If you should call upon him he can tell you all particulars respecting me as I have had a long conversation with him to that effect **sufice** to say that I am now after many ups and downs turned Sugar Planter. (James Egleston to his father, 18 January 1845).
- (12) I now take this **propper** Opportunity to write to you Concerning my affairs. I am now in good health I hope you the same. (James Egleston to his parents, 2 August 1833).

At the same time, the errors may have been due to haste, as indeed James writes in a letter to his parents:

- (13) you must excuse the mistakes my **being in Great Haste**. (James Egleston to his parents, 2 August 1833).

Other forms found represent contemporary spelling variants rather than mistakes (Osselton 1998), such as *allways*, found in letters by James Egleston as well as William Goldsmith:

- (14) and **allways** thought if the **all wise** disposer of events should permit us to see that time which thanks be to Him has come to pass. (James Egleston to his mother, 18 February, no year mentioned).
- (15) I am very glad to see the greate improvement you have made in writing. your writing is quite plain. I hope ^{you} will take pain and try to learn all you can as you never can learn too much, and it is **allways** of use to you" (William Goldsmith to his sister Mary, 17 February 1843).

The variant is only recorded in the *Oxford English Dictionary* (OED) down to the eighteenth century, but the context of (14) suggests what may have given rise to the spelling concerned. Neither James Egleston nor William Goldsmith appear to have had much formal education, but despite these unusual spellings their letters show that these men were far from illiterate. Awareness of the need to spell correctly may be seen in the following self-correction in a letter from the Fish family collection:

- (16) The^{ir} visit was town talk. Their mother could trust no one (Susan Fish to her brother Simeon, 31 December 1862).

Other evidence that confirms that the letters were from the hand of the authors themselves rather than being cloned from a letter-writing manual is the occurrence of instances of what in Standard English today would be considered faulty grammar, such as the use of *done* for *did*, *you was* and *have wrote*:

- (17) I wrote you soon after Mr Goldsmith Died he had nothing to leave me and I did not expect any thing for he **done** a greate deal for me (William Goldsmith to his sister Mary, 9 May 1846).
- (18) I was glad to hear that **you was** well and learning but am sorry to hear that you are by Captain morgan account very mischivious (William Goldsmith to his sister Mary, 22 October 1838).
- (19) It gave me great pleasure to hear that **you was** enjoying good health. (James Egleston to his parents, 10 February 1847).
- (20) I have almost given myself up to despair as to ever writing again to you as I **have wrote** almost innumerable letters to you and have never had the peasure of hearing from you nor any other of the Companions

of my Childhood and I think that this will be the Last time that I shall write to you (James Egleston to his father, 18 January 1845).

In the course of the English standardisation process, such grammatical features were relegated to dialect grammar, and they still characterise many non-standard varieties of American (and also British) English today (see Wolfram 2004: 144 on *done* as a past tense form and *you was*, and Tieken-Boon van Ostade – Kostadinova 2015 on participial forms like *have wrote*). Today, they receive considerable criticism when encountered in writing or even in relatively informal spoken English, and letter writing-manuals from the period would not have contained such forms. Actual errors are *greate* in (17), the absence of genitival marking in *Captain morgan account* and *mischivious* in (18), and *peasure* in (20) – compare James Egleston's use of *pleasure* in (19).

But there is some evidence showing that some of the letter writers may have had access to a letter-writing manual after all, either directly or indirectly, and this evidence derives from the use of letter writing formulas. Letter writing formulas, opening or closing ones, functioned as anchor points for less experienced, lower-class letter writers with the help of which they could start and finish their letters. Frances Austin (1973, 2004) analysed this phenomenon in her study of the letters of the Clift family, a poor, working-class family from the late eighteenth, first half of the nineteenth century that originated from Bodmin, Cornwall, in England. Though most members of this family were barely literate, they kept in touch by letter, and many of their letters have come down to us (Austin 1991). Austin identified a use of epistolary formulas that appeared to derive from sample letters in letter-writing manuals, and she discovered that the family did indeed possess a copy of one. One such formula is *I take this opportunity*, which is often used to start a letter, though with slight variations. Austin also found the formula in sailors' letters, so this may be how formulas like this crossed the Atlantic, if not directly through the publication of letter-writing manuals in nineteenth-century America.

The formula already occurred in example (12), in a letter by James Egleston to his parents, and other instances are:

- (21) **I now take the present opportunity** to inform you that I am well and hope you the same. (James Egleston to his father, 15 July 1835).
- (22) **I take the opportunity of a few leisure moments** (and how could I employ them better) to write to you by Mr Woods who returns to the USA by the ship Erie. (James Egleston to his father, 2 March 1848).

If we look at the dates of these letters, 1833, 1835 and 1848, we see from the expansion of the formula in (22) that in the course of time James learnt to express himself more freely; in other words, his epistolary literacy had grown in the intervening years. Ten years later, he dropped the formula altogether. *Writing a few lines* is another conventional opening formula according to Austin, and we find it not only in James Egleston's letters, but also in those of William Goldsmith:

- (23) I now sit down **to write you a few lines** to let you know that I am in good health except my leg which is some lame. (James Egleston to his parents, 1 October 1835).
- (24) **I write you a few lines** to let you know that I am well and hope this may find you and mrs Morgan and family all well (William Goldsmith to his sister, 29 January 1849).
- (25) **I write you a few lines** to let you know that I have not forgotten you and hope **these few lines** may find you in good health (William Goldsmith to his friend Ebenezer Morgan, 7 March 1844).

The formula is usually followed by the words *to let you know*, as in the above examples.

Examples (12), (21) and (23)-(25) continue with what Austin (1973: 328) calls "the health formula", which obviously contained essential information to be shared between distant correspondents, and another example from the Goldsmith letters is example (9) above. Its opening, "I have the pleasure to inform you", is unusually formal in a letter from a fifteen-year-old boy to a sister who was only nine at the time, so he must have picked it up somewhere. His reference to his parents as "pa and ma" in (9) is more in line with his age (and hers), as is his spelling of *allso* already discussed and his abrupt phrasing. The health formula is part of the letters of the more highly placed families in the collection as well, though their letters are usually preceded by a reference to having received the correspondent's letter:

- (26) **I have just received thy letter** and am glad to hear of thy welfare (Sarah Morgan to Charles Morgan, 18 June 1859).⁷

⁷ The Morgans were Quakers – hence the use of *thy* rather than *your* here (cf. Görlach 1991: 85 and Fens-de Zeeuw 2011: 103-108).

- (27) **I have just received a letter** from your Ma by [which?] I learn that you are all in tolerable health which I am glad to hear. (Nathan Fish to Simeon Fish, 18 November 1849).
- (28) **We have just received your good letter** (Hattie) of July 19th with five others (Isabel Wood to her brother- and sister-in-law, 5 August 1874).

It is a formula that James Egleston also eventually adopted, which confirms his increased letter-writing skills:

- (29) **I had the pleasure of recieving your letter** dated 18 Dec 45 and it gave me great satisfaction to hear from you all and that you was well. (James Egleston to his brother, 9 June 1846).⁸

Such increased epistolary confidence is also described by Frances Austin in her analysis of the letter writing habits of Julia Miles (1805-1852), the wife of the English poet William Barnes (1801-1886) (1990: 41).⁹

An opening formula that is more typical of less experienced letter writers according to Austin is the following phrase:

- (30) **I hold pen in hand** to write to you because I desire to inquire after your health. (Frederick Egleston to his grandmother, 8 December 1861).

Frederick was James Egleston's Peruvian son, born in 1842. He was Spanish-speaking, but the letter shows that he had learnt English as well – though not from his father, as he explained to his aunt Ellen:

- (31) In the letter above I have told you something about my improvements in English language Such a language as English is much necessary and it requires to be learned by practises. On that way I am learning it but still I subject myself to a good many trials. The first is keeping on writing and the second is to talk to English or American people nevertheless I believe that I want to spend a good time on learning it. I made a acquaintance with **a young fellow natural of Pennilvania** (United States) who got a good education in his own country and

⁸ Note the spelling of *receive* in this example: not uncommon at the time, but “now” nonstandard, according to the *OED*.

⁹ Analysing the language of the letters, Austin found that Julia Miles, though of higher social standing than her husband to be, was less linguistically literate.

whom I owe a good many favors and **taught me all which I know about English language** (Frederick Egleston to his aunt Ellen, 14 June 1862).

Because Frederick's mother didn't know any English, he was the one who kept in touch with his American relatives, informing them of the Egleston family's welfare:

- (32) Mother says that she wants to know you and as **it is difficult for her to write to you because she does either to write nor to read English** and that is the reason why she requests of me to tell you that she remembers you, your husband and children (Frederick Egleston to his aunt Ellen, 5 August 1862).

The formula in (30), according to Austin (1973: 325-326), does not have a letter-writing manual as its source but is typically found with members of the Clift family. As Austin argues, it was probably passed on from one generation to another. Perhaps Frederick copied it – as the self-correction in (30) suggests – from a letter he had received from his American grandmother.

Closing forms, too, are often formulaic, and we regularly find references to haste as an excuse for poor writing, as in example (13) – cf. Austin (1990: 40). The formula as such is not characteristic of the writers' social class:¹⁰

- (33) This very unintelligible letter is **written in great haste**. (Horace Fish to Simeon Fish, 18 June 1863).

Sometimes, the excuses give us little glimpses into the lives of the letter writers, as in:

- (34) **Benny stands right at my elbow** so that I cannot write. (Phebe Fish to her brother Simeon, 23 April 1863).
- (35) Hope you will excuse the many mistakes I make. **Oscar has so many questions to ask about the steamers** that I cannot have only a part of

¹⁰ It also occurs in an out-letter from the Morgan collection that was not included in the present analysis (this study only includes the Morgans' in-letters – see Table 1): **"In great haste** which must be my apology for this writing" (Sarah Morgan to her son Samuel, 27 January 1850).

my mind on my letter. (Isabel Wood to her brother- and sister-in-law, 21 September 1879).

Benny must have been Simeon's little nephew, and about Oscar – eleven at the time and home-schooled on board the ship during his parents' Transatlantic voyages – Aaron wrote three years later that he was "old enough to need school & more to do than has now" (Aaron Wood to his brother and sister-in-law, 19 September 1882). Eventually, Aaron and his wife gave up their seafaring life, possibly settling down in California. References like those in (34) and (35) testify to full epistolary literacy with these members of the Fish and Wood families, since the writers did not need to resort to age-old letter-writing formulas in their letters.

Asking to be remembered to friends and relatives at home was common among distant letter writers. Austin found such closing formulas in the Clift family correspondence (1973: 345), and they occur in the present letter collections, too, with spelling errors and all:

- (36) **Remember me to all my friends relations** and believe **your afftionate brother** James W. Egleston. (James Egleston to his brother, 9 June 1846).
- (37) **Remeber me to all those** that appear to have any anxiety of feeling for me (James Egleston to his mother, 31 May 1846).
- (38) Give my love to all the children and **remember me to Bridget** (Nathan Fish to his son Simeon, 29 May 1851).

3.2 Practicing letter writing

While it is impossible to prove how James Egleston acquired his letter writing skills, whether from a letter-writing manual or by learning from the letters he received, as his son Frederick did (see example [30]), it must have been through continued practice that his epistolary literacy evolved over the years, and that he learnt to express himself more freely. Some of the letter writers among the families studied here received active help and were encouraged to apply themselves to the job of writing a letter so that they would be able to keep in touch with their relatives. This appears from the letters of Charles and Sarah Morgan's eldest daughter, Emily (1821-1861), Melinna Fish (b. 1869), and Mary, William Goldsmith's little sister.

There are three letters in the Morgan family collection that were written by a certain A. E. Brastow, who appears to have been a teacher but who has

otherwise remained unidentified. Towards the end of 1828, Brastow looked after the Morgans' two children, seven-year-old Emily and her younger brother Samuel (b. 1824), while their parents were away in New York, and she kept the Morgans informed of their children's well-being by letter. Some of her letters were unfortunately lost in the post, a regular mishap in those days, as we learn from the complaints of several of the letter writers in the collections. As Brastow wrote regretfully:

- (39) I understand that **but one of my epistles have reached you**. I have written & sent **four letters** to you. splendid productions they were. They had this redeeming quality in the eyes of a fond & anxious mother & this makes me regret most sincerely that **they have not reached you**, they contain a most favorable & true account of your dear children. (A. E. Brastow to Sarah Morgan, 14 November 1828).

Two of her other letters that did come down to us, dated 24 October and 3 November 1828, included letters in Emily's own hand, which, Brastow insists, were "entirely original and written **with but little direction** of any kind" (24 October) and "written **almost entirely without aid** this afternoon" (3 November). The letters show that even at a very young age, Emily knew how to write a letter, what formulas to adopt, and that she had to use capitalisation and punctuation:

- (40) My dear mother I hope you will come home. **Are you all well** we are all very well (Emily Morgan to her mother Sarah, 3 November 1828).
- (41) **your afetionate child** Emily Morgan (Emily Morgan to her mother Sarah, 3 November 1828).¹¹

The contents of the letter are of a narrative nature that fits the girl's age: she informs her parents that "Miss Brastow has had the tooth ache because she went out without any thing on her head", reminds her parents to bring presents when they come home and tells them about visits she had made. Emily had evidently been well instructed in how to write a letter by her caretaker.

The need to practice letter writing and the realisation that epistolary skills will benefit from frequent practice are stressed throughout the Fish

¹¹ Note the spelling error in the second letter.

family correspondence. Father Nathan regularly admonished his children to engage in letter writing:

- (42) I don't hear particulars from [you?] unless I get a letter sometimes hear that you are well probably but I want Susan **to write occasionally** and John & Horace can write **it will help them to improve in writing**. (Nathan Fish to his son Simeon, 7 March 1850).
- (43) I want you to get in **the habit of writing frequently** it is quite an accomplishment to write letters with ease and facility I am now sorry that I practised so little when I was young practice alone makes perfect in anything. (Nathan Fish to his sons Simeon and John, 15 April 1852).

Simeon and John, the joint recipients of the letter in (43), were only fifteen and thirteen respectively at the time, but, as their father explained, being able to write a letter "with ease and facility" was considered "quite an accomplishment". Two years previously their older sister Susan (b. 1834) was already an accomplished letter writer, while ten-year-old Horace and John at the age of eleven still needed practice. Education was considered important in the Fish family, for the collection includes an undated letter from Nathan to his sons stressing the importance of acquiring "a good a knowledge of Latin [...] & French may be useful to you in business Geometry is certainly important astronomy and arithmetic also". To these subjects we may thus add letter writing as well. Simeon, apparently, passed on the need to acquire the skill of letter writing to his daughters Melinna (b. 1869) and Helen (b. 1877), for the Fish family collection includes letters which were written by these two girls at the ages of eleven and nine, respectively. Melinna's letter starts with acknowledging the receipt of her father's letter and apologises for not having written any earlier ("Dear Papa, I got your letter yesterday. I would have written before but did not have time"). She adds a last-minute message about the family's well-being:

- (44) I don't think of any thing else to write now. Love to all. **all well**. Linnie (Melinna Fish to her father Simeon Fish, 26 May 1881).

Helen's letter, dated 21 April 1886, was in Melinna's hand, but the collection also includes a letter in her own hand, written two years later when she was eleven, and thus more epistolary literate herself by that time. Within three generations, letter writing had become an established skill within the family, and the need to practice no longer needed to be explicitly stressed, as Nathan had done earlier.

Reconstructing the family history of William and Mary Goldsmith is complicated because of their different adoption histories. The earliest letter in the collection dates from 1838, when William was only fourteen and his sister eight. He urges her to keep in touch by asking her adoptive mother, a Mrs Morgan, a captain's wife (not related to the Morgan family in this study), to write a letter for her:

- (45) I have not heard from you for a long time and would be very much pleased to have a letter, **request Mrs Morgan to write**, and let me know how you are, and how you like mystic and the Family you live with and if you go to school and gain all the information you can (William Goldsmith to his sister Mary, 5 April 1838).

Five years later, Mary was thirteen and able to write to him herself:

- (46) I am very glad to see the **great improvement you have made in writing**. your writing is quite plain. I hope ^{you} will take pain and try to learn all you can as you never can learn too much, and it is allways of use to you (William Goldsmith to his sister Mary, 17 February 1843).

But though Mary's epistolary skills evidently developed to such an extent that brother and sister were able to communicate independently, contact remained irregular, as the following letter shows:

- (47) I should like to Know **why you have not wrote to me**, you mite ask me the same question, **the Reason I have not wrote to you** is that I have been waiting an answer to my last letter hope you have not forgot your Brother, no that canot be. (William Goldsmith to his sister Mary, 23 December 1845).

Unfortunately, none of Mary's letters appear to have come down to us, so it is impossible to assess whether her linguistic skills had become any better than her brother's, and whether their different social situations as they developed over the years might have been the reason for their eventual loss of contact. William's last letter to her dates from 1849, though the last letter in the Goldsmith collection, addressed to Mary's husband Nelson Lamb upon a request for information about her brother, described William as being "in good health and doing well" a year previously (William Hendley to Nelson Lamb, 23 May 1854). William was 29 at the time.

If the case of Melinna Fish discussed above illustrates the importance which letter writing had acquired in the course of three generations of the

Fish family, what we see in the Egleston family is even more striking. There were four generations of Eglestons who, after James's emigration to Peru, continued to communicate with each other. That they did so in English is not self-evident, for after settling down in his new country and starting a family, James claims to have given up his native language:

- (48) I must here tell you ^{the} same as all the rest that is that **scarcely or never writing English** except makeing Poetry and at that I am verry deficient, but **in the Spanish Language I verry often write** for a Public paper that is printed in Lima (James Egleston to his sister Henrietta, 31 May 1846).

Eventually, he even adopted a Spanish name, Santiago Flores. From the letter quoted in example (31), it appears that James's son Frederick, whose mother tongue was Spanish, no longer spoke fluent English, but that he had learnt the language from a friend. In the same passage, he explains that for him letter writing served as a means to practice his English. By writing letters to his relatives in Connecticut, he was thus not only developing his epistolary literacy, but it was epistolary literacy *in English* that he was seeking to acquire. Frederick's daughter Henrietta (b. 1870) continued the family tradition by writing to her great-aunt Ellen in Connecticut, who had been the recipient of her father's letters as well. Henrietta, too, wrote in English. One of her letters in the collection, written when she was only fourteen, shows her already as an experienced letter writer. Later letters suggest that she had fallen out with her parents, and at the age of nineteen she expressed her wish to go to America, invoking her great-aunt's assistance in that project. If she had managed to do so, it would have been thanks to the continuing importance attached to letter writing within four generations of the Egleston family that ties with the former mother country could be kept alive, combined with the determination of James's children and grandchildren to acquire epistolary literacy in the native language of their ancestors.

For the Wood family, the situation was different from that of the members of the three families discussed here since there was no need for Aaron or Isabel to improve their epistolary skills through practice; their letter-writing skills were already sufficiently established (like those of the Morgan parents). As a former schoolteacher (Dooling 2014: 122), Isabel was fairly well educated. Moreover, letter writing was an important means for her to pass the time while she and her husband were at sea. At times, as when her husband suffered from an eye problem, she even wrote letters for him in his name:

- (49) I am to **write his business letters** to day (dictated by him of course)
(Isabel Wood to her brother- and sister-in-law, 27 May 1874).

Isabel's letters are of additional significance because they illustrate what life on board a ship was like for a New England family and the role that letter writing played in their seafaring lives. This was why she urged her brother- and sister-in-law to continue writing to them, even if only a few lines at a time:

- (50) I really thought that for some reason unknown to us you did not want to write. I do not doubt that you are real busy and in such cases **four lines would satisfy us**. just tell us how you all are and why it is inconvenient for you to write then we should feel satisfied for the time and hope for more when you were more at leisure. (Isabel Wood to her brother- and sister-in-law, 8 February 1874).

It also explains why the letters were preserved:

- (51) All of yours that we rec'd in S.F. [San Francisco] and many of our other letters [...] we quite **enjoy reading them over at sea** (Isabel Wood to brother- and sister-in-law, 12 April 1874).

Letters were not only written to maintain family relations, but also formed important reading material to while away the time.

4. Conclusion

As case studies, the five sets of family correspondence analysed in this paper all illustrate different ways in which epistolary literacy was acquired – and maintained – at the time. What is more, these differences correlate with the families' social positions in life. For Charles Morgan and his wife, the most highly placed family in this study, it was self-evident that they were kept informed by letter by the woman they had employed to look after their children while they were away from home. This woman, a schoolteacher, encouraged their daughter, who was only eight at the time, to write to her parents herself, thus passing on the notion to her ward that letter writing was an important skill to have. Aaron Wood's wife was also a schoolteacher, and for her and her husband letter writing was a common activity, serving to pass the time on board ship as well as to communicate with relatives

and to conduct business; for their son Oscar, travelling with his parents as a child, epistolary literacy would consequently have been a natural part of his life. Simeon Fish's father emphasised the importance of particular school subjects in his correspondence with his sons, stressing at the same time that being able to communicate by letter represented a valuable accomplishment. Within the Fish family, this accomplishment was passed on from parents to children and grandchildren in turn, something which we also see within the Egleston family, for whom we have evidence of four generations of letter writers. For the Egleston children and grandchildren in Peru, moreover, letter writing also served as an important means to practice their English linguistic skills.

For William Goldsmith, a truly heart-breaking case, letter writing was the only means by which he could try and keep the tie with his far-away sister alive. His attempts in this respect date from when he was only fifteen, and were eventually rewarded by communications from his sister in her own hand. Both thus acquired the necessary epistolary literacy to keep their relationship going, though not for very long. Less than fifteen years later, all contact appears to have ceased. William's letters, as they have come down to us, and along with James Egleston's letters, also show the greatest amount of linguistic insecurity. In William's case, this appears to have been due to his lack of formal schooling, and in that of James, who was probably somewhat better educated, to the fact that he had emigrated to Peru and lived in a Spanish-speaking environment. James's children and grandchildren had to learn their father's language from others in order to be able to communicate with their American relatives, and, as his granddaughter Henrietta's letters show, to try and look for a better future in her grandfather's native country. It was her acquired epistolary literacy that enabled her to make this attempt. Both James Egleston and William Goldsmith, as well as James's son Frederick, made the greatest use of epistolary formulas, usage that can be traced to the influence of letter-writing manuals. Whether they actually drew on such manuals to develop their epistolary literacy is impossible to prove; they might equally well have picked up the formulas from letters they received. This is as far as the interpretation of the evidence presented by the five letter collections goes. For all that, it is clear that having or indeed acquiring epistolary literacy was regarded as an important skill among nineteenth-century New Englanders. The five families represent much larger epistolary networks than those consisting of the relatives and friends referred to in this study alone, so it is of considerable interest to know that there is a great deal more material to be analysed where these letter collections came from.

REFERENCES

Primary sources

Austin, F. (ed.)

- 1991 *The Clift Family Correspondence*. Sheffield: CECTAL.
The G. W. Blunt White Library (BWL) collections
https://research.mysticseaport.org/mss/?post_type=mystic_collections, accessed December 2023.

Eighteenth Century Collections Online (ECCO)

<http://www.gale.com/EighteenthCentury/>, accessed January 2023.

Fisher, G.

- 1735? *The Instructor: Or, Young Man's Best Companion*. London:
 A. Bettesworth, C. Hitch, R. Ware, J. Clark, S. Birt, S. Birt, J. Hodges.

Le Faye, D. (ed.)

- 2011 *Jane Austen's Letters* (4th edn.). Oxford: Oxford University Press.

Oxford English Dictionary (OED online)

www.oed.com, accessed January 2023.

Webster, N.

- 1784 *Grammatical Institute of the English Language*, Part II. Hartford: Hudson & Goodwin.

Secondary sources

Alston, R. C.

- 1965 *A Bibliography of the English Language from the Invention of Printing to the Year 1800*. Vol. 1. Leeds: Arnold and Son.

Austin, F.

- 1998[1973] "Epistolary conventions in the Clift family correspondence".
 In: M. Rydén – I. Tieken-Boon van Ostade – M. Kytö (eds.) *A Reader in Early Modern English*. Frankfurt am Main: Peter Lang, 319-347.
- 1990 "Text as social indicator: The letters of Julia Miles Barnes". In: J. Fisiak (ed.) *Historical Linguistics and Philology*. Berlin: De Gruyter Mouton. 29-43.
- 2004 "Heaving this Importunity: The survival of opening formulas in letters in the eighteenth and nineteenth centuries". *Historical Sociolinguistics and Sociohistorical Linguistics* 4. http://www.let.leidenuniv.nl/hsl_shl/heaving_this_importunity.htm, accessed 26 May 2022.

Baker, F. (ed.)

- 1980 *The Works of John Wesley*. Vol. 25: *Letters, I (1721-1739)*. Oxford: Clarendon.

Barton, D. – N. Hall (eds.)

- 2000 *Letter Writing as a Social Practice*. Amsterdam: John Benjamins.

- Butts, R. F. – L. A. Cremin
1953 *A History of Education in American Culture*. New York: Henry Holt. & Co.
- Carr, J. F.
2003 "Nineteenth-century girls and literacy". In: J. Greer (ed.) *Girls and Literacy in America: Historical Perspectives to the Present*. Santa Barbara, CA/Denver, CO/Oxford, England: ABC CLIO, 51-77.
- Carr, J. F. – S. L. Carr – L. M. Schultz
2005 *Archives of Instruction: Nineteenth-century Rhetorics, Readers, and Composition Books in the United States*. Carbondale: Southern Illinois University Press.
- Charvat, W.
1959 *Literary Publishing in America, 1790-1850*. Philadelphia: University of Pennsylvania Press.
- Davis, L. E. – R. E. Gallman – K. Gleiter
1997 *In Pursuit of Leviathan: Technology, Institutions, Productivity, and Profits in American Whaling, 1816-1906*. Chicago: University of Chicago Press.
- Dierks, K.
2009 *In My Power: Letter Writing and Communications in Early America*. Philadelphia: University of Pennsylvania Press.
- Dooling, M. C.
2014 *Seaworthy Timber, The Life & Times of New England Sea Captain Aaron H. Wood*. Carrollton: The Carrollton Press.
- Dossena, M. – G. Del Lungo Camiciotti (eds.)
2012 *Letter Writing in Late Modern Europe*. Amsterdam: John Benjamins.
- Dossena, M. – I. Tieken-Boon van Ostade (eds.)
2008 *Studies in Late Modern English Correspondence: Methodology and Data*. Bern: Peter Lang.
- Fens-de Zeeuw, L.
2008 "The letter-writing manual in the eighteenth and nineteenth centuries: From polite to practical". In: M. Dossena – I. Tieken-Boon van Ostade (eds.) *Studies in Late Modern English Correspondence: Methodology and Data*. Bern: Peter Lang, 163-192.
2011 *Lindley Murray (1745-1826), Quaker and Grammarian*. Utrecht: The Netherlands: LOT
- Fens-de Zeeuw, L. – R. Straaijer
2012 "Long-s in Late Modern English manuscripts", *English Language and Linguistics* 16 (2), 319-338.
- Garrard, J.
2002 *Democratisation in Britain: Elites, Civil Society, and Reform since 1800*. Houndmills: Palgrave Macmillan.
- Görlach, M.
1991 *Introduction to Early Modern English*. Cambridge: Cambridge University Press.

- Grover, K.
2001 *The Fugitive's Gibraltar: Escaping Slaves and Abolitionism in New Bedford, Massachusetts*. Amherst and Boston: University of Massachusetts Press.
- Hall, N.
2000 "The materiality of letter writing: A nineteenth-century perspective". In: D. Barton – N. Hall (eds.) *Letter Writing as a Social Practice*. Amsterdam: John Benjamins, 83-108.
- Henkin, D. M.
2006 *The Postal Age: The Emergence of Modern Communications in Nineteenth-century America*. Chicago: The University of Chicago Press.
- Hewitt, E.
2016 "Prologue: Networks of nineteenth-century letter-writing". In: C.-M. Bernier – J. Newman – M. Pethers (eds.) *The Edinburgh Companion to Nineteenth-Century American Letters and Letter-Writing*. Edinburgh: Edinburgh University Press, 1-10.
- Hickey, R. (ed.)
2019 *Keeping in Touch: Emigrant Letters across the English-Speaking World*. Amsterdam: John Benjamins.
- Kaestle, C. F.
1983 *Pillars of the Republic: Common Schools and American Society 1780-1860*. New York: Hill and Wang.
- Monaghan, J. E.
2005 *Learning to Read and Write in Colonial America*. Amherst and Boston: The University of Massachusetts Press.
- Osselton, N. E.
1998[1984] "Informal spelling systems in Early Modern English: 1500-1800". In: M. Rydén – I. Tieken-Boon van Ostade – M. Kytö (eds.) *A Reader in Early Modern English*. Frankfurt a. M.: Peter Lang, 33-45.
- Sairio, A.
2009 *Language and Letters of the Bluestocking Network. Sociolinguistic Issues in Eighteenth-century Epistolary English*. Helsinki: Société Néophilologique.
- Schieg, M. – J. Huber (eds.)
2023 *Intra-Writer Variation in Historical Sociolinguistics*. Oxford: Peter Lang.
- Schultz, L. M.
1999 *The Young Composers: Composition's Beginnings in Nineteenth-Century Schools*. Carbondale: Southern Illinois University Press.
- Tieken-Boon van Ostade, I.
1987 *The Auxiliary Do in Eighteenth-Century English: A Sociohistorical Linguistic Approach*. Dordrecht: Foris.
1996 "Two hundred years of Lindley Murray: An introduction". In: I. Tieken-Boon van Ostade (ed.) *Two Hundred Years of Lindley Murray*. Münster: Nodus, 9-25.

- 2009 *An Introduction to Late Modern English*. Edinburgh: Edinburgh University Press.
- 2013 "Flat adverbs and Jane Austen's letters". In: M. van der Wal – G. Rutten (eds.) *Touching the Past. Studies in the Historical Sociolinguistics of Ego-Documents*. Amsterdam: John Benjamins, 91-106.
- 2014 *In Search of Jane Austen: The Language of the Letters*. Oxford: Oxford University Press.
- Tieken-Boon van Ostade, I. – V. Kostadinova
- 2015 "Have Went – An American usage problem", *English Language & Linguistics* 19 (2), 293-312.
- Whyman, S. E.
- 2009 *The Pen and the People: English Letter Writers 1660-1800*. Oxford: Oxford University Press.
- Wolfram, W.
- 2004 "Rural and ethnic varieties in the Southeast: Morphology and syntax". In: K. Burridge et al. (eds.) *A Handbook of Varieties of English*. Vol. 2. *Morphology and Syntax*. Berlin: De Gruyter Mouton, 281-302.
- Yáñez-Bouza, N. – M. E. Rodríguez-Gil
- 2013 "The ECEG database", *Transactions of the Philological Society* 111 (2), 143-164.

Address: INGRID TIEKEN-BOON VAN OSTADE, Faculty of Humanities, Leiden University
Centre for Linguistics, Postbus 9515 – 2300 RA Leiden, The Netherlands.

I.M.Tieken@hum.leidenuniv.nl

ORCID code: <https://orcid.org/0000-0003-4362-6188>.

REVIEWS

**Michael Skiba, *Participial Prepositions and Conjunctions in the History of English*, Munich: utzverlag, 2021, 235 pp.
(Reviewed by RAFAŁ MOLEŃCKI, University of Silesia
in Katowice, Poland)**

The subject of medieval and modern English participial prepositions and conjunctions has not hitherto received much attention on the part of historical linguists, so Michael Skiba's study is a welcome contribution to this largely neglected field. The book is a modified version of the author's doctoral dissertation of 2019 successfully defended at Ludwig-Maximilians-Universität in Munich.

The participial prepositions and conjunctions (e.g. *including, considering/considered, providing/provided (that), seen/seeing (that)*) are a relatively recent phenomenon in English. However, some of them, in particular *during, except, past*, are so common and well-established as grammatical words that most native speakers do not realize their deverbal origin. In the case of *during* there is no synchronic basis for treating the word as a participle after the verb *dure(n)* became obsolete in Early Modern English.

Michael Skiba rightly associates the appearance of participial prepositions and conjunctions with the language contact situation between English, French and Latin in late medieval England. He provides convincing arguments and evidence showing that the words had been originally grammaticalized in Old French (both continental and Anglo-Norman) with some symptoms of the new status found already in Latin. Then bilingual French and English speakers, and indeed writers, simply copied the Romance forms and uses first into their Middle English legalese and other educated varieties from the fourteenth century onwards. This paved the ground for the subsequent coining of new grammatical words also from native Germanic elements: some of them already calqued according to the Romance pattern in Late Middle English, e.g. *outtaking, outtaken, (not) withstanding*. And the process is still productive in Present-day English, as the author shows in Chapter Seven in the discussion of the new items which arose in the 20th century. The process appears to have become a consequence of the typological change from synthetic to analytic, which

affected both Romance and Germanic languages. The external factor was the increasing number of argumentative texts written in Late Middle and Early Modern English, which required more and more precise exponents of abstract relationships (cf. Kortmann 2012).

The language material for Skiba's study comes from both modern and diachronic corpora of English checked against the definitions and examples found in historical dictionaries. The first chapters introduce the theoretical concepts and the methodology of the study. The author goes back to ancient and medieval grammar books in order to show how the categories of prepositions, conjunctions and participles were presented there and later points to the difficulties of their identification, categorization and classification in modern linguistic research. In Chapter Three the status of these categories is discussed in detail in old and modern Germanic and Romance languages against the Indo-European background with special attention paid to the participial constructions in Old English and Early Middle English and also in Old French and Anglo-Norman. The author shows that the main source for the prepositional uses of participles were absolute constructions found in classical languages.

The most interesting data and the author's original analyses are found in Chapters Five and Six. The general tendencies are discussed in Chapter Five on the basis of a relatively limited Helsinki Corpus: we find here quantitative analyses presented in numerous graphs and charts. The results are then verified for each item by a detailed survey of the data, including the first attestations, from the lexicographic databases, especially the *Oxford English Dictionary* and the *Middle English Dictionary*, and historical corpora such as the *Corpus of Middle English Prose and Verse*, the *Parsed Corpus of Early English Correspondence* and the *Corpus of Late Modern English Texts*. The discussion is lavishly illustrated with numerous examples showing how the participial prepositions and conjunctions functioned in the language.

The author has made references to most of the relevant and significant works published on the subject. His critical overview of the state-of-the-art literature in the first chapters of the book is most pertinent and exhaustive. For example, Skiba (pp. 26-28) does not hesitate to question the validity of the classifications offered in some standard reference grammars of English (Huddleston – Pullum 2002) and academic manuals of linguistics (e.g. Herbst – Schüller 2008). To the author's extensive bibliographical list one could add at least two more studies written by the present reviewer (Molencki 2011, 2012), where some of the participial prepositions and conjunctions are discussed. Another thing that might facilitate the reader in interpreting the results of the author's research would be including a subject and author

index at the end of the book, which could be done by means of common software these days.

As I said at the beginning of this review, the book by Michael Skiba is a very solid corpus-based study of an emerging new type of grammatical words in Middle and Early Modern English. Particularly useful are the summary tables illustrating the first occurrences of participial prepositions and conjunctions in the history of English. I believe that this valuable work will certainly become an important point of reference for future studies in the history of English, language contact, word formation and grammaticalization. As signalled by the author in Chapter Seven, where recent developments are taken account of, Present-day English participles still have the potential to grammaticalize into prepositions and conjunctions. Following this, the methods of research developed by Michael Skiba may prove valid in future studies of new participial prepositions and conjunctions.

REFERENCES

- Herbst, Th. – S. Schüller
 2008 *Introduction to Syntactic Analysis. A Valency Approach*. Tübingen: Narr.
- Huddleston, R – G. Pullum
 2002 *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.
- Kortmann, B.
 2012 *Adverbial Subordination. A Typology and History of Adverbial Subordinators Based on European Languages*. Berlin/Boston: De Gruyter.
- Molencki, R.
 2011 "New prepositions and subordinating conjunctions of Romance origin in Middle English". In: J. Fisiak – M. Bator (eds.) *Foreign Influences on Medieval English*. Frankfurt am Main-New York: Peter Lang. 9-24.
 2012. "Providing/provided that: Grammaticalization or loan translation?". In: I. Hegedűs – A. Fodor (eds.) *English Historical Linguistics 2010. Selected Papers from The Sixteenth International Conference on English Historical Linguistics (ICEHL 16), Pécs, 22-27 August 2010*. Amsterdam-Philadelphia: John Benjamins. 197-214. <https://doi.org/10.1075/cilt.325.09mol>.

Address: RAFAŁ MOLENCI, Institute of Linguistics, University of Silesia in Katowice, ul. Gen. St. Grot-Roweckiego 5, 41-205 Sosnowiec, Poland.

ORCID code: <http://orcid.org/0000-0001-8400-1577>.



Printed and bound by
Jan Kochanowski University of Kielce Press.