

*Introduction: Different Ways of Saying
Different Things*
Non-Canonical Syntax in Registers of English

Heidrun Dorgeloh and Anja Wanner

7.1 Introduction

Research on registers has to account for two kinds of variation, register-based variation and text-internal variation. In this chapter, non-canonical constructions will be dealt with, first, from the point of view of register-based variation and, second, as text-internal variation due to language processing. Speakers can switch into and out of registers (Schilling-Estes 2004: 375; Dorgeloh & Wanner 2010: 7) depending on the situation of use, whereas choices within texts tend to be motivated by the ways these are structured and the material is presented, that is, by conditions of information structure, complexity, or weight. Choices on structure and form have an impact on language processing, which means that register variation also has a psycholinguistic dimension.

Textual variation poses a challenge for the assumption of equivalence in meaning, as posited in the definition of (non-)canonicity in the Introduction to this volume. While *interspeaker*, such as social or regional, variation at least in principle relates to alternative ways of expressing ‘the same thing’ (Labov 1972: 188), *intraspeaker* variation automatically extends itself to ‘different ways of saying different things’ (cf. Halliday 1978: 35). For example, when comparing speech to writing, or literary to everyday, informal language, limiting the analysis to categories that are (nearly) semantically equivalent means one would miss out on those lexical and grammatical features that result from differences in content or topic.

Within the field of register-based variation, some overlap and potential for confusion of the term ‘register’ with the concepts of ‘genre’ and ‘style’ must be noted. What, in traditional sociolinguistic research, is often referred to as ‘speech styles’ (e.g., Jucker 1992), with an interest in factors such as formality, social identity, and social practices (e.g., Eckert & Rickford 2002), is nowadays part of a broader field of study, in which

'register' has become the more established term. It covers all kinds of functional relationships between linguistic form and situational context. Still, genre and style as concepts in their own right remain relevant for explaining textual variation: while the analysis of genres (e.g., Swales 1990) typically highlights conventions and generally considers aspects of the rhetorical organisation of texts, their stylistic analysis tends to focus on choices that result from aesthetic, often literary, preferences or linguistic preferences associated with individual authors (e.g., Jeffries & McIntyre 2010). Genres are not necessarily concerned with pervasiveness but rather with what is typical; for example, the genre of news may well contain a non-canonical structure, like a verbless clause in a headline, precisely once. Styles, by contrast, are generally seen as aesthetically, not functionally, motivated, and can therefore not be relied on to account for systematic register variation. For example, authors or newspapers may have their own style, but individual authorship will usually cause less variation than the overall situation in which a text is produced. While these approaches can be integrated as different 'perspectives for analysing text varieties' (Biber & Conrad 2019: 15), it is mostly registers that cover what is frequent and pervasive in texts, and thus match the quantitative approach to (non-)canonicity in this volume.

We have structured this introduction on textual variation as follows: Section 7.2 will introduce the concept of register and key aspects of register analysis as well as its classic methodology in more detail. Section 7.3 then (re)turns to the issue of (non-canonical) syntax, covering three patterns of possible variation: reduction, expansion, and placement variation. Each of these patterns corresponds to one of the case studies that will follow in this part of the volume. Throughout the chapter, we will take special care to point out methodological issues, in particular the existence of the two approaches that arise from the field being rooted in both variationist linguistics and register studies. In Section 7.4, we will also point out some more recent trends and open questions in the field, including a look at related issues in other fields (e.g., psycholinguistics) and at text varieties such as online registers and AI-generated text.

7.2 Key Aspects of Register Analysis

7.2.1 The Concept of Register

Register is a concept that traditionally exists at an intersection of several linguistic disciplines, such as sociolinguistics, discourse studies, literary and

linguistic stylistics, text-linguistics, applied linguistics, and the study of language for specific purposes. All these disciplines share an interest in the study of language use, addressing linguistic variation with a focus on the discourse situation, for example its formality, purpose, preferences or traditions of style, or specific topics and audiences. With the concept of register, linguists thus aim to recognise, for instance, that 'people speak differently depending on whether they are addressing someone older or younger, of the same or opposite sex, of the same or higher or lower status . . .; whether they are speaking on a formal occasion or casually, whether they are participating in a religious ritual, a sports event, or a courtroom scene' (Ferguson 1994: 15).

For more than a decade now, the study of register variation has also been recognised as a discipline of its own. Aiming at the precise and systematic description of linguistic features associated with different situations of language use, register studies nowadays propose a systematic framework for the linguistic analysis of textual variation. The starting point is a comprehensive set of situational parameters that provide the template for register classification. Following Biber and Conrad's (2019) *Register, genre, and style*, this set comprises six major situational characteristics: the discourse participants and their relations, the channel, circumstances, and setting (i.e., time and place) of both language production and comprehension, and the purpose and topic of the discourse produced (Biber & Conrad 2019: 40). Based on this framework, registers are text varieties with specific linguistic characteristics arising from these core components of the discourse situation. For identifying and classifying registers, the properties of the situation are thus 'more basic' than their linguistic characteristics (Biber & Conrad 2019: 9), which are both frequent and pervasive because they are functional for the situation. This means that this understanding of register variation is closely associated with the frequency-based approach to (non-)canonicity laid down in the Introduction to this volume. However, since some functional associations are also described as non-basic patterns (e.g., as non-basic patterns of word order resulting from information structure), the approach also relies on theory-based assumptions.

Biber and Conrad's approach highlights that linguistic features, which include non-canonical constructions, are conditioned by the context, not the other way round. For example, the real-time production mode of spoken, conversational registers is typically associated with the occurrence of forms of ellipsis. However, as the work by Biber, Wizner, and Reppen in Chapter 8 of this volume shows, structurally reduced clauses also occur

in settings with mixed characteristics of speech and writing. The language of news broadcasts they deal with is mostly planned and scripted but received in real time, which are properties that account for the mix of reduction and complexity features they observe. Situational characteristics are thus logically prior to linguistic features, and register analysis focuses on such ‘functional associations’ (Biber & Conrad 2019: 10).

Registers can be explored at very different levels of specificity, ranging from quite general text varieties, like conversation, news, and academic prose, to more specific sub-registers, such as scripted face-to-face conversation, for instance the dialogue in TV shows (e.g., Quaglio 2009), news in social media (e.g., Liimatta 2019; see also Clarke 2022 or Scheffler et al. 2022), or subtypes of academic discourse, which can be as diverse as introductions to research articles or office hour consultations. Two studies in this part of the volume, Chapter 8 by Biber et al. and Chapter 9 by Pham, deal with more general registers (news, reviews) but ultimately look into more specific sub-registers (television news broadcasts; printed, spoken, and online reviews). Often, a complex interplay of situational factors is required for pinning down functional associations. Pham’s study of clefts in evaluative language, looking at reviews from different media, refers to the situational parameter of channel of communication, but also suggests that the parameter of purpose is particularly relevant.

It needs to be emphasised that this framework of register analysis does not easily deal with all register constellations. Problems of register definition arise from cases (1) where there is a lot of variation among texts within one register, (2) where texts from different registers share many characteristics, or (3) where texts possibly do not belong to a register at all (cf. Biber & Egbert 2023). An example of the first category that has recently been discussed in the literature is student academic writing (Goulart et al. 2022; Biber & Egbert 2023: 10). The texts discussed in these studies typically involve more than one communicative purpose; they contain an almost equal amount of text passages that either explain or argue. In a similar vein, the register of conversation typically includes discourse varieties as diverse as joking around, engaging in conflict, or giving advice (Biber et al. 2021). A classic example of case (2), cross-register similarities, is the register of fiction, with novels, in particular, typically containing a mix of narration and speech. It has been an issue both of debate and empirical analysis in register analysis if fiction is one register or (at least) two (Egbert & Mahlberg 2020). As for case (3), the issue of ‘texts-with-no-register’ is still open for discussion. Some scholars find that such texts are ‘more prevalent than we might believe’ (Biber & Egbert 2023: 18; also Biber & Egbert

2018) but, since corpora are usually based on pre-classified registers, corpus-based register analysis tends not to target this issue in particular. However, Biber and Egbert's (2018) study of web documents finds many 'hybrid' texts that users did not identify as belonging to a particular register, for example, due to their containing an inseparable mix of description and both personal and commercial persuasion. In experimental studies participants are typically not asked to produce a specific register (corresponding to a real-life situation). The study on particle placement by Günther in this part of the volume shows that the choice between continuous and discontinuous particle verbs results from cognitive complexity, thus highlighting that sentence- or discourse-internal parameters are also relevant for syntactic variation.

7.2.2 *Non-Canonical Syntax and Discourse*

Registers are *text* varieties (i.e., units of language use), which turns the syntax within them into *utterances* rather than mere units of grammar. As an utterance, a (canonical or non-canonical) sentence must be seen as tied to its discourse in two possible ways. Since discourse is, formally, any element larger than the sentence and, functionally, language use with a given purpose (Dorgeloh & Wanner 2023: 16), any construction is bound to its discourse by the surrounding text (the *co-text*) as well as by the discourse situation (the proper *context*).

As for the role of co-text, syntactic choices can be explained by factors such as information packaging, topic, focus, or processing load – all factors in which the text surrounding a construction plays a role. For example, Günther in Chapter 10 of this volume looks at particle placement in sentences in isolation. In longer texts, the placement of the object NP in front of or behind the particle also depends on the co-text since it will impact on the information status of the NP (Lohse et al. 2004). In addition, aspects of the context are also relevant for particle placement; for example, there is an effect of speakers' intentions, like emphasising what is important to them (Dehé 2002). Other non-canonical constructions are often primarily explained by context rather than the co-text, for instance, *that*-omission, or the passive voice. For example, zero-*that* clauses (*I heard you were sick*) are more likely in speech than in writing (Biber 2012), while the passive voice, especially the long passive, is more common in writing (Biber et al. 2021: 929). However, properties of the matrix clause, such as a first- or second-person pronoun and certain verbs (e.g., verbs of sensation), also play a role in *that*-omission (Thompson & Mulac

1991), and the givenness of information in the subject vs. the agent *by*-phrase is also relevant for the use of the passive voice. According to Biber et al. (2021: 932), 90% of the *by*-phrases in long passives express new information, while Birner (1996, 2018) shows that English long passives are subject to similar information-structuring constraints as, for example, inversion (e.g., *Above his head hung a massive seagull with its beak open*; Birner 2018: e159). So, ultimately, we need to consider both co-text and context together to understand the use of non-canonical constructions as utterances within discourse.

7.2.3 *The Study of Non-Canonical Syntax in the Context of Register*

The relevance of discourse, in the form of co-text and context, motivates another core distinction for the study of syntax and the role register plays therein. There are two distinct approaches to exploring this role, which differ fundamentally in how register is conceptualised. One approach is rooted in so-called *variationist* linguistics and looks at syntactic variation as different ways of ‘accomplishing the same function’ (Szmrecsanyi 2019: 277). For example, the variationist approach aims to understand the syntactic choice between *that*- and *that*-less complement clauses, or between active and passive voice, by focusing on ‘constraints’ that govern which variant will be chosen (Szmrecsanyi 2019: 78). Register will be one crucial factor here, which means it is a *predictor* variable in that kind of work (Dorgeloh & Wanner 2023: 32–41).

Register has an alternative role in the so-called *text-linguistic* approach, which is based more directly on the register analysis framework. As we have explained in the previous section, this approach looks at the frequency of linguistic characteristics and explains them with reference to their ‘functional correspondences’ with the discourse situation (Biber & Egbert 2023: 4). For example, the passive is relatively more frequent in news and academic texts because these are texts that focus on events or generalisations rather than individual agents (Biber et al. 2021: 930). In that way, the approach uses syntactic characteristics for describing registers, which turns the role of register into one of a proper ‘object of investigation’ (Biber 2012). Here, the description of registers is based on counting features, with a focus on those features that have a higher rate of occurrence in one register compared to others. A non-canonical construction thus becomes a register *feature* when it has a higher frequency (relative to text length or corpus size) in one specific register than in other registers. When certain features do not occur in any other register, they become

register *markers* (Biber & Conrad 2019: 51). The studies by Biber et al. and Pham in this part of the volume mainly use the text-linguistic approach, presenting (normalised) frequencies across several registers, while Günther's experimental study exemplifies the variationist approach, comparing the participants' reactions to different syntactic variants.

Both approaches require usage-based evidence, typically in the form of corpus-based frequencies. These corpora need to be both representative and comparative (Biber & Conrad 2019: 10): As we emphasised in the discussion of register classification, register differences are often only gradient, since many registers are only more (or less) alike or different. For example, the study by Biber et al. in this part of the volume, dealing with the complexity of reduced structures in news broadcasts, touches on similarity with spoken discourse (reduction in general) but, in focusing on multiple reduction, there are also similarities with other written 'economy' registers. From a methodological viewpoint, syntactic features become register features if they are found to occur more frequently in the given register compared to at least one other (cf. Biber & Conrad 2019: 215).

The necessity for a comparative approach explains that corpus linguistics is the primary method in register studies. Depending on the specificity of the target registers, people either use freely available corpora or design their own to match a specific research question. Corpus size will vary depending on the frequency with which features occur. The study by Pham in this part of the volume looks at clefting, which is relatively rare, and uses a corpus of 310,000 words from six registers. By contrast, the TV news broadcast corpus compiled for Biber et al.'s study has a size of approximately 50,000 words covering three networks, further split into four text sections for an analysis of sub-registers. This corpus size is sufficient because reduced structures occur throughout the texts, which means they are not only more frequent than elsewhere but also pervasive in that register (cf. Biber & Conrad 2019: 9). As we pointed out earlier, experimental work typically lacks a specific register context, but instead usually strives for controlling for the co-text as variable. In this way, Günther in this part of the volume explores conditions of cognitive complexity and their effect on particle placement using the methods of a self-paced reading and a split rating task.

A conceptually and methodologically more complex approach to register analysis is the so-called multi-dimensional (MD) analysis, first introduced in Biber (1991). The method originally started out with a group of 67 lexical and grammatical features, which, by way of a computational factor analysis, were turned into a set of five dimensions of textual variation, identified in a large

corpus of texts. Subsequently, these dimensions were labelled for reflecting certain functional properties of the texts, such as being ‘involved’ or ‘informational’, or ‘abstract’ or ‘non-abstract’. For example, as for some syntactic features we looked at before, Dimension 1 (involved vs. informational discourse) contains a higher density of *that*-deletion, while it is marked by a low density of agentless passives and of what is here called ‘deletions’ (Biber 1991: 104–8). The main point of the approach is to provide quantifiable and generalisable descriptions of registers (Biber & Conrad 2019: 216), and the approach has established a broad research area over the last thirty years. The method goes beyond describing registers by individual features, but focuses on their patterns of *co*-occurrence. Within a dimension, syntactic characteristics are always part of a group of positive or negative features: for example, in an MD analysis of university registers, *that*-omission co-occurs with other ‘oral’ features such as contractions or *wh*-questions and with a negative dimension score for agentless passives (Biber & Conrad 2019: 228). In this way, the method manages to cover a larger set of linguistic features and to reveal similarities and differences across registers. Biber et al. (Chapter 8 in this volume) also discuss their findings on multiple reduction in the context of other ‘phrasal’ features belonging to the Dimension 1 characteristics of radio broadcasts.

7.3 Non-Canonical Syntactic Phenomena Studied

Throughout this volume, it is shown that syntactic constructions that may be judged as unusual, questionable, or even ungrammatical in isolation may still be acceptable in, and even characteristic of, specific registers in a language. We will now turn to an overview of non-canonical constructions that have been looked at from a register angle. There are different ways in which these could be systematised. For example, one could go by sentence-level vs. phrase-level vs. word-level phenomena, or we could arrange this section by the factors of the discourse that are conducive to the use of non-canonical syntactic patterns, such as modality (with spoken registers drawing differently on cognitive resources of the speaker than written registers) or function of the text. Instead, we will use *form* as our starting point. With regard to form, a non-canonical clause can deviate from the canonical clause in one of three ways: (a) the non-canonical clause is a shorter, reduced version of the canonical clause (such as in cases of ellipsis or deletion), (b) it is an expanded, more complex version (such as in cleft constructions), or (c) the word order of the canonical clause is rearranged in the non-canonical clause (such as in the case of particle placement, argument alternations or other non-default form–function mappings,

including *it*-extraposition).¹ In looking at discussions of these non-canonical syntactic patterns and their relationship to register and genre, we will also keep in mind the two perspectives introduced earlier: studies focused on explaining the constructional make-up of a register (text-linguistic approach) and studies focused on explaining the choice between two competing constructions (variationist approach).

7.3.1 *Reduction*

Grammatical reduction, the simplification of grammatical structure (and with it the reduction of number of words per utterance), is a pattern characteristic of spoken registers, particularly conversation. Biber et al. (2021: 1037) list pronouns, other proforms, and ellipsis among the most common forms of grammatical reduction occurring in this register. The use of reduction here is mainly due to situational parameters. Firstly, since conversation happens in real time and is interactive, it draws significantly on cognitive resources. Grammatical reduction can be seen as a strategy to reduce the demand on those resources. Secondly, the shared situational context of a situation allows speakers to underspecify what is said explicitly and to rely on the shared experience of the situation (and often also on shared knowledge). As a result, conversation includes many utterances that are non-clausal units, which is one of the reasons that the syntactic make-up of conversations looks very different from that of written registers. In their chapter on ‘the grammar of conversation’, Biber et al. (2021: 1065) found that 38.6% of units in a sample of American and British English conversation (1,000 turns, 5,369 words) were non-clausal. Below is an example that shows how they count speech units (separated by a double line) and classify them as clausal (<Cl>) or non-clausal (<NCl>).

- (1) A: II So do you think an alligator would like salt water? <Cl> II
 B: II It would probably kill him wouldn't it? <Cl> II
 A: II That one on the news has been out in the ocean for a while. <Cl> II
 B: II Really? <NCl> II
 C: II What are you talking about? <Cl> II I didn't hear. <Cl> II
 A: II The alligator in the ocean. <NCl> II I was asking him how he thought it liked saltwater. <Cl> II
 C: II Oh. <NCl> II (Biber et al. 2021: 1065)

¹ Our classification is not quite the same as the one proposed in the Introduction to this volume, which, apart from reordering, additions, and subtractions, proposes two more categories of non-canonicity: one for a lack of default formal marking (e.g., lack of number agreement) and one for the realisation of clause elements by expletives (as with *it*-clefts or existential clauses).

In this example, all non-clausal units are treated the same, but in their contribution to this volume, Biber et al. make the distinction between non-clausal units without any syntactic structure, which include discourse markers like speaker C's last utterance in the example above (*Oh*), and non-clausal units with internal syntactic structure, such as speaker A's response *The alligator in the ocean*. They subsume the latter, as well as units that include a verb, under the category of 'non-canonical reduced structures'. In their study, they look at such structures specifically in news broadcasts, looking into the parameters of the register that favour such expressions. Another example of reduced structures is the case of subject ellipsis, which is characteristic of spoken registers like conversation (Narimaya 2004) and written registers like diary writing (Haegeman 2013), texting (van Dijk et al. 2016), and blog writing (Teddiman & Newman 2007). Other ways to reduce or shorten an utterance include auxiliary contractions (*hasn't*, *we'll*) and acronyms. Both typically occur in registers that prize short forms due to high interactivity.

In most of these cases, it is clear which syntactic material has been omitted, and the non-canonical, reduced form can be considered a variant of the canonical, non-reduced form, which opens the path for a variationist design. However, as Biber et al. point out in their contribution to this volume, there are also reduced units for which it can be hard to say what the non-reduced form would have been, especially in cases where an utterance is constituted by a single noun phrase or adjective phrase. Structures reduced in this way do not lend themselves to a variationist approach, since there is no clear comparison of two variants. In example (1) above, we cannot say with certainty which canonical sentence C's utterance *Really?* is the reduced version of. The variationist research design is best applied when it is clear which elements exactly have been reduced or deleted, as in the case of subject or object ellipsis. Since this does not hold for all of the data discussed here by Biber et al., their analysis follows the text-linguistic design.

7.3.2 *Expansion*

Non-canonical syntactic constructions that arise from making a canonical sentence longer and/or more complex include, for example, *it*-extraposition (2), *it*-clefts (3), *wh*-clefts (4), and left dislocation (5).

- (2) a. It is bad to have such sharply diverging classes. (*Corpus of Contemporary American English*, COCA, News)
- b. To have such sharply diverging classes is bad.

- (3) a. It was Democrats who killed the DREAM act. (COCA, Spoken)
b. Democrats killed the DREAM act.
- (4) a. What you should not do is prescribe carrot juice (COCA, Blog)
b. You should not prescribe carrot juice.
- (5) a. This guy, . . . he is an extra brand of crazy. (COCA, Spoken)
b. This guy is an extra brand of crazy.

The distribution of these constructions is different from the reduced structures just discussed. They are found in all types of registers, but their uses are quite register-specific and, seen across registers, they are quite rare. For example, subject *it*-extraposition, as in (2a), is found to occur, at a rate of about two times per 1,000 words, in both academic and popular writing (Zhang 2015), while left dislocation, as in (5a), occurs almost exclusively in conversation (around two times per 10,000 words) and only ‘occasionally’ in fictional dialogue or written prose (Biber et al. 2021: 948). The choice of a longer or more complex form is also discussed as affecting processing. For example, psycholinguistic work on reading comprehension shows that readers utilise *it*-clefts to anticipate the way in which information will be packaged next (e.g., Alemán Bañón & Martin 2019).

Sentences that lend themselves to *it*-extraposition (esp. sentences with a subject clause and an adjectival predicate, as in (2b)) occur more often in writing than in spoken registers (Kaltenböck 2005), but, from a variationist angle, it is true for both written and spoken registers that in the vast majority of syntactic environments that allow *it*-extraposition the construction is applied. The reasons for choosing the construction usually lie in the co-text. Based on 1,701 instances of examples extracted from the British component of the *International Corpus of English (ICE)*, Kaltenböck showed that in almost three out of four occurrences of *it*-extraposition the extraposed subject clause contains new information, which aligns with general ideas about information packaging.

Register-based research on cleft sentences is usually built on a small set of examples, and researchers look at the syntactic category and the information status of the foregrounded element (e.g., Hedberg 1990). In Chapter 9 of this volume, Pham takes a different approach, focusing on how the use of cleft sentences may be influenced by situational factors, particularly the communicative purpose of evaluation. She found that, while most cleft constructions themselves are evaluative (a concept that is not as easy to code as syntactic categories like ‘clausal’ above), clefts do not occur more frequently in texts that clearly have an evaluative purpose. She

hypothesises that clefts, which are non-canonical sentences by syntactic criteria, could be considered 'less non-canonical' if one looks into the function of the construction. This shows again that a syntactically non-canonical pattern can be the functionally canonical one.

7.3.3 *Placement Variation*

English is considered a strict word order language in which the arguments of the verb (like agent, theme, goal) are mapped onto syntactic positions predictably. Non-canonical sentences include those in which the expected arrangement is not followed. For example, it is generally assumed that canonically a direct object will directly follow the verb (Huddleston & Pullum 2002: 247), as in (6a). Sentences in which the direct object occurs after an adjunct, as in (6b), a case of heavy NP-shift, are non-canonical.

- (6) a. He praised me for mastering the trills quickly. (*COCA*, Fiction)
- b. He was by all accounts a prodigy and mastered quickly the hard-earned lessons that most shipwrights spent a lifetime accumulating. (*COCA*, Magazine)

Another example of placement variation are passive constructions, which break the expected alignment of agent and subject position, resulting either in a sentence in which the agent is not realised at all (short passive) or a sentence in which a non-agent is placed in subject position and the agent becomes an adjunct inside a *by*-phrase (long passive). Much of the research on the occurrence of the passive as a non-canonical construction is text-linguistic in nature. The use of passive constructions is known to vary considerably across registers; for example, they are a frequent and pervasive characteristic of academic writing (Biber et al. 2021). The reasons for that register specificity lie, to some extent, in the purpose and genre conventions of academic writing. Texts tend to be about processes, data, and discoveries rather than about the people who make them. The passive is one way to place a non-agent in the subject position of a sentence, that is, in the syntactic position most clearly linked to the discourse function of topic. Long passives are quite rare, even in academic writing. The main reason for that is that there is a long tradition, especially in the natural sciences, of presenting research as a disembodied endeavour. The passive provides the option to leave an agent unexpressed, and academic writing is a register in which this option is often considered desirable, especially if the agent is the author of the text, despite modern style manuals explicitly taking the position that the use of the first person (*I will show* instead of *It will be shown*) is clearer and intellectually more honest. As pointed out

above, if passives do occur with *by*-phrases, the *by*-phrase tends to express new information or at least information that is less given than the subject (Biber et al. 2021: 933).

Like most syntactic constructions, non-canonical or not, the use of passive constructions can be examined from a variationist or a text-linguistic perspective. Biber et al. (2021), for example, take a text-linguistic approach and measure the use of the passive by tokens per X number of words in a corpus. A different approach is chosen by Seoane (2006), who studies the frequency of passives from a variationist perspective, counting the percentage of transitive verbs realised in active vs. passive voice.

Like reductions, the use of placement alternations and the preference for a non-canonical pattern in a given discourse situation is also explained from a cognitive angle. Following the influential work by Hawkins (1994, 2004), it is generally assumed that speakers, when given a choice, prefer constructions that can be processed more efficiently. Hawkins' principle of 'minimising domains' allows him to calculate which structures are more efficient in that regard. Simply put, the question is how much syntactic material (the 'domain') must be processed before the overall structure of the verb phrase is determined. Changing the word order in a sentence can be a way to minimise that domain.

Building on Hawkins' work, Günther (Chapter 10 in this volume) looks at word order variation in the case of particle verbs (*look up the information/look the information up*). Generally, the continuous word order (verb-particle) is considered the canonical pattern, the one that minimises the 'domain' that must be processed to figure out the meaning of the verb. If that is the case, it is not immediately clear why the non-canonical pattern, the one that is supposed to be less efficient to process, is more common in spoken language, a register that is supposed to favour the minimising of processing burdens, and that the canonical form cannot be chosen at all if the object of the verb is a pronoun (**look up it*). Günther makes the important observation that the cognitive load for language production is not necessarily the same as for language perception. She chooses experimental data to explore the difference between minimising one's own cognitive load as speaker vs. reducing the cognitive load for the hearer.

7.4 Trends and Open Questions

As we see throughout this part of the volume, register studies is a dynamic field that reflects trends in the development of registers as well as in

linguistics as a discipline. Register is increasingly considered as one of the core factors that regulate syntactic variation (e.g., Szmrecsanyi 2019) and language change (e.g., Biber et al. 2016). With respect to the former, the categorisation and analysis of digital modes of communication, which never really fit into the written/spoken dichotomy and are increasingly multi-modal, has fuelled new branches of research into registers (for an overview, see Biber & Egbert 2018 and Page et al. 2022). Such studies look both into how digital modes have expanded the inventory of linguistic forms and how they provide the ecosystem for new registers or transforming existing ones. For example, Zappavigna (2018) has offered a classification of hashtags that includes the function of register-specific topic markers, Bohmann (2016) has looked into posts on Twitter, now known as 'X', as the potential origin of a new, non-canonical use of *because*, and Zhang (2023) has shown that the linguistic profile of printed news is influenced by the now dominant register of digital news, a case that illustrates how a change in extra-linguistic behaviour (people increasingly consuming news in digital environments) impacts a register.

On the methodological side, the go-to approach to studying register both from the text-linguistic and variationist experience is still corpus-based, but there is growing awareness of the fact that corpus data are mainly production data, and that any analysis that looks into cognitive factors like processing cost as a criterion for explaining the use of non-canonical constructions should ideally also include data from language processing (as done by Günther in this part of the volume). It is no surprise, therefore, that there is a growing trend to rely on converging evidence, that is, by combining corpus data and experimental data.

An interesting situation for register-based work arises when criteria from different domains for what we consider 'canonical' do not align. In the case of particle verbs, for example, based on structural criteria, the discontinuous variant is considered to be non-canonical (see Günther, this volume), but data from language acquisition show that the vast majority of particle verb constructions produced by young children exemplify the discontinuous order (Diessel & Tomasello 2005), a fact that, at first sight, is not easy to reconcile with the assumption that canonical constructions occur more often in the input and are easier to acquire. In a similar vein, it is not immediately obvious why *it*-extraposition, a non-canonical sentence pattern by structural criteria, when looked at from a variationist perspective, is used much more frequently than its canonical, non-extraposed competitor, both in speech and writing (Kaltenböck 2005).

Additionally, statistical analyses have become more sophisticated and have moved from monofactorial to complex multifactorial models. Examples are Gries' (2003) groundbreaking study on particle verbs and Grafmiller's (2014) study of the realisation of the genitive in English in terms of an interaction of processing-related factors with register conditions across modality and genre.

As we move into a future in which we will increasingly encounter texts generated by artificial intelligence tools, we predict that one topic the field of register studies will have to wrestle with is the production of texts not conditioned by human communicative needs or processing constraints. Language models that power tools like ChatGPT (launched in late 2022) are trained on 'large, uncured, static datasets from the Web' (Bender et al. 2021: 615). Not only do such models not use language to encode meaning, they also tend to 'encode hegemonic views that are harmful to marginalized populations' (Bender et al. 2021: 615). Texts generated by such models may sound plausible and formally in line with register expectations, but since they were not generated to express meaning in a specific speech situation, it is not immediately clear if they fall within the purview of register as a variable as discussed in this part of the volume. If, without any outward markers, a sizeable number of texts are generated by 'stochastic parrots' (Bender et al. 2021: 617) rather than by human speakers with human communicative needs, a discipline that relies on large databases of text samples may have to rethink from the ground what it means to analyse syntactic variation by speech situation.

REFERENCES

- Alemán Bañón, José & Clara Martin (2019). Anticipating information structure: An event-related potentials study of focus assignment via the *it*-cleft. *Neuropsychologia*, 134, 107203.
- Bender, Emily, Timnit Gebu, Angelina McMillan-Major, & Shmargaret Shmitchell (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of FAccT 2021*, 610–23.
- Biber, Douglas (1991). *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Biber, Douglas (2012). Register as a predictor of linguistic variation. *Corpus Linguistics and Linguistic Theory*, 8(1), 9–37.
- Biber, Douglas & Susan Conrad (2019). *Register, genre, and style*. Second edition. Cambridge: Cambridge University Press.
- Biber, Douglas, with Jesse Egbert, Bethany Gray, Rahel Oppliger, & Benedikt Szmeccsanyi (2016). Variationist versus text-linguistic approaches to grammatical change in English: Nominal modifiers of head nouns. In Merja Kytö

- & Päivi Pahta, eds, *The Cambridge handbook of English historical linguistics*. Cambridge: Cambridge University Press, 351–75.
- Biber, Douglas & Jesse Egbert (2018). *Register variation online*. Cambridge: Cambridge University Press.
- Biber, Douglas & Jesse Egbert (2023). What is a register? Accounting for linguistic and situational variation within – and outside of – textual varieties. *Register Studies*, 5(1), 1–22.
- Biber, Douglas, Stig Johansson, Geoffrey N. Leech, Susan Conrad, & Edward Finegan (2021). *Grammar of spoken and written English*. Amsterdam: John Benjamins.
- Birner, Betty J. (2018). On constructions as a pragmatic category. *Language*, 94(2), e158–e179.
- Bohmann, Axel (2016). Grammatical change because Twitter? Factors motivating innovative uses of *because* across the English-speaking Twittersphere. In Lauren Squires, ed., *English in computer-mediated communication: Variation, representation, and change*. Berlin: Mouton de Gruyter, 149–78.
- Clarke, Isobelle (2022). Register and social media. *Register Studies*, 4(2), 133–7.
- COCA *The Corpus of Contemporary American English* (520 million words, 1990–present) (2008–). Compiled by Mark Davies. Retrieved from <http://corpus.byu.edu/coca/>.
- Dehé, Nicole (2002). *Particle verbs in English: Syntax, information structure, and intonation*. Amsterdam: John Benjamins.
- Diessel, Holger and Michael Tomasello (2005). Particle placement in early child language: A multifactorial analysis. *Corpus Linguistics and Linguistic Theory*, 1(1), 89–112.
- Dijk, Chantal N. van, Merel van Witteloostuijn, Nada Vasić, Sergey Avrutin, & Elma Blom (2016). The influence of texting language on grammar and executive functions in primary school children. *PLoS ONE*, 11(3), e0152409. Retrieved from <https://doi.org/10.1371/journal.pone.0152409>.
- Dorgeloh, Heidrun & Anja Wanner (2023). *Discourse syntax: English grammar beyond the sentence*. Cambridge: Cambridge University Press.
- Dorgeloh, Heidrun & Anja Wanner, eds (2010). *Syntactic variation and genre*. Berlin: Mouton de Gruyter.
- Eckert, Penelope & John R. Rickford (2002). *Style and sociolinguistic variation*. Cambridge: Cambridge University Press.
- Egbert, Jesse & Michaela Mahlberg (2020). Fiction – one register or two? Speech and narration in novels. *Register Studies*, 2(1), 72–101.
- Ferguson, Charles A. (1994). Dialect, register, and genre: Working assumptions about conventionalization. In Douglas Biber & Edward Finegan, eds, *Sociolinguistic perspectives on register*. Oxford: Oxford University Press, 5–30.
- Goulart, Larissa, Douglas Biber, & Randi Reppen (2022). *In this essay, I will ...: Examining variation of communicative purpose in university writing*. *Journal of English for Academic Purposes*, 59, 101159.
- Grafmiller, Jason (2014). Variation in English genitives across modality and genres. *English Language and Linguistics*, 18(3), 471–96.

- Gries, Stefan Th. (2003). *Multifactorial analysis in corpus linguistics: A study of particle placement*. New York: Continuum.
- Haegeman, Liliane (2013). The syntax of registers: Diary subject omission and the privilege of the root. *Lingua*, 130, 88–110.
- Halliday, Michael A. K. (1978). *Language as social semiotic*. London: Edward Arnold.
- Hawkins, John A. (1994). *A performance theory of order and constituency*. Cambridge: Cambridge University Press.
- Hawkins, John A. (2004). *Efficiency and complexity in grammars*. Oxford: Oxford University Press.
- Hedberg, Nancy (1990). *Discourse pragmatics and cleft sentences in English*. PhD thesis, Minneapolis: University of Minnesota Press.
- Huddleston, Rodney & Geoffrey K. Pullum, eds (2002). *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.
- ICE International Corpus of English (1990–). Retrieved from www.ice-corpora.uzh.ch/en.html.
- Jeffries, Lesley & Daniel McIntyre (2010). *Stylistics*. Cambridge: Cambridge University Press.
- Jucker, Andreas H. (1992). *Social stylistics: Syntactic variation in British newspapers*. Berlin: Mouton de Gruyter.
- Kaltenböck, Gunther (2005). *It*-extraposition in English: A functional view. *International Journal of Corpus Linguistics*, 10(2), 119–59.
- Labov, William (1972). *Sociolinguistic patterns*. Philadelphia: University of Philadelphia Press.
- Liimatta, Aatu (2019). Exploring register variation on Reddit: A multi-dimensional study of language use on a social media website. *Register Studies*, 1(2), 269–95.
- Lohse, Barbara, John A. Hawkins, & Thomas Wasow (2004). Domain minimization in English verb-particle constructions. *Language*, 80(2), 238–61.
- Nariyama, Shigeko (2004). Subject ellipsis in English. *Journal of Pragmatics*, 36(2), 237–64.
- Page, Ruth, David Barton, Carmen Lee, Johann Wolfgang Unger, & Michele Zappavigna (2022). *Researching language and social media: A student guide*. London: Routledge.
- Quaglio, Paulo (2009). *Television dialogue: The sitcom Friends vs. natural conversation*. Amsterdam: John Benjamins.
- Scheffler, Tatjana, Lesley-Ann Kern, & Hannah Seemann (2022). The medium is not the message: Individual level register variation in blogs vs. tweets. *Register Studies*, 4(2), 171–201.
- Schilling-Estes, Natalie (2004). Investigating stylistic variation. In Jack K. Chambers, Peter Trudgill, & Natalie Schilling-Estes, eds, *The handbook of language variation and change*. Oxford: Blackwell, 375–401.
- Seoane, Elena (2006). Changing styles: On the recent evolution of scientific British and American English. In Christiane Dalton-Puffer, Dieter Kastovsky, Nikolaus Ritt, & Herbert Schendel, eds, *Syntax, style and grammatical norms: English from 1500–2000*. Bern: Peter Lang, 191–211.

- Swales, John (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.
- Szmrecsanyi, Benedikt (2019). Register in variationist linguistics. *Register Studies*, 1(1), 76–99.
- Teddiman, Laura & John Newman (2007). Subject ellipsis in English: Construction of and findings from a diary corpus. Paper presented at the 26th International Conference on Lexis and Grammar, Bonifacio, France, 2–6 October. Retrieved from https://www.researchgate.net/publication/255608296_Subject_Ellipsis_in_English_Construction_of_and_Findings_from_a_Diary_Corpus.
- Thompson, Sandra A. & Anthony Mulac (1991). The discourse conditions for the use of the complementizer *that* in conversational English. *Journal of Pragmatics*, 15(3), 237–51.
- Zappavigna, Michele (2018). *Searchable talk: Hashtags and social media metadiscourse*. Amsterdam: John Benjamins.
- Zhang, Difei (Lynn) (2023). *Flipping those pages, swiping that screen: A corpus-based analysis of the digital transformation of the news register*. PhD thesis, University of Wisconsin-Madison. ProQuest Dissertations Publishing, 30316227.
- Zhang, Guiping (2015). *It is suggested that ... or it is better to ...?* Forms and meanings of subject *it*-extraposition in academic and popular writing. *Journal of English for Academic Purposes*, 20, 1–13.