

CHAPTER I

*Introduction: New Perspectives on Non-Canonical English Syntax**Teresa Pham and Sven Leuckert*

The characterisation generally provided of word order in English is that the subject typically precedes the verb, which in turn precedes any obligatory complementation. Thus, the sentence *I met Sigrid* exemplifies the default composition and ordering of clause elements in Present-Day English (PDE). Yet, if we study actual language use, we find frequent deviations from this default: *Sigrid I met*; *It was Sigrid that I met*; *Sigrid, I met her*; or even *Met Sigrid*. Similarly, at the phrasal level the characterisation can be summed up as: what belongs together is usually placed together. A relative clause, for example (as in *I met Sigrid, who was very happy*), typically directly follows its antecedent. Yet again, in actual language use, we may encounter discontinuous structures in multi-word verbs, prepositional phrases, or even noun phrases containing a postmodifying relative clause (e.g., *Some options were considered that allow for more flexibility*, Francis & Michaelis 2017: 332).

If the characterisation provided represents a default and perfectly acceptable way of expressing a particular meaning, why would an alternative ever be used and why would alternatives even exist in the first place? What are the factors that condition the use of these alternative constructions? And, perhaps even more importantly, what are the factors that lead one variant to be identified as the default? Indeed, if we broaden our perspective, we realise that the supposed default option suddenly becomes a moving target. For example, if we do not consider language use in its entirety, but only *to*- and *that*-clauses functioning as subjects (e.g., *That I missed Sigrid is a pity*), then these are more frequently extraposed (e.g., *It is a pity that I missed Sigrid*) than not in PDE (cf. Quirk et al. 1985: 1049, 1062). Furthermore, there are studies showing that ellipses of subject pronouns (e.g., *Met Sigrid*) are particularly common in spoken language samples (cf. Wilson 2000: 62). Subjectless clauses may even be frequent, as they are in synchronic computer-mediated communication (Hård af Segerstad 2002: 245; Bieswanger 2016). Additionally, in conversational English, more than one third of all units are non-clausal (cf. Biber et al. 2021: 1065). If we

extend our view further to include varieties of English in use worldwide, we have to acknowledge that modifications of the order of clause elements are much more frequent in some varieties, as for example in Indian English, than in others (cf. Lange 2012). And finally, in earlier periods, when English was more synthetic, word order was less rigid. In Old English (OE), a predominantly verb-second language (cf. Fischer & van der Wurff 2006: 185), the finite verb typically preceded the subject in main clauses with fronted constituents, but the pattern SOV is found both in main clauses and in subordinate clauses. Consequently, one might ask: does it even make sense to assume the existence of one default grammar of English for all situations of language use, for all varieties, for all modes, and for all registers?

It is observations and considerations like these that have puzzled and fascinated the contributors to this volume for many years, and which have led to the compilation of this volume, which brings together research projects shedding light on these questions from a range of different perspectives. What has emerged is that expected default constructions, or what could be called the syntactic ‘canon’, are neither static nor persistent and that thus no universally valid description of this canon can be provided. Rather, the syntactic canon is influenced by an interplay of various factors, such as idiolectal preferences of language users, register, discursive context, time, etc., and needs to be redefined for each specific communicative situation (or group of recurring situations).

This introduction aims to shed light on syntactic (non-)canonicity from a morphological and etymological perspective, outlining different basic approaches to the notion in order to provide context for the subsequent research, which explores specific instances of non-canonical English syntax. Section 1.2 then provides definitions of the concepts of *canonical* and *non-canonical syntax*, which underlie the contributions to this volume. Section 1.3 situates these terms in relation to ‘syntactic variation’. Section 1.4, finally, describes the structure of the volume and gives an overview of the individual contributions.

1.1 Approaching Syntactic (Non-)Canonicity

The terms ‘canonical’ and ‘non-canonical’ have been used in linguistic and, more specifically, syntactic studies for decades, but not in a uniform way, as will be shown in the paragraphs below. Therefore, a promising way of accessing these terms is to look at their morphology and etymology. Since the negative prefix {non-} is transparent, the main focus of our

discussion needs to be on what may be considered *canonical* in syntax: in accordance with the meaning of its etymon in Greek and Latin, two basic (partially overlapping) interpretations of the term ‘canonical’ can be identified in present-day linguistics.

The noun *canon* was adopted into OE from Latin *canon* (ultimately from Greek *κανών* ‘rule’). Originally predominantly used in the ecclesiastical domain, it underwent an extension and generalisation of meaning in the late sixteenth century to refer to an individual, general non-ecclesiastical ‘rule, fundamental principle’ (*OED*) or collectively to ‘a body of principles, rules, standards, or norms’ (*Merriam-Webster*). The corresponding adjective ‘canonical’, formed by the addition of the adjectival suffix {-al}, can refer to what conforms either to *one* fundamental general principle or to a *collection* of accepted rules or norms. Thus, clearly, the interpretation of the adjective will vary depending on which rule(s), norm(s), or standard(s) are – explicitly or implicitly – assumed.

In non-scientific PDE usage, what underlies canonicity is usually an intuitive judgment of frequency, so that ‘canonical’ is used synonymously with adjectives such as ‘common’, ‘usual’, or ‘normal’. This frequency-based interpretation of the adjective can also be found in linguistic publications with an empirical foundation, for example in the *Grammar of spoken and written English*, where *yeah* is treated as the canonical positive response form in conversational English due to its being ‘considerably more frequent than *yes*’ (Biber et al. 2021: 1084).

In other linguistic publications, however, the interpretation of *canonical* is (or seems to be) based on a specific linguistic model. Huddleston and Pullum, for example, introduce *canonical* as a synonym of ‘basic’ (2002: 45), mentioning negation, interrogation, passivisation, subordination, and main-clause coordination as examples of non-canonical procedures or constructions. This implies a definition of a canonical clause as a ‘minimally complete grammatical structure’. Prescriptive approaches, by contrast, regard as canonical all constructions which conform to a set of established norms or rules of a specific standard variety. Many of these norms were imposed upon the English language by eighteenth-century grammarians (cf. Ebner 2017), and strong reservations over certain constructions like split infinitives or preposition stranding survive to this day, sometimes even among linguists.

In language typology, finally, much effort has been put into the classification of languages based on ‘word order’, more specifically the *basic* order of syntactic constituents at the clause level (cf. Song 2012: 14), that is, the most frequent constituent order found ‘in stylistically neutral,

independent, indicative clauses with full noun phrase ... participants, where the subject is definite, agentive and human, the object is a definite semantic patient, and the verb represents an action, not a state or an event' (Siewierska 1988: 8). This clearly shows that the classification of SVX as the canonical word order in English is both theory- and frequency-based, to a certain degree reconciling, in fact, the two approaches described above. The (implicit or explicit) classification of the SVX pattern as canonical also underlies information-structural approaches to word order variation, like that of Birner and Ward (1998; also Ward & Birner 2004), in which non-canonical constructions, that is, constructions deviating from the canonical pattern, are motivated by the need to arrange constituents according to an increasing degree of informativity. Other discourse functions may, however, also play a role in classifying constructions as (non-)canonical.

Ultimately, this shows that the meaning of the adjectives 'canonical' and 'non-canonical' varies depending on whether a frequency-based and/or a theory-based approach to (non-)canonicity is chosen and depending on the theoretical approach that underlies the latter. It is important to acknowledge that there are several interpretations or usages of the adjectives *canonical* and *non-canonical* and several approaches to syntactic (non-)canonicity.

1.2 Defining Syntactic (Non-)Canonicity

Following Halliday and Matthiessen's influential conceptualisation of a Theme (2004: 73),¹ and based on the observations outlined at the beginning of this introduction, we define as a 'canonical' syntactic construction a default structure, that is, an ordering, composition, formal marking, or realisation of elements, which under general circumstances will be chosen with the highest likelihood by a speaker or writer unless there are good reasons for choosing a different syntactic structure. By contrast, a 'non-canonical' syntactic construction is defined as a deviation from the default ordering, composition, formal marking, or realisation of elements which in the production process is motivated by one or several factors.

In order to be able to consider non-canonical syntax in its full diversity, we intend these definitions to subsume, and in fact go beyond, both frequency-based and theory-based approaches to syntactic (non-)canonicity, which, despite apparent overlaps, have to date seemed

¹ The relevant passage is: 'The Subject is the element that is chosen as Theme unless there is good reason for choosing something else' (Halliday & Matthiessen 2004: 73).

irreconcilable. All contributions to this volume test and discuss the applicability of either of the two approaches to the respective areas of non-canonical syntax on which they focus and with regard to their specific methodology.

Furthermore, the above definition of non-canonicity is intended to embrace, first, modifications of the default order of elements. These include manipulations of the order of constituents, such as topicalisation (*Sigrid I met*) and inversion (*Equally important is the fact that . . .*), but also stranded prepositions, split infinitives, and other discontinuous phrases. Second, the definition also covers additions to the default composition of elements within a clause, which might make constructions more than minimally complete, such as dislocation (*Sigrid, I met her*), the introductory-*it* pattern, also called ‘*it*-extraposition’ (*It is a pleasure to meet you*), and cleft constructions (*This is what I said*). Third, subtractions from the canonical composition of elements may result in constructions which are less than minimally complete. Amongst these are main clauses without an explicit subject (*Met Sigrid*) or lexical verb (*Sigrid happy*), structurally reduced clausal units (*Why not?*), and noun phrases without a determiner (*It’s true story*), that is, constructions characterised by the omission of elements at the phrase or clause level, which would otherwise be expected to occur in PDE. Fourth, the definition subsumes structures which do not manifest a default formal marking of internal relationships, such as a lack of number agreement (*there are many group; four cup*). Fifth, it also accommodates cases where the usual syntactic form–function correlations are not adhered to and where clause elements are realised by semantically empty expletives with a purely syntactic function. This is the case, for example, in *it*-clefts (*It was Sigrid that I met*) and clauses with existential or presentational *there* (*There is a problem*). The aforementioned examples of non-canonical constructions show that changes to the ordering, the composition, the formal marking, and the realisation of elements alike may affect each formal and/or functional level of the syntactic hierarchy. Finally, we consider the concepts ‘canonical’ and ‘non-canonical’ to be gradable antonyms. This becomes evident from the fact that numerous combinations of the aforementioned deviations from the default are possible, such as in *there’s five people*, featuring the non-referential pronoun *there* typical of the existential *there*-construction as well as a singular verb in combination with a plural notional subject. The existence of such combinations is also implied in Huddleston and Pullum’s discussion of canonical and non-canonical constructions (2002: 1365). Consequently, each degree of (non-)canonicity may (but need not) be represented by one or several

syntactic structures and there is no one-to-one correspondence between canonical and non-canonical constructions.

1.3 Delimiting Syntactic (Non-)Canonicity

Since calling a construction ‘non-canonical’ requires comparison with at least one alternative, syntactic non-canonicity is, of course, related to syntactic variation. In fact, we maintain that most constructions which have been or can be studied as syntactic variants may also be conceptualised as canonical and non-canonical alternatives. Consequently, the total of what we classify as canonical and non-canonical syntactic constructions can be regarded as a subset of those phenomena traditionally studied under the umbrella of syntactic variation. The main difference consequently may not lie in *what* is being studied, but in *how*, that is, in the perspective from which it is being studied. To better understand this difference, a brief and necessarily simplified outline of the concept of syntactic variation is expedient.

Linguistic variation is commonly defined as the existence (usually in actual language use) of ‘two or more formal alternatives which can be considered optional variants’ and which are ‘nearly equivalent in meaning’ (Biber et al. 2021: 14). It may be challenging, however, to decide how similar alternatives have to be with regard to form to count as variants and, more importantly, when alternatives can be claimed to have nearly the same meaning. These challenges arise particularly in the context of syntactic variation.

Interest in linguistic variation was triggered by William Labov’s groundbreaking studies on phonetic variation from the early 1960s (Labov 1963, 1966). These and later studies (on phonetic as well as, later, (morpho-) syntactic variation) demonstrated that linguistic variation is typically not random or ‘free’ as previously assumed, but rather tends to pattern systematically, conditioned by the interplay of a multitude of factors, both intra-linguistic (e.g., structural or lexical factors) and extra-linguistic (e.g., cognitive, psycholinguistic, discourse-related, pragmatic, or social factors; Dorgeloh & Wanner 2010: 5–6). The objective in research on linguistic and syntactic variation has thus been to explain why one option is chosen over another in a given communicative situation and, in more recent multifactorial, probabilistic approaches, also to estimate the effect (and interaction) of these factors in language use by corpus studies (De Cuypere et al. 2017: 2). Constructions or phenomena of English (morpho-)syntax which have been studied extensively as syntactic variants (or under the

label ‘syntactic variation’) from various theoretical and methodological perspectives include the following: the genitive alternation (e.g., Heller et al. 2017), the analytic vs. synthetic comparative constructions (e.g., Mondorf 2009), particle placement in phrasal verbs (e.g., Gries 2003), zero vs. *that*-complementiser (e.g., Shank et al. 2016), inversion (e.g., Kreyer 2006), the double object construction or dative alternation (e.g., Goldberg 1995; Bresnan and Ford 2010), double modals (Hasty 2014), pied piping vs. preposition stranding (e.g., Hornstein & Weinberg 1981; Hoffmann 2005), and negative concord (Blanchette 2013, 2017).

As claimed before, all of these constructions may also be analysed from the perspective of syntactic (non-)canonicity when the focus is on the fact that one or some of the syntactic alternatives are more expectable than others in specific communicative situations. In this volume, we go beyond Standard British or American English in our analyses of syntactic (non-)canonicity and include as databases of our studies other varieties, such as Learner Englishes, English as a Lingua Franca (ELF), L2 English, and various forms of World Englishes. We agree with the variationist assumption that, in a specific communicative situation, a range of factors and the interaction between them may condition the use of a non-canonical (or less canonical) syntactic variant. We further acknowledge that, depending on the approach to (non-)canonicity, the canonical variant(s) may be defined differently and that, in frequency-based approaches to (non-)canonicity, depending on the nature and scope of the database, these factors condition the placement of all alternatives available in a specific communicative situation on the gradient between syntactic canonicity and syntactic non-canonicity. It is in these frequency-based approaches to (non-)canonicity that the syntactic canon is elusive and, as stated before, needs to be redefined relative to each specific communicative situation (or group of recurring situations).

With regard to these challenges, we understand equivalence in meaning as equivalence in propositional meaning, while discourse-pragmatic meanings enter the general picture as part of the aforementioned extra-linguistic factors. Further, constructions of different degrees of formal similarity may present themselves as candidates for a study as to syntactic (non-)canonicity. Thus, a *wh*-cleft and a corresponding non-cleft sentence may be regarded as obvious syntactic alternatives (*This book is what I wanted* vs. *I wanted this book*), but in a given communicative situation even a clause with a fronted constituent (*This book I wanted*) or a clause with existential *there* (*There is this book that I wanted*) may fulfil similar functions and may thus present themselves as further syntactic alternatives, although they

have more in common with the non-cleft sentence. Moreover, we claim that in theory-based approaches a canonical alternative does not necessarily have to be available in a specific communicative situation for one construction to be considered non-canonical (and vice versa). For example, as mentioned before, a theory-based approach may define *it*-extraposition as a non-canonical deviation from SVX, the canonical order of sentence constituents. In a given context, however, structural factors such as sentence type, verb complementation, or weight may make extraposition obligatory and thus preclude the canonical alternative (cf. Quirk et al. 1985: 1049, 1062). Similarly, not all clauses with *there*-insertion have a corresponding structurally related counterpart without *there* (cf. Quirk et al. 1985: 1406). In frequency-based approaches, by contrast, while constructions can be described as syntactic variants if a wide range of contexts is considered, only one viable option may remain if all relevant factors are taken into consideration. Thus, for example, particle placement in transitive phrasal verbs is known to be influenced by factors such as the stress pattern and idiomaticity of the phrasal verb, the form of the object, its length, complexity, information status, register, and mode (spoken vs. written) (Gries 2003: 161–2). It is well known that with pronominal objects the V-O-particle ordering is the only option (**Sigrid looked up it*) but, similarly, the aforementioned factors and their interaction might also render the V-O-particle ordering unacceptable in other contexts. This shows that considering syntactic constructions as more or less canonical options not only places the emphasis on what is more or less expectable from the perspective of language users; it also permits the researcher to define the scope of individual studies more flexibly, in a broader or narrower way depending on their perspective or research interest. Finally, our broad and flexible definition of (non-)canonicity also permits us to unite in this volume theory- and frequency-based approaches, as well as their methodologies. We believe that this will lead to a better understanding of what constitutes syntactic (non-)canonicity and syntactic variation, establish new and interdisciplinary ways of approaching syntactic variation in English in all its forms and functions, and inform our understanding of syntactic structure and the nature of (non-)canonicity in general.

1.4 Structure of the Volume

The contributions to this edited collection assess the merits of applying frequency-based and theory-based approaches to (non-)canonicity by contextualising the concepts within empirical case studies related to

(a) historical varieties, (b) register-based varieties, and (c) non-native varieties of English. The linguistic features investigated in the chapters match the above broad definition of what may count as canonical and our outline of possible deviations from the canon (cf. Section 1.2). As the chapters illustrate, the factors that influence to what extent a given syntactic feature may be considered ‘non-canonical’ are manifold. The structure of the volume illustrates three of the most important factors: time (what is canonical changes over time), register (what is canonical changes based on textual variables), and the interplay between region and acquisition (what is canonical changes based on where and how English is acquired and used). Thus, canonicity serves as a heuristic that brings together different case studies set in individual compositions of intra- and extra-linguistic factors at work; at the same time, the case studies explore the value of applying a framework with a canonical–non-canonical continuum at its core to these different scenarios.

Following this Introduction, **Sven Leuckert** and **Sofia Rüdiger** continue to set the terminological and theoretical stage by investigating usages of terms such as ‘canonical’/‘non-canonical’ and ‘standard’/‘non-standard’ in six linguistic journals. This approach to investigating the frequency and usage of linguistic terminology is still rare but fills a relevant gap, since the empirical chapters that make up the rest of the volume continue to build on what ‘canonical’ and ‘non-canonical’ may offer conceptually in usage-based analyses of syntactic constructions.

Part I of the volume offers empirical studies of non-canonical syntax in historical varieties of English. The syntax of English has changed dramatically from its early days to the present day, and this also means that what would have been considered ‘canonical’ in OE or Middle English (ME) may not necessarily be considered canonical in one of the later periods. This part of the volume begins with an introduction to key directions and open questions in the field by **Marianne Hundt**. In the next chapter, **Gea Dreschler** investigates full-verb inversion as a potential continuation of the late subject pattern from OE and ME. A particularly interesting question posed by this chapter is to what extent this construction was already ‘non-canonical’ in earlier stages of English. Next, **Claudia Lange** uses the *Old Bailey Corpus 2.0* to analyse existential *there*-constructions in Late Modern English. Her case studies focus on tokens which do not manifest the default formal marking of internal relationships, first on default singulars with plural notional subjects and then on coordinated noun phrases (e.g., *bread and cheese*) and so-called notional plurals. These are nouns such as *crowd* which are singular in number but have more than one referent,

which leads to variation in their verbal agreement patterns. Finally, **Louise Mycock** and **Sharon Glaas** study ProTags such as *that* in *Spooky, that*, *that*, *that* is, pronouns which are attached to the right of a clause (or more generally, C-Unit) and which, unlike right-dislocated elements, do not have a clarificatory function. Their study focuses on Early Modern English, using the *Chadwyck–Healey English Drama Collection* as database to study a pattern which is non-canonical both from a theory- and from a frequency-based perspective.

Part II of the volume offers empirical studies of non-canonical syntax in register-based varieties of English. Research on non-canonicity in this context is strongly linked to text types, which, as the contributions to this part show, play a major role in the realisation of syntactic features. This part of the volume begins with an introduction by **Heidrun Dorgeloh** and **Anja Wanner**, who outline key directions and open questions in register-based varieties of English. The next chapter, by **Douglas Biber**, **Stacey Wizner**, and **Randi Reppen**, is an investigation of Non-Canonical Reduced Structures (NCRSs) in television news broadcasts, illustrated by sentences such as *First, to the Mueller interview itself*, in which various expected constituents may be missing. Based on a corpus analysis, the authors show that NCRSs are significantly more frequent in TV news broadcasts than in other registers, which leads to a noticeable increase in textual complexity and permits us to question their status as non-canonical in this particular register. In the next chapter, **Teresa Pham** considers different types of clefts in evaluative language as represented in a diverse corpus of reviews published, for instance, in Lonely Planet travel guides and on Airbnb. Clefts, as illustrated by *The thing that Sigrid loves is linguistics*, are an important structuring tool of evaluative language. The statistical analysis, which includes cleft-related and evaluation-related variables, reveals that a range of factors influence the type of cleft construction chosen for a given purpose. In fact, most clefts are explicitly evaluative, which suggests that they should be regarded as part of an extended set of lexico-grammatical stance constructions. In the final chapter of this part of the volume, **Christine Günther** considers particle placement and links cognitive complexity and non-canonicity. The study follows an approach typical of cognitive and psycholinguistics and uses two experiments to investigate the non-canonical feature illustrated by the discontinuous variant *They looked the topic up*, which contrasts with the canonical, continuous variant *They looked up the topic*. Günther finds that non-canonical particle placement may have the additional function of reducing cognitive complexity, which takes it beyond what is commonly assumed in the literature.