

Genre-dependent Discourse-Topic Persistence in English Inversion

HEIDRUN DORGELOH

HEINRICH HEINE UNIVERSITY - DÜSSELDORF

Submitted: 17/11/2023 Revised version: 31/07/2025

Accepted: 03/08/2025 Published: 13/11/2025



Articles are published under a Creative Commons Attribution 4.0 International License (The authors remain the copyright holders and grant third parties the right to use, reproduce, and share the article).

Abstract

Variation in English constituent order is typically explained by information packaging and, with that, by the relation of the construction to the discourse preceding it. This paper presents an empirical approach to the study of the construction of English inversion, focusing on its role in subsequent discourse and its relation to the discourse topic. Building on work on other constructions that are commonly considered to be topic-marking, a dataset of inversions from two genres of the COCA corpus was investigated. The research questions addressed are (i) whether, in the sequence of the two constituents that are reversed by way of an inversion, the fronted, non-subject element or the grammatical subject of the sentence should be considered the topic of the sentence, and (ii) to what extent the topical structure following the inversion varies with the register or genre. Since genres typically differ in text structure, the second research question is also a reality check for the first. Outcomes show that looking at inversion from the point of view of discourse-level topic persistence rather than information packaging adds an important insight to an understanding of its function. In addition, an analysis with respect to discourse- rather than information structure helps to account for genre differences in the use of inversion across discourse types. On a methodological level, the paper illustrates an approach that operationalizes the retrieval of inversions from a corpus and the analysis of discourse-topic persistence.

Keywords: subject-verb inversion; information structure; topic persistence; genre; COCA corpus.

1. Introduction

In a language with rigid SV-constituent order, such as English, inversion refers to a sentence pattern in which the grammatical subject is positioned after the verb, while an element of the verb phrase is found in the clause-initial position (Ward, Birner & Huddleston 2002: 1385). In English, one of the few typical properties of the subject function is its position before the verb. As a result, this deviant constituent order is attributed the status of a non-canonical construction (Dorgeloh 1996, 1997; Huddleston & Pullum 2002: 46). Deviating from the basic, i.e., canonical order, where the subject is placed before the verb, leads to a non-basic way of packaging information, which “can only be fully accounted for by taking the surrounding discourse into consideration” (Dorgeloh & Wanner 2023: 3).

To start with an example of the construction, the inversion in (1) can be contrasted with the corresponding sentence in basic (SV) order in (2). Both sentences are equally grammatical in English, but (1) suggests the context of a text passage in which the fronted constituent *the fence* is likely to directly connect to the previous discourse. In all likelihood, the preceding text will have contained an element such as *backyard* or *garden*, i.e., some element inducing the contextual accessibility (Ariel 1990) of the fence.

(1) I looked up. *Peering over the fence was the new boy who had just moved with his family into the house behind ours.* (COCA, Fiction, 2009)

(2) I looked up. The new boy who had just moved with his family into the house behind ours was peering over the fence.

Constructions with non-canonical word order, like the inversion in (1), are standardly explained by information packaging (e.g., Ward, Birner & Huddleston 2002), i.e., by the principle that given information tends to precede new information in a sentence. When this principle is applied to inversion, the prediction would be that it has a distribution of information where the fronted constituent represents information that is recoverable from the preceding discourse, while the postposed subject will contain information that is new (e.g., Birner & Ward 2009; Ward & Birner 2019). This prediction has been discussed in various work on inversion (e.g., Birner 1994; Dorgeloh 1997; Kreyer 2006), although in a hedged version: at a minimum, an

inversion does not occur if the information of the fronted constituent is less familiar than the information of the postposed subject (Birner 1994: 244).

While such analyses of inversion take into account its relation to the previous discourse, there is also work emphasizing that inversion in English is basically a presentative construction and that its discourse function is therefore one that is also relevant for the subsequent discourse (Dorgeloh 1997; Kreyer 2006; Prado-Alonso 2011, 2014). In that respect, looking back at (1), it is not so much the (relative) discourse familiarity of *the fence* that makes the inversion an adequate choice here with view to its surrounding discourse but, in all likelihood, it is also the relation of the postposed subject (*the new boy*) to the upcoming discourse. As becomes obvious when looking at the larger context of (1), given in (3) below, the narration after the inversion turns to the referent of the postverbal NP, which here recurs as the subject of the sentence following the non-canonical construction.

(3) Pushing his brown hair out of his eyes, he said, "Don't you hate getting dirty like that?"

Example (1), together with its context in (3), illustrates a typical pattern of using inversion where the construction induces a change of interest towards a new referent, here of a new character. This function of inversion is described in the literature as one of "focus management," which results from topic shifts (Dorgeloh 1997: 105-116; also, Kreyer 2006, 2010). However, what is not made explicit in these accounts of inversion is the nature of topicality that is being addressed. For example, a standard reference grammar of English (Biber et al. 2021: 888) equals given information with "topical" information, which means the NP in the non-subject constituent of an inversion (*the fence* in (1)) would be its (non-subject) topic. Other grammars standardly identify topic as the discourse-oriented notion of the subject, of "what we are talking or writing about" (e.g., Börjars & Burridge 2019: 226), i.e., they follow the quite standard approach of equating the topic with the referent the proposition is about (e.g., Reinhart 1981; Gundel 1988; Lambrecht 1994). Following that kind of analysis, the topic of the inversion in (1) would be the syntactic subject (*the new boy...*).

With the study presented here I will argue that inversion in English is particularly prone to the factors concealed by the concept of topic. Shifting away from the view of topicality as a unitary and universal phenomenon (Jacobs 2001, and other papers

in this special issue) opens the way to examining the specific functions of diverse constructions and uncovering the sources of the effects associated with topicality. Instead of applying a pre-empirical, theoretical notion of topicality to language-specific data, we explore directly the properties of English inversion analysing the structure together with its discourse context. This allows for contrasting and interrelating the sentence-level and discourse functions of inversion, as both are evoked in different definitions of topicality and its analysis.

The analysis that will be presented focuses on the status of both elements of the inversion by investigating their relation to the subsequent discourse. This analysis aims to explore whether the inversion operates as a sentence-level backgrounding, “topicalizing” device for the initial material (as suggested, e.g., by Biber et al. 2021), and/or whether its function is rooted in attributing discourse properties to the referent of the postverbal grammatical subject. As such, it simultaneously teases apart and links the sentence-level information packaging processes and discourse structure (RQ1). The analysis also focuses on genre differences, since genres most typically differ in their text structures (Biber & Conrad 2019). It will be argued here that specific genres are the adequate level for categorizing discourse when looking at grammatical and discourse patterns closely. Since the literature on inversion finds its occurrence to vary considerably with the type of discourse, it is an open question to what extent the function of inversion varies by the genre, as some studies might suggest (e.g., Kreyer 2006; Prado-Alonso & Acuña-Fariña 2010) (RQ2). On a more general level, the paper argues for a relatively fine-grained look at effects associated with topicality, highlighting that both the construction of inversion, the syntactic types included in the analysis as well as the genres the data are taken from require the choice of an adequate “degree of specificity” (Dorgeloh & Kunter 2015) for the analysis.

2. Background

2.1. Inversion in English

2.1.1 Definition and types of inversion

English inversion results from an argument reversal, i.e., from the displacement of two sentence constituents. While a non-subject, often semantically locative argument

is placed sentence-initially, the syntactic subject (as judged by its cross-reference on the verb) ends up in sentence-final position (Birner & Ward 1998; Ward & Birner 2006; Webelhuth 2011). The resulting English sentence pattern is variably described in the literature as "locative" or "stylistic" inversion (Bresnan 1994; Culicover & Levine 2001; Webelhuth 2011) and, apart from the non-canonical word order, has also some other grammatical characteristics: there must be an intransitive verb, which is most typically the verb *be*, the sentence cannot carry negation, and the grammatical subject must not be an anaphoric pronoun (cp. Webelhuth 2011: 83).

Depending on the phrase type of the sentence-initial element, there are in principle five syntactic patterns of inversion: inversion following a fronted *prepositional phrase* (PP-inversion), an *adjective phrase* (AdjP-inversion), a *verb phrase* (VP-inversion), a *noun phrase* (NP-inversion), or an *adverb phrase* (AdvP-inversion). Examples (4) – (8) are attestations that illustrate each of these types:

- (4) Below the aristocracy were various dependents, including employees [...] and slaves. (COCA, Academic, 1992) (PP-inversion)
- (5) More painful were the personal things - like the photo albums that recorded his parents lives and his own. (COCA, Fiction, 2003) (AdjP-inversion)
- (6) Following the directions was a scale representing the Likert choices, 1 to 7. (COCA, Academic, 1993) (VP-inversion)
- (7) Now is the time to cut spending. (COCA, Spoken, 2013) (AdvP-inversion)
- (8) An exception to this is the Test of Narrative Language. (COCA, Academic, 2017) (NP-inversion)

Two of the sentence patterns illustrated in (4) through (8) need to be excluded for the analysis applied to the corpus data below. In (7), the reversal of subject and verb follows a fronted deictic adverb (*now*), not a fully specified, lexical verbal argument. This sentence pattern is the “presentative prototype” of the construction (Dorgeloh 1997: 67), and as such it cannot be analysed as containing a potential topic. Adv-inversion is a case of discourse-topic introduction (Kreyer 2006: 165ff.) typical ofthetic sentences, to which presentatives belong.

By contrast, NP-inversion, like the one in (8), is excluded because it contains two NPs: an ascriptive NP (*an exception to this*) and the syntactic subject (*the Test of Narrative Language*). Such NP-inversion is “equative” rather than predicative (cp., e.g., Mikkelsen 2005; Geist 2007; Haspelmath 2017), i.e., the presence of two NPs and the

copula verb *be* means that the position of subject and initial element could also be reversed, i.e., the initial phrase could be equally analysed as the subject. Without closer analysis, taking into account properties of the NPs such as their specificity or definiteness, NP-*be*-NP sequences could thus variably be analysed as a case of inversion or not, which would in turn raise the question whether NP-inversion links a topicality effect to the use of a marked construction at all. For this reason, only PP-inversion, AdjP-inversion, and VP-inversion were included in the analysis of competing topics for this study.

2.1.2 Inversion as a discourse phenomenon

All patterns of full inversion are described as serving the discourse function of "information-packaging" in the literature (e.g., Ward, Birner & Huddleston 2002; Ward & Birner 2019; Biber et al. 2021; Dorgeloh & Wanner 2023). Since a basic principle of constituent ordering in an English sentence is that known information typically precedes new information, the initially placed constituent of an inversion can be expected to contain information that is more familiar, either from the discourse, or for the hearer or reader, than the information expressed by the subject. This assumption has been confirmed in a wide range of corpus work (e.g., Birner 1994; Kreyer 2006): by far the most common information structure of inversion is one where the information of the fronted constituent is recoverable, while the postposed subject contains information that is relatively new. For example, the PP-inversion in (4) was preceded by two sentences on *the aristocracy*, as shown in (9):

- (9) Wealth was enough to guarantee influence. Nobles acted as managers of large firms and controlled most resources [...]. Below the aristocracy were various dependents, including employees [...] and slaves.

The literature also points out that there are cases where both the fronted constituent and the postposed subject contain some kind of given information. However, it is highly unlikely that the information in the fronted constituent is less familiar than the information of the postposed subject (Ward, Birner & Huddleston 2002: 1387; Dorgeloh & Wanner 2023: 75). In addition, there is a difference in information structure between the so-considered "locative" types of inversion, i.e., PP- and VP-inversion, such as (4) and (6), and "non-locative" inversion, like AdjP-inversion. Non-

locative inversion, like the one in (5), does not only follow the constraint of 'no new before given information', but requires that the entire sentence, except for the postposed subject, expresses at least contextually inferable, and preferably given, information. This familiar rest is considered an "open" proposition, which must be, as a whole, derivable from the preceding discourse (Ward, Birner & Huddleston 2002: 1388). For example, in (5), which is shown in (10) together with some of its preceding text, the proposition *something was painful* is already given information, as derivable from the discourse preceding the inversion.

(10) But that was only the superficial loss. More painful were the personal things...

This difference in what is given information in locative and non-locative inversion types also indicates that the relative discourse familiarity of individual constituents from an inversion cannot directly suggest what could be considered as its topic. It therefore makes the construction a useful candidate for disentangling topicality effects from givenness, which is what the remainder of this paper will aim for.

2.2. Topicality and topicalizing constructions in English

2.2.1 Inversion and topicality

Generally, the concept of "topic" is variably be used to refer (i) sentence-level topics, or (ii) to discourse-topics, while in the former case the term often conflates both pragmatic referents and their linguistic expression. Moreover, the pragmatic, sentence-level notion has a broad range of definitions, rooted in diverse types of interrelation between a unit of information and the proposition. Finally, a sentence-topic and discourse-topic can be different within the same example. In particular, a case relevant to this study is that a new discourse topic can be introduced or established as such by a focal constituent on a sentence level. Resulting from this range of definitions that variably refer to the sentence or discourse, the "topicality" of a sentence constituent can either mean *being* the topic or *having a certain likelihood* of being the topic (cp. Orita, Vornov & Feldman 2021: 745). In the latter sense, the notion becomes measurable and, with that, something that can be subjected to an empirical investigation beyond local sentence interpretation. This clearly comes with

the caveat that the actually measured property is not topicality of the referent itself, but its contextual accessibility as its proxy.

In a language with subject-first order, inversion and the resulting complement verb subject ordering always comes at a certain price (Perez-Guerra 2012). Placing the subject non-canonically after the full verb in principle allows for two directly competing analyses. One is to make the fronted constituent function as the topic-expression. The other view is to identify the topic in an inversion as the non-canonically placed grammatical subject, thus suggesting that this is the referent the proposition is about. As aboutness is an elusive interpretive effect, operable limits of its application remain vague. However, if we want topic to be a meaningful notion, it should be rooted in an independent factor of communicative structure, and not be merely derived from its position or grammatical category. As a result, none of these two views of what constitutes the topic in an inversion contributes to a proper understanding of the actual phenomenon, beyond labelling it with an available term. As a result, the case of inversion highlights that, in a language with rigid word order like English, argument reversal can plausibly only be approached in the sense of (ii), i.e., investigating the likelihood of being a topic in the current proposition due to the discourse state. Topicality does then not refer to a property of a construction or one of its components, but is a discourse phenomenon identifiable through pragmatic properties of information in actual language data.

2.2.2 Other topic-marking English constructions

There are additional constructions in English typically treated as topic-marking, or topicalizing: (NP-)fronting, also often referred to as "preposing", left-dislocation, or left-detachment (e.g., Prince 1984; Ward 1988; Geluykens 1992; Lambrecht 1994; Gregory & Michaelis 2001; Leuckert 2017, 2019). These constructions can be described as placing the topic in front of the comment, i.e., as showing a specific "pragmatic relation" (Lambrecht 2001: 1066) of an initially placed constituent to the rest of the predication.

Similar to inversion, fronting and left-dislocation are non-canonical in the English syntactic system, in that the preposed NP is a non-subject verbal argument preceding the subject. In fronting, it is just the NP that is fronted while, in left-dislocation, an anaphoric pronoun fills the gap resulting from the movement of the NP. Both constructions are described as *topicalizing*, again reflecting either that given

information is equalled with being topical, or that fronted information is found to be associated with highlighting a referent (Geluykens 1992: 158). However, research on the two constructions similarly shows that they do not only mark topics in the sense of given, identifiable information, but possess more varied discourse functions (e.g., Prince 1981; Ward & Birner 2019; Leuckert 2019; Cimmino 2023; Cimmino & Saccone this volume).

Examples of fronting and left-dislocation that serve topicalization are given in (11) and (12), constructed examples derived from (8) as their imagined prior discourse. Both might be sentences that could easily follow the inversion in (8). By contrast, (13) and (14) are both quite unlikely, indicating already the role of the postposed subject in an inversion as a potential new discourse topic.

- (11) The Test of Narrative Language, I don't know. (NP-fronting/topicalization)
- (12) The Test of Narrative Language, I don't know it. (left-dislocation)
- (13) ?An exception to this I don't know. (NP-fronting/topicalization)
- (14) ??An exception to this I don't know it. (left-dislocation)

It is pointed out in the literature that fronting, in contrast to left-dislocation, requires a particularly close link to its prior discourse. The construction either places an NP with given information in sentence-initial position, or the fronted element carries new information, in which case the link to prior discourse exists via an "open proposition" (Huddleston & Pullum 2002: 35), which means that everything in the clause, except for the fronted element, is inferable or given information. Example (11) would thus only be felicitous if the fronted NP was just mentioned before, or it could occur in a context where it is already clear that the speaker denies knowing something, with only *The Test of Narrative Language* being highlighted through fronting¹. By contrast, left-dislocation is described in the literature as a construction that results from a more general need to negotiate or clarify a topic (Geluykens 1992). So, for example, in (12) the speaker might take up again ("re-activate") the topic of some *Test of Narrative Language*, which could either have been mentioned a while ago or could otherwise be "to some extent inferable" (Dorgeloh & Wanner 2023: 63). There is thus no overall, common function of the sentence-initial, non-canonically placed constituent in fronting and left-dislocation which could easily be related to topicality.

¹ Cases like these are also described as focus preposing.

Both preposing and left-dislocation are primarily constructions of spoken language use and, in that respect, differ from inversion, which typically occurs in written and "displaced", i.e., narrative or descriptive, discourse (e.g., Dorgeloh 1997, 2006; Kreyer 2006, 2010; Dorgeloh & Wanner 2023: 77-80). Corpus studies on preposing and left-dislocation have typically sought for evidence of topic- and discourse-management functions of the two constructions in dialogue discourse. Pioneering work of that kind is a study by Gregory & Michaelis (2001), whose methodology is applied to the study of inversion based on the data presented below.

In their study of preposing ("topicalization") and left-dislocation, Gregory & Michaelis (2001) equal topic with the status of persistent referents in the sense of topic continuity. They define topic persistence referring to the "cataphoric discourse status of the denotatum of the pre-clausal NP" (Gregory & Michaelis 2001: 21), i.e., as the recurrence of a referent in the subsequent discourse. This approach also takes note of the condition whether the recurrence of a referent takes place in the form of a full NP or a pronoun, since discourse varies widely in how forms of referring expressions recur within it (e.g., Gundel et al. 1993). Following their approach, topic persistence as a measurement looks at whether the referent of a pre-clausal NP recurs within a given number of subsequent sentences, which could be anything from ten, as in Givón (1984), to five, as in the study by Gregory & Michaelis (2001)². On a conceptual level, topic persistence in this approach reflects that the referent of an NP has been established in discourse, i.e., the approach manages to substantiate the status of referent in discourse, making it measurable. However, in that referential continuity and topic structure can be syntactically expressed in a variety of ways, the actual realization of those patterns is still variable and likely to vary with the register or genre.

2.3. Topicality and Genre/Register Variation

While register variation is generally associated with functional, situation-dependent variation in language use, genre variation focuses on features that are conventional and/or typical, including features that one only finds by looking at an entire text (cf. Biber & Conrad 2019; Dorgeloh & Wanner 2020: 655). Resulting from that distinction, register characteristics are typically features that are pervasive, whereas genres also

² For reasons explained below, for this study the number of subsequent sentences was reduced further to only two.

have features belonging to their rhetorical structure or textual organisation (Biber & Conrad 2019). When going into differences among texts in terms of topic structures, it is likely that genre characteristics, rather than pervasive lexico-grammatical properties of a register as a whole, turn out to be relevant.

For example, the kind of narration and handling of the topic shift exemplified by (1) is less likely a feature of fiction as a whole, but it is plausibly a characteristic of a certain kind of novel, including its perspective, narrative technique, etc. Although corpora such as the COCA corpus (Davies 2008), which is the source of the set of inversions investigated here, are in principle structured along registers (with sections such as fiction, spoken, or academic³), looking at them in terms of their properties of genre will mean to focus on their text organisation.

Information structuring processes can vary across spoken and written registers (Cimmino 2023). Referents in speech are likely to be negotiable and negotiated, often introduced by a separate turn marked by intonation, or co-constructed. By contrast, written genres require other strategies and inversion is known to be among them (e.g., Kreyer 2006: 172-188; Prado-Alonso & Acuña-Fariña 2010; Prado-Alonso 2014). However, within the written register, there is still a lot of syntactic, genre-driven variation (e.g., Dorgeloh & Wanner 2010, 2020), and a common place of variation is the sentence-initial position. For example, in an analysis of that position as "proxy for register variation", Pérez-Guerra (2021) finds that any first-element in the sentence, rather than any preverbal element, is the best metric for register distinctions. As a result of this analysis based on sentence beginnings, two main clusters of registers turn out to be particularly relevant: popular, informal ones and learned, formal ones. Each of these clusters is also addressed by the two genres chosen for the corpus study presented below: academic discourse representing the learned/formal register and fictional discourse the popular/informal one.

2.4. Hypotheses

This study aims to highlight that both the construction of inversion, rather than the sentence-initial position as such, as well as the register or genre have their role in topicality effects. While in preposing and left-dislocation the front position of a constituent may indicate (or even trigger an interpretation of) its givenness,

³ This applies to all corpora on [English.corpora.org](https://english.corpora.org).

aboutness, or status as an established discourse topic, inversion involves the non-canonical position of two sentence elements, of which the placement of the syntactic subject is possibly the more relevant one. In addition, the register or genre is an important variable since it is plausible to assume that factors commonly regarded as constituting topic conditions (such as givenness, prominence, or animacy) are likely to vary, for example, in academic compared to fictional texts. Similar to differences across languages, disentangling topicality effects therefore also means choosing the right “degree of specificity” when defining a construction and its discourse function, to arrive at suitable generalizations (Dorgeloh & Kunter 2015).

More concretely, the analysis of inversion will test the following hypotheses:

- i. In the sequence of the two constituents affected by the argument reversal of inversion, the referent of the sentence-final syntactic subject, rather than the one of the fronted verbal complements, provides the sentence-level, "aboutness" topic. In analogy with other constructions dubbed topic-marking, such as fronting and left-dislocation, which have been described as "Left Marked Structures" (LMSs), PP-inversion, AdjP-inversion, and VP-inversion could in such case be considered Right Marked Structures.
- ii. The information structuring function of inversion varies with the genre as it plays an important role in discourse organisation. Effects associated with the notion of topicality are likely to differ in syntactic realization (e.g., noun vs. pronoun) as well as in the nature of the semantic relation among persistent referents (e.g., co-reference, part-whole relations, hyponymy).

3. Method

The dataset of inversions analysed for the present study was obtained from the Corpus of Contemporary American English COCA (Davies 2008), with an equal number of attestations coming from the two genres of ‘fiction’ and ‘academic discourse.’ The dataset retrieved consists of 530 validated observations of inversions following fronted prepositional phrases (PP-inversion), adjective phrases (AdjP-inversion) and verb phrases (VP-inversion) (Otte 2021). The dataset was retrieved using various lexical shortcuts for PP-inversion and AdjP-inversion, based on a limited set of prepositions (e.g., *above*, *between*, and *in front of*) as well as adverbs modifying

adjectives (e.g., *equally*, *also*, and *more/most*)⁴. The study was limited to the straightforwardly searchable verb *be*, excluding other intransitive verbs with which inversion occasionally occurs. The resulting search strings were a combination of these lexemes and a corresponding word-class tag for an NP, or adjective, followed by *be*. For VP-inversion, the search was limited to sequences of a present participle followed by *the* + noun followed by a form of *be*. All these sequences were further limited, using punctuation, to their occurrence at sentence beginnings in order to avoid too many false positives, e.g., of a sequence of PP + *be* within the sentence, but not with an inversion, as illustrated by (15)). All attestations were subsequently cleaned manually.

- (15) The specific paths among the variables are depicted in the structural model in Figure 2. (COCA, Academic, 2014)

The attestations were coded for the recurrence of the referent either of an NP within the fronted element or of the finally placed subject in the discourse following the inversion. The analysis equalled "topic status" with "topic persistence", as a measure that was operationalized as the number of times a referent recurs in subsequent discourse (Givón 1984: 908; see also Gregory & Michaelis 2001). In Givón's original work, that condition was defined as a recurrence "within subsequent 10 clauses following the current clause" (Givón 1984: 908), but this was reduced here to only two sentences, since the sentences from the two genres turned out to be substantially longer. This analysis was applied to the total of the 530 validated inversions (200 for PP-inversion, 200 for AdjP-inversion, and 130 for VP-inversion).

Examples (16) through (20) illustrate the four degrees of topic persistence for which the dataset was coded, taking into account that the choice of a (pronominal vs. lexical) NP reflects its accessibility in discourse (Ariel 1988). The recurrent referent that followed the inversion could either be a co-referential pronoun, as in (16) (persistence SCORE 3), a co-referential NP, as in (17) and (18) (persistence SCORE 2), or an anchored NP (persistence SCORE 1). The category of co-referential NPs also included referents in a part-whole relation, such as *parent* and *mother* in (18). Anchored NPs involved the lowest degree of topic persistence in that a given lexeme

⁴ For a detailed description of how to search for marked constructions with lexical shortcuts, i.e., using specific lexemes, in a corpus and, specifically, how to retrieve a set of inversions, see Dorgeloh & Wanner (2023: Ch. 3).

or concept, though not necessarily the same referent, could recur in the subsequent discourse, as in (19). Example (20) exemplifies a case of no topic persistence. Note that the referential relation was found either in the head of the two elements, like in (16) (*various dependents* - *they*), or in the relation of any dependent element within the two phrases, like in (17) (*a Mozart opera* - (*a portion of*) *the Mozart recording*), (18) (where the NP *his parents* occurs within a postmodifying PP), and (19) (*sales* being the complement of a postposed subject clause):

- (16) Below the aristocracy were various **dependents**, including employees [...] and slaves. The status of each depended on their position in the larger system, each was attached to a specific noble or noble section, and **they** had varying degrees of freedom. (COCA, Academic, 1992) [TOPICALITY SCORE 3: **co-referential pronouns**]
- (17) Among the albums is **a Mozart opera** [...]. He plays a portion of **the Mozart recording**, setting the stylus on a duet sung by two women. (COCA, Academic, 1997) [TOPICALITY SCORE: 2: **co-referential NPs**]
- (18) More painful were the personal things - like the photo albums that recorded **his parents** lives and his own. **The mother** he hardly remembered. (COCA, Fiction, 2003) [TOPICALITY SCORE 2: **coreferential NPs in part-whole relation**]
- (19) Also uncertain is whether premiums and other insurance-related income constitute "**sales**" for this purpose. Apparently, start-up companies that do not yet have **any sales** (...) are also excluded. (COCA, Fiction, 2007) [TOPICALITY SCORE 1: **anchored NP**]
- (20) Among the boulders are 30 caves inhabited by Buddhist monks beginning in the third century B.C. The final ascent to the top began at the Lion Staircase. The site's name derives from Simha-giri or Lion Mountain [...]. (COCA, Academic, 2002) [TOPICALITY SCORE 0: **no recurrence**]

Although we use numbers for the score assignment, the used scale is ordinal and not a ration scale. In other words, score 3 is higher than score 1, but is not thrice higher.

For the case of AdjP-inversion, a difficulty in coding was that, almost always, the fronted AdjP did not contain a proper NP for which a referent could be determined, as also obvious in (18) and (19). Still, due to the semantics of the modifying adverb preceding these inversions, the fronted AdjP could be analysed as containing an

implicit referent. Like in (18) and (19), the inversion here follows a sequence of linking or comparing adverb (*also, more*) + adjective, which is a pattern that has been observed in AdjP-inversion in general (cf. Dorgeloh & Kunter 2015). Such a linking or comparing AdjP carries a meaning of 'likeness,' which can be interpreted as being referential: 'a thing cannot just be 'like'; it must be 'like something' (Halliday & Haisan 1976: 78; see also Dorgeloh 1997: 114). For example, in (18), there is something known from the previous discourse that is less *painful*; in (19), there is something that is also *uncertain*. In view of these implicit referents, it was in principle possible to determine the persistence of such topics implicitly contained in the fronted AdjPs even if, in practice, none of these implicit referents derivable from AdjPs in the dataset recurred as topic within the subsequent discourse.

4. Results

4.1. Topic persistence of both inverted elements

With respect to the first hypothesis, the analysis first looks at the likelihood of the two constituents of the inversion to become a discourse topic in the subsequent text. Table 1 and Figure 1 provide the overall distribution, signalling two main findings. First, it becomes obvious that, based on the overall dataset of 530 attestations, 452 fronted constituents (85.3%) do not show any persistence of an (explicit or implicit) referent from the fronted phrase in the subsequent discourse. By contrast, a comparatively higher likelihood of persistence is observable for the subject NPs in the dataset. 62.5% of the syntactic subjects that are placed in clause-final position contain some kind of referent recurring in the subsequent discourse. There is thus only a small likelihood of the fronted constituent to become the topic of the subsequent discourse and a significantly higher likelihood for the postposed subject.

Topic persistence	Fronted NP	Subject NP
as pronoun	3	77
as NP	99	155
anchored	26	99
none	452	199

Table 1: Four values of topic persistence in 530 inversions from COCA.

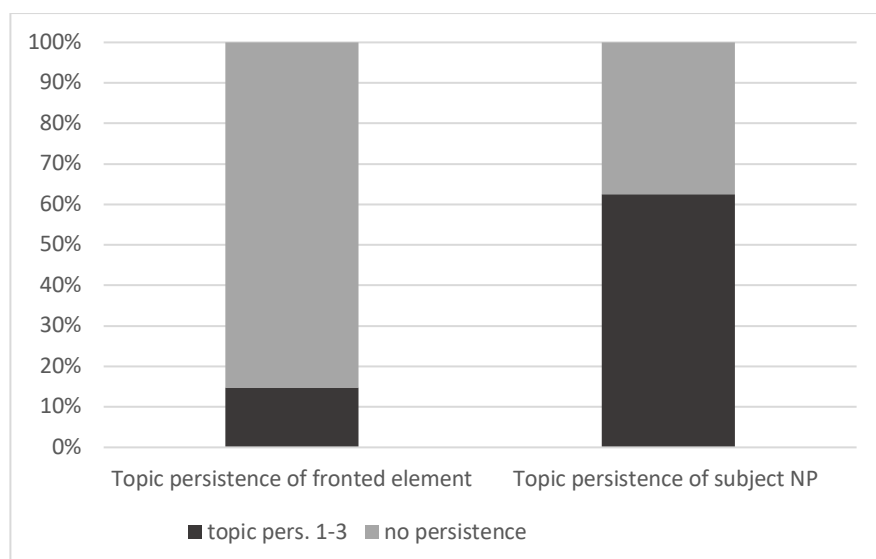


Figure 1: Topic persistence of Subject vs. Fronted NP in 530 inversions from COCA ($p < .05$, chi-square value at 254.83).

Interestingly, the picture does not get much different when looking at the three syntactic types of inversion separately. As Figure 2 indicates, the likelihood of recurring in the subsequent discourse does not differ substantially when comparing PP-inversion, AdjP-inversion, and VP-inversion. The proportion of instances with no topic persistence is slightly higher for VP-inversion (for a possible explanation, see Dorgeloh 2023), but that difference is clearly below statistical significance. In sum, all the postposed subjects in all types of inversion can be noted to have a substantially higher likelihood of persisting as discourse topics than the sentence-initial elements of the corresponding sentences.

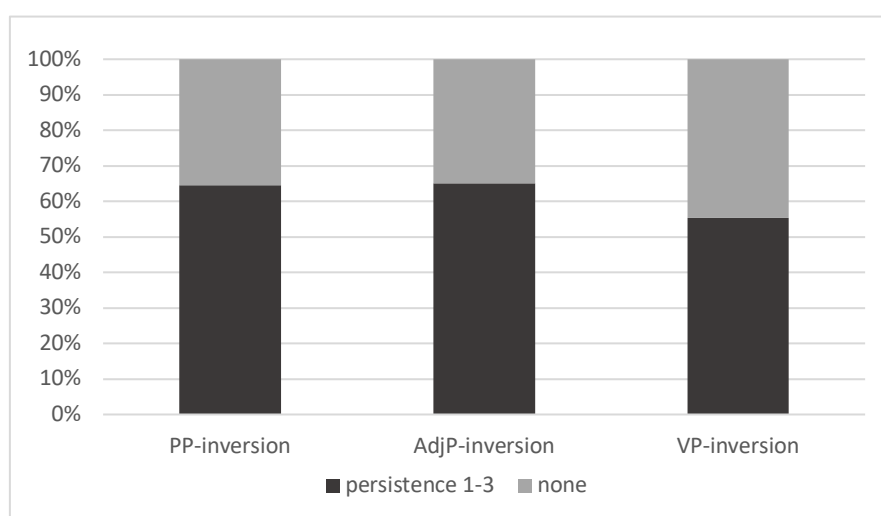


Figure 2: Topic persistence of subject NP in three types of inversion from COCA ($p > .05$, chi-square value at 3.6805).

4.2. Genre-dependent topic persistence

This section looks at two more variables that may reveal discourse persistence effects in the use of English inversion. One is the effect of genre per se, i.e., the question whether the function of inversion related to the discourse topic is dependent on the genre. The other aspect of inversion explored here is *how* the observed discourse persistence is realized.

The overall effect of the corresponding corpus sections is surprisingly streamlined: there is almost no difference between the two genres of academic texts and fiction when it comes to the likelihood of the non-canonically placed syntactic subject to recur in the subsequent discourse. As the absolute numbers in Table 2 as well as the proportional frequencies in Figure 3 indicate, the likelihood of the subject NP to become a persistent discourse topic is near to identical in each of the two genres (61% in fiction; 64% in academic texts).

Topic persistence	Academic	Fiction
as pronoun	15	62
as NP	76	79
anchored	71	28
none	103	96

Table 2: Topic persistence of subject NP in 530 inversions from two genres.

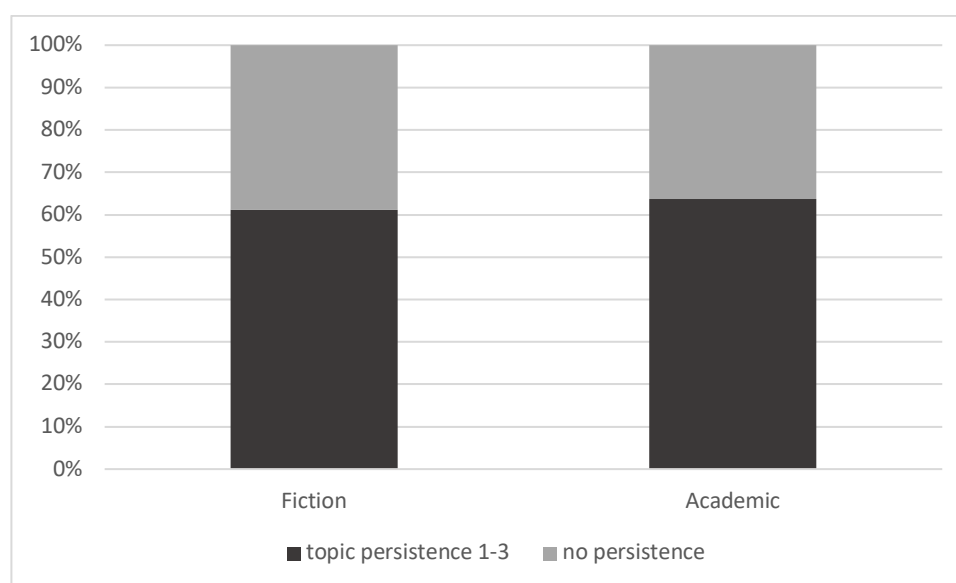


Figure 3: Topic persistence of subject NP in two genres from COCA (n = 265 per genre).

In addition, Figure 3 again shows that the likelihood of the subject NP in an inversion to become a discourse topic is overall two to one. Given the effort of a non-canonical construction, this may still seem remarkably low. In that light, the fact that we find no effect of the genre is all the more surprising, given that academic and fictional discourse are sure to vary considerably in the nature of the referents they deal with and in their resultant text structures. This is why a closer look at the semantic relations across referents and at the syntactic realizations of the persistent discourse topics is needed.

Figure 4 is based on the different types of referential relations that exist between a recurring referent and the preceding subject NP (as described in Section 3). In contrast to the outcomes in Figure 3, a comparison of these different relations shows a pronounced genre effect. While the amount of co-referential NPs, like in the examples (16) and (17) above, is almost identical in the two genres, the realization of discourse persistence in the form of an NP that is semantically anchored and of a co-referential pronoun differs substantially between genres. Academic discourse favors a relation of anchored NPs, while in fiction pronouns predominate. This finding confirms the expectation that there is a clear difference in how discourse topics are handled in fiction compared to academic texts: while the former, typically narrative texts, tend to possess a substantial referential continuity, e.g., a continuity of characters and places, academic texts typically build on more implicit semantic relations, such as hyponymy or co-hyponymy.

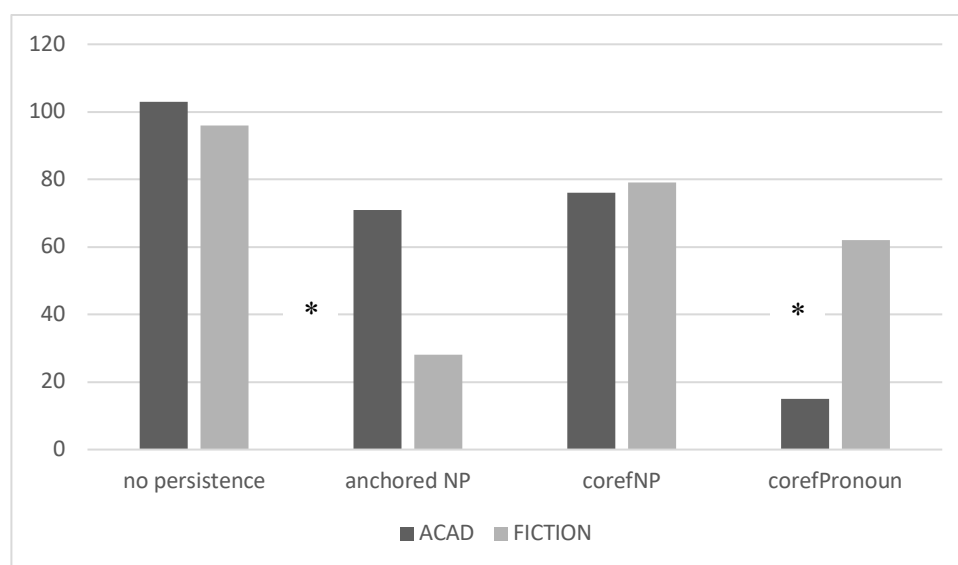


Figure 4: Referential relations for topic persistence of subject NPs from inversion in two genres
(* = significant at $p < .05$).

The fact that we first found no genre difference when looking at topic persistence in the two corpus sections indicates that it is crucial to find the right level of analysis for discourse persistence effects. We observed inversions in both genres to foreshadow discourse persistence of the expressed referent. Inversion is thus obviously right-, rather than left-, marking. What varies with the genre is the precise semantic and syntactic relation of the construction to its surrounding discourse. The differences that can be observed here make its discourse-structuring effects perhaps less obvious, but also highlight its dependence on the nature of the discourse organisation.

5. Discussion and conclusion

The close analysis of a comparatively large set of natural attestations of inversion retrieved from a corpus has shown that the discourse function of the construction is related to topicality in the sense of a documented likelihood of its subject to remain a discourse topic in the subsequent text. In that sense, the construction of inversion shapes the notion of topicality in the discourse perspective, which could be measured as the persistence of referents from the subject NP of an inversion in subsequent discourse. Despite the reversal of two verbal arguments, the discourse function of inversion does not primarily have to do with the sentence-initial, or theme, position, and it is right- rather than left-marking.

While this function is supported by a discourse persistence of subject NP referent that is clearly above 50% and, with that, beyond a random distribution, the significantly low persistence of preposed elements (slightly above 10%) is also telling. Their low likelihood of recurring after the inversion indicates that the inversion by itself signals low persistence, thus a discourse topic shift. The fact that both locative and non-locative inversions show similar patterns in that respect suggests that the construction, as it was defined here, is the right level of specificity for studying these effects, commonly associated with topicality. Despite recent findings on similarities with PP-fronting, or "PP-topicalization" (Bruening 2022), the analysis presented here thus reveals important differences rather than similarities of the two constructions.

Still, similar to topicalization in left-marking constructions, the syntactic realization of discourse topics is also varied if performed by a right-marking construction and shows plausible differences in terms of register specificity (cf. Pérez-Guerra 2021). As the syntactic subject of an inversion must always be a full NP, the

construction is likely to function as a referent introduction structure, signalling the referent's non-trivial contribution in the forthcoming discourse. However, the variability in the further realization of this referent is considerable. This variability is where we expect to see genre characteristics of referential continuity.

To deepen our understanding on the role of inversion in discourse processing, further research could be conducted on a more fine-grained variety of text types within the macro-categories of "Fiction" and "Academic writing". In that respect, other types of descriptive and narrative discourse are certainly "natural habitats" for the use of English locative inversion (Webelhuth 2011: 99; also, Dorgeloh 2006). Moreover, varieties of World English could be taken into account; language contact may indeed lead to alternative patterns or trends, revealing a different variation in terms of register and genre.

Overall, an analysis of English inversion beyond its information structure does not question the relevance of information-packaging accounts. Both the relation of the construction to the previous and to the subsequent discourse plausibly belong to the presentative mechanism that has long been known to connect full, lexical inversion to a deictic presentative construction of the type *here is...* or *now comes...* (Dorgeloh 1997: 67-68). Being able to substantiate both accounts with corpus data, applying an adequate operationalization and categorization, helps to fully explore the work of inversion in discourse processing and its role as a device of information management.

Acknowledgements

Many thanks go to Kendra Otte for providing the original dataset, to Ann-Sophie Haan and Chiara Webers for their help with the coding of the inversions, and to the reviewers for their comments and suggestions. My special thanks go to the two editors of this volume for their patience and for their support.

Abbreviations

COCA = Corpus of
Contemporary American
English

AdvP = adverb phrase
AdjP = adjective phrase
NP = noun phrase

PP = prepositional phrase
VP = verb phrase

References

- Ariel, Mira. 1988. Referring and accessibility. *Journal of Linguistics*, 24(1). 65–87.
- Ariel, Mira. 1990. *Accessing NP antecedents*. London: Routledge.
- Biber, Douglas & Susan Conrad. 2019. *Register, Genre, and Style*. Cambridge: Cambridge University Press.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 2021. *Grammar of Spoken and Written English*. Amsterdam & Philadelphia: John Benjamins.
- Birner, Betty J. 1994. Information status and word order: An analysis of English inversion. *Language*, 70(2). 233–259.
- Birner, Betty J. & Gregory Ward. 1998. *Information Status and Noncanonical Word Order in English*. Studies in Language Companion Series. Amsterdam: John Benjamins.
- Birner, Betty J. & Gregory Ward. 2009. Information Structure and Syntactic Structure. *Language and Linguistics Compass*, 3(4). 1167-1187.
- Börjars, Kersti & Kate Burridge. 2019. *Introducing English Grammar*. London: Routledge.
- Bresnan, Joan. 1994. Locative inversion and the architecture of Universal Grammar. *Language*, 70(1). 72–131.
- Bruening, Benjamin. 2022. Locative inversion, PP Topicalization, and Weak Crossover in English. *Journal of Linguistics*, 58(4). 739-757.
- Cimmino, Doriana. 2023. On the topic-marking function of Left Dislocations and Preposings. Variation across spoken and written Italian and English. In Simone Mattioli & Alessandra Barotto (eds.), *Discourse phenomena in typological perspective*, 337-368. Amsterdam & Philadelphia: John Benjamins.
- Cimmino, Doriana & Valentina Saccone. 2025 (this volume). Beyond the topic-marking discourse function of Left Dislocations. Evidence from Italian, Spanish, and English spoken corpora. *Linguistic Typology at the Crossroads*.
- Culicover, Peter W. & Robert D. Levine. 2001. Stylistic inversion in English: A reconsideration. *Natural Language & Linguistic Theory*, 19(2). 283-310.
- Dalrymple, Mary & Irina Nikolaeva. 2011. *Objects and Information Structure*. Cambridge: Cambridge University Press.
- Davies, Mark. 2008. *The Corpus of Contemporary American English (COCA): One billion words, 1990–2009*. www.english-corpora.org/coca/.

- Dorgeloh, Heidrun. 1996. The cognitive status of marked syntax: English inversions. In Kazimierz A. Sroka (ed.), *Kognitive Aspekte der Sprache. Akten des 30. Linguistischen Kolloquiums, Gdansk 1995*, 55-60. Berlin, Boston: Max Niemeyer Verlag.
- Dorgeloh, Heidrun. 1997. *Inversion in Modern English: Form and Function*. Amsterdam & Philadelphia: John Benjamins.
- Dorgeloh, Heidrun. 2006. Inversion in descriptive and narrative discourse: A text-typological account following functional principles. *Cahiers de Recherche*, 9. 101–114.
- Dorgeloh, Heidrun. 2023. Who's leading the pack? Verb phrase inversion beyond information packaging. In Jasmin Pfeifer, Sabine Arndt-Lappe, Heidrun Dorgeloh, Gero Kunter & Christian Uffmann (eds.), *INGO 6.0: The Proceedings. New empirical Insights about language*, 113-127. In press.
- Dorgeloh, Heidrun & Gero Kunter. 2015. Modelling adjective phrase inversion as an instance of functional specialization in non-locative inversion. In Christina Sanchez-Stockhammer (ed.), *Building bridges into the future: Can we predict linguistic change?* Studies in Variation, Contacts and Change in English 16. Helsinki: VARIENG.
- Dorgeloh, Heidrun & Anja Wanner. 2010. *Syntactic Variation and Genre*. Topics in English Linguistics 70. Berlin/New York: Mouton de Gruyter.
- Dorgeloh, Heidrun & Anja Wanner. 2020. Genre Variation. In Aarts Bas, Jill Bowie & Gergana Popova (eds.), *The Oxford Handbook of English Grammar*, 654-672. Oxford: Oxford University Press.
- Dorgeloh, Heidrun & Anja Wanner. 2023. *Discourse Syntax: English Grammar Beyond the Sentence*. Cambridge: Cambridge University Press.
- Geist, Ljudmila. 2007. Predication and equation in copular sentences: Russian vs. English. *Existence: Semantics and Syntax*. Dordrecht: Springer Netherlands. 79-105.
- Geluykens, Ronald. 1992. *From Discourse Process to Grammatical Construction*. Amsterdam: John Benjamins.
- Givón, Talmy. 1984. *Syntax: A functional-typological introduction*. Volume I. Amsterdam/Philadelphia: John Benjamins.
- Gregory, Michelle L. & Laura A. Michaelis. 2001. Topicalization and left-dislocation: a functional opposition revisited. *Journal of Pragmatics* 3(11). 1665–1706.

- Gundel, Jeanette K. 1988. Universals of topic-comment structure. In Moravcsick Hammond (ed.), *Studies in Syntactic Typology*, 209-242. Amsterdam/Philadelphia: John Benjamins.
- Haspelmath, Martin. 2017. Equative constructions in world-wide perspective. *Similitive and equative constructions: A cross-linguistic perspective*. 9-32.
- Halliday, Michael A.K. & Hasan Ruqaiya. 1976. *Cohesion in English*. English Language Series. London: Longman.
- Huddleston, Rodney & Geoffrey K. Pullum. 2002. *The Cambridge Grammar of the English Language*. Cambridge: CUP.
- Jacobs, Joachim. 2001. Two dimensions of topic- comment. *Linguistics*, 39(4). 641– 681.
- Kreyer, Rolf. 2006. *Inversion in Modern Written English: Syntactic Complexity, Information Status and the Creative Writer*. Tübingen: Narr.
- Kreyer, Rolf. 2010. Syntactic constructions as a means of spatial representation in fictional prose. In Heidrun Dorgeloh & Anja Wanner (eds.), *Syntactic Variation and Genre*, 277-303. Berlin/New York: Mouton de Gruyter.
- Lambrecht, Knud. 1994. *Information structure and sentence form: Topics, focus, and the mental representations of discourse referents*. Cambridge: Cambridge University Press.
- Lambrecht, Knud. 2001. A Framework for the Analysis of Cleft Constructions. *Linguistics*, 39. 463-516.
- Leuckert, Sven. 2017. Typological interference in information structure: The Case of Topicalization in Asia. *Zeitschrift für Anglistik und Amerikanistik*, 65(3). 283-302.
- Leuckert, Sven. 2019. *Topicalization in Asian Englishes: Forms, Functions and Frequencies of a Fronting Construction*. London: Routledge.
- Mikkelsen, Line. 2005. *Copular clauses: Specification, predication and equation*. Vol. 85. Amsterdam: John Benjamins.
- Orita, Naho, Eliana Vornov & Naomi H. Feldman. 2021. Informativity, topicality, and speech cost: comparing models of speakers' choices of referring expressions. *Discourse Processes*, 58(8). 743-765.
- Otte, Kendra. 2021. Beyond information-packaging: Inversion and topicality. Unpublished MS.
- Ozerov, Pavel. 2021. Multifactorial Information Management (MIM): summing up the emerging alternative to Information Structure. *Linguistics Vanguard*, 7. 402-418.
- Pérez-Guerra, Javier. 2012. Discourse status and syntax in the history of English. In Anneli Meurman-Solin, Maria Jose Lopez-Couso & Bettelou Los (eds.), *Information*

- Structure and Syntactic Change in the History of English*, 122-138. Oxford: Oxford University Press.
- Pérez-Guerra, Javier. 2021. Theme as a proxy for register categorization. In Douglas Biber & Elena Seoane (eds.), *Corpus-based Approaches to Register Variation*, 85-110. Amsterdam: John Benjamins.
- Prado-Alonso, Carlos. 2011. Text structuring in written English: The role of inversion. *English Studies*, 92(4). 449–463.
- Prado-Alonso, Carlos. 2014. Verb phrase inversion in fictional and non-fictional written English. *English Studies*, 95(2). 177-194.
- Prado-Alonso, Carlos & Carlos J. Acuña-Fariña. 2010. A comprehensive account of full-verb inversion in English. *Folia Linguistica*, 44(2). 509-554.
- Prince, Ellen, 1984. Topicalization and left-dislocation: A functional analysis. *Annals of the New York Academy of Sciences*, 433. 213-225.
- Reinhart, Tania. 1981. Pragmatics and Linguistics: an analysis of Sentence Topics. *Philosophica*, 27. 53-94.
- Ward, Gregory. 1988. *The Semantics and Pragmatics of Preposing*. New York: Garland.
- Ward, Gregory, Betty J. Birner & Rodney D. Huddleston. 2002. Information packaging. In Rodney D. Huddleston & Geoffrey K. Pullum (eds.), *The Cambridge Grammar of the English Language*, 1363–1449. Cambridge: Cambridge University Press.
- Ward, Gregory & Betty J. Birner. 2006. Information Structure and Non-canonical Syntax. In Laurence R. Horn & Gregory Ward (eds.), *The Handbook of Pragmatics*. 152-174. Malden: Blackwell Pub.
- Ward, Gregory & Betty J. Birner. 2019. *Discourse effects of word order variation*. In Paul Portner, Claudia Maienborn & Klaus von Stechow (eds.), *Semantics - Sentence and Information Structure*, 413-449. Berlin: De Gruyter Mouton.
- Webelhuth, Gert. 2011. Motivating non-canonicity in Construction Grammar: The case of locative inversion. *Cognitive Linguistics* 22(1). 81-105.

CONTACT

dorgeloh@hhu.de