Cartography and Locality in German: a quantitative study with dependency structures

Giuseppe Samo

Beijing Language and Culture University

Abstract. This paper investigates theoretical considerations on cartography and locality (Rizzi & Cinque 2016) with respect to grammatical structures in German, adopting quantitative tools and large-scale syntactically annotated corpora. German has represented an interesting case in the literature allowing restrictions (bottleneck effect) and freedom of movements (scrambling) of the constituents in different areas of the structures. Both phenomena can be explained in terms of locality (Samo 2019), which is quantitatively tested here. After having presented a model for mapping cartographic projections into universal dependencies (Nivre 2015), a quantitative study is carried out. Results confirm that the more dissimilar in terms of classes of features (Rizzi 2004) two elements are, the higher the frequency of an element lower in the structure in crossing an intervener. These results aim to add a quantitative dimension to the qualitative descriptions provided in cartographic studies.

Keywords: Cartography, Locality, German, Quantitative, Dependencies

1 Introduction

In this work, I will adopt quantitative tools to investigate the interactions between frequency and locality principles adopting the guidelines of the Cartography of Syntactic Structures (Cinque & Rizzi 2010, Rizzi & Cinque 2016). This paper targets two goals. The first aim is to investigate whether the predictions provided by cartographic studies concerning the loci of generation result in frequencies extracted from syntactically annotated treebanks. The second research question focuses on how locality principles interact with the computation of grammatical structures naturally occurring in corpora.

To verify these dimensions, a mapping of cartographic functional projections to syntactic annotation schemata is required. In this work, the reorderings of syntactic constituents are retrieved from a set of treebanks annotated under the guidelines of the Universal Dependencies (henceforth UD, Nivre 2015, Zeman, Nivre & Abrams 2020).

Following Belletti (2018), the displacement of syntactic units is observed as a phenomenon involving a dependency relation whereby a constituent is interpreted simultaneously in two different positions. The hierarchically lower position is the referred to as the (base-)generation position, whereas the hierarchically higher one, tendentially overt at phonological form, is labelled as a landing site for movement. Let us consider the *canonical* case of a wh-question in English, given in (1).

(1) *Which book are you reading <which book>?*

The sentence in (1) contains an interrogative element *which book* occurring at the very beginning of the clause, guarantying that the clause is to be interpreted as a question on the object in English; at the same time, the "copy" of the interrogative element, in hook brackets, is interpreted as an element of the argumental structure. The constituent *which book* is externally merged in the vP layer and then internally merged in a dedicated left peripheral position hosting interrogative elements to allow the interpretation "for which x, you are reading". Functional projections hosting moved elements are referred to as scope-discourse phenomena (e.g. topic structure like, relative clauses, etc.) and discussed in detail in the cartographic literature since earlier seminal works (Rizzi 1997).

Studies in syntactic cartography (Cinque & Rizzi 2010, Rizzi & Cinque 2016) have provided formal tools to understand reorderings in the light of syntacticization (Rizzi 2013). The derived syntactic maps, beyond their heuristic capacity of describing linguistic variability are designed to offer transparent configurations for interpretative routines to be applied on, ultimately calling forth to the syntacticization of semantics, pragmatic and prosodic properties (Rizzi 2013).

This paper outlines a system of mapping cartographic projections to UD schemata[1], in order to add a quantitative dimension to the established qualitative descriptions. The quantitative results are to be read in the spirit of Quantitative Computational Syntax (Merlo 2016): differentials in counts are the expression of underlying grammatical properties (Merlo 2016, Samo & Merlo 2019). The analysis here is developed on a series of treebanks of German of different genres and sizes, annotated under the guidelines of UD.

1. Similar mappings of UD has been proposed, such as the transformation of UD into logical forms for semantic parsing, Reddy et al. 2016).

Dimensions of variation in the analysis of the reordering of constituents shall be observed: quantitative syntax is able to provide further insights on the nature of theoretical proposals, in terms of maps and principles.

In §2, I shall present the tools I adopt for the analysis, namely cartographic maps and definitions of reorderings and locality. In §3, I develop a model of transformation of syntactic dependencies into a series of cartographic projections. In §4, I present the empirical data extracted from a series of UD-treebanks from German. Finally, §5 summarizes and concludes.

2 Reordering and cartographic maps in German

2.1 Reordering, cartography and locality

One of the most easily observable distinguishing features of natural languages is word order (Greenberg 1975 *inter alia*), the order in which the verb and its core arguments occur[2]. Typological results[3] assume that languages adopt one of the factorial combinations of the core elements Verb (V), Subject (S) and Object (O) as "canonical"[4]. Following Kayne (1994) and Cinque (1999), I add also non-core constituents like adverbials and modifiers as part of the "canonical" order of natural languages.

With the term "canonical", I refer to as the standard ordering of constituents in which informational properties are clause related or about the subject (Rizzi 2015; Belletti & Rizzi 2017).

Every reordering of the "canonical" order leads to (i) an ungrammatical sentence or (ii) a *grammatical* (semantico-pragmatic) marked option. Let us observe the case of a SVO language like English: while a structure as in (1) is a case of a *grammatical* reordering triggered by discourse

2. I will not discuss the notion of *unfixed* word order (Dryer 2005).

3. Relevant literature concerning word orders could be found in the relevant pages of the website of *The World Atlas of Language Structure (WALS)* (https://wals.info/, Dryer & Haspelmath 2013, accessed 19.06.2020)

4. I propose that in every language, for every canonical word order, constituents undergo two main types of movements from their generation locus. A first type of movement is a *meaningless movement* (Cinque 2018), that I here interpret as the movement to reach the canonical orders. A second type of movement is a *meaningful movement*, triggered for discourse requirements (Rizzi 1997, e.g. Topic, Focus). I will refer to as this type of movement as criterial movement. There are also intermediate types of movement, such as the movement of verbal chunks in the IP (Belletti & Rizzi 2012) and or movements in relation with other phenomena such as smuggling (Collins 2005). The analysis here focuses only on criterial movements.

requirements (a question on the object), a reordering having the object displaced between the subject and the inflected verb (e.g. **You which book are reading?*) results in an ungrammatical sentence.

The function of the displaced element, the loci involved in the displacement and other principles have been crosslinguistically investigated in syntactic theory.

The cartography of syntactic structures (Cinque & Rizzi 2010, Rizzi & Cinque 2016) observes reorderings in terms of syntacticization (Rizzi 2013): functional projections provide instructions at the interfaces with the systems of sound and meaning (Rizzi 2013, Bocci 2013). In other words, cartographic maps describe a typology of functional projections within the syntactic architecture. The analysis here focuses on two types of functional positions. Positions where syntactic elements are generated and positions where syntactic elements can move to.

Here, *generation positions* represent the loci where syntactic elements are generated[5] (cf. the position in hook brackets in 1), whereas *landing sites* describe those functional projections where elements move to satisfy criterial requirements (Rizzi 2007, 2015) and give instructions to the systems of meaning and sound for interpreting the moved element in terms of the appropriate notion and/or through the assignment of the appropriate intonational contour (Bocci 2013). The *grammatical* orders resulting from this displacement are labelled here as ”uncanonical”[6].

Following a the criterial approach to scope-discouse semantics (Rizzi 2007), a syntactic element respectively externally and internally merges in at least two positions dedicated to two kinds of interpretative properties, as given in (2).

(2) Merging positions in criterial configurations
  i. A *generation position*, where
    a. the element is “generated”,
    b. the verb assigns an argument role to the ”generated” element;
  ii. a *landing position*, where
    a. the element moves to,
    b. the properties of scope-discourse semantics of the *moved* element are interpreted at the interfaces with the systems of sound and the systems of meaning (Bocci 2013).

5. Discussions on how elements enter syntax are given in Rizzi (2015, 2016, 2017) in terms of, respectively, labelling, merge systems and parameters.
6. Throughout this work, I shall use the term ”uncanonical” over ”marked” (see Haspelmath 2006 for a discussion on the role of the term *markedness*).

I here adopt a series of functional sequences that will turn relevant for the analysis of the theoretical prediction presented in §3 and the empirical data discussed in §4.

As for the Complementizer Phrase (CP), I will focus only on the positions hosting topicalized elements (TOPIC), focalized constituents (FOCUS) and "highlighted" adverbials (MOD, Rizzi 2004), discussed in the map of the fine structure of the Left Periphery (henceforth LP) discussed in Rizzi (1997 2001, 2004) and Rizzi & Bocci (2017), given in (3).

(3) The fine structure of the Left Periphery (Rizzi & Bocci 2017)
[FORCE [TOPIC [INT [TOPIC [FOCUS [TOPIC [MOD [TOPIC [QEMB [FIN [IP [...]]]]]]]]]]]]

Different types of subjects target different positions in the syntactic architecture. The highest position of the Inflectional Phrase (IP) hosting subjects (Rizzi 2007, Rizzi 2015) is "obligatory" realized. This positions satisfies the subject criterion, which expresses the classical EPP feature (Rizzi 2007) in cartographic terms. As for the landing site(s) of "canonical" subjects, I will discuss SUBJ and EPP adopting a cartography of subject positions proposed by Cardinaletti (2004), given in (4). Furthemore, EPP has also been analysed as a locus of generation of expletives (cf. Vikner 1995).

(4) Cartography of subject positions (Cardinaletti 2004)
[SUBJ [EPP [AGRS]]]

Questions of movement and interpretation of criterial heads are related to other domains of the syntactic architecture, such as the functional projections within the IP related to tense, mood and aspect of the utterance. As for this work, I adopt the map of invariant ordering clausal functional projections in (5), drawn from cross-linguistic evidence presented in Cinque (1999). Due to the nature of this work (which cannot provide an in-depth analysis of the semantic functions of the functional projections), the hierarchy will be discussed in §3 and §4 under the label ADV.

(5) Cinque (1999)'s hierarchy of adverbs
[MoodSpeechAct [MoodEvaluative [MoodEvidential [ModEpistemic [TPast [TFuture [ModIrrealis [ModNecessity [Modpossibility [AspHabitual [AspRepetetive [AspFrequentative(i) [ModVolitional [AspCelerative(i) [TAnterior [AspTerminative [AspContinuative [AspPerfect [AspRetrospective [AspProximative [AspDurative [AspGeneric/Progressive [AspProspective [AspSg.Completive(i) [AspPl.Completive [Voice [AspCelerative(ii) [AspRepetetive(ii) [AspFrequentative(ii) [AspSg.Completive(ii)]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]

Schweikert (2005) proposed an enrichment of Cinque (1999)'s hierarchy, adding a layer of complements and prepositional phrases, which is here presented in (6). Similarly to the hierarchy of adverbs, I will refer to these elements under a unique label Pp.

(6) Hierarchy of complements (Schweikert 2005: 132)
Temporal > Locative > Comitative > Benefactive > Reason > Source > Goal > Malefactive > Instrumental > Matter > Manner

Belletti (2004) has shown the existence of a further layer of scope-discourse (Topic, Focus) positions in the periphery of the vP area / low IP area, as given in (7). I will refer to these positions as LowIP.

(7) Low IP area (Belletti 2004)
[Top [… [Foc [… [Top]]]]]

Finally, I adopt a classical architecture of the layer vP, the locus of generation of arguments, as given in (8). As mentioned with respect to the hierarchies, the quantitative nature of this work does not discuss in detail the cartography of arguments within the vP (Ramchand 2008, Si 2019)[7] or the nature of the verb (e.g. intransitive). Here, only a minimal typology (subjects, objects and indirect objects) is provided. When the class of the element is not specified, I refer to these positions as ArgvP.

(8) vP
[External Argument(s) [v [... [V [Internal Argument]]]]]

---

[7] Similarly, fine-grained analysis of case marking have been provided in Caha (2009) and Zompì (2017).

Functional Projection	Type of locus
TOPIC (Rizzi 1997)	Landing Site
FOCUS (Rizzi 1997)	Landing Site
MOD (Rizzi 2004)	Landing Site
SUBJ (Rizzi 2007)	(Obligatory) Landing Site
EPP (Cardinaletti 2004)	Generation Site / Landing Site
ADV (Cinque 1999)	Generation site
PP (Schweikert 2005)	Generation site
LOWIP (Belletti 2004)	Landing Site
ARGVP	Generation Site

Table 1: Nature (generation site or landing site) of functional projections (and related references).

The maps presented from (3) to (8) are reconsidered in terms of loci of generation or landing sites in Table 1.

Given the relevant discourse requirements such as several topics in one sentence targeting different position in the tree, one might expect that languages would not allow "ungrammatical" patterns. Indeed, principles, such as locality (Rizzi 1990, Belletti 2018), reduce the "freedom" of movement syntactic elements in the map.

The theory of locality in terms of featural Relativized Minimality (henceforth, fRM Rizzi 1990, 2004; Starke 2001) has played a major role in the interaction between cartographic maps and basic operations of grammar such as movement[8]. In a nutshell, the landing position turns inaccessible, or less accessible, for an element to move to, if a similar element hierarchically intervenes between the generation site and the landing position. According to the theory of intervention in terms of fRM, the crucial property is not the amount of material that can be considered as intervener, but rather its quality.

If the syntactic constituent that has to be displaced and the intervener share sets features relevant for locality, ungrammatical structures (Rizzi 1990) or slower parsing effects crosslinguistically arise in adult grammars (Frauenfelder, Segui & Mehler 1980).

Interventions effects also show clear asymmetries in the comprehension and production of structures in typical development (Friedmann, Belletti & Rizzi 2009), atypical development (Durrleman et al. 2015) and in language pathology (Grillo 2008). Similar asymmetries are found in studies in quantitative computational syntax: Samo & Merlo (2019) extracted grammatical object relative clauses in large-scale corpora of

8. The interaction between maps and locality has been debated in the literature (Abels 2012, Rizzi 2013). The study of locality here might provide a "further explanation" for the functional sequences (Rizzi 2013: 213).

Italian and English showing that the observed counts of mismatch configurations of a set of morphosyntactic features were higher than the expected counts based on sentences where movement was not involved, while the observed counts of matching configurations were lower than expected.

Here, the notion of similarity is based on classes of features provided in Rizzi (2004: 243: 61). Due to the nature of this paper, I only focus on two classes of features, namely *argumental* and *modifier* features. The generation sites can therefore be classified in terms of features: ARGVP bear *argumental* features, whereas ADV and PP bear *modifier* features.

To sum up, languages do not exhibit random word orders, as *quantitatively* shown in classical "free word order" languages such as Latin and Ancient Greek on the basis of large-scale annotated treebanks (Gulordava & Merlo 2015). Natural languages have "canonical" orders and trends for a subset of "uncanonical" orders. These trends are regulated, among others, by the principles of locality.

I shall investigate whether it is possible to quantify this level of analysis in a series of treebanks from one language. German is a suitable candidate, since it has been described in the literature as a language extremely conditioned by locality effects: strict restrictions on multiple elements in the "left" of the clause (*bottleneck effect*), but a high degree of freedom in the "middle" and in the "right" of the clause (*scrambling*). These notions will be presented in §2.2.

2.2 Restrictions and freedom of constituents in German

Word orders in German have been investigated early on in the history of linguistics (since Erdmann 1886) and in-depth in generative syntax (Haider 2010 *inter alia*).

German is described as a verb second language (henceforth V2; Samo 2019 for a cartographic approach) in the sense that the inflected verb fills the "second" linear position in main clauses[9] following exactly one constituent. In the presence of a complex verbal forms (e.g. auxiliaries, modals, separable particles, etc.) only one part (the hierarchically higher) targets the *second* position, while the other remains in a position at the end of the structure. Different types of constituents (subjects, arguments, adverbials) bearing relevant discourse features may undergo movement to the left of the inflected verb creating the V2 structure, as given in (9).

9. A pure "linear" position second element can only be referred to as a byproduct of subjacent syntactic phenomena. As noted by Zwart (1992: 76), grammars are not sensitive to notions like "first" or "second".

(9) a. *Hänsel hat gestern das Buch gelesen.*
Hansel has yesterday the book read
'Hansel read the book yesterday.'

b. *Das Buch hat gestern Hänsel gelesen.*
the book has yesterday Hansel read

c. *Gestern hat Hänsel das Buch gelesen.*
Yesterday has Hansel the book read

The position of the verb has played an important role in the history of the descriptions of the syntax of German. Indeed, the functional projection that the verb targets (a position in the CP, since Den Besten 1983) represents an hallmark for the notions of restrictions and freedom of movement in German: the part of the structure before (higher than) the inflected verb (*Vorfeld* 'prefield') seems inaccessible to more than one syntactic constituent, whereas the section right after the inflected verb (*Mittelfeld* 'Middlefield') is depicted as a locus of extreme flexibility for the movement of syntactic elements.

As for the *Vorfeld*, only one constituent can target the syntactic position in front of the inflected verb. In other words, V3 orders, in which the verb targets the "third" linear position as those given in (10), are ruled out by the grammar of German[10].

(10) a. **Das Buch Hänsel hat gestern gelesen.*
The book Hansel has yesterday read

b. **Gestern Hänsel hat das Buch gelesen.*
Yesterday Hansel has the book read

c. **Das Buch gestern hat Hänsel gelesen.*
The book yesterday has Hansel read

Earlier accounts (Roberts 2004) expressed this violations in terms of a "bottleneck effect". The fronted element bearing a generalized EPP feature blocks the movement of further elements in the LP. On (literally) the other side (of the verb), the *Mittelfeld* is described as a locus of a phenomenon called *scrambling* (Lenerz 1977, Frey 2004, Hinterholzl 2006.), according to which syntactic constituents seems to be *freely* placed, as shown in (11), with different degrees of acceptability[11].

10. V3 orders discussed for German by Müller (2013) can be considered as superficial, since only one constituent has been fronted. According to Samo (2019: 95-101), superficial V3 orders in German could be the result of (a) one element extracted and the higher belonging to a higher clause, (b) a "big" DP, (c) a chunk of functional projections respecting Schweikert (2005)'s hierarchy moving as a unique XP to the left of the inflected verb.

11. I do not mark here a focus features on the item *dem professor* 'to the professor'

(11) a. *Der Student hat dem Professor das Buch gegeben.*
the.NOM student has the.DAT professor the.ACC book given
'The student gave the handbook to the professor.'

b. *Der Student hat das Buch* dem *Professor gegeben.*
the.NOM student has the.ACC book the.DAT professor given

c. *Das Buch hat der Student dem Professor gegeben.*
the.ACC book has the.NOM student the.DAT professor given

d. *Das Buch hat dem Professor der Student gegeben.*
the.ACC book has the.DAT professor the.NOM student given

e. *Dem Professor hat der Student das Buch gegeben.*
the.DAT professor has the.NOM student the.ACC book given

f. *Dem Professor hat das Buch der Student gegeben.*
the.DAT professor has the.ACC book the.NOM student given
(Samo 2019: 60-62; 30-35)

Recently, Samo (2019) proposes that these two phenomena, "bottleneck effect" and scrambling, might be explained by the same principle of locality. This (criterial) approach claims that the moved element targets a criterial position in both the *Vorfeld* and the *Mittelfeld*.

The fronted element in the *Vorfeld* targets its criterial position in the Left Periphery (SpecTopicP, SpecFocusP, SpecModP) according to the features it bears and creates a Spec-Head configuration with the inflected verbal head.

Scrambled elements share properties with criterial configurations, since they are not felicitous answers in out-of-the-blue contexts (Lenerz 1977) and, as noted by Müller & Sternefeld (1993), are strictly clause-bound. This latter point is translated in terms of criterial freezing in Samo (2019).

Following a criterial approach, *scrambled* elements may target criterial positions both in the Left Periphery to the right of the inflected verb and in the Low IP area discussed in Belletti (2004). A criterial approach to V2 and *scrambling* predicts restrictions on the "freedom" of movement of constituents. The ungrammaticality of certain reorderings are due to standard fRM effects in both *Vorfeld* (building on the lack of multiple topics in English, Rizzi 2013) and the *Mittelfeld*. As shown in (11), reorderings seems however accepted given the right featural configurations of *scrambled* elements.

This work focuses on the frequencies of reorderings. Is it possible to have quantitative insights on scrambling? Does similarity play a role in

and the relative negative tags. An in-depth analysis of these patterns is given in Samo (2019: 60-63).

”uncanonical” orders? Are frequencies correlated with the loci of generation of syntactic constituents?[12].

A transformation of cartographic maps into UD queries in order to use large-scale syntactically annotated corpora is presented in §3, while §4 shall investigate empirical data extracted from syntactically annotated treebanks of German.

3 Mapping cartography to dependency structures

3.1 Towards a *Quantitative Cartographic Syntax*

This work is inspired by the framework of Quantitative Computational Syntax (Merlo 2016 and related works) which uses large-scale resources and simple computational models in order to answer quantitative linguistic questions and provide new linguistic insights. Frequency, intended here as differential in counts, is representative of underlying grammatical properties. It is expected that the observed counts (frequency) in grammatical structures with intervention effects involving similar elements should appear with a lower frequency than the observed counts of reordering involving dissimilar elements. In other words, the closer the similarity, the less frequent an element will cross another element which is higher in the structure.

To gather data from large-scale resources, a translation of cartographic configurations in syntactic annotation schemata is required. In this work, I translate cartographic maps in terms of Universal Dependencies (henceforth UD, Nivre 2015, Zeman, Nivre & Abrams 2020). UD offer a rich set of treebanks from hundreds of languages, within a variety of text genres, annotated under the same guidelines. UD represent thus a tool for both intralinguistic investigations (in terms of registers, e.g. Haegeman 1990) and crosslinguistic analyses.

In this respect, cartographic results might shed light on discourse elements in corpora, only relying on syntactic clues without any prosodic information, as it is the case of reorderings of ”canonical” orders.

Using large scale corpora adds a quantitative dimension to the qualitative analysis of functional projections provided in cartographic studies.

12. I leave aside a discussion on expected and observed counts in intervention effects in the spirit of Samo & Merlo (2019). This type of analysis would add a computational side in future research on reorderings, cartography and locality in German.

3.2 Mapping Universal dependencies to cartographic projections

The UD syntactic annotation (based on De Marneffe et al. 2014) presents 37 Universal syntactic relations[13]. The guidelines suggest that only overtly realized strings of texts should be annotated. That means that, for example, in a null subject language like Italian (Rizzi 1982), no subject dependency will appear if the annotated sentence does not have an overt subject.

Only the subset of nominal and modifier relations at the clausal level are considered here. A preliminary analysis on the language under investigation in §4 (German) restricted our set of investigation to six dependencies[14], namely active subjects[15] (in UD annotation labelled as *nsubj*), objects (*obj*), further arguments and complements (*iobj*, *obl*), adverbial elements *advmod* and expletive (*expl*).

Syntactic dependencies have three main components. A governor, a dependent and a direction. For example, a SVO structure in English has the following configuration: two arrows leaves from the root represented by the lexical verb. The relation *nsubj* is dependent to the left, while the relation *obj* is dependent to the right.

If the "canonical" word order changes, the dependencies vary their directions. In the case that the internal argument (object) targets the LP as in (1), both arguments *nsubj* and *obj* are dependent to the left. The combinations of these three features (governor, dependent and direction) in one language provide "canonical", "uncanonical" and ungrammatical orders.

However, there is no direct one-to-one mapping between cartographic functional projections and syntactic universal dependencies. Not every functional projection corresponds to a specific dependency: a syntactic dependency might represent sets of functional projections (e.g. *expl* or *iobj* as it will be discussed in this section) or no dependency represents non-overt elements of grammar (e.g. null subject Rizzi 1982).

I here present the six syntactic relations and their interactions with cartographic maps. I start with argumental elements (*obj*, *iobj*), then move to modifiers (*obl*, *advmod*) and finally discuss the dependencies related to the subject criterion (*nsubj*, *expl*). Table 2 shows a preliminary mapping in terms of generation and landing sites. The discussion here

13. https://universaldependencies.org/u/dep/index.html (accessed 20.06.2020)
14. After a preliminary analysis, due to the very small frequency in the treebanks under investigation in §4, the dependencies of *dislocated*, *vocative* and *discourse* were not analysed further.
15. Here only active structures are investigated, since passive structures imply a further layer of analysis involving the movement of the object and the verb under a smuggling approach (cf. Collins 2005).

Dep	Generation	Landing site
obj	ARGVP	LOWIP, TOPIC, FOCUS
iobj	ARGVP	LOWIP, EPP, TOPIC, FOCUS
obl	PP	LOWIP, TOPIC, FOCUS, MOD
advmod	ADV	TOPIC, FOCUS, MOD
nsubj	ARGVP	LOWIP, (obligatory) SUBJ, TOPIC, FOCUS
expl	EPP	

Table 2: Universal dependencies syntactic relations (Dep) and functional projections according to the generation or landing sites.

is based on the grammar of German.

obj The syntactic dependency *obj*[16] represents internal arguments in transitive/ditransitive verbs. The same dependency is used both for objects in their generation position[17], where minimal discourse properties take place, and for displaced objects. Mapping *obj* to the cartographic maps, it represents one VP argument(ARGVP), a Topic or Focus in the Low IP area (LOWIP) or in the LP (TOPIC, FOCUS).

iobj The relation *iobj*[18] represents other nominal phrases part of the argumental structure of the verb than subjects and objects (e.g. recipient of ditransitive verbs). Theoretically speaking, *iobj* elements represent other argumental features generated in the vP layer (ARGVP). Movement to both peripheries is allowed if the element bears relevant features to target the positions of Topics of Focus (LOWIP, TOPIC, FOCUS). Other elements annotated as **iobj** are, for example, experiencers of impersonal psych-verb constructions (Mohr 2005), as the natural occurring example extracted from one of the treebanks which will be presented in §4, given in (12).

(12) *Mir drängt sich aber der Eindruck auf*
me.DAT imposes REFL but the impression PARTICLE.
'But I get the impression' (Treebank: HDT, ID: hdt-s178959)

In case of experiencers annotated as (*iobj*), this element targets the EPP position (cf. Samo 2019: 169), while the subject targets a low IP area projection.

16. https://universaldependencies.org/u/dep/obj.html (accessed 20.06.2020)
17. Beyond the scope of this paper is the further challenge represented by the interpretative properties of those cases of movement of objects with the verb within the IP (cf. Belletti & Rizzi 2012).
18. https://universaldependencies.org/u/dep/iobj.html (accessed 20.06.2020)

oblique The relation oblique *obl*[19] has been adopted for phrases functioning as a non-core arguments of the verbal domain ad represent the set of complements, modifiers and prepositional phrases (temporal, locative, etc.) discussed in Schweikert (2005). Mapping *obl* to the cartographic architecture, these elements are generated in PP and can move to the peripheries, if they bear Focus and Topic features (LOWIP, TOPIC, FOCUS) or they are "highlighted" (cf. Rizzi 2004) in ModP (MOD).

advmod The adverbial modifier labelled as *advmod*[20] denotes adverbial phrases modifying predicates. UD differentiates adverbials realized as adverbs (*advmod*) and adverbials realized by maximal projections (*obl*), but it does not differentiate adverbials from a subset of modifier at the DP level. Thus, a query adding limits to the retrieval is required. Adopting (Cinque 1999)'s hierarchy, *advmod* represent the functional projections where adverbs are generated (ADV). Most of adverbs can move to the Left Periphery if they bear Focus and Topic features (TOPIC, FOCUS) or they are "highlighted" in ModP (MOD).

nsubj The syntactic relation (*nsubj*)[21] represents the active subject of a sentence[22]. UD however does not annotate absent subjects, such as the cases of sentences with null subjects (cf. Rizzi 1982). Subjects are generated as arguments (external and internal according to the nature of the verb) in the vP layer (ARGVP), but syntactically obliged to move to SpecSubjP (SUBJP). Naturally, they might also be topicalized/focalized in the LP (TOPIC, FOCUS). Subjects can also target the low IP area (cf. Frey 2004) if a *iobj* or an expletive satisfies the subject criterion (LOWIP).

expl The relation *expl*[23] captures all form of expletives on arguments. Indeed, the term expletives in the UD schemata is used also for expletives of other arguments and reflexives. Here, I only investigate those expletives satisfying the subject criterion. Since expletives cannot be moved to position in the LP, I predict that they can only be located in their generation (criterial) position (EPP).

Merging the results of Table 2 with the syntactic configurations presented in §2.1, we can envisage a reduced cartographic structure with loci of generation and movement for labels of syntactic dependencies, as

19. https://universaldependencies.org/u/dep/obl.html (accessed 20.06.2020)
20. https://universaldependencies.org/u/dep/advmod.html (accessed 20.06.2020)
21. https://universaldependencies.org/u/dep/nsubj.html (accessed 20.06.2020)
22. UD has a specific query for passive structures (*nsubj: pass*), which are not investigated in this work.
23. https://universaldependencies.org/u/dep/expl.html(accessed 20.06.2020)

Layer	Functional projections	Dependency
CP	TOPIC/FOCUS/MOD	*nsubj, advmod, obl, iobj, obj*
IP	SUBJ, EPP	**nsubj**, <expl>, *iobj*
	ADVP	<advmod>
	LOWIP	*obl, iobj, obj, nsubj,*
	PP	<obl>
vP	ARGVP	<iobj>, <nsubj>, <obj>

Table 3: Translating universal dependencies labels into functional projections in the syntactic tree, with hook brackets indicating generation loci, bold used for obligatory landing sites and italic for landing sites.

given in Table 3 (with hook brackets indicating <generation loci>, bold **obligatory landing sites** and italic *landing sites*).

3.3 Hypotheses

A series of predictions can be made according to the results given in Table 3. As a matter of fact, the orders of generation loci (represented by hook brackets) should give us ”canonical” orders of these constituents.

The highest generation locus is represented by expletives, while the lowest are the arguments. Because of the subject criterion in German, subjects obligatory move to a high functional projection in the IP. Leaving aside impersonal constructions of the type of (12), the dependencies *nsubj* and *expl* will be analyzed under the unique label *subj* (when a clearer differentiation is not required) since they both represent the obligatory criterial position.

A map is given in (13, > meaning precedence), representing a definition of ”canonical” order.

(13) subj > advmod > obl > iobj > obj

From a quantitative point of view, a first research question investigates whether orders on the basis of generation are more frequent than reorderings. To have a better understanding and higher counts, ”canonical” orders are given as the occurrences of pairs of elements with respect to the subject, which is always realized, being German a non-null-subject language. This hypothesis can be stated in H_1.

> H_1: For every pair of constituents, the frequencies of the relations with respect to subjects constituting the ”canonical order” should be more frequent than the frequencies of the reverse pattern.

A second hypothesis is based on locality. An element which is located lower in the structure should be able to cross an intervener with a higher frequency, if the two elements do not share classes of features predicted in fRM. Elements generated as arguments of the vP bear features belonging to the argumental class, while *advmod* and *obl* bear features of the modifier class. Naturally, the presence of further features such as Topic or Focus (quantificational) reduces intervention effects.

The classes of features, according to the nature of the elements, are given in (14).

(14) a. Argumental: *nsubj, obj, iobj*
b. Modifier: *advmod, obl*

The second hypothesis H_2 takes similarity into consideration.

> H_2: For every pattern of pair of constituents belonging to "uncanonical" orders, frequencies are reduced if the two constituents are similar in terms of classes of features.

In other words, I expect that the frequencies (in terms of proportions) of an element crossing an element higher in the structure would increase if there is mismatch of class of features.

The quantitative hypotheses presented here are to be contrasted to an H_0 hypothesis that would predict that grammatical properties are uncorrelated to observed frequencies (Merlo 2016, Samo & Merlo 2019).

H_1 and H_2 shall be explored in §4, since empirical data extracted from a set of treebanks in German annotated under the guidelines of UD are provided.

4 A quantitative study on German

The collection of frequencies of syntactic structures is operated on four syntactically annotated treebanks of German. Subsection 4.1 presents materials and methods, while subsection 4.2 will show the results and the relative discussion.

4.1 Materials and Methods

Materials The syntactic configurations are extracted from four syntactically annotated treebanks following the guidelines of Universal Dependencies (Zeman, Nivre & Abrams 2020) annotation scheme: the Hamburg Dependency Treebank[24](Borges Völker et al. 2019), the GSD tree-

24. http://match.grew.fr/_meta/UD_German-HDT@2.6_table.html (accessed 20.06.2020).

Treebank	Trees	Tokens	Genres
UD_German-HDT@2.6	189928	3589318	news, nonfiction, web
UD_German-GSD@2.6	15590	308378	news, reviews, wiki
UD_German-PUD@2.6	1000	22329	academic, fiction, nonfiction
UD_German-LIT@2.6	1922	42362	nonfiction

Table 4: Treebanks, size and genres.

bank[25], the German version of a parallel universal dependencies (henceforth, PUD) treebank[26] and finally a treebank gathering texts of the literary history in modern German, LIT[27]. Some statistics about their size and genres are given in Table 4. The collection of four different treebanks of different genres is intended to provide a better understanding and a better representativity of the language. Therefore, data will be collapsed into one single German treebank.

Methods All the materials is extracted with the Grew-match tool maintained by Inria in Nancy (http://match.grew.fr). The query[28] retrieved sentences for which a variable *x* annotated with a dependency *dep_x* precedes a variable *y* annotated with another syntactic label *dep_y*. The results will be presented here as frequency and inputed counts of frequencies[29]. An analysis on a sample of sentences has been manually conducted to detect the accuracy of the queries. Some natural occurring examples of the 30 combinations of patterns involving two constituents are given in Table 5.

The total number of dependencies (*obj*, *iobj*, *obl*, *advmod*, *expl*, *nsubj)* are also automatically retrieved from the treebanks, as given in Table 6.

25. http://match.grew.fr/_meta/UD_German-GSD@2.6_table.html (accessed 20.06.2020).
26. http://match.grew.fr/_meta/UD_German-PUD@2.6_table.html (accessed 20.06.2020).
27. http://match.grew.fr/_meta/UD_German-LIT@2.6_table.html (accessed 20.06.2020)
28. The query is as follows (with *x* and *y* indicating syntactic relations): *pattern* { verb –[dep_x]→ *x*; verb -[dep_y]→ *y*; *x* < *y*; }
29. The tool adopted in the investigation provided only the first 1000 occurrences of the query and a proportion of exploitation of the treebank. Here frequency is calculated as a coefficient on the basis of these two data. This coefficient is calculated to provide a better understanding of a predictive tool. Given *I* as the input count, *F* the frequency of the pattern, and *C* the percentage of the exploitation of the corpus, *I* is derived with the formula $I = (F \times (1 - C))/C + F$.

Pattern	TB	ID	Sentence
nsubj > obj	PUD	n01057014	*Die Trümmer formten einen Ring um den Erdäquator*
			'The debris formed a ring around the earth's equator'
obj > obl	HDT	hdt-s122	*dass sie die Musik mit Hilfe der [...] verkommen lassen.*
			'that she had let the music degenerate with the help[...]'
expl > nsubj	HDT	hdt-s108219	*Es liegen alle Fakten auf dem Tisch*
			'All the facts are on the table'

Table 5: Examples of naturally occurring examples extracted from the treebanks (TB) and relative ID sentence in a specific pattern.

Treebank	nsubj	obj	iobj	obl	advmod	expl
HDT	221680	154653	6088	221735	258465	3114
GSD	17412	8408	1276	17726	17332	397
PUD	1481	1344	95	895	1103	90
LIT	3274	1618	202	1598	4127	170
Tot.	243847	166023	7661	241954	281027	3771

Table 6: Total numbers of syntactic relations for treebank.

4.2 Results and Discussion

A first interesting result is that subjects cooccur with expletives only in the order [expl > subj] and the reduced size of the findings (712 occurrences representing a small portion of the the corpus under investigation equal to the three *per mille* 0.003, of the total of subjects[30]) shows that this order is a clear case of a reordering for scope-discourse semantics reasons, since the subject targets a Low IP peripheral position.

The frequencies of the syntactic relations cooccuring with the active subjects are given in table 7.

As expected, proportions of frequencies strongly support that subjects in "canonical" order precede the other arguments (88% *obj*, 81% *iobj*). As for the elements belonging to the featural class of modifiers, subjects precede both *obl* (73%) and *advmod* (69%).

The proportions with respect to the reverse pattern is adopted for providing a first mapping between frequencies and cartographic prediction.

In (15), the proportion figures show the percentage of the relevant element crossing the subject. The results seem linearly similar to the

30. As for the reversed order [subj > expl], the query retrieved those cases of the expletive *es* in copular (Moro 1997) constructions introducing an embedded CP (e.g. treebank: PUD, ID: n01055008, *Das Ziel dieser CRTC - Sitzungen ist* ***es****, Reaktionen von Teilhabern aus der Industrie und von der Öffentlichkeit zu bekommen* 'The goal of these CRTC sessions is [**expl**,] to get responses from industry and public stakeholders.'). Copular sentences might be treated as classical transitives and therefore the *expl* represents an object expletive.

Pattern	freq.	%	%subj	Pattern	freq.	%	%subj.
nsubj > obj	123766	0.88	0.51	obj > nsubj	17659	0.12	0.07
nsubj > iobj	4466	0.81	0.02	iobj > nsubj	1045	0.19	0.01
nsubj > obl	113485	0.73	0.47	obl > nsubj	41957	0.27	0.17
nsubj > advmod	74699	0.69	0.31	advmod > nsubj	33588	0.31	0.14

Table 7: Frequencies of patterns involving subjects, % their proportion with respect to the reverse pattern and %subj the proportion with the number of subjects. The patterns in bold confirm the hypothesis.

Pattern	freq.	%ce	%obj	Pattern	freq.	%	%obj
obj > iobj	1219	0.19	0.01	**iobj > obj**	5185	0.81	0.03
obj > obl	45996	0.46	0.28	**obl > obj**	55047	0.54	0.33
obj > advmod	30761	0.40	0.19	**advmod > obj**	46007	0.60	0.28

Table 8: Frequencies of patterns involving objects, % their proportion with respect to the reverse pattern and %obj the proportion with the number of objects. The patterns in bold confirm the hypothesis.

predicted canonical order in (13).

(15) adv (0.31) > obl (0.27) > iobj (0.19) > obj (0.12).

However, H_1 cannot be fully confirmed. To confirm whether (15) is the proposed "canonical" order, further evidence is required, such as the syntactic patterns involving the other core element *obj*.

The observations of the patterns with respect to objects are given in Table 8.

Similar asymmetries are found with objects. Objects, which are generated lower, can cross indirect objects in a restricted set of grammatical sentences when they cooccur (19%). The situation is totally reversed in the case of objects crossing elements whose class of features belong to the modifier group. Elements annotated as *obj* precede *advmod* (40%) and *obl* (46%) at a higher proportions than *iobj* and *nsubj* (12%).

These results provide further evidence for a theory of locality, in which a mismatch of class of features seems suggesting a higher degree of "freedom" of movement in the structure.

To detect the last layer of the "canonical" order, I investigate whether the pattern *advmod* preceding *obl* is more frequent than the reverse pattern.

As shown by Table (9), *advmod* seems preceding *oblique*. These results are in the direction of confirming H_1, suggesting that (a subset of) adverbs are clearly generated higher than oblique. However, a finer-grained search in treebanks of the label ADV and PP with more detailed

Pattern	freq.	%	%obl	Pattern	freq.	%	%obl
advmod > obl	62400	0.61	0.25	obl > advmod	40167	0.39	0.17

Table 9: Frequencies of patterns involving adverbs and obliques, % their proportion with respect to the reverse pattern and % the proportion with the number of oblique (the biggest set). The pattern in bold confirms the hypothesis.

Pattern	%CE	Match	Landing Sites
obj > nsubj	0.12	Argumental	TOPIC, FOCUS
iobj > nsubj	0.19	Argumental	TOPIC, FOCUS, EPP
obj > iobj	0.19	Argumental	TOPIC, FOCUS, LOWIP
obl > nsubj	0.27	Mismatch	TOPIC, FOCUS, MOD
advmod > nsubj	0.31	Mismatch	TOPIC, FOCUS, MOD
obl > advmod	0.39	Modifier	TOPIC, FOCUS, MOD
obj > advmod	0.40	Mismatch	TOPIC, FOCUS
obj > obl	0.46	Mismatch	TOPIC, FOCUS

Table 10: Frequencies of patterns involving adverbs and obliques, % their proportion with respect to the reverse pattern and of co-occurring elements (CE); Match = type of match or mismatch; Landing Sites = available landing sites for crossing the intervener.

queries is required to provide further insights concerning the locality issues within the IP.

The data in Table (9) show another asymmetry. The featural class of modifiers represents a weaker intervener if it is taken as a single class with respect to the argumental class. But this conclusion needs to be revised, as already predicted in Rizzi (2004), since the class of modifier is an heterogeneous group based on a open class of subtypes of features in the spirit of the functional labels discussed in (Cinque 1999), such as Modality, Tense, Mood, Aspect, Voice, and so on (Rizzi 2013: 218).

Finally, let us discuss H_2. To do so, we compare the proportions of those patterns involving an element crossing an intervener higher in the clause. Furthermore, another dimension seems related to the sets of possible landing sites where the element can land to in order to cross the intervener (given that it bears the relevant features to target criterial positions also in the LP).

Table (10) summarizes the results. Comparing the proportions of the co-occurring patterns in Table (10), the lowest scores (*obj* > *nsubj* 12%, *iobj* > *nsubj*, *obj* > *iobj* 19%) belong to the combinations in which a lower argument crosses a higher argument.

The highest scores, on the other hand, show a mismatch in the class of feature (*obj* > *advmod* 40%, *obj* > *obl* 46%).

The dependencies *advmod* and *oblique* crossing the subject in its criterial position require a further dimensions of features to target the LP,

therefore their proportion partially dropped (*obl* > *nsubj* 27%, *advmod* > *nsubj* 31%). The class of modifier features has similar frequencies (*obl* > *advmod* 39%), confirming that this group is less homogeneous than the class of argumental features.

To sum up, the results presented here support an hypotesis that reorderings are limited by locality. Translating these results in theoretical terms, the data presented here suggest that *scrambling* is related to locality, as predicted by a criterial approach to the Cartography of German (Samo 2019).

Further research should take into consideration fine-grained tools to detect issues of locality, with respect to morphosyntactic features. The marginality observed in grammaticality judgments should be confirmed by frequencies in large-scale treebanks, in the spirit of the Quantitative Computational Syntax approach (Merlo 2016).

5 Conclusions

In this paper, I carried out a quantitative analysis on the reorderings in German, a language which has represented an interesting case in the literature allowing restrictions and freedom of movements of the constituents in different areas of the structures.

After having provided a model mapping syntactic dependencies into cartographic maps, I have investigated four treebanks annotated with UD. The results show that frequencies, in terms of proportion for the two orders given by the relations of precedence of two constituents, confirm the predicted generation order as ”canonical”.

The results also show that locality effects play a role: the more dissimilar two elements, the more a lower element can cross an intervener higher in the structure.

Cartographic studies are thus providing an important formal tool with a strong heuristic capacity: analyses on large-scale treebanks may add a quantitative dimension to the already established qualitative dimension, providing further insights.

6 Acknowledgements

This research was supported by the grant 18ZDJ06 *Studies on Language Acquisition and Cognition from a Cartographic Perspective* of ”the Fundamental Research Funds for the Central Universities” (Science Foundation of Beijing Language and Culture University).

Thanks to Paola Merlo, Luigi Rizzi, Fuzhen Si, Francesco-Alessio Ursini and the audience of the *Séminaire de Recherche en Linguistique* (University of Geneva) for helpful insights and comments in different stages of this paper. I gratefully thank Tián Yīnghuì for her help in the collection of preliminary data.

References

Abels, Klaus. 2012. The italian left periphery: a view from locality. *Linguistic Inquiry* 43(1). 229–254.

Belletti, Adriana. 2004. Aspects of the low ip area. In Luigi Rizzi (ed.), *The structure of cp and ip. the cartography of syntactic structures*, vol. 2, 16–51. Oxford, New York: Oxford University Press Oxford.

Belletti, Adriana. 2018. *Locality in syntax*. Oxford, New York.

Belletti, Adriana & Luigi Rizzi. 2012. Moving verbal chunks in the low functional field. In Laura Brugé, Anna Cardinaletti, Giuliana Giusti, Nicola Munaro & Cecilia Poletto (eds.), *Functional heads [the cartography of syntactic structures, volume 7]*, 129–137. New York: Oxford University Press Oxford.

Belletti, Adriana & Luigi Rizzi. 2017. On the syntax and pragmatics of some clause-peripheral positions. In Joanna Blochowiak, Cristina Grisot, Stephanie Durrleman & Christopher Laenzlinger (eds.), *Formal models in the study of language*, 33–48. Cham: Springer.

Bocci, Giuliano. 2013. *The syntax prosody interface: a cartographic perspective with evidence from italian*. Vol. 204. Amsterdam: John Benjamins Publishing.

Borges Völker, Emanuel, Maximilian Wendt, Felix Hennig & Arne Köhn. 2019. Hdt-ud: a very large universal dependencies treebank for german. In *Proceedings of the third workshop on universal dependencies (udw, syntaxfest 2019)*, 46–57. Paris, France: Association for Computational Linguistics. `https://doi.org/10.18653/v1/W19-8006`. `https://www.aclweb.org/anthology/W19-8006`.

Caha, Pavel. 2009. *The nanosyntax of case* dissertation.

Cardinaletti, Anna. 2004. Toward a cartography of subject positions. In Luigi Rizzi (ed.), *The structure of cp and ip. the cartography of syntactic structures*, vol. 2, 115–165. Oxford, New York: Oxford University Press Oxford.

Cinque, Guglielmo. 1999. *Adverbs and functional heads: a cross-linguistic perspective*. Oxford, New York: Oxford University Press.

Cinque, Guglielmo. 2018. *Some notes on "meaningless" movement, (Paper presented at 44th Incontro di Grammatica Generativa, University of Roma Tre, 03.02.2018)*.

Cinque, Guglielmo & Luigi Rizzi. 2010. The cartography of syntactic structures. In Bernd Heine & Heiko Narrog (eds.), *Oxford handbook of linguistic analysis*. Oxford, New York: Oxford University Press.

Collins, Chris. 2005. A smuggling approach to raising in english. *Linguistic Inquiry* 36(2). 289–298.

De Marneffe, Marie-Catherine, Timothy Dozat, Natalia Silveira, Katri Haverinen, Filip Ginter, Joakim Nivre & Christopher D Manning. 2014. Universal stanford dependencies: a cross-linguistic typology. In *Lrec*, vol. 14, 4585–4592.

Den Besten, Hans. 1983. On the interaction of root transformations and lexical deletive rules. *On the formal syntax of the Westgermania*. 47–131.

Dryer, Matthew. 2005. Order of subject, object and verb. *The world atlas of language structures*. 330–333.

Dryer, Matthew S. & Martin Haspelmath. 2013. *Wals online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. `https://wals.info/`.

Durrleman, Stephanie, Loyse Hippolyte, Sandrine Zufferey, Katia Iglesias & Nouchine Hadjikhani. 2015. Complex syntax in autism spectrum disorders: a study of relative clauses. *International Journal of Language & Communication Disorders* 50(2). 260–267.

Erdmann, Oskar. 1886. *Grundzüge der deutschen syntax nach ihrer geschichtlichen entwicklung*. Vol. 1. Stuttgart: Cotta.

Frauenfelder, Ulrich, Juan Segui & Jacques Mehler. 1980. Monitoring around the relative clause. *Journal of Verbal Learning and Verbal Behavior* 19(3). 328–337.

Frey, Werner. 2004. A medial topic position for german. *Linguistische Berichte*. 153–190.

Friedmann, Naama, Adriana Belletti & Luigi Rizzi. 2009. Relativized relatives: types of intervention in the acquisition of a-bar dependencies. *Lingua* 119(1). 67–88.

Greenberg, Joseph H. 1975. Research on language universals. *Annual Review of Anthropology* 4(1). 75–94.

Grillo, Nino. 2008. *Generalized minimality* dissertation.

Gulordava, Kristina & Paola Merlo. 2015. Diachronic trends in word order freedom and dependency length in dependency-annotated corpora of latin and ancient greek. In *Proceedings of the third international conference on dependency linguistics (depling 2015)*, 121–130.

Haegeman, Liliane. 1990. Non-overt subjects in diary contexts. *Grammar in Progress, GLOW essays for Henk van Riemsdijk*. 167–174.

Haider, Hubert. 2010. *The syntax of german*. New York: Cambridge University Press.

Haspelmath, Martin. 2006. Against markedness (and what to replace it with). *Journal of linguistics* 42(1). 25–70.

Hinterholzl, Roland. 2006. *Scrambling, remnant movement, and restructuring in west germanic*. Oxford, New York: Oxford University Press on Demand.

Kayne, Richard. 1994. *The antisymmetry of syntax*. Vol. 25. Cambridge MA: MIT press.

Lenerz, Jürgen. 1977. *Zur abfolge nominaler satzglieder im deutschen*. Vol. 5. Tübingen: Verlag Narr.

Merlo, Paola. 2016. Quantitative computational syntax: some initial results. *Italian Journal of Computational Linguistics* 2(1). 11–30.

Mohr, Sabine. 2005. *Impersonal constructions in english and german*. Amsterdam, Philadelphia: John Benjamins Publishing.

Moro, Andrea. 1997. *The raising of predicates: predicative noun phrases and the theory of clause structure*. Vol. 80. Oxford: Cambridge University Press.

Müller, Gereon & Wolfgang Sternefeld. 1993. Improper movement and unambiguous binding. *Linguistic inquiry* 24(3). 461–507.

Müller, Stefan. 2013. Datensammlung zur scheinbar mehrfachen vorfeldbesetzung. Freie Universität Berlin. Available at http://hpsg. fu-berlin. de/s̃tefan/PS/mehr-vf-daten. pdf, accessed in June 2020.

Nivre, Joakim. 2015. Towards a universal grammar for natural language processing. In *International conference on intelligent text processing and computational linguistics*, 3–16.

Ramchand, Gillian. 2008. *Verb meaning and the lexicon: a first-phase syntax*. Vol. 116. Cambridge: Cambridge University Press Cambridge.

Reddy, Siva, Oscar Täckström, Michael Collins, Tom Kwiatkowski, Dipanjan Das, Mark Steedman & Mirella Lapata. 2016. Transforming Dependency Structures to Logical Forms for Semantic Parsing. *Transactions of the Association for Computational Linguistics* 4. 127–140. `http://sivareddy.in/papers/reddy2016transforming.pdf`.

Rizzi, Luigi. 1982. *Issues in italian syntax, foris, dordrecht*. Dordrecht: Foris.

Rizzi, Luigi. 1990. *Relativized minimality*. Cambridge MA: The MIT Press.

Rizzi, Luigi. 1997. The fine structure of the left periphery. In Liliane Haegeman (ed.), *Elements of grammar*, 281–337. Dordrecht: Kluwer Academic publisher.

Rizzi, Luigi. 2001. On the position “int (errogative)” in the left periphery of the clause. In Guglielmo Cinque & Giampaolo Salvi (eds.), *Current studies in italian syntax: essays offered to lorenzo renzi*, vol. 59, 287–296. Amsterdam: Elsevier.

Rizzi, Luigi. 2004. Locality and left periphery. In Adriana Belletti (ed.), *Structures and beyond: the cartography of syntactic structures*, vol. 3, 223–251. Oxford, New York: Oxford University Press Oxford.

Rizzi, Luigi. 2007. On some properties of criterial freezing. *Studies in linguistics* 1. 145–158.

Rizzi, Luigi. 2013. Notes on cartography and further explanation. *Probus* 25(1). 197–226.

Rizzi, Luigi. 2015. Notes on labeling and subject positions. In Elisa di Domenico, Cornelia Hamann & Simona Matteini (eds.), *Structures, strategies and beyond: studies in honour of adriana belletti*, vol. 223, 17. Amsterdam: John Benjamins Publishing Company.

Rizzi, Luigi. 2016. Monkey morpho-syntax and merge-based systems. *Theoretical Linguistics* 42(1-2). 139–145. `https://doi.org/https://doi.org/10.1515/tl-2016-0006`. `https://www.degruyter.com/view/journals/thli/42/1-2/article-p139.xml`.

Rizzi, Luigi. 2017. On the format and locus of parameters: the role of mophosyntactic features. *Linguistic Analysis* 41(3). 159–191.

Rizzi, Luigi & Giuliano Bocci. 2017. Left periphery of the clause: primarily illustrated for italian. *The Wiley Blackwell Companion to Syntax, Second Edition*. 1–30.

Rizzi, Luigi & Guglielmo Cinque. 2016. Functional categories and syntactic theory. *Annual Review of Linguistics* 2. 139–163.

Roberts, Ian. 2004. The c-system in brythonic celtic languages, v2, and the epp. In Luigi Rizzi (ed.), *The structure of cp and ip*, vol. 2, 297–328. Oxford, New York: Oxford University Press Oxford.

Samo, Giuseppe. 2019. *A criterial approach to the cartography of v2*. Vol. Linguistic Aktuelle 257. Amsterdam, Philadelphia: John Benjamins Publishing Company.

Samo, Giuseppe & Paola Merlo. 2019. Intervention effects in object relatives in english and italian: a study in quantitative computational syntax. In *Proceedings of the first workshop on quantitative syntax (quasy, syntaxfest 2019)*, 46–56.

Schweikert, Walter. 2005. *The order of prepositional phrases in the structure of the clause*. Vol. 83. Amsterdam: John Benjamins Publishing.

Si, Fuzhen. 2019. A cartographic study of light verb constructions. *Yuwen Xuexi (Linguistic Research)* (1). 1–20.

Starke, Michal. 2001. *Move dissolves into merge: a theory of locality*. Université de Genève dissertation.

Vikner, Sten. 1995. *Verb movement and expletive subjects in the germanic languages*. Oxford: Oxford University Press.

Zeman, Daniel, Joakim Nivre & Mitchell et al. Abrams. 2020. *Universal Dependencies 2.6*. LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics

and Physics, Charles University. http://hdl.handle.net/11234/1-3226.

Zompì, Stanislao. 2017. *Case decomposition meets dependent-case theories* dissertation.

Zwart, Jan-Wouter. 1992. Subject initial verb second in west flemish: a reply to haegeman. *GAGL: Groninger Arbeiten zur germanistischen Linguistik* (35). 72–91.