

# CONSTITUENT ORDER IN MALTESE: A QUANTITATIVE REEVALUATION<sup>1</sup>

Slavomír Čéplö

## Abstract

This paper examines the question of constituent order in Maltese in light of major approaches to it and previous descriptions of Maltese. Using a syntactically annotated corpus (treebank), a quantitative analysis of constituent order in various clause types is performed. This analysis confirms that the default order in Maltese is SVO (with VS in existential clauses as the only exception). Furthermore, it is found that the constituent order in Maltese is quite rigid, more akin to English than – as has been previously argued – to languages with pragmatically determined order.

Dan l-istudju jeżamina l-ordni tal-kostitwenti fil-Malti fid-dawl tal-approċċi ewlenin u tad-deskrizzjonijiet tal-Malti s'issa. Permezz ta' korpus annotat sintattikament (treebank), issir analiżi kwantitattiva tal-ordni tal-kostitwenti f'diversi tipi ta' sentenzi. Din l-analiżi tikkonferma li l-ordni tipika fil-Malti hija SVO (bl-unika eċċezzjoni ta' VS f'sentenzi eżistenzjali). Barra minn hekk, turi li l-ordni tal-kostitwenti fil-Malti hija pjuttost riġida, u aktar tixbah lill-Ingliz milli – kif hemm min argumenta – lil-lingwi, li fihom l-ordni tal-kostitwenti hija determinata b'mod pragmatiku.

1 This paper is a revised and condensed version of chapters 1, 2, 3 and 7 of my dissertation (Čéplö 2018).

## 1. Introduction

### 1.1 General

Constituent order, i.e. the order of the verb (V) and its main arguments – the subject (S) and the direct object (O) – within a clause or sentence,<sup>2</sup> is one of the fundamental elements of syntactic description. Its importance is evidenced by the fact that it is often the only piece of information available on the syntax of a language; indeed as Dixon (2009: 73) notes, since most of the world’s languages are under-described, it is often the only piece of information on the grammar of a language available. Comprehensive overviews of the world’s languages such as *Ethnologue* (Lewis et al. 2016) are the best witness to this. To pick two random examples: the *Ethnologue* entry for Swedish (ISO 639-3 code “swe”), a relatively small but well-described language, lists the following under “Typology”:

SVO; prepositions; noun head final; gender (common, neuter); definite and indefinite articles; passives (active, middle, passive); comparatives; 19 consonant and 17 vowel phonemes; tonal (2 tones).

For Övdalian (ISO 639-3 code “ovd”), also spoken in Sweden, a close relative of Swedish and thus hardly an exotic language, the same section contains only the following:

SVO; 24 consonants, 9 vowels, 6 diphthongs and 1 triphthong.

The noticeably frequent appearance of constituent order in even the most rudimentary language descriptions is likely due

- 2 In what follows, I will use the term “constituent order” as defined above. The term “word order” is often used in this sense as well, but for clarity’s sake, I will define “word order” as the order of elements within a phrase (e.g. the order of nouns and adjectives) and use it in this sense throughout.

to two factors: first, constituent order is typologically associated with a number of other syntactic and even morphological features and can thus serve as a microcosm of a language's grammar. Secondly, constituent order is one of those properties of a language that are conspicuous (especially when different from what one is used to) and thus seem relatively easily discernible, much like its phonological inventory (again, see the Övdalian example above).

The former is arguably correct, at least to some extent (on which see 1.2 below); the latter, however, is not entirely so and the answer to the question of what the constituent order of a particular language is will almost invariably be a complex one. In this paper, I will attempt to provide it for Maltese, considering its context within both general and Maltese linguistics.

## 1.2 Constituent order and typology

The undoubtedly most influential work on constituent order in modern linguistics is Joseph H. Greenberg's 1963 paper titled *Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements* (cited from the second edition, Greenberg 1966). Greenberg expanded relatively trivial observations on how languages differ in the order of "modifying or limiting elements" (Greenberg 1966: 76) into a full-fledged typological classification of languages based on a list of so-called universals. The fundament on which these rest is his basic order typology: Greenberg takes the observation that "languages have several variant orders but a single dominant one" (Greenberg 1966: 76) to its logical conclusion and establishes a six-way typology of dominant orders of subject, verb and object: SVO, SOV, VSO, VOS, OSV and OVS. He immediately notes, however, that three of those – VOS, OSV and OVS – "do not occur at all, or at least are rare" (Greenberg 1966: 76) and proceeds to draw from this his first universal:

Universal 1. In declarative sentences with nominal subject and object, the dominant order is almost always one in which the subject precedes the object.

Greenberg combines the remaining three configurations – referred to as Type I (VSO), Type II (SVO) and Type III (SOV) – with two additional binary criteria (whether a language has prepositions or postpositions and whether an adjective of quality follows the noun it modifies or precedes it) and investigates the correlations between these syntactic properties in a sample of 30 languages (Greenberg 1966: 74-75):

Basque, Serbian, Welsh, Norwegian, Modern Greek, Italian, Finnish (European); Yoruba, Nubian, Swahili, Fulani, Masai, Songhai, Berber (African); Turkish, Hebrew, Burushaski, Hindi, Kannada, Japanese, Thai, Burmese, Malay (Asian); Maori, Loritja (Oceanian); Maya Zapotec, Quechua, Chibcha, Guarani (American Indian).

Using these correlations as the starting point, Greenberg postulates 45 implicational universals, 15 of which relate to constituent order or at least the position of the verb and its arguments, including question words.

Greenberg's universals were met with almost immediate acceptance and despite substantial criticism (on which see below) and some empirical evidence to the contrary (like the case of OVS order in Hixkaranya described by Derbyshire 1977), Greenberg's six-way typology continues to be the dominant paradigm in the cross-linguistic study of constituent order variation. Works like Payne (1997: 71-74), Song (2011b), the *Ethnologue* (see the entries above) and *The World Atlas of Language Structures* (WALS; Dryer and Haspelmath 2013) are but a few of the most prominent examples of Greenberg's enduring legacy.

### 1.3 The problem of ‘basic’ constituent order

As with any new paradigm, criticisms of Greenberg began to appear almost immediately. One of the primary issues that emerged as a major point of contention is the problem of basic (default) word order. Greenberg’s original formulation of his universal does not actually define what qualifies as ‘basic’, merely assumes it: “If a language has verb-subject-object as its basic word order in main declarative clauses...” (Greenberg 1966: 74). Greenberg is aware that this presupposes, at the very least, the existence of a subject-predicate structure in all languages under investigation. He acknowledges the problems with this assumption, but proceeds without resolving this issue, since doing so would have “prevented me from going forward to those specific hypotheses, based on such investigation, which have empirical import and are of primary interest to the non-linguist” (Greenberg 1966: 74). In other words, Greenberg was primarily interested in the universals (and their correlations) and was willing to sacrifice accuracy in determining the basicness of a particular constituent order in a particular language to achieve his goal.

This is obviously a problem and one that is related to a larger issue in linguistics: if the ostensible goal of linguistics (or at least its descriptive and typological branches) is to provide a description of one or more languages, then the primary question becomes what it is one is actually describing. For example, a linguist who is a native speaker of a language could base their description of that language on their own knowledge. Such an approach to linguistic description, commonly referred to as introspective or intuitive (Itkonen 2005), is (or at least was) typical for generative linguistics; one infamous instance involves Noam Chomsky arguing that the English verb ‘perform’ cannot take mass nouns as objects<sup>3</sup> and insisting he is correct because “I am a native speaker of the English language” (Harris 1995: 97). Whether such an

3 He was, as is often the case with such pronouncements in general, wrong.

approach to the analysis of anything is truly scientific is best left for another time and venue; what matters is that there is an obvious practical issue with this approach: what happens if another native speaker disagrees, as one immediately has in the case described above (Harris 1995: 97)?

The only other option available to a linguist is to collect data, i.e. the empirical approach. Within modern linguistics, there are two major ways of doing this: the first one is elicitation, which essentially involves asking many native speakers, thus hopefully at one point arriving at a consensus or at least clearly defined variation. This is a tried and true method, but it often brings with itself not only practical challenges (e.g. how much is many, how one gets cooperative respondents etc.), but also entails problems of epistemological nature: human beings have all types of ideas and preconceptions about language; chances are, therefore, that asking them about their language and their use thereof will yield information that is not objective, reflecting the respondents preconceptions, rather than the actual linguistic reality.

The other route to take is to use a corpus, i.e. a collection of texts (whether they originated in writing or they came about as transcriptions of speech) in a particular language. The corpus approach, often taken to be synonymous with the empirical approach, is nothing new in principle – grammarians and lexicographers have been using collections of texts to do their work for centuries. Modern corpus linguistics, however, does differ from those in two ways, both thanks to the relatively recent advances in computing: first, modern-day corpora are by orders of magnitude larger than those available to anyone in history. The size of corpora, along with the fact that texts typically contain spontaneously produced language, is the main advantage of corpus linguistics over elicitation, as it eliminates the epistemological issues associated with the latter described above. Second, the use of computers to store and query those corpora has inevitably put

large emphasis on quantitative measures, especially frequency, which has led to some surprising insights, such as the Menzerath's Law (Milička 2014).

This, once again, is nothing new in principle: Greenberg's work is, after all, all about statistics. The problem described above lies in the fact that the proper statistical considerations (sampling, sample size, representativeness etc.) are only applied to the universals, not to the analysis of individual languages. Many of Greenberg's successors and critics have attempted to correct this, but Matthew Dryer is by far the most successful and thus most influential. Dryer's work on constituent order typology began as a criticism of Greenberg's sampling methods and a test of hypotheses raised by Greenbergian universals (Dryer 1989b) and included a large follow-up study of the universals using a larger and more balanced sample of languages (Dryer 1992). This work led Dryer to renounce Greenbergian six-way typology and propose a new typology, based on two independent but interacting binary parameters, SV/VS and VO/OV (Dryer 1997, Dryer 2013b). Dryer lays out a complex case for this, the chief arguments being that "some word order parameters correlate with both the order of the object and the verb and with the order of the subject and the verb" (Dryer 2013b: 295) and that a typology based on these two parameters is more fundamental than the six-way typology, as it is "based on clause types that occur much more frequently" (Dryer 1997: 70). The latter illustrates Dryer's focus on frequency as an important element in linguistic description and explanation: Dryer recognizes that "speakers store grammatical knowledge independent of frequency", but argues that "frequency plays a pervasive role in explaining why languages — and grammars — are the way they are" (Dryer 2013b: 292). Consequently, Dryer's concept of basic order is based solely on frequency where, admirably, Dryer is aware of the inherent dangers of inadequate sampling (Dryer 1997: 72, *italics in the original*):

If a particular order is more common in most or all texts, then we can justifiably describe that order as most frequent. If no order is most frequent over most texts, however, or if the order varies from genre to genre or text to text, we should probably not describe any particular order as the basic order (in the sense of most frequent order) and we should say that the language is one that lacks a basic word order [...]. In short, while it may be relatively easy to identify a most frequent order in a single text or in a small body of texts, it is necessary to examine a wide variety of texts before one can decide with confidence that a particular order is most frequent in the language *as a whole*.

In typological studies of word and constituent order, Dryer's work has become the standard reference, as evidenced not only by his contribution to general discussions on the state of the question (see the special issue of *Linguistic Typology* 15), but also his authorship of chapters on word order in such overviews of language typology as Shopen 2007 (Dryer 2007) or WALS (Dryer 2013a and 2013c). And while the latter work also uses Greenbergian six-way typology in its description of constituent order typology (though not exclusively), it is here that Dryer provides the ultimate definition of basic or – in Dryer's terminology – dominant order defined in terms of frequency (Dryer 2013a):

The expression *dominant order* is used here, rather than the more common expression *basic order*, to emphasize that priority is given here to the criterion of what is more frequent in language use, as reflected in texts. ... The rule of thumb employed is that if text counts reveal one order of a pair of elements to be more than twice as common as the other order, then that order is considered dominant, while if the frequency of the two orders is such that the more frequent order is less than twice as common as the other, the language is treated as lacking a dominant order for that pair of elements. For sets of three elements, one order is considered dominant if text counts reveal it to be more than twice as common as the next most frequent



order; if no order has this property, then the language is treated as lacking a dominant order for that set of elements.

This definition (applicable to both word order and constituent order, and both pairs and triads) is specific, empirically founded, without any theoretical baggage, cross-linguistically applicable, and clearly actionable (step 1: get texts; step 2: count); as such, it constitutes a significant improvement to previous definitions of “basic” constituent (and word) order; it will therefore be adopted in what follows under the name Dryer’s 2:1 method.

#### 1.4 The problem of ‘free’ constituent order

The typological classification of languages by basic constituent order assumes that such a basic order exists in all languages. It has, however, long been known that there exist languages with seemingly endless variation in their constituent order, also known as “free word/constituent order” languages; in fact, it is probably the oldest classification of languages by constituent order, dating at least as far back as Weil (1844: 25). Weil’s observations focused on flexibility of the order of constituents in classical Greek and Latin compared to the relative rigidity in modern languages such as French and German and were thus somewhat of a surprising revelation. To other linguists, such as those of the Prague Linguistic Circle almost a century later, the fact that some languages are very flexible when it comes to constituent order was no surprise, since their own native language – Czech – was one. The relatively free constituent order in Czech led Vilém Mathesius to the fundamental insight that in some languages, constituent order and pragmatics (i.e. the context in which a sentence is produced and the purpose for which it is produced) are intrinsically linked and “[t]he functional analysis of a sentence must be juxtaposed to its formal analysis” (“Aktuální členění věty je třeba klásti proti jejímu členění formálnímu.” Mathesius

1939: 171; see Firbas 1992: 22 for the English terminological choice). Expanding on previous work by Weil (1844) and von der Gabelentz on the distinction between grammatical subject and “psychological subject” (“das psychologische Subjekt”, von der Gabelentz 1869: 378), Mathesius establishes a two-way division of sentence in terms of its communicative effect: the “theme”, defined as “a thing about which we assert something” (“to, o čem něco tvrdíme”, Mathesius 1961: 91) and “what we say about the theme is the nucleus or the enunciation” (“to, co o základu tvrdíme, je jádro výpovědi neboli vlastní výpověď”, Mathesius 1961: 92). This division, for which Mathesius’ successors (Firbas 1957) established the terms “theme” and “rheme”, is the cornerstone of what has become known as the Functional Sentence Perspective (FSP). And while FSP as a theory of communication is largely unknown outside of Czech linguistics, its foundational works by Mathesius (1961 in its English translation) and Firbas (1964) are credited with establishing the subfield of information structure (Féry and Ishihara 2016b: 3). Its basic terminology, redressed and redefined multiple times – typically as ‘topic’ and ‘comment’ or ‘topic’ and ‘focus’ – and its fundamental ideas like context-boundness (Krifka and Musan 2012) have become a firm part of modern linguistic terminology (Féry and Ishihara 2016a).

Mathesius was far from the only one to notice the relationship between constituent order and pragmatics. Even Chomsky, despite his focus on structural description formulated as transformation rules, recognizes the importance of pragmatics (or, in his words, “stylistic factors”; Chomsky 1965: 11) for the variation of constituent order, noting that “grammatical transformations do not seem to be an appropriate device for expressing the full range of possibilities for stylistic inversion” (Chomsky 1965: 126). He resolves this conundrum by claiming that the rules of pragmatically determined variation in constituent order “are not so much rules of grammar as rules of performance” and while interesting, they have “no apparent bearing, for the moment, on the theory of grammatical

structure” (Chomsky 1965: 127). The moment in question did not last long and soon generativist works began to appear dealing with “the annoying problem that languages differ from one another” (Carnie 2013: 27) in the ordering of the constituents. John R. Ross’ 1967 PhD dissertation devotes some attention to the problem of free word order in Latin and other languages in the context of node deletion or tree pruning, i.e. reducing the complexity of sentences generated by existing theories of generative grammar (Ross 1967: 41). In the analysis of the various possible configurations of constituents and even components of noun phrases in Latin, Ross proposes the Scrambling Rule (Ross 1967: 75) which permits the seemingly unlimited surface variation of words in Latin sentences. Since Ross’s day, two approaches have developed to account for scrambling: the base-generation approach argues that variation in constituent order is a syntactic phenomenon, i.e. it is generated randomly at the D-structure level (Corver and van Riemsdijk 1994b: 1). The distinction made here is between configurational languages which do not allow this random generation of constituents and non-configurational languages (also termed “flat languages” by Hale 1983: 10, since they do not have a unitary Verbal Phrase) which do. In contrast, the movement approach (Corver and van Riemsdijk 1994b: 2) explains variation in constituent order by different types of movements, such as object shift (e.g. Broekhuis 2008 for Germanic languages) or VP fronting (Zubizarreta 1998). Both approaches have produced much literature (see Corver and van Riemsdijk 1994a for an overview), but so far, without any consensus in sight.

While the generativist discussion of scrambling seems to be dominated by the base-generation and movement approaches, there is still a third school of thought harkening back to Chomsky 1965 and Ross 1967 which considers constituent order variation from the point of view of pragmatics. This school, best represented by Kiss (1995a), has surveyed a number of languages very different from Standard Average European (Kiss 1995b: 4) and observed that “the structural role that the grammatical subject plays in the

English sentence may be fulfilled by a constituent not restricted with respect to grammatical function or case in other languages” (Kiss 1995b: 3). In simple terms, this school of thought argues that languages fall into two groups: subject-prominent languages where the surface constituent order is Subject – Verbal Phrase and topic-prominent languages, where the place of the Subject can be taken by an arbitrary element bearing a particular discourse (or pragmatic) function (Kiss 1995b: 4). These languages are termed discourse-configurational and their fundamental properties are as follows (Kiss 1995b: 6):

- A. The (discourse-)semantic function ‘topic,’ serving to foreground a specific individual that something will be predicated about (not necessarily identical with the grammatical subject), is expressed through a particular structural relation (in other words, it is associated with a particular structural position).
- B. The (discourse-)semantic function ‘focus,’ expressing identification, is realized through a particular structural relation (that is, by movement into a particular structural position).

One crucial aspect of the theory behind discourse configurationality is the empirical distinction between categorical andthetic statements (Kiss 1995b: 7-8). The distinction is based on Marty’s (1897) observation that there exist two types of sentences: those that do not express judgments (in the philosophical sense), like interrogative or imperative sentences (Marty 1897: 189), and those that do. Furthermore, the latter group can be divided into two types: the first type is referred to as compound or categorical judgments which actually contain two judgments, one about the existence of the subject and the other about a property of the subject. The second type is referred to as pseudo-categorical orthetic judgments (Marty 1895: 298) and they contain a single judgment only; these typically include existential, impersonal and universal sentences (Kiss 1995b: 7). A language can be discourse-

configurational with the property A only if it differentiates between categorical andthetic sentences syntactically.

Kiss goes on to argue that while sometimes properties A and B go hand in hand, they are not interdependent and so some discourse-configurational languages can display only type A characteristics, whereas others only show the type B properties (Kiss 1995b: 6). It should be noted, however, that while the fundamentals of this subset of generativist theory are framed in terms of pragmatic function, much of the explanation offered by its proponents still depends on movements (Choe 1995), such as the Focus Movement (focalization) and the Topic Movement (topicalization). And as with literature on scrambling, there seems to be no consensus in generativist literature on the general properties and nature of discourse configurationality. The term, however, is often used as nearly synonymous with “pragmatically determined word/constituent order” or its equivalents and, by extension, “free word/constituent order”.

## 2. Studies of constituent order in Maltese

For a numerically small and geographically and culturally marginal language, Maltese boasts a remarkably long and rich tradition of scholarly interest. This is evidenced, *inter alia*, by the fact that the first grammatical description of Maltese worthy of the name, de Soldanis’ 1750 *Nuova scuola di grammatica per agevolmente apprendere la lingua punica – maltese* (published in de Soldanis 1750), predates the first actual printed book in Maltese (Francesco Wzzino’s translation of the Catholic Catechism titled *Taġħlim Nisrani* published in Rome) by two years. In the intervening 270 years, many grammars of Maltese have been written, some of which addressed the question of constituent order in one form or another. A detailed analysis would require more space than is available here (and in any case, I have provided it in Čéplö 2018: 31-49), so Table 1 below summarizes their findings.

<b>Work</b>	<b>Classification</b>
Vella 1831: 224-225	- SV - VS in relative clauses
Sutcliffe 1936: 210	- VSO with variation “for euphony or emphasis” - VS in subordinate clauses
Aquilina 1959: 341	- SVO as the default - VS in “emphatic or high-flown literary language” and in subordinate clauses
Vella 1970: II.98	- VS “as is the Semitic custom” - SV “[d]ue to foreign influence”
Krier 1976: 79	- SV with “liberté de position est due à la mise en valeur stylistique (variation is due to stylistic emphasis)”
Kalmár and Agius 1983: 336-337	- SV - pragmatically determined VS
Fabri 1993: 7, 131	- “relative freie Wortstellung (a relatively free word order)” - “eine konfigurationale Sprache (a configurational language)” - considerable variation
Borg and Azzopardi-Alexander 1997: 57	- SVO(I) as “neutral order”
Fabri and Borg 2002: 362	- SV - VS with stress on V - SVO - OVS with stress on O
Fabri 2010: 793-794	- “a topic-oriented language” - “relatively free” - “SVO”
Borg and Fabri 2016: 417	- “a discourse configurational ... language, especially in its spoken form”

Table 1: Overview of previous descriptions of constituent order in Maltese

As Table 1 shows, two constant themes are interwoven throughout the history of the study of Maltese constituent order:

First, there is the question of what is the default (unmarked, basic, dominant) constituent order in Maltese. This has been answered in at least two different ways: verb-first, as argued by Sutcliffe 1936 and Vella 1970; or subject-first, as described by Aquilina 1959, Kalmár and Agius 1983, Borg and Azzopardi-Alexander 1997 and others.

The other theme is that of classifying Maltese constituent order as ‘free’ (e.g. Fabri 1993: 7, 131 and Fabri 2010: 793), including its near-synonyms like “discourse-configurational” (Fabri and Borg 2002, Borg and Fabri 2016) and “topic-oriented” (Fabri 2010: 793, Fabri and Borg 2017: 83). All those terms describe Maltese as a language where “constituent order, at sentence level is strongly influenced by pragmatic factors, in particular topic and focus, contrast and emphasis, more than by syntactic factors” (Fabri and Borg 2017: 83). In this context, a number of authors note a great deal of variation in Maltese constituent order (Sutcliffe 1936: 211, Krier 1976: 79, Fabri and Borg 2002) and attempt to account for it (Borg and Azzopardi-Alexander 1997, Fabri and Borg 2002).

Additionally, a number of works (e.g. Borg and Azzopardi-Alexander 1997, Borg and Azzopardi-Alexander 2009 and Čéplö 2014) devote a significant amount of attention to topicalization of direct and indirect objects, i.e. the placement of the object before the verb, typically also accompanied by a resumptive clitic and a phonological break. This phenomenon, which according to Borg and Azzopardi-Alexander (1997: 126) “is such a wide spread characteristic of Maltese, that it even features in Maltese English”, is related to both the question of the default constituent order in Maltese, as it at the very least assumes VO as the default, as well as to the question of the influence of pragmatic factors on the same.

All these analyses can be shown to have serious shortcomings: for the question of the default (unmarked, basic, dominant), the chief one is obviously the lack of general agreement. Additionally, there are multiple methodological issues, ranging from the lack of a meaningful definition of “default (unmarked, basic, dominant)” constituent order, through the lack of detailed studies on clause-type level (with Borg and Azzopardi-Alexander 1997 as sole attempt to do so in a systematic manner), all the way to the fact that most such studies have been introspective at best, impressionistic at worst. Even those that employed some sort of empirical approach (which is the case for Krier 1976 and Kalmár

and Agius 1983) did so more than imperfectly, rendering their conclusions tentative at best. Much of this also applies to works which describe Maltese constituent order as free or pragmatically determined; additionally, these have problems of their own. And so for example even those studies that provide a detailed account of the possible variation based on pragmatic (information structure) factors (Borg and Azzopardi-Alexander 1997, 2009; Fabri and Borg 2002) essentially only described potentiality, i.e. what options are available to speakers of Maltese, but did not (except in the broadest terms, e.g. Borg and Azzopardi-Alexander 1997: 126) provide a description of how those possibilities are instantiated.

In what follows, I will try to remedy those shortcomings by using an empirical analysis.

### **3. Quantitative analysis of constituent order in Maltese**

#### **3.1 Methodology and data**

Having reviewed the major ways of analyzing and classifying constituent order (section 1) and how they have been applied to the analysis of constituent order in Maltese (section 2), we can now proceed with the actual analysis. For the methodology, I will use Dryer's 2:1 method to analyze both the Greenbergian six-way classification, as well as Dryer's two-way classification.

As advertised above, the analysis I am about to conduct is empirically founded, i.e. corpus-based. Such an analysis, however, requires a syntactically annotated corpus (also known as a treebank); neither of the two large corpora already available for Maltese (MLRS and *bulbulistan*; Gatt and Čéplö 2013) contain such annotation. The solution is to compile a treebank of Maltese, which I have done. For the annotation scheme, I have chosen that employed by the Universal Dependencies project (UD; Zeman,



Nivre, Abrams et al. 2020), a de facto standard in syntactic annotation of corpora for NLP purposes. As with most languages, the UD annotation scheme had to be adapted to Maltese. The process is somewhat complicated as it amounts to compiling a sketch of Maltese syntax and there is, sadly, no space here for the full description; those interested are welcome to consult my dissertation (Čěplö 2018: 83-171).

What does require further elaboration, however, is the extent and composition of the Maltese UD treebank (henceforth: MUDT; at the time of writing in the version 2.7, hence MUDT v2.7). As this is the first effort in compiling a syntactically annotated corpus of Maltese, the vast majority of the annotation would have to be done manually, and so a balance had to be found between the desire to end up with as much data as possible and the practicality of what could be achieved with a manageable amount of effort within a reasonable time frame. In the final count of 44,162 tokens in 2074 sentences, MUDT is comparable to UD treebanks for such languages as Vietnamese, Wolof or Hungarian, each of which has many more speakers than Maltese.

The issue of the composition of MUDT is directly related to its size and by extension to the problem of whether corpus data accurately reflect the language under investigation. In corpus linguistics, this is a critical issue and several solutions have been adopted (McEnergy and Hardie 2011: 6-10). Considering the fact that the treebank had to be drawn from the existing corpora which are opportunistic by nature (i.e. based on the “we take all we can get” principle) and are composed of roughly four different genres, the solution I adopted for MUDT was to create a balanced treebank where the four genres – or text types – would be represented more or less equally in terms of sentence counts. This ensures that any description of Maltese based on MUDT is not just a description of a single genre, say, the journalistic language, which is very well known to differ substantially from other genres (Suter 1993).

Table 2 summarizes the composition of MUDT. The text type and subtype descriptions are self-explanatory, save perhaps for

the ‘quasi-spoken’ text type: I have chosen this label because while the texts in that group do originate from spoken language (interviews and parliamentary debates), they have undergone some form of editorial processing and as such cannot be considered transcriptions of speech.

Text type	Subtype	Sentence count
newspaper	news	239
	op-eds	240
	<i>Subtotal</i>	<i>479</i>
quasi-spoken	newspaper interviews	280
	parliament: debates and Q&A	294
	<i>Subtotal</i>	<i>574</i>
fiction	short stories	246
	novel chapters	251
	<i>Subtotal</i>	<i>497</i>
non-fiction	humanities	249
	science, encyclopedic and instructional	275
	<i>Subtotal</i>	<i>524</i>
	<b>Total</b>	<b>2074</b>

Table 2: The composition of MUDT v2.7 by genre

### 3.2 The analysis

Having established our methodology and the data set, we can now proceed to data collection and analysis. For the former, I have opted to import MUDT (in its most recent version v2.7) into an instance of the corpus management software ANNIS3 (Krause and Zeldes 2016) available at <https://bulbul.sk/annis-gui-3.6.0> (item *MUDT\_v27*). I then ran a number of queries to obtain the data in question, such as this one:

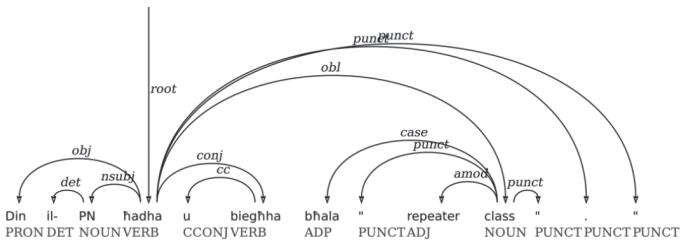
```
tok ->dep[deprel=/nsubj/] tok & #1 ->dep[deprel=/obj]obl:arg/]
tok & #3 .* #2 .* #1
```

The query searches for:

- any token (first *tok*)<sup>4</sup>
- which has a nominal subject token (*dep[deprel=/nsubj/ tok*) as a dependent
- while at the same time (first &), it (#1) also has an object or a non-canonical object<sup>5</sup> token (*dep[deprel=/obj|obl:arg/] tok*) as a dependent,
- with the added condition (second &) that the three tokens must appear in a specified order, i.e. the *obj|obl:arg* (#3) token, *nsubj* token (#2), the first token (#1) and the.

In other words, this query will retrieve all OSV clauses, such as the one in (1).<sup>6</sup>

- (1) *Din il- PN hadha u*  
 this.F DEF PN take.PAST.3SGM-ACC.SGF and  
*bieghha bhala "repeater class".*  
 SELL.PAST.3SGM-ACC.SGF as "repeater class"  
 ‘This the PN took and sold it as “repeater class”.’



[MUDT v2.7, file 22\_02J03]

- 4 Since only verbs or pseudoverbs can have both a subject and an object, the first token will always be one of these parts of speech; we could specify the parts of speech we’re looking for directly, e.g. by replacing the first *tok* with *pos = /VERB/*.
- 5 See Čéplö 2018: 127-128. These include, for example, prepositional objects, such as the objects of the verb *nduna* “to notice” introduced by the preposition *b’*.
- 6 In the following examples, I will include punctuation with the glossed word, whereas in dependency graphs, punctuation is considered a separate token.

Using this and equivalent queries, the following data was obtained:<sup>7</sup>

Configuration	Count	%
SVO	445	94.08%
SOV	0	0.00%
VSO	3	0.63%
VOS	11	2.33%
OSV	3	0.63%
OVS	11	2.33%
<i>Total</i>	<i>473</i>	<i>100%</i>

Table 3: Constituent order in MUDT v2.7 – Greenbergian analysis

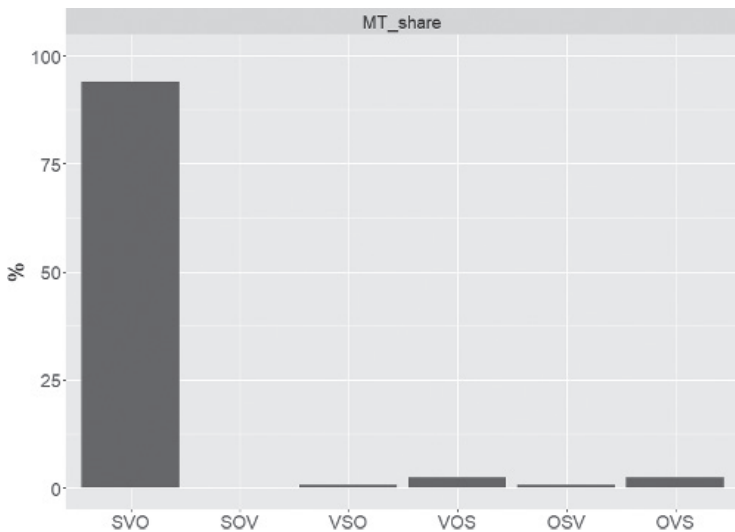


Figure 1: Constituent order in MUDT v2.7 – Greenbergian analysis

7 The data and code used to produce the analysis below can be downloaded from <https://bulbul.sk/jms2020>.

## CONSTITUENT ORDER IN MALTESE

Configuration	Count	%
VO	1830	95.11%
OV	94	4.89%
<i>Total V+O</i>	<i>1924</i>	<i>100%</i>
SV	1697	76.34%
VS	526	23.66%
<i>Total S+V</i>	<i>2223</i>	<i>100%</i>

Table 4: Constituent order in MUDT v2.7 – Dryerian analysis

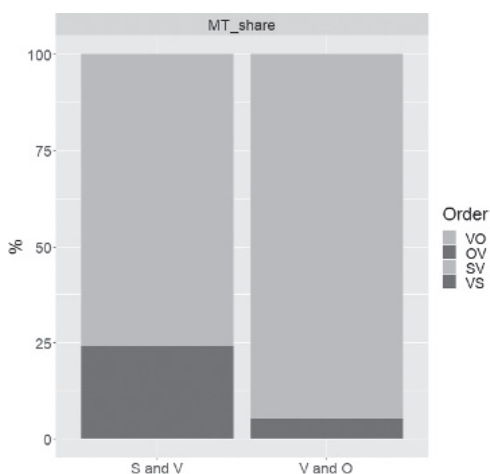
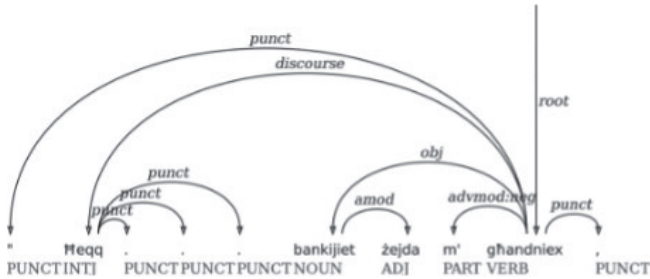


Figure 2: Constituent order in MUDT v2.7 – Dryerian analysis

This data provides a clear picture of both the Greenbergian and the Dryerian classifications of Maltese as, respectively, an SVO and SV/VO language. In fact, Dryer’s 2:1 need not even be employed; in all cases, the dominant configuration occurs at least three times as often as the other one.

Of particular interest here is the share of OV clauses, like the one in (2).

- (2) “Heqq           ... bankijiet        zejda            m’                ghandniex,  
 INTJ            ... bench-PL       additional-PL   NEG            have.PRES-1 PL-NEG  
 ‘Yeah ... we don’t have any more benches,’



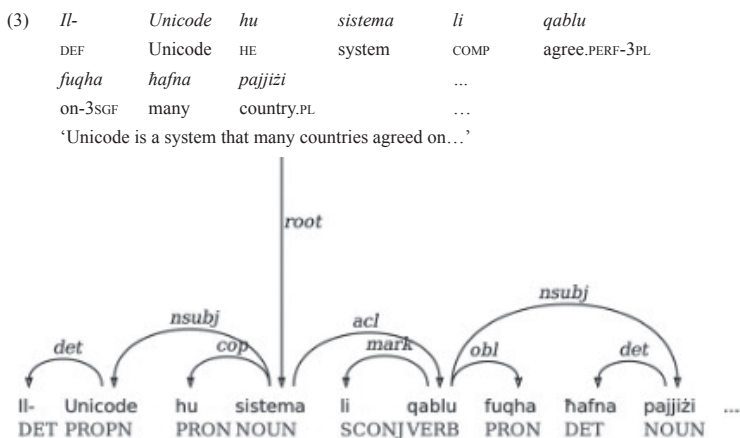
[MUDT v2.7, file 49\_03F09]

As noted in section 2 above, these constructions have been described as “a wide spread (sic) characteristic of Maltese” (Borg and Azzopardi-Alexander 1997: 126). And yet in MUDT v2.7, only <5% of all direct objects fall into that group (and that is assuming that all of them represent topicalization which is – what Borg and Azzopardi-Alexander refer to – which is far from certain), a figure which certainly does not represent a widespread phenomenon.

Furthermore, the data offers a clear case for Dryer’s SV/VS and VO/OV typology over the Greenbergian one: with MUDT v2.7, Greenbergian typology only has 487 data points to work with; using Dryerian typology, the data sample expands four-fold for both subjects (2223 total) and objects (1924).

This analysis is of course a rough one and can be refined. One way to do it would be to consider the full spectrum of clause types. Those can be first divided into main and subordinate clauses, which come in several types; of special interest here would be relative clauses (or *acl* in the UD nomenclature, see example 3 below) and adverbial clauses (*advcl*), which some authors (see Table 1 above) described as having VS as the default order.

CONSTITUENT ORDER IN MALTESE



[MUDT v2.7, file 57\_04N11]

Figure 3 below plots the data for both clause types in MUDT v2.7.

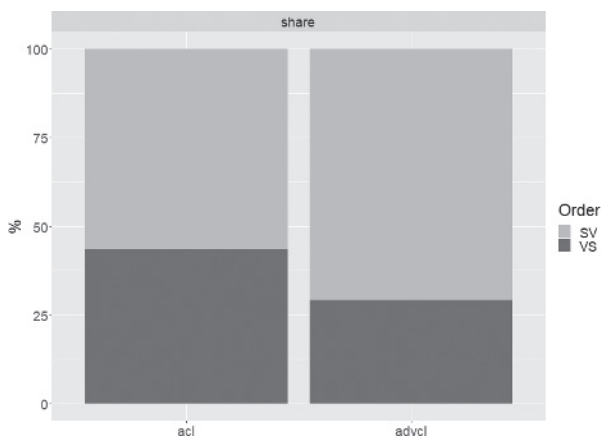


Figure 3: Constituent order in MUDT v2.7 – acl and advcl

As is evident from the plot, the dominant order for *advcl* is SV; no dominant order can be established for *acl* with the distribution of both configurations nearly equal. My preliminary investigation suggests that the VS order in *acl* is positively associated with the heaviness of the subject (i.e. its syntactic complexity and length),

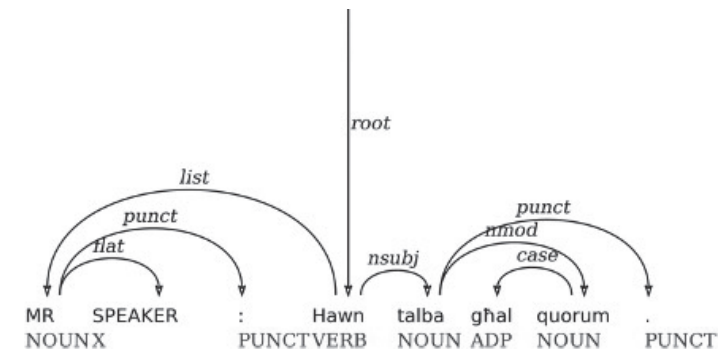
while the SV order is positively associated with the clause length (Čéplö 2018: 199-203). However, more data and a more detailed analysis are required to provide a definitive answer.

Another subdivision of clause types is that by the word class of the predicate (or root in the terminology of the dependency grammar). Greenbergian analysis is obviously limited to transitive verbs (and pseudo-verbs) only; Dryerian analysis can also take into account intransitive verbs, as well as copular clauses and other clause types. A preliminary analysis (Čéplö 2018: 218-225) has revealed that there is one clause type (defined by the word class of its root) where the default order is VS: existential clauses. These are clauses with the pseudo-verb *hemm* “there is” (and its synonym *hawn*) as the root/predicate, as in (4).

- (4) MR      SPEAKER:    *Hawn*    *talba*    *ghal*    *quorum*.  
 Mr      Speaker      EXIST    request    for      quorum  
 ‘Mr Speaker: There is a request for quorum.’

Figure 4 plots the distribution of the two possible configurations in existential clauses in MUDT UD v2.5.

This finding confirms an observation by Kalmár and Agius regarding the Maltese constituent order (Kalmár and Agius 1983: 343-344), and also a general cross-linguistic trend: as has been noted on many occasions (e.g. Givón 2001: 257), VS appears to



[MUDT v2.7, file 38\_02P06]



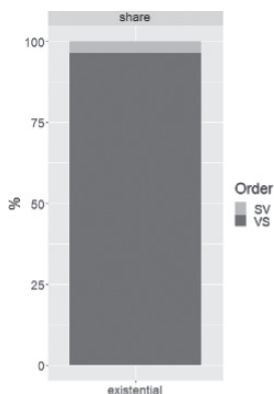


Figure 4: Constituent order in MUDT v2.7 – existential clauses

be the preferred order in existential clauses even in languages which otherwise show clear preference for SV. There are various explanations for this, for which there is little space here. For the purposes of this chapter, it suffices to conclude that Maltese is one of those languages where VS order is the dominant one in existential clauses and at the same time, existential clauses are the only clause type (defined by root) which exhibits this particular configuration as the dominant one.

Even such fine-grained analysis is far from the complete picture of Maltese constituent order, let alone its relationship to clause structure, complex sentence structure, word order, verbal valency and many other problems in Maltese syntax. It is but the first step, and the data provided by the treebank can be used to expand on it and to accomplish much more.

### 3.3 The problem of ‘free’ constituent order revisited, or: a k t fadatb zis reg nye

As we have seen in section 2, Maltese has repeatedly been described as a discourse-configurational language, either explicitly (Fabri and Borg 2002 and Borg and Fabri 2016, both citing Kiss 1995a), or implicitly: Fabri 2010 and Fabri and Borg 2017 describe Maltese

as “a topic-oriented language” (Fabri 2010: 793, see also the almost identical phrasing in Fabri and Borg 2017: 83). Considering the imprecise nature of the terminology, I take this to be a synonym of “topic-prominent language” (Kiss 1995b: 4-5), a term which in the strictest sense designates a subset of languages falling under the “discourse-configurational” umbrella, the so-called type A discourse-configurational languages, where any topicalized constituent can assume the preverbal position typically reserved for the subject (Kiss 1995b: 6-7). In type B discourse-configurational languages, focus-prominent languages, the same is true of focus (Kiss 1995b: 15-24); discourse-configurational languages can be type A, type B or both, depending on the interaction between topic and focus and on inter-language variation. Those works that describe Maltese as discourse-configurational do not elaborate on that particular aspect of this property, but judging from description of focus provided by Fabri 1993 and Fabri and Borg 2002, if Maltese is a discourse-configurational language, it is both type A and type B. This, however, is ultimately irrelevant: Maltese has been described at least twice as discourse-configurational without any elaboration or qualification and it is this description that is the focus of this section.

The framework-dependent reasoning behind this classification is not of interest here. What is, however, is the classification itself, i.e. the claim that Maltese is a discourse-configurational language; more specifically, what I want to focus on is the fact that this claim can be (to some extent) tested. The line of thinking that leads me here is the following:

1. Hungarian is considered the paragon of a discourse-configurational language (cf. Kiss 1995a), i.e. a member of a class of languages defined by a shared property involving constituent order.
2. Maltese has also been described as a discourse-configurational language.

3. Ergo, if one were to investigate the distribution of constituent order configurations in both, one would find that it is at the very least quite similar.

One might also expect that in any discourse-configurational language (and thus both Maltese and Hungarian under assumptions 1 and 2 above), the distribution of SV and VS on one hand and VO and OV on the other would be approximately the same, i.e. 50-50 for both pairs. This is, of course, not realistic, as the theory behind the classification of discourse-configurational languages makes clear: the ordering of constituents is not random<sup>8</sup> but based on pragmatic (and possibly other) criteria. Additionally, the subject is more likely to be the topic (as there is a “close correspondence between the topic and the grammatical subject”, Kiss 1995b: 10) and in any case, there are inter-language differences in how far discourse-configurationality goes. Nevertheless, the hypothesis above stands and with the Maltese UD v2.7 (Zeman, Nivre, Abrams et al. 2020) and Hungarian UD v2.5 treebanks (Zeman, Nivre, Abrams et al. 2019), there is a way to test it quantitatively.<sup>9</sup>

To conduct the actual analysis, I replicated the queries used in section 3.2 for both the Maltese UD v2.7 and the Hungarian UD v2.5 treebank. The data obtained is plotted in Figure 5 below.

The data sets underlying these two plots are, needless to say, not the same or even similar. To employ Dryer’s 2:1 method (see section 1.3), two different classifications would have to be applied here: Maltese (as represented in UD v2.5) is a language with SVO as the dominant constituent order; Hungarian (as represented in UD v2.5) is a language with no dominant constituent order.

- 8 On the other hand, both Maltese (Fabri 2010: 793) and Hungarian (Puskás 2000: 41) have been described as having “free word order”, so a case could be made that the constituent order in such languages is indeed random (in statistical terms).
- 9 Hence the subtitle of this section, best translated as “a tale of two treebanks”. Having failed to find a commonly used (or indeed any) Hungarian translation of “treebank”, I came up with my own, a portmanteau of *fa* “tree” and *adatbázis* “database”.

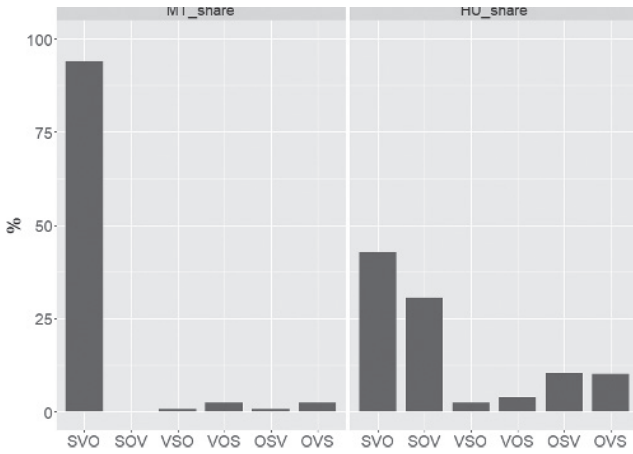


Figure 5: Maltese vs Hungarian – a Greenbergian comparison

A Dryerian analysis provides a more complicated picture (Figure 6):

The primary takeaway here is that both Maltese and Hungarian could safely be classified as SV languages. This, however, does not mean that they behave identically: as we’ve seen above, no dominant order can be established for Maltese *acl* clauses (and the distribution of the two configurations is almost equal),

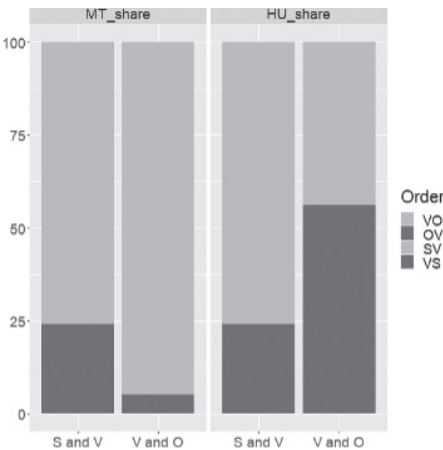


Figure 6: Maltese vs Hungarian – a Dryerian comparison

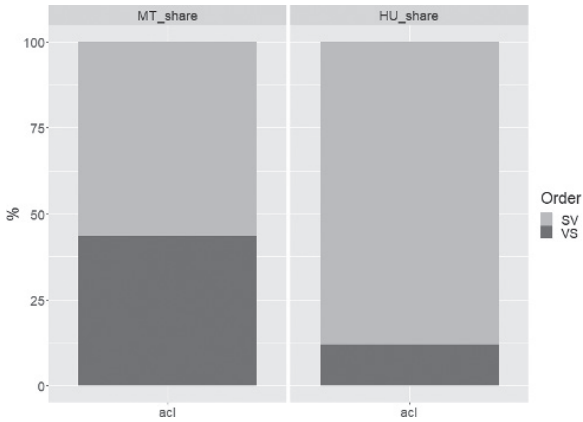


Figure 7: Maltese vs Hungarian – *acl* clauses

whereas in Hungarian, the share of the VS configuration in *acl* clauses is only half the share of the VS configuration across the board (Figure 7).

As such, Maltese *acl* clauses – and only these clauses – are much more flexible in their ordering of subject and predicate than all other clause types in Maltese; and, conversely, Hungarian *acl* clauses are much more rigid in their ordering of S and V than all other clause types in Hungarian. Whatever this means for the syntax of each respective language, the conclusion one must reach is that despite the numerical similarity, the two languages are actually not that similar in the distribution of the SV and VS constituent order configurations.

A plot of the distribution of VO/OV configurations in both languages (Figure 6) tells a much simpler story. The sharp difference between Maltese (as represented in MUDT v2.7) and Hungarian (as represented in UD v2.5) once again clearly shows that the two languages are not even similar, let alone the same, when it comes to their constituent order. Furthermore, while the data for Hungarian shows that Hungarian (as represented in UD v2.5) cannot be classified as either a VO or an OV language,

it also conforms to the naive expectation regarding constituent order variation in discourse-configurational languages expressed above: the roughly 50-50 distribution of VO and OV is what one would expect if the position of the object were only determined by pragmatic (or, more specifically, information structure) considerations: with only two options (the object is either a topic or it is not), the distribution of VO and OV really should be 1:1.

One might argue that this little comparison does not prove very much: for one, both treebanks are relatively small and thus hardly representative of the language as a whole, especially seeing as the Hungarian UD v2.5 treebank only includes journalistic texts (Zeman, Nivre, Abrams et al. 2019). Additionally, Fabri (2010: 793) may very well be correct in arguing that spoken Maltese is different from written Maltese when it comes to constituent order and so a treebank consisting of spoken materials only might offer a different picture.

As a rebuttal to the second objection, I offer this back-of-the-envelope calculation: MUDT v2.7 contains 1924 clauses featuring a *obj* or a *obl:arg*, of which 94 are OV, for a rate of 5%; the rate of OV in the Hungarian UD v2.5 treebank is 53.1%. If one were to increase the number of OV clauses in MUDT v2.7 five-fold, thus raising the total count of OV clauses to 500 (rounding up), the overall OV share in MUDT v2.7 would climb to only 21% and it would still not even approach the level of OV in the Hungarian UD v2.5 treebank. It would therefore seem more likely that MUDT v2.7 represents this particular aspect of Maltese as a whole rather faithfully (in other words, spoken Maltese may very well be different from written Maltese, but it surely isn't that different), and that this difference between the two treebanks really does represent a real difference between the two languages.

And, to answer the first objection, the composition of the Hungarian v2.5 treebank only underscores this: journalistic texts are typically written in a dry and formal style driven by desire

for clarity and brevity and produced under time crunch, which encourages the use of canned constructions (“journalese”, Suter 1993: 63-68). The fact that even when compared to a relatively balanced MUDT v2.7, the Hungarian UD v2.5 treebank is so different when it comes to the distribution of VO and OV configurations then cannot be explained away by sampling issues. This is doubly true in light of the fact that – as evident from Table 5 below – if one were to compare journalistic texts only, the difference would be even more pronounced: in those types of texts in MUDT v2.7, the share of the OV configuration (3.37%) is even lower than the average in MUDT v2.7 (5%).

Order	newspaper	quasi-spoken	fiction	non-fiction
SV	74.03%	68.57%	78.22%	85.41%
VS	25.97%	31.43%	21.78%	14.59%
VO	96.63%	94.12%	93.70%	95.57%
OV	3.37%	5.88%	6.30%	4.43%

Table 5: Constituent order in MUDT v2.7 – Dryerian analysis by genre

Consequently, there are two conclusions to be drawn here: first, Maltese (at least as represented in MUDT v2.7) really is fundamentally different from Hungarian (as represented in the Hungarian UD v2.5 treebank) when it comes to the distribution of constituent order configurations and ipso facto, the two languages cannot belong to the same class defined by a shared property related to constituent order. If one chooses to describe Hungarian as a discourse-configurational language based on the description of its constituent order, it does not seem appropriate to do the same for Maltese. By extension, neither does applying the label “topic-prominent”.

The second conclusion to be drawn from the calculations above is essentially the same as the first one, except broader and methodological rather than descriptive: Borg and Fabri (2016) use

the label “discourse-configurational” as a typological one which is itself somewhat problematic. The real problem, however, is that they do so without considering the entire theory it is based on.<sup>10</sup> As a part of a generative framework, discourse-configurationality is inexorably tied to its fundamental theory of sentence production and its complex conceptual apparatus including base generation, movements and functional projections (cf. Kiss 1995b: 9-10). And even if they were to argue that they only borrow the name and the descriptive information structure concepts behind it (as opposed to the theory of sentence generation), Borg and Fabri fail to consider one crucial property of discourse-configurational languages as defined by Kiss (1995b); the empirical distinction between categoral and thetic statements. In Kiss’s wider definition, “[a] language is identified as topic-prominent, more precisely, as a discourse configurational language with property A, if it realizes categoral and thetic judgements in different syntactic structures” (Kiss 1995b: 7-8, see also Chapter 2). Their work does not take this into account and this further invalidates their description of Maltese as a discourse-configurational or a topic-prominent language: such a label, after all, only makes sense within the context of the theory.

Ironically, I’ve shown here that Maltese actually does employ a different syntactic structure for at least one type of thetic judgments, existential clauses, so taking this into account would support Fabri and Borg’s description of Maltese as discourse-configurational as defined in the theory. This argument could be used to make a renewed case for this classification. One could, for example, extend the comparison provided here to other languages and consider the plot in Figure 8, produced from UD v2.7 for Maltese (a putative discourse-configurational language), UD v2.5 for Hungarian (a discourse-configurational language, cf. Kiss

10 This is not the case with Fabri (1993: 140) who describes Maltese as a configurational language, citing the exact definition established in generative literature (see section 1.4).



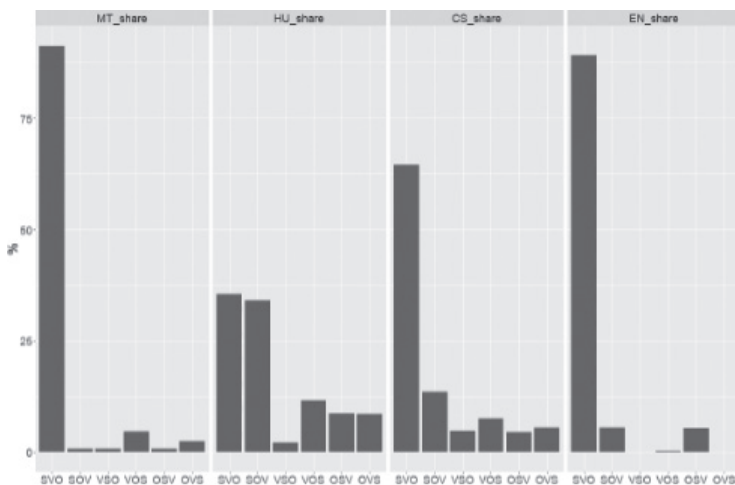


Figure 8: Greenbergian comparison of Maltese, Hungarian, Czech and English

1995b), UD v2.5 for Czech<sup>11</sup> (classified as a language with free or pragmatically determined constituent order, cf. Siewierska and Uhlířová 1998: 109-110) and UD v2.5 for English<sup>12</sup> (a language with a rigid SVO constituent order, cf. Kiss 1995b: 5, 8).

Upon reviewing this data, one could observe that Hungarian and English behave quite differently, as expected from their respective typological classifications. One could also note that Czech is quite different from English and also not that similar to Hungarian. Consequently, one could argue that discourse-configurationality (or indeed topic-orientedness or pragmatical determination of constituent order) is a scale, with Hungarian on one end and English on the other. Whether that would be consistent with the theory is beside the point, what is important is that based on the data above, Maltese (at least as represented in MUdT v2.7) looks much more like a strict SVO language like English, rather than a discourse-configurational language like Hungarian, or a language with pragmatically determined constituent order like Czech.

11 More specifically, the Czech-PDT UD treebank in version 2.5.

12 The English GUM treebank in version 2.5.

## Abbreviations

1, 2, 3	first, second, third person	INTJ	interjection
ACC	accusative	M	masculine
COMP	complementizer	NEG	negative
DEF	definite article	PAST	past
EXIST	existentials	PL	plural
F	feminine	SG	singular

## Acknowledgements

The work underlying this paper was partly financed by grant number APVV-15-0030 from the Slovak Research and Development Agency (APVV), and by the ERC Starting Grant 679083 from the European Research Council, whose support is hereby gratefully acknowledged.

## References

- Aquilina, Joseph (1959) *The structure of Maltese: A study in mixed grammar and vocabulary*. Malta: The Royal University of Malta.
- Borg, Albert & Azzopardi-Alexander, Marie (1997) *Maltese. Lingua descriptive grammars*. London & New York: Routledge.
- Borg, Albert & Azzopardi-Alexander, Marie (2009) Topicalisation in Maltese. In: Comrie, Bernard, Fabri, Ray, Hume, Elizabeth, Mifsud, Manwel, Stolz, Thomas & Vanhove, Martine (eds.), *Introducing Maltese linguistics*. Amsterdam: John Benjamins, 71–81.
- Borg, Albert & Fabri, Ray (2016) Semantic functions of complementizers in Maltese. In: Boye, Kasper & Kehayov, Peter (eds.), *Complementizer semantics in European languages*. Berlin & Boston: De Gruyter Mouton, 413–447.
- Broekhuis, Hans (2008) Derivations and evaluations: Object shift in the Germanic languages. *Studies in Generative Grammar 97*. Berlin & New York: Mouton de Gruyter.
- Carnie, Andrew (2013) *Syntax: A generative introduction*. Chichester: Wiley-Blackwell.
- Čeplö, Slavomír (2014) An overview of object reduplication in Maltese. In: Borg, Albert, Caruana, Sandro & Vella, Alexandra (eds.), *Perspectives on Maltese linguistics*. Berlin: Akademie Verlag, 201–223.

- Čéplö, Slavomír (2018) *Constituent order in Maltese: A quantitative analysis*. Prague: Charles University dissertation.
- Choe, Hyoon Sook (1995) Focus and topic movement in Korean and licensing. In: Kiss, Katalin É. (ed.), *Discourse Configurational Languages*. Oxford: Oxford University Press, 269–334.
- Chomsky, Noam (1965) *Aspects of the theory of syntax*. Cambridge, Mass: MIT Press.
- Corver, Norbert & van Riemsdijk, Henk (1994a) Introduction: approaches to and properties of scrambling. In: Corver, Norbert & van Riemsdijk, Henk (eds.), *Studies on scrambling: movement and non-movement approaches to free word-order phenomena*. Berlin & New York: Mouton de Gruyter, 1–13.
- Corver, Norbert & van Riemsdijk, Henk (1994b) *Studies on scrambling: Movement and non-movement approaches to free word-order phenomenon. Studies in generative grammar*. Berlin & New York: Mouton de Gruyter.
- de Soldanis, Giovanni Pietro Francesco Agius (1750) *Della lingua punica presentemente usata da maltesi*. Rome.
- Derbyshire, Desmond C. (1977) Word order universals and the existence of OVS languages. In: *Linguistic Inquiry* 8.3, 590–599.
- Dixon, Robert M. W. (2009) *Basic linguistic theory- vol 1: Methodology*. Oxford: Oxford University Press.
- Dryer, Matthew S. & Haspelmath, Martin (eds.) (2013) *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info> (checked 02/29/2020)
- Dryer, Matthew S. (1989a) Discourse-governed word order and word order typology. In: *Belgian journal of linguistics* 4, 69–90.
- Dryer, Matthew S. (1989b) Large linguistic areas and language sampling. In: *Studies in Language* 13, 257–292.
- Dryer, Matthew S. (1992) The Greenbergian word order correlations. In: *Language* 68.1(1), 81–138.
- Dryer, Matthew S. (1997) On the 6-way word order typology. In: *Studies in Language*, 69–103.
- Dryer, Matthew S. (2007) Word order. In: Shopen, Timothy (ed.), *Language typology and syntactic description. Second edition- vol. I: Clause Structure*. Cambridge: Cambridge University Press, 61–131.
- Dryer, Matthew S. (2013a) Determining dominant word order. In: Dryer, Matthew S. & Haspelmath, Martin (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology, n/a.
- Dryer, Matthew S. (2013b) On the six-way word order typology, again. In: *Studies in Language* 37.2, 267–301.
- Dryer, Matthew S. (2013c) Order of subject, object and verb. In: Dryer, Matthew S. & Haspelmath, Martin (eds.) *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Fabri, Ray (1993) *Kongruenz und die Grammatik des Maltesischen* (Agreement and the Grammar of Maltese). Tübingen: Niemeyer.
- Fabri, Ray (2010) Maltese. In: *Revue belge de philologie et d'histoire* 88.3, 791–816.
- Fabri, Ray & Borg, Albert (2002) Topic, focus and word order in Maltese. In: Abderrahim, Youssi, Benjelloun, Fouzia, Dahbi, Mohamed & Iraqui-Sinaceur, Zakia (eds.), *Aspects of the dialects of Arabic today. Proceedings of the 4<sup>th</sup> Conference of*

- the International Arabic Dialectology Association (AIDA). Marrakesh, Apr. 1-4.2000. In Honour of Professor David Cohen. Rabat: Amapatriil, 354–363.*
- Fabri, Ray & Borg, Albert (2017) Modifiers and complements within the Maltese verb sequence. In: Saade, Benjamin & Tosco, Mauro (eds.), *Advances in Maltese linguistics*. Berlin & New York: De Gruyter Mouton, 67–86.
- Féry, Caroline & Shinichiro Ishihara (2016a) Introduction. In: Féry, Caroline & Shinichiro Ishihara (eds.), *The Oxford handbook of information structure*. Oxford: Oxford University Press, 3–15.
- Féry, Caroline & Shinichiro Ishiharaeds (eds.) (2016b) *The Oxford handbook of information structure*. Oxford: Oxford University Press.
- Firbas, Jan (1964) On defining the theme in functional sentence analysis. In: *Travaux Linguistiques de Prague* 1, 267–280.
- Firbas, Jan (1992) *Functional sentence perspective in written and spoken communication*. Cambridge: Cambridge University Press.
- Gatt, Albert & Čéplò, Slavomír (2013) Digital corpora and other electronic resources for Maltese. *Talk at the Corpus Linguistics 2013 conference at Lancaster University*.
- Givón, Talmy (2001) *Syntax. An introduction- vol. II*. Amsterdam & Philadelphia: John Benjamins.
- Greenberg, Joseph H. (1966) Some universals of grammar with particular reference to the order of meaningful elements. In: Greenberg, Joseph H. (ed.), *Universals of language*. Mass: MIT Press, 73–113.
- Hale, Ken (1983) Warlpiri and the grammar of non-configurational languages. In: *Natural Language & Linguistic Theory* 1.1, 5–47.
- Harris, Randy Allen (1995) *The linguistics wars*. Oxford: Oxford University Press.
- Itkonen, Esa (2005) Concerning the synthesis between intuition-based study of norms and observation-based study of corpora. In: *SKY Journal of Linguistics* 18, 357–377.
- Kalmár, Ivan & Agius, Dionisius (1981) Verb-subject order in Maltese. In: *Journal of Maltese Studies* 14, 22–31.
- Kiss, Katalin É. (ed.) (1995a) *Discourse configurational languages*. Oxford: Oxford University Press.
- Kiss, Katalin É. (1995b) Discourse configurational languages: Introduction. In: Kiss, Katalin É. (ed.), *Discourse configurational languages*. Oxford: Oxford University Press, 3–27.
- Krause, Thomas & Zeldes, Amir (2016) ANNIS3: A new architecture for generic corpus query and visualization. In: *Digital Scholarship in the Humanities* 31.1, 118–139.
- Krier, Fernande (1976) *Le maltais au contact de l'italien. Etude phonologique, grammaticale et sémantique*. Hamburg: Helmut Buske Verlag.
- Krifka, Manfred & Musan, Renate (2012) Information structure: Overview and linguistic issues. In: Krifka, Manfred & Musan, Renate (ed.), *The Expression of Information Structure*. Berlin/Boston: De Gruyter Mouton, 1–44.
- Lewis, Paul M., Simons Gary F. & Fennig, Charles D. (2016) *Ethnologue: Languages of the world*, Nineteenth edition. Dallas, Texas: SIL International.
- Marty, Anton (1897) Über die Scheidung von grammatischem, logischem und psychologischem Subjekt resp. Prädikat. In: *Archiv für Systematische Philosophie* 3, 174–189.

- Marty, Anton (1895) Über Subjektlose Sätze und das Verhältniss der Grammatik zu Logik und Psychologie [Sechster Artikel]. In: *Vierteljahrsschrift für wissenschaftliche Philosophie* 19, 19–87, 263–334.
- Mathesius, Vilém (1939) O tak zvaném aktuálním členění věty. In: *Slovo a slovesnost* 5, 171–174.
- Mathesius, Vilém (1961) Obsahový rozbor současné angličtiny na základě obecně lingvistickém. Praha: Nakladatelství Československé akademie věd.
- McEnery, Tony & Hardie, Andrew (2011) *Corpus linguistics: Method, theory and practice*. Cambridge: Cambridge University Press.
- Milička, Jiří (2014) Menzerath's Law: The whole is greater than the sum of its parts. In: *Journal of Quantitative Linguistics* 21(2), 85–99.
- Payne, Thomas E. (1997) *Describing morphosyntax: A guide for field linguists*. Cambridge: Cambridge University Press.
- Puskás, Genoveva (2000) *Word order in Hungarian. The syntax of  $\bar{A}$ -positions*. Amsterdam: John Benjamins.
- Ross, John R. (1967) *Constraints on variables in syntax*. PhD thesis. Massachusetts Institute of Technology.
- Siewierska, Anna & Uhlířová, Ludmila (1998) An overview of word order in Slavic languages. In: Siewierska, Anna (ed.), *Eurotyp: Typology of languages in Europe - vol 1: Constituent order in the languages of Europe. Empirical approaches to language typology* 20.1. Berlin & New York: Mouton de Gruyter.
- Song, Jae Jung (ed.) (2011a) *The Oxford handbook of linguistic typology*. Oxford: Oxford University Press.
- Sutcliffe, Edmund (1936) *A grammar of the Maltese language. With chrestomathy and vocabulary*. Oxford: Oxford University Press.
- Suter, Hans-Jürg (1993). *The wedding report. A prototypical approach to the study of traditional text types*. Amsterdam: John Benjamins.
- Vella, Francis (1831) *Maltese grammar for the use of the English*. Leghorn: Glaucus Masi.
- Vella, Joseph (1970) *A comparative study in Maltese and Libyan (Benghazi Dialect). Phonetics, morphology, syntax & lexicon*. Malta: The Royal University of Malta (Unpublished PhD thesis).
- von der Gabelentz, Georg (1869) Ideen zu einer vergleichenden Syntax. In: *Zeitschrift für Völkerpsychologie und Sprachwissenschaft* 6, 376–384.
- Weil, Henri (1844) *De l'ordre des mots dans les langues anciennes comparées aux langues modernes*. Paris: Joubert.
- Zeman, Daniel, Nivre, Joakim, Abrams, Mitchell et al. (2019) *Universal dependencies 2.5*. 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/11234/1-3105>
- Zeman, Daniel, Nivre, Joakim, Abrams, Mitchell et al. (2020). *Universal dependencies 2.7*. 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/11234/1-3424>
- Zubizarreta, Maria Luisa (1998). *Prosody, focus, and word order*. Cambridge, Massachusetts & London, England: The MIT Press.