

**Univerzita Karlova
Filozofická fakulta**

Ústav srovnávací jazykovědy
Filologie – matematická lingvistika

Slavomír Čéplö

**Constituent order in Maltese:
A quantitative analysis**

**Vetný slovosled v maltčine:
Kvantitatívna analýza**

Autoreferát disertační práce

Vedoucí práce: doc. Petr Zemánek, CSc.

2018

1 Theory and goals

1.1 Approach

This thesis is, as apparent from the title, a study of constituent order in Maltese and as such, it is a work on Maltese syntax and (to a smaller extent) pragmatics. The first chapter of a thesis is normally the place for setting the research questions and describing the data and methodology employed in answering them, but before that, there are broader issues to be discussed, issues of fundamental importance that are often taken for granted or downright ignored, like the nature of linguistics, its goals and its methods.

The general approach I employ in this thesis is best described using the adjectives “descriptive” and “empirical”. What follows is the definition of those terms and the reasoning behind them. It is my view that the primary task of linguistics is to describe a language i.e., in Haspelmath’s (2009: 344) definition, to provide a “characterization of grammatical regularities” of a language. In much of linguistic literature, the descriptive approach is contrasted with the theoretical approach, where the latter is rooted in a particular framework, i.e. “a sophisticated and complex metalanguage for linguistic description intended to work for any language” (Haspelmath 2009: 343). It is primarily within the context of that dichotomy that I wish to characterize my approach as descriptive and framework-free: I aim to provide a description of a part of the grammar of a particular language while doing so outside of any existing theoretical framework, i.e. considering the language on its own, without any conscious preconceptions or biases. In this sense, I prefer to speak of a description that accounts for the data. The traditional adjective applied in these circumstances is empirical. I will gladly accept it in the context of its contrast to introspective or intuitive approaches (Itkonen 2005).

1.2 Metalanguage

While this work eschews the use of any particular framework and strives to describe its object of study on its own terms, a metalanguage is nevertheless necessary for the description of the phenomena it sets out to study. In this thesis, I employ a compromise and use terms that are largely familiar to anyone who has ever read a grammar, but with their meaning extended or narrowed as necessary, a solution practiced by linguists for centuries. This section defines the terms I use and the concepts behind them, starting with two fundamental ones, “syntax” and “Maltese”. As for the former, syntax is discussed here in terms of the principles of dependency syntax (Tesnière 1959: 11-15, Tesnière 2015: 3-6) as implemented by the Universal Dependencies project (UD, de Marneffe et al. 2014, Nivre, Ginter, et al. 2014, Nivre, Marneffe, et al. 2016 and Nivre, Ginter, et al. 2016). The latter, the primary subject of this thesis, is then defined as “the written language produced by native speakers of Maltese in the first two decades of the 21st century as represented by the texts contained in the two corpora used as the source of data for the analysis herein (Chapter 5 and 6)”.

Using dependency grammar as the starting point, this section goes on to provide the definitions to a number of concepts fundamental to the purposes of this work, chief among them “sentence”, “predicate”, “clause” and especially “constituent order”. In this thesis, a strict distinction between constituent order and word order is maintained. This is a contradiction to the dependency-based analysis employed here, as dependency grammars deal with relationships between words and thus anything that involves the order and the sentence is, by definition, word order. I have established the distinction for reasons of clarity: this thesis is only concerned with the order of the predicate and its core arguments (as defined in UD) and does not address the order of elements within phrases (word order), save in passing.

2 Approaches to the study of constituent order

Constituent order 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 the most of the world's languages are under-described, it is often the only piece of information on the grammar of a language available. This chapter examines a number of approaches to the study of constituent order current in modern linguistics: the typological approach, the generative approach, pragmatic approaches and the quantitative approach. These approaches have been selected primarily to inform the following discussion of previous works on constituent order in Maltese, as well as to provide the basis for the research questions.

In this context, the typological approach stands out as the primary reason for the ubiquitous appearance of constituent order in even the most rudimentary descriptions of languages (see e.g. Lewis, Simons, and Fennig 2016 or Song 2011). This is largely due to the enormous influence of Greenberg 1966 and his six-way typology (SVO, SOV etc.) which has been shown to correlate with a number of grammatical features. Greenberg's typology continues to be the dominant paradigm in linguistic typology and descriptive linguistics, whether unchanged or modified. One of those modifications, Dryer's binary SV/VS and VO/OV typology (Dryer 1997, Dryer 2013b), offers compelling arguments in its favor as an alternative to Greenberg's typology: first, as Dryer notes, such typology "is based not only on clauses containing both a nominal subject and a nominal object but also on clauses containing just one of these" (Dryer 2013b: 269). This is not only appropriate in light of the existence of transitive and copular clauses and the fact that clauses featuring only one of the core verbal arguments "occur much more frequently" (Dryer 1997: 70), but it is also particularly relevant for languages like Maltese where the nominal subject is not obligatory in verbal clauses. Secondly, the binary typology allows for fine-grained analysis and better visualization, especially when multiple objects of analysis (i.e. various types of clauses) are involved. Dryer's binary typology is therefore of particular relevance for the analysis attempted here.

3 Maltese constituent order: state of the question

Chapter 3 examines previous works on Maltese constituent order, from the earliest grammatical descriptions which devote some attention to constituent order (F. Vella 1831) to the most recent ones. Two constant themes are interwoven throughout this literature: 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 J. 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: 131 and Fabri 2010: 793), including synonyms like "discourse-configurational" (Fabri and Borg 2002, Borg and Fabri 2016) and "topic-oriented" (Fabri 2010: 793, Fabri and Borg 2017: 83), all of which 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).

Both 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 (Krier 1976, 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, which additionally 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, Borg and Azzopardi-Alexander 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 broad-

est terms, e.g. Borg and Azzopardi-Alexander 1997: 126) provide a description of how those possibilities are instantiated.

In what follows, I set out to remedy these shortcomings.

4 Research questions

4.1 Introduction

This work seeks to provide the answers to the following questions:

1. What is the dominant constituent order in Maltese?
2. What is the variation in dominant constituent order in Maltese?
3. What are the deviations from the dominant constituent order in Maltese?
4. What are the determinants of variation in Maltese constituent order?

4.2 Research Questions

4.2.1 Research Question 1

In addition to general descriptive considerations, Research Question 1 is motivated by previous research into Maltese constituent order, most of which assumes (to some extent) the existence of default (unmarked, basic, dominant) order. Consequently, the main task at hand is to determine whether there is a default (unmarked, basic, dominant) constituent order configuration in Maltese and what it is.

In light of the typological nature of this investigation, constituent order in Maltese is primarily analyzed in terms of Dryer's binary SV/VS and VO/OV typology (Dryer 1997, Dryer 2013b). The concept of "dominant constituent order" is likewise borrowed from Dryer 2013 who defines the dominant order (whether constituent order or word order) as follows:

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[.]

To answer Research Question 1, a syntactically annotated corpus (treebank) of Maltese is examined to determine the distribution of SV/VS and VO/OV orders in clauses contained therein. The syntactic annotation used in the treebank is based on the UD standard and accordingly, the quantitative analysis of constituent order configurations is performed not only across all clauses, but also separately for main clauses and various types of dependent clauses as defined by UD.

This type of analysis, along with the concept of dominant constituent order used here, necessitates establishing and defining two types of variation from dominant constituent order: the first, termed "variation" proper, is used for situations where one or several types of clauses (however defined) display dominant constituent order different from that in other clause types or across all clauses. The second term, "deviation", is used for the non-dominant configuration: recall that the dominant constituent order is defined here in statistical terms as the configuration which is twice as frequent as the other option; that other option is then referred to as "deviation" or "deviant order".

4.2.2 Research Question 2

This question seeks to address the second part of the assumption underlying Research Question 1, i.e. the existence of alternative dominant constituent order(s) in certain types of clauses, including situations when the dominant one cannot be established.

The answer to Research Question 2 is provided using the same type of quantitative analysis employed in answering Research Question 1, as a complement to it.

4.2.3 Research Question 3

This question focuses on contexts where the dominant constituent order could be established, but the clauses in question exhibit the non-dominant order. The primary purpose of this question is to check a number of observations made regarding the topicalization of the direct and indirect object in the literature on Maltese, best summarized by Borg and Azzopardi-Alexander who describe it as "such a wide spread characteristic of Maltese, that it even features in Maltese English" (Borg and Azzopardi-Alexander 1997: 126).

The answer to this question is provided as a complement to the answer to Research Question 1.

4.2.4 Research Question 4

Having established what the variation in dominant constituent order is, the focus shifts to those types of clauses that exhibit said variation, analyzing their structure and attempting to determine what causes said variation.

The issue of to what extent is the variation (and deviation) in constituent order a phenomenon rooted in grammar (i.e. syntax) and to what extent it is a pragmatic (information structure) phenomenon is one of the major problems in the study of constituent order. That both are involved is now taken for granted: schools of thought that started out arguing for the former now recognize the role of pragmatics and included information structure concepts in their theory (e.g. generative linguistics and the concept of discourse-configurationality or the Cartographic Project); those schools of thought that do focus on the role pragmatics have always explicitly recognized the role that syntactic constraints play in constituent order variation (e.g. Firbas 1992: 118). In my efforts to map and account for constituent order variation and deviation in Maltese, I therefore focus on two major players: syntax and information structure.

The former is obvious and in line with the descriptive approach employed here; same goes for the practical aspect of it and the syntactic factors are examined in terms of dependency relations. Additionally, concepts relating to description of language in quantitative terms are introduced into the discussion, some relatively straightforward, like clause length (Köhler 2012: 142-146), some less so, like "heaviness" (Arnold et al. 2000) which has repeatedly been found to influence the ordering of constituents (Arnold et al. 2000: 51, Stolz 2011 for Maltese).

As for information structure, this is primarily discussed in terms of describing constituent order variation, as well as while addressing previous typological classifications of Maltese. The analysis of constituent order deviation (assuming any is found) where information structure plays a role is, for the most part, outside of the scope of this work.

The clauses which exhibit variation in dominant constituent order are analyzed either computationally or manually as to their syntactic, semantic and pragmatic properties, primarily as compared to those clauses that exhibit the dominant order.

4.3 Data and methodology

4.3.1 Data

The analysis as outlined above is performed using corpus data. These come in two forms: the primary source of data for the quantitative analyses is the Maltese Universal Dependencies Treebank v1 (MUDTv1). This treebank, annotated according to the Universal Dependencies annotation standard, version 1 (UD v1; Nivre, Ginter, et al. 2014; Nivre, Marneffe, et al. 2016), is the first ever compiled for Maltese; I have created it myself for the purpose of this thesis and it will be made available to the public upon its defense with the hopes that should I fail to achieve the goals set herein, the treebank will at least be of some use to someone. Chapter 6 describes in detail the composition of the treebank, the annotation decisions and the reasoning behind them.

The other data source is the general corpus of Maltese which I have also compiled myself (*bulbulistan maltiv3*, *BCv3*), described in Chapter 5. *BCv3*, despite being only annotated with the bare minimum of linguistic information and thus incapable of serving as data source for the actual analysis of constituent order in Maltese, nevertheless plays a crucial role here: first, it is the primary source of texts for MUDTv1. Secondly, it provides material for the analysis of syntactic phenomena that are being described as a part of the annotation of syntactic relations in MUDTv1 from the fundamentals of linguistic analysis like part-of-speech tags all the way to verbal valency, which is crucial for the phenomena under study. And finally, it is used to check and test information obtained from the analysis of MUDTv1. As such, it is an integral part of this work.

4.3.2 Methodology

The primary tools employed here are those of descriptive statistics. Using the data from MUDTv1, I provide the basic statistics on the distribution of the orders of the subject and the predicate (SV/VS) and the object and the predicate (VO/OV), including visualizations thereof, to determine the dominant constituent order both across the entirety of MUDTv1, as well as per clause type. Wherever applicable, I also apply methods for the testing of statistical significance, primarily to determine whether the differences encountered (such as the ratio of one configuration versus another) are real or only due to chance. And finally, in scenarios such as those where no dominant order can be established, I use statistical modeling to account for it.

5 *BCv3*: A corpus of written Maltese

Next to the *Korpus Malti* (a part of the Maltese Language Resource Server)¹ developed by Albert Gatt at the University of Malta, the *bulbulistan* corpus² is one of the two major digital corpora of Maltese (see Gatt and Čéplö 2013 for a preliminary description of both). Originally conceived as independent projects, their most recent versions (*Korpus Malti v3.0* and *bulbulistan maltiv3*, henceforth *MLRSv3* and *BCv3* respectively) have taken a step towards the eventual integration of both corpora into a single resource by sharing data, adopting standardized processing methods, expanding their reach to over 200 million word tokens and developing a common part-of-speech tagging scheme.

¹ mlrs.research.um.edu.mt (last consulted on February 28th 2018)

² bulbul.sk/bonito2

Text type	Documents	Sentences	Tokens	%
newspaper			113,120,057	52.31%
parliament			98,333,466	45.47%
fiction			2,437,831	1.13%
non-fiction			2,351,730	1.09%
Total	313,499	9,769,815	216,243,084	100%

Tab. 1: Text types in *BCv3*

While *MLRSv3* serves as the focal point of Maltese corpus linguistics, *BCv3* continues its existence as a separate entity for technical reasons, legacy reasons and as the data source for a number of special projects, of which this thesis is the primary one. This chapter describes in detail the composition (Table 1) of *BCv3* as an opportunistic corpus tailored to the purposes of this thesis, its processing and enrichment, particularly part-of-speech tagging which is shared with *MLRSv3* and provides a background to the syntactic analysis and annotation of the treebank.

6 Maltese Universal Dependencies Treebank v1

This chapter describes a Maltese treebank based on the Universal Dependencies standard, version 1 (UD v1; Nivre, Ginter, et al. 2014, de Marneffe et al. 2014). UD v1 was chosen as the annotation scheme for MUDTv1 for many reasons, chief among them the fact that UD v1 has been adopted as the de-facto industry standard for dependency parsing (as evidence by its use by Google in Andor et al. 2016) and the fact that it is a remarkably well organized and managed project which meanwhile includes over 60 languages.

The core of the chapter contains a detailed description of the preparation of MUDTv1, particularly the process of syntactic annotation. This amounts to the compilation of a sketch of Maltese syntax and while a large part of the annotating decisions was based on previous descriptions of Maltese (primarily Borg and Azzopardi-Alexander 1997 and Vanhove 1993), some aspects of annotation required a detailed treatment of underdescribed phenomena. This includes the starting point of the analysis which involved the classification of clauses by root into verbal, copular, non-verbal verbless, existential and non-expletive subjectless clauses where, inter alia, previous descriptions of copular clauses were refined and a new analysis of existential clauses was provided.

Primarily, however, the syntactic analysis involves the application of the UD relations (Table 2) to Maltese and the detailed discussion of these and related phenomena.

	Nominals	Clauses	Modifier words	Function Words
Core arguments	nsubj nsubjpass dobj iobj <u>nmod:obj</u>	csubj csubjpass ccomp xcomp		
Non-core dependents	nmod * nmod:agent <u>nmod:advmod</u> vocative expl dislocated	advcl	advmod* discourse	aux auxpass cop mark neg <u>part</u>
Nominal dependents	nmod* <u>nmod:poss</u> appos nummod	acl	amod advmod*	det case <u>case:det</u>
Coordination	MWE	Loose	Special	Other
conj	compound	list	foreign	punct
cc	mwe name	parataxis	goeswith remnant reparandum	root dep

Tab. 2: UD v1 relations adapted to and extended for Maltese

In this context, three problems take center stage: the issue of verbal auxiliaries (where the decision was made to only annotate the copular verb *kien* as such, contrary to previous analyses), the related issue of `xcomp` clauses (which present a significant problem in the syntactic analysis of Maltese, see e.g. Maas 2009, Stolz 2009 and Fabri and Borg 2017, and where only a moderately satisfactory solution could be achieved) and finally, the issue of core verbal dependents. This problem is central to the primary purpose of the thesis and has so far gone largely untreated in works on Maltese grammar; Borg and Azzopardi-Alexander 1997: 55-57 briefly discuss the distinction between direct and indirect objects and even highlight the existence of object-like verbal arguments introduced by prepositions, but do not take the matter any further.

Chapter 6 therefore seeks to put the analysis of core verbal dependents on a more solid footing by using Tesnière’s concept of valency, especially as implemented in VALLEX (Lopatková et al. 2017). As the first step, non-core arguments (free verbal dependents) are identified and, in some cases, special subtypes are broken off the respective UD relation, such as the possessive nominal (`nmod:poss`), the passive agent (`nmod:agent`) and the nominal adverbial (`nmod:advmod`) where for the latter, the VALLEX criteria for free dependents were the basis of identification. As for the core dependents, the three standard ones – nominal subject (`nsubj`), direct object (`dobj`) and indirect object (`iobj`) – are identified largely by morphosyntactic criteria (e.g. agreement, whether in terms of affixes or clitics). To them, a new type of core dependent is added, the so-called non-canonical object (`nmod:obj`), a dependent semantically equivalent to either a direct or an indirect object (in terms of VALLEX functors PAT, EFF or ADDR) or a dependent that is obligatory, regardless of whether it is realized as a noun phrase or a prepositional phrase. In practical terms, this entailed carefully examining the valency frame of each verb using data from *BCv3* and classifying their dependents based on a simple decision tree:

- I. A core verbal dependent is a verbal dependent that
 - a. is obligatory (as evidenced in *BCv3*); and
 - b. is not an adverbial (with the list of free dependents in VALLEX as a rough guideline); and
 - c. can only appear once with a single verb instance.
- II. A verb may not take an obligatory dependent, but if it does (as evidenced in *BCv3*) and the dependent
 - a. is of a specific type; and
 - b. fulfills a particular semantic role (VALLEX actants); and

- c. can only appear once,
that dependent is also considered core.

The first main branch is rather self-explanatory, if somewhat complicated in its second sub-branch which involves semantic analysis. Consider, for example, the imperfect verb *ddependa* "to depend" which in BCv3 primarily occurs in 3rd person imperfect (11,712 hits, 54.16 per million) and takes only one type of nominal dependent, an obligatory prepositional phrase introduced either by *min* "from" or *fuq* "on". Both are normally locative prepositions, yet in this context, the phrases they feature in can hardly be interpreted as either directional or locative adverbials. They could be interpreted as adverbials of origin, but the semantics of the verb do not support this, as there is no movement denoted in the verb and in any case, this would only apply to those phrases introduced by *min*. Such a prepositional phrase is therefore considered a core dependent.

As for the second main branch, it is best illustrated by the example of the verb *nduna* "to notice" which often occurs with a verbal dependent prepositional phrase introduced by *bi* "with". When such a construction is encountered, BCv3 is queried for all forms of the verb to find that this verb typically takes no nominal dependents save for the subject, but if it does take other nominal dependents, they are overwhelmingly introduced by *bi* (subbranch IIa). Moreover, such dependents typically fulfill the semantic role of a direct object (patient, VALLEX PAT) in that they denote the person, object or phenomenon observed (subbranch IIb). As such, they cannot be considered either facultative or adverbials and must therefore be core dependents.

Based on the description above, one might have the suspicion that this type of analysis only investigates verbal government (Rektion) or a special subclass of verbs (so-called prepositional verbs, Aquilina 1976: 67-80), but this is not the case: first, the primary purpose here is to determine which prepositional phrases can be considered core dependents and which are adverbials; this requires a more careful analysis than just "this verb takes dependents marked with this preposition". Secondly and more importantly, since Maltese verbs can take adverbials consisting of noun phrases unmarked for case (the UD v1 relation), such an investigation requires a careful analysis of all dependents of a particular verb, i.e. its entire valency frame, and cannot be reduced to the deciding whether a particular prepositional phrase is an adverbial or a non-core dependant. This path lead to a number of surprising and hitherto unexamined nooks in the study of Maltese: one is the classification of certain verbs as trivalent where the third core dependent (along with subject and direct object) is not a traditionally conceived indirect object, such as the verb *tkellem* "to speak" or *nforma* "to inform" which both take a direct object and an obligatory (in terms discussed above) prepositional phrase introduced by *dwar* or *fuq*. The other is the issue of core dependents of verbs denoting change like *mela* "fill" which take an (unmarked) direct object indicating the patient and another unmarked core dependent indicating the effect; in passive clauses, this dependent must be introduced by a preposition.

The full account of these phenomena is well beyond the scope of this work; chapter 6 provides a basic analysis of valency frame in Maltese which, along with the principles and rules of UD v1, can be used to make quick and consistent decisions for the purposes of syntactic annotation of core verbal dependents.

In terms of composition, MUDTv1 seeks to offset the opportunistic nature of BCv3 and sets balance in terms of text type as its primary goal; the secondary goal is then a size exceeding that of the smallest UD v2.1 treebanks (Table 3).

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

Tab. 3: MUDTv1 composition

For visualization and analysis purposes, ANNIS3 (Krause and Zeldes 2016) was selected for its flexibility, adaptability and ease of use; the particular installation is available publicly at bulbul.sk/annis-gui-3.4.4. The full set of files was also converted to the standard format for UD visualization and the individual HTML files are available at bulbul.sk/bonito2/treebank (login name: guest, password: Ghilm3), as well as in Appendix B, directory HTML.

7 Dominant constituent order and its variations in Maltese: A quantitative analysis

7.1 Introduction

With the metalanguage (Chapter 1), data (Chapter 5 and 6) and methodology (Chapter 4) established, this chapter proceeds to answering the research questions in the order in which they were asked, starting with the question of what is the statistically dominant constituent order in Maltese, both in general terms, as well as in specific types of clauses. The variation in constituent order is analyzed as well and the chapter concludes with a typological characterization of the constituent order in Maltese.

7.2 Basic statistics

7.2.1 Sentence length and complexity

The analysis begins with a brief description of MUDTv1 in terms of sentence length (Table 4) and sentence branching (Figure 1). The former underscores the special nature of journalistic texts (echoing the findings in Fenech 1978) and thus the need for balance in any Maltese corpora.

Text type	Total sentence count	Mean sentence length (tokens)	SD	SE
newspaper	479	26.8	13.25	0.60
quasi-spoken	574	21.6	16.18	0.67
fiction	497	16.9	11.07	0.49
non-fiction	524	24	13.19	0.57
MUDTv1	2074	22.3	14.09	0.31

Tab. 4: MUDTv1: Mean sentence length by text type

The analysis of sentence branching (in terms of order of sentence root and dependent clauses), while only a subset of the typology of branching established by Dryer 1992, nevertheless provides a preliminary classification of Maltese as right-branching.

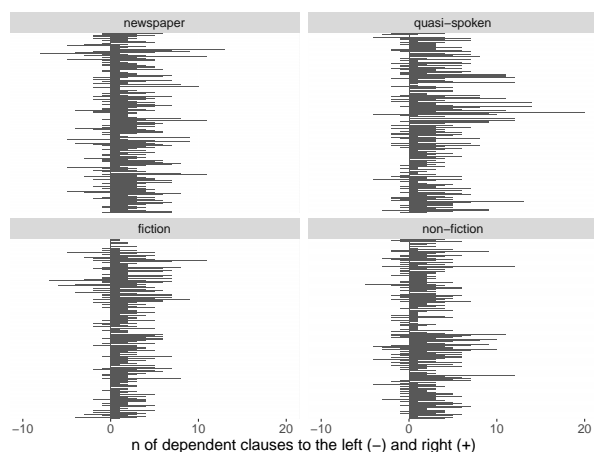


Fig. 1: MUDTv1: Sentence complexity by text type

Additionally, all these findings provide definitive justification for including as many and as varied types of texts in MUDTv1 as possible: relying only on one text type – say, journalistic texts, which are shown to be quite different from other text types in many ways – would provide a very skewed picture of Maltese. And while representativeness may be a pipe dream for Maltese corpus linguistics, balance should be a priority.

7.2.2 Clause types

7.2.2.1 General

The following analysis considers constituent order from the point of view of two classifications of clauses employed here so far UD clause types and clause types by root. In this section, basic statistics for both classifications are provided.

7.2.2.2 UD clause types

Table 5 below contains an overview of UD clause types in MUDTv1 (as per UD v1) and the count of their occurrences.

Clause type	Number
Main clauses	2074
acl	1318
advcl	843
xcomp	1375
ccomp	684
parataxis	185
conj	871
csubj	23
Total clauses	7373

Tab. 5: MUDTv1: UD clause types

This list includes all main clauses (root) and their direct or indirect dependents with the labels listed above, save for conj; in their case, only those catenae with a head marked as conj were taken into account that were verbal or existential predicates or had a subject and/or a copula as a dependent.

7.2.2.3 Clause types by root

The second classification of clauses used here is that by their structure, i.e. their root. Since it is the order of the predicate (root) and its core nominal dependents (nsubj, nsubjpass, dobj and nmod:obj) that is the primary focus of this analysis, only those clauses that contain at least one of the latter group are considered for the purposes of the analysis attempted here. Table 6 contains an overview of all such clauses and their counts by UD clause type.

UD clause type	Verbal nsubj	Verbal nsubjpass	Verbal dobj	Verbal nmod:obj
main	699	73	408	74
acl	146	35	246	61
advcl	201	31	214	21
xcomp	40	6	371	75
ccomp	278	44	136	28
parataxis	67	1	23	6
conj	143	25	215	36
csubj	8	0	4	0
Total	1582	215	1617	301

UD clause type	Copular nsubj	Existential nsubj
main	231	54
acl	23	13
advcl	47	25
xcomp	7	10
ccomp	99	44
parataxis	10	3
conj	45	20
csubj	1	0
Total	463	169

Total nsubj	2205
Total nsubjpass	215
Total dobj/nmod:obj	1918

Tab. 6: MUDTv1: Clauses containing core dependents by root (columns) and UD clause type (rows)

7.3 Constituent order in MUDTv1 by the numbers

This section examines the constituent order in MUDTv1. Before doing so from the point of view of the SV/VS and VO/OV dichotomies, an overview of constituent order according to the Greenbergian six-way classification is provided: there are only 472 verbal clauses with both a subject and the distribution of configurations is laid out in Table 7.

Configuration	Count	%
SVO	443	93.86%
SOV	0	0.00%
VSO	3	0.63%
VOS	11	2.33%
OSV	4	0.85%
OVS	11	2.33%
Total	472	100%

Tab. 7: MUDTv1: Constituent order – Greenbergian classification

Figures 2 and 3 below then provide a detailed breakdown by clause type (by root) and UD clause type.

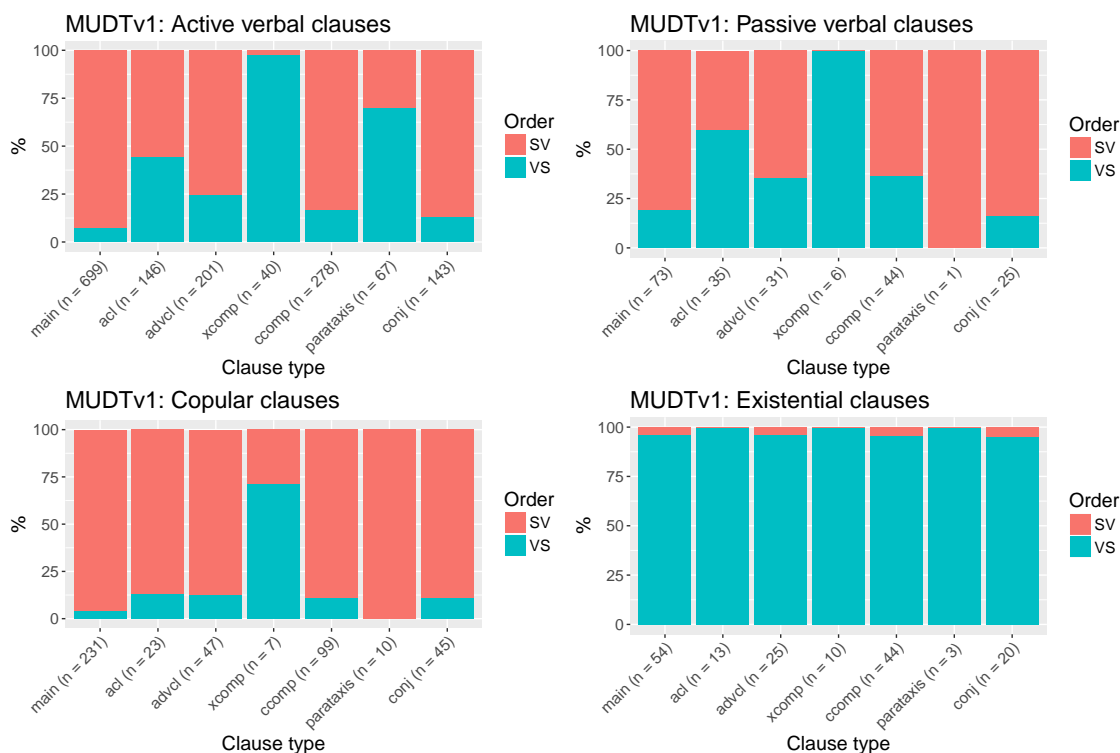


Fig. 2: MUDTv1: Order of predicate and subject by root and by UD clause type

The data for the order of subject and predicate requires further elaboration in two cases: first, for *xcomp* clauses, the numbers are actually misleading, as the queries used to obtain them actually represent two different and not at all complementary types of information: for SV, looking for a *nsubj* preceding an *xcomp* will match only those verbal *xcomp* that are not a part of a verbal chain. For VS, on the other hand, looking for a *nsubj* following a *xcomp* only collects verbs that are the last in a verbal chain and have a *nsubj* after them. The

numbers for VS *xcomp* clauses above therefore represent entire verbal chains and their actual syntactic role (UD relation) is that of the first verb in the chain. With the figures adjusted this way, one actually finds that there is a solitary active *xcomp* clause exhibiting the SV order and there is no passive *xcomp* clause proper.

Secondly, no dominant order could be established for active and passive *acl* clauses and passive *advcl* and *ccomp* clauses. While the data set is too small to provide a definitive answer, an analysis of active *acl* clauses conducted by means of a linear mixed effect model suggests that subject heaviness is positively associated with the VS order, whereas clause length is positively associated with SV order.

As for the order of object and predicate, Figure 3 provides a relatively straightforward picture.

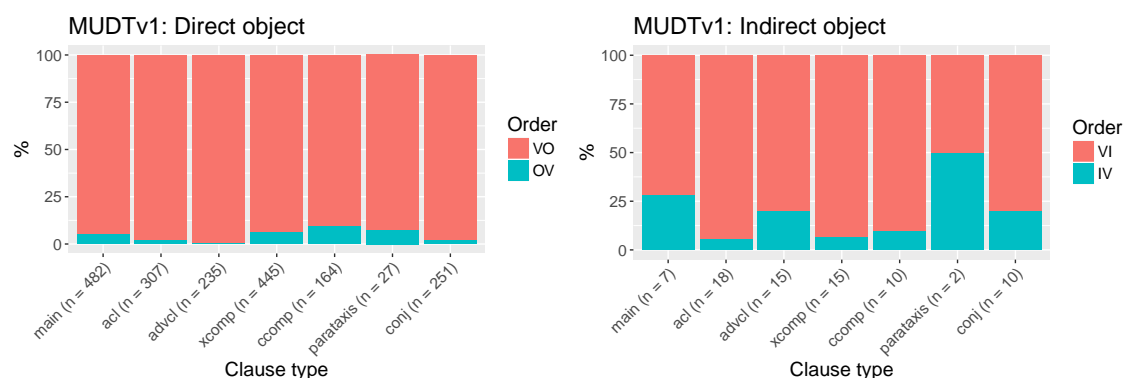


Fig. 3: MUDeTv1: Order of predicate and direct object by UD clause type

For both types of objects, a quick analysis of the role of information structure was performed, largely to check previous observations on the role of topicalization in OV order (Borg and Azzopardi-Alexander 1997: 126). The findings indicate that this is indeed the primary function of OV order, however, it is far from being widespread: in total, only 90 of direct objects in MUDeTv1 occur before their predicate and only 48 of them are instances of topicalization.

The data above shows that Maltese as represented in MUDeTv1 is a SV/VO language with the deviations representing 23.7% of the clauses examined for VS and 4.7% for OV. In the final step in this analysis, I provide a breakdown of the deviant orders by text type (Table 8).

Text type	% of VS	% of OV
newspaper	25.5 %	2.8%
quasi-spoken	29.7 %	4.2%
fiction	23.9 %	4.8%
non-fiction	15.5 %	4.9%

Tab. 8: MUDeTv1: Ratio of VS and OV across text types

This underscores the differences between text types highlighted in section 7.2.1 where again the text type newspaper stands out, this time as the text type with the lowest rate of OV (2.8%). Surprisingly, it is also the non-fiction text type that is the odd one out here, with the share of VS much lower than the average across MUDeTv1.

The former defies an easy explanation, but is an important fact in and of itself. As for the latter, comparison across UD clause types would suggest that this is largely due to the combined effect of *parataxis* and *advcl* clauses (see the highlighted parts of the graph): non-fiction is the only text type where SV *parataxis* clauses predominate (albeit only 14 to 10 in absolute numbers) which is hardly surprising considering that this is the

only text-type that does not prominently feature reported speech parataxis clauses. This combines with *advcl* where in non-fiction, the ratio of VS *advcl* is much lower (15.5%) than in other text types (28%-39%).

7.4 A brief comparison

On a number of occasions, Maltese has 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: 793 and Fabri and Borg 2017: 83 describe Maltese as “a topic-oriented language”, cf. Kiss 1995b: 4-5). The framework-dependent reasoning behind this classification is of not if interest here. What is, however, is the classification itself, i.e. the claim that Maltese is a discourse-configurational language; a claim can be tested quantitatively using both MUDTv1 and the Hungarian UD v2.1 treebank (Nivre, Agić, et al. 2017). 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.

For the purposes of this comparison, I imported the Hungarian UD v2 treebank (henceforth: HUUDv2)³ into the same instance of ANNIS3 where MUDTv1 resides.⁴ Using the ANNIS3 interface, I ran the queries I used for the quantitative analysis of MUDTv1 in section 7.3 above (taking into account the changes from UD v1 to UD v2 and the Maltese specific UD relations and excluding the UD *xcomp* clause) on HUUDv2. I then plotted the two sets of numbers against each other (Figure 4).

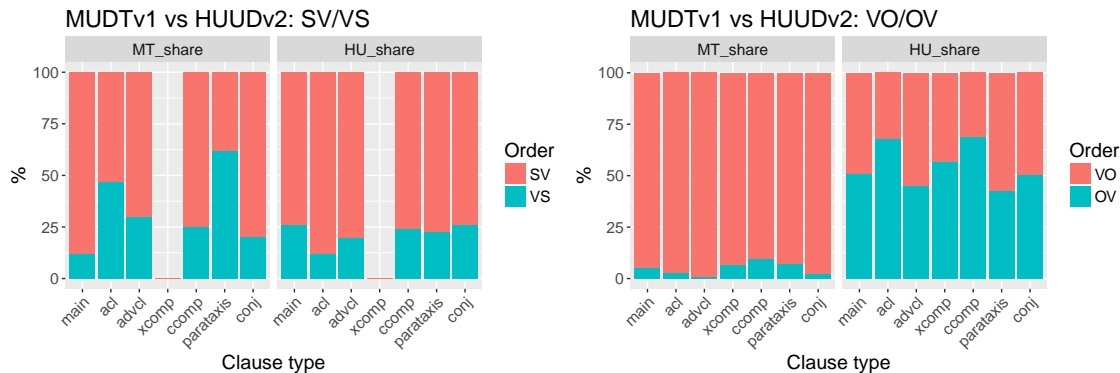


Fig. 4: MUDTv1 vs HUUDv2: Order of predicate and direct object by UD clause type

This data shows that while both Maltese (as represented in MUDTv1) and Hungarian (as represented in HUUDv2) can be safely classified as SV in Dryer’s binary typology, the two languages do not behave similarly when considered more closely: the distribution of the VS configuration is much smoother in Hungarian whereas in Maltese, *acl* clauses allow much more freedom in the way it orders the subject and the predicate. More important, however, is the stark difference between the two treebanks when it comes to the order of object and predicate, a difference that is too large to be explained away by sampling issues or by treebank

³ In terms of size, MUDTv1 and HUUDv2 are very similar: 2047 sentences in MUDTv1, 1800 in HUUDv2; 44,162 tokens in MUDTv1, 42,032 in HUUDv2.

⁴ bulbul.sk/annis-gui-3.4.4

composition, i.e. the fact that HUUDv2 consists only of journalistic texts. If one were to compare only those text types, the difference would be even more pronounced, as in the journalistic component of MUDTv1, the ratio of OV is even lower (Table 8).

This leads me to conclude that Maltese (at least as represented in MUDTv1) really is fundamentally different from Hungarian (as represented in HUUDv2) 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".

7.5 Summary: Answers to Research Questions

7.5.1 Answer to Research Question 1: What is the dominant constituent order in Maltese?

The dominant constituent order in Maltese is SV (except for the variation described in the next section) and VO. The dominant constituent order in Maltese can also be described as SVO.

7.5.2 Answer to Research Question 2: What is the variation in dominant constituent order in Maltese?

In Maltese, the dominant order of subject and predicate is VS rather than SV in all existential clauses and in verbal parataxis clauses; no dominant order could be established for verbal acl clauses (both active and passive), passive advcl clauses, passive ccomp clauses and copular xcomp clauses. Table 9 contains a summary with the variation highlighted in bold.

UD clause type	Verbal (active)	Verbal (passive)	Copular	Existential
main	SV	SV	SV	VS
acl	no dominant order	no dominant order	SV	VS
advcl	SV	no dominant order	SV	VS
xcomp	SV	n/a	no dominant order	VS
ccomp	SV	no dominant order	SV	VS
parataxis	VS	SV	SV	VS
conj	SV	SV	SV	VS

Tab. 9: MUDTv1: Variation in dominant constituent order

In contrast, the dominant order of predicate and object is VO in all clause types.

7.5.3 Answer to Research Question 3: What are the deviations from the dominant constituent order in Maltese?

The deviations from the dominant constituent order are recorded in plots and tables in section 7.3.

Of particular interest is the share of OV clauses which fall into a number of distinct groups, one of which involves what previous descriptions of constituent order of Maltese refer to as topicalization of objects (Borg and Azzopardi-Alexander 1997, Borg and Azzopardi-Alexander 2009). These constructions have been described as "a wide spread characteristic of Maltese" (Borg and Azzopardi-Alexander 1997: 126) and yet in MUDTv1, only 48 clauses, i.e. 2.5% of all direct objects, fall into that group (interestingly, the ratio for topicalized indirect objects is nearly identical at 2.6%). One might once again justifiably question the balance and representative-

ness of MUDTv1 or invoke the special nature of spoken language in explaining the discrepancy between the MUDTv1 data and the descriptions such as the one above, but until more data is available, the conclusion to be drawn here is that in Maltese, topicalization of objects is not as wide-spread as generally thought.

7.5.4 Answer to Research Question 4: What are the factors that cause variation in dominant constituent order?

Existential clauses are the only type of clause that consistently exhibits dominant order opposite to that of the other clauses and Maltese as a whole. The dominant predicate-first order reflects what has been observed regarding existential clauses cross-linguistically (cf. Givón 2001: 257, McNally 2011: 1833). Existing hypotheses explain this in terms of information structure or in terms of diachronic development (Givón 2001: 259); whether that holds for Maltese remains to be seen.

For those clause types where the dominant order could not be determined (Table 9), preliminary research indicates that heaviness of the subject and length of the clause play a role with the former being associated with VS and the latter with SV. Whether this is in fact true and if, what other factors there are and how they interact with heaviness and length, is a question that remains to be answered.

8 Summary

It should go without saying that the conclusions presented in this work are far from the last word on the subject. In fact, the exact opposite is true: the data and the insights drawn from them offered here should be viewed as nothing but a first step towards the detailed description of constituent order in Maltese and its variation (and deviation). As evident from the many instances of "beyond the scope of this work" and its synonyms dispersed throughout the dissertation, many of the issues involved are complex and requiring extensive treatment, preferably on more and more varied data. Chief among them is the issue of actual spoken Maltese and to what extent it is (if at all) different from the written language represented in MUDTv1. Answering this question would, naturally, involve building and annotating a treebank of spoken Maltese and while some steps towards that goal have been taken, such a treebank is still in our future; so is the expansion of MUDTv1 with more and more diverse texts which would make a follow-up confirmation study possible.

Let us therefore focus on the many shortcomings of this work and how they can be addressed using the data already available. Firstly, the analysis here largely remains silent on constituent order in sentence types classified by modality (imperative, exhortative, interrogative etc.). Secondly, while Chapter 7 also attempts to account for the OV deviation in terms of information structure, this analysis should be first put on a more solid theoretical footing by providing clear and actionable definitions of the information structure concepts in question and then expanded to the VS (or SV, in case of existential clauses) deviation. And of course with further effort at annotating semantic properties of core dependents (e.g. animacy), analyses like the one conducted for *acl* clauses can be extended to all types of clauses to determine the full set of factors influencing constituent order variation and deviation.

Going beyond the analysis of constituent order, the very next project MUDTv1 should be used for is the analysis of word order and its relationship to constituent order, clause structure and complex sentence structure, including, but not limited to, a full classification of branchedness in Maltese. The question of the order of elements within a noun phrase is a particularly fascinating one, as it touches upon the mixed nature of Maltese morphology and syntax. Such an analysis can then be immediately extended to that of valency of nouns and adjectives while reviewing and expanding the work begun here on verbal valency and providing a more generally grounded description of non-canonical objects, including their passive diathesis. This, naturally, ties to the further development of the treebank where a number of areas not involving constituent or word order need to be revisited; these include paratactic clauses and their further subdivision, comparative constructions, numerals, compounds (especially Light Verb Constructions), *x_{comp}* clauses featuring pseudoverbs, the concept

of “auxiliary verb” in Maltese, as well as further refinement and classification of the generic *nmod* relation used for nominal dependents of noun phrases.

And finally, this thesis has pointed out a number of issues of general descriptive import. Chapter 6 highlighted several lacunae in the description of Maltese, such as the copular and existential clauses, which both (but especially the latter) lack a comprehensive synchronic and diachronic description, as well as an analysis of their areal aspects (Arabic varieties for type (ii) and especially type (iii) copular clauses; both Arabic and Romance neighbors and ancestors of Maltese for existentials). The same is of course true of non-copular verbless clauses and the synchrony and diachrony of the expletive, as well as of non-expletive subjectless clauses. Chapters 6 and 7 also demonstrated that verbal chains and verbal complementation in Maltese still lack a satisfactory description; the same applies to various dislocation phenomena, coordination and ellipsis.

The data provided and described in this work can be used to accomplish these goals and even more.

References

- Andor, Daniel et al. (2016). “Globally Normalized Transition-Based Neural Networks”. In: *CoRR* abs/1603.06042.
- Aquilina, Joseph (1959). *The Structure of Maltese: A Study in Mixed Grammar and Vocabulary*. Malta: The Royal University of Malta.
- (1976). *Maltese Linguistic Surveys*. Malta: The University of Malta.
- Arnold, Jennifer E. et al. (2000). “Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering”. In: *Language* 76.1, pp. 28–55.
- Borg, Albert and Marie Azzopardi-Alexander (1997). *Maltese*. London: Routledge.
- (2009). “Topicalisation in Maltese”. In: *Introducing Maltese Linguistics*. Ed. by Bernard Comrie et al. Amsterdam: John Benjamins, pp. 71–81.
- Borg, Albert and Ray Fabri (2016). “Semantic functions of complementizers in Maltese”. In: *Complementizer Semantics in European Languages*. Ed. by Kasper Boye and Petar Kehayov. Berlin/Boston: De Gruyter Mouton, pp. 413–447.
- de Marneffe, Marie-Catherine et al. (2014). “Universal Stanford Dependencies: a Cross-Linguistic Typology”. In: *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*. Ed. by Nicoletta Calzolari (Conference Chair) et al. Reykjavik, Iceland: European Language Resources Association (ELRA), pp. 26–31.
- Dixon, R. M. W. (2009). *Basic Linguistic Theory Volume 1: Methodology*. Oxford: Oxford University Press.
- Dryer, Matthew S. (1992). “The Greenbergian Word Order Correlations”. In: *Language* 68.1 (1), pp. 81–138.
- (1997). “On the 6-way Word Order Typology”. In: *Studies in Language*, pp. 69–103.
- (2013a). “Determining Dominant Word Order”. In: *The World Atlas of Language Structures Online*. Ed. by Matthew S. Dryer and Martin Haspelmath. Leipzig: Max Planck Institute for Evolutionary Anthropology, n/a.
- (2013b). “Order of Subject, Object and Verb”. In: *The World Atlas of Language Structures Online*. Ed. by Matthew S. Dryer and Martin Haspelmath. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Fabri, Ray (1993). *Kongruenz und die Grammatik des Maltesischen*. Tübingen: Max Niemeyer Verlag.
- (2010). “Maltese”. eng. In: *Revue belge de philologie et d'histoire* 88.3, pp. 791–816.
- Fabri, Ray and Albert Borg (2002). “Topic, Focus and Word Order in Maltese”. In: *Aspects of the Dialects of Arabic Today*. Ed. by A. Youssi and Alii. AIDA. Rabat: AMAPATRIL, pp. 354–363.
- (2017). “Modifiers and complements within the Maltese verb sequence”. In: *Advances in Maltese Linguistics*. Ed. by Benjamin Saade and Mauro Tosco. Berlin: De Gruyter, pp. 67–86.
- Fenech, Edward (1978). *Contemporary journalistic Maltese*. Leiden: E. J. Brill.
- Gatt, Albert and Slavomír Čéplö (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: *Universals of language*. Ed. by Joseph H. Greenberg. Cambridge, Mass: MIT Press, pp. 73–113.
- Haspelmath, Martin (2009). “Framework-Free Grammatical Theory”. In: *The Oxford Handbook of Linguistic Analysis*. Ed. by Bernd Heine and Heiko Narrog. Oxford: Oxford University Press, pp. 341–365.
- Itkonen, Esa (2005). “Concerning the Synthesis between Intuition-based Study of Norms and Observation-based Study of Corpora”. In: *SKY Journal of Linguistics* 18, pp. 357–377.
- Kalmár, Ivan and Dionisius Agius (1983). “Verb-Subject Order and Communicative Dynamism in Maltese”. In: *Zeitschrift für Dialektologie und Linguistik* 50.3, pp. 335–354.
- Kiss, Katalin É., ed. (1995a). *Discourse Configurational Languages*. Oxford: Oxford University Press.

- Kiss, Katalin É. (1995b). "Discourse Configurational Languages: Introduction". In: *Discourse Configurational Languages*. Ed. by Katalin É. Kiss. Oxford: Oxford University Press, pp. 3–27.
- Köhler, Reinhard (2012). *Quantitative Syntax Analysis*. Berlin/Boston: De Gruyter Mouton.
- Krause, Thomas and Amir Zeldes (2016). "ANNIS3: A new architecture for generic corpus query and visualization". In: *Digital Scholarship in the Humanities* 31.1, pp. 118–139.
- Krier, Fernande (1976). *Le maltais au contact de l'italien. Etude phonologique, grammaticale et sémantique*. Hamburg: Helmut Buske Verlag.
- Lewis, M. Paul, Gary F. Simons, and Charles D. Fennig (2016). *Ethnologue: Languages of the World, Nineteenth edition*. Dallas, Texas: SIL International.
- Lopatková, Markéta et al. (2017). *Valenční slovní českých sloves VALLEX*. Praha: Karolinum.
- Maas, Utz (2009). "Complex predicates in Maltese: From a Neo-Arabic perspective". In: *Introducing Maltese Linguistics*. Ed. by Bernard Comrie et al. Amsterdam: John Benjamins, pp. 113–132.
- McNally, Louise (2011). "Existential sentences". In: *Semantics: an international handbook of natural language meaning*. Ed. by Klaus von Heusinger, Claudia Maienborn, and Paul Portner. Vol. 2. Handbooks of linguistics and communication science; 33.2. Berlin/Boston: De Gruyter Mouton, pp. 1829–1848.
- Nivre, Joakim, Željko Agić, et al. (2017). *Universal Dependencies 2.1*. LINDAT/CLARIN digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University.
- Nivre, Joakim, Filip Ginter, et al. (2014). *Online documentation for Universal Dependencies, version 1 (2014-10-01)*.
— (2016). *Online documentation for Universal Dependencies, version 2 (2016-12-01)*.
- Nivre, Joakim, Marie-Catherine de Marneffe, et al. (2016). "Universal Dependencies v1: A Multilingual Treebank Collection". In: *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. Ed. by Nicoletta Calzolari (Conference Chair) et al. Portorož, Slovenia: European Language Resources Association (ELRA).
- Song, Jae Jung (2011). "Word Order Typology". In: *The Oxford Handbook of Linguistic Typology*. Ed. by Jae Jung Song. Oxford: Oxford University Press, pp. 253–279.
- Stolz, Thomas (2009). "Splitting the verb chain in modern literary Maltese". In: *Introducing Maltese Linguistics*. Ed. by Bernard Comrie et al. Amsterdam: John Benjamins, pp. 133–179.
- (2011). "The possessive relative clause in Maltese". In: *Variation and Change. The Dynamics of Maltese in Space, Time and Society*. Ed. by Sandro Caruana, Ray Fabri, and Thomas Stolz. *Studia Typologica* 9. Berlin: Akademie Verlag, pp. 183–232.
- Sutcliffe, Edmund (1936). *A Grammar of the Maltese Language. With Chrestomathy and Vocabulary*. Oxford: Oxford University Press.
- Tesnière, Lucien (1959). *Éléments de syntaxe structurale*. Paris: Librairie C. Klincksieck.
- (2015). *Elements of structural syntax*. Amsterdam: John Benjamins.
- Vanhove, Martine (1993). *La langue maltaise. Etudes syntaxiques d'un dialecte arabe »périphérique«*. Wiesbaden: Harrassowitz Verlag.
- 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).

Appendix: Candidate's publication list

- Burke, Tony and Slavomír Čéplö (2016a). "Legend of the Thirty Pieces of Silver". In: *New Testament Apocrypha: More Noncanonical Scriptures. Volume 1*. Ed. by Tony Burke and Brent Landau. Grand Rapids, MI: Eerdmans, pp. 293–308.
- (2016b). "The Syriac Tradition of the Legend of the Thirty Pieces of Silver". In: *Hugoye* 19.1, pp. 35–121.
- Čéplö, Slavomír (2012). "Testament of Solomon and Other Pseudepigraphical Material in 'Aḥkām Sulaymān". In: *The Canon of the Bible and the Apocrypha in the Churches of the East*. New York: Peter Lang, pp. 21–38.
- (2013). "Secular Šarḥ? The Curious Case of the Marquis' Daughter". A paper read at the "Variation within and across Jewish Languages" conference, June 26th – June 28th, University of Antwerp.
- (2014). "An overview of object reduplication in Maltese". In: *Perspectives on Maltese Linguistics*. Ed. by Albert Borg, Sandro Caruana, and Alexandra Vella. Berlin: De Gruyter, pp. 201–222.
- (2017a). "Focus particles in Maltese: A corpus survey". In: *Advances in Maltese Linguistics*. Ed. by Benjamin Saade and Mauro Tosco. Berlin: De Gruyter, pp. 87–120.
- (2017b). "On Herod and John the Baptist: An edition and translation of a previously unknown New Testament apocryphon". In: *Arabic and Islamic Studies In Honour of Ján Pauliny*. Ed. by Zuzana Gažáková and Jaroslav Drobný. Bratislava: FiF UK.
- Čéplö, Slavomír, Ján Bátora, et al. (2014). "Standardizing Functional Testing of Mutual Intelligibility in Arabic Dialectology: Methods and Tools". A paper read at the "Methods in Dialectology XV" conference, August 11th – August 15th, University of Groningen.
- (2016). "Mutual Intelligibility of Spoken Maltese, Libyan and Tunisian Arabic Functionally Tested: A Pilot Study". In: *Folia Linguistica* 50.2, pp. 583–628.
- Čéplö, Slavomír and Chris Lucas (2017). "Lanqas, negative concord and predicate negation in Maltese". A paper read at the Sixth International Conference on Maltese Linguistics, June 8th – June 9th, Comenius University in Bratislava.
- Čéplö, Slavomír and Lonneke van der Plas (2017). "Light verb constructions in Maltese: Evidence from corpora". A paper read at the Sixth International Conference on Maltese Linguistics, June 8th – June 9th, Comenius University in Bratislava.
- Tišliar, Pavol and Slavomír Čéplö, eds. (2014). *Studies in the Population of Slovakia II*. Kraków: Spolok Slovákov v Poľsku.