The interaction of word order entropy and verb agreement: A token-based approach*

Annika Spirgath
ACLC, University of Amsterdam

This paper presents the first token-based approach to investigate the interaction between word order entropy and verbal agreement. By analyzing 18 languages from the Multi-CAST corpus, the study reveals a unidirectional implication between word order entropy and agreement rate: Languages with high agreement rates tend to have lower word order entropy, meaning they exhibit more rigid word order. The observed implication supports the assumption that both flexible word order and agreement are used to manage information structure in discourse. However, the lack of statistical significance, likely due to the limited sample size, highlights the need for further research with more languages. Overall, the study contributes to our understanding of how languages balance different argument-marking strategies.

1 Theoretical background

Nearly all languages employ some strategies for marking arguments. According to Lestrade (2015), there are three basic argument-marking options: i) Head Marking, specifically referring to verbal agreement[1] in this paper, ii) Dependent Marking, also known as case marking/flagging[2], and iii) word order variation. The interaction between these argument-marking strategies, particularly those involving word order, has interested many typologists over the past two decades. This interest arises from the observation that these strategies interact systematically across languages, due to the overlapping functions they serve: rigid word order and case marking both serve the function of argument disambiguation (Levshina, 2019), while flexible word order and agreement are

* I am grateful to Pegah Faghiri for her dedicated supervision and guidance throughout this thesis. Moreover, I want to thank Eva van Lier, Natalia Levshina and Jelke Bloem for their valuable comments and feedback on earlier versions of this paper.

[1] The notion of agreement used in this paper embraces only syntactic agreement in the sense outlined by Siewierska (2004) and Hengeveld (2012). For further explanation, see section 2.5.

[2] Case marking and flagging are used interchangeably in this paper.

Linguistics in Amsterdam, 2025: 30-58
© Annika Spirgath

primarily used to manage information structure in discourse (Lestrade 2015; Simpson 2012). Thus, the interaction is driven by different types of functional overlap (disambiguation or information structure management), depending on which word order type (rigid or flexible) co-occurs with which morphological marking type (case marking and/or agreement).

In the following sections of the Theoretical Background, I will focus on previous findings regarding the interaction of argument-marking strategies, distinguishing between type-based approaches (section 1.1) and token-based approaches (section 1.2) to word order. Type-based approaches treat languages as single data points with fixed word order values (e.g., SVO or VOS). In contrast, a more recent token-based approach (section 1.2) is gaining popularity because it accounts for the variability of word order patterns within a language. For example, Levshina (2019) used this gradient measure in one of her case studies to explore the correlation between word order variability and case marking, assuming they share the function of argument disambiguation. The current paper builds on Levshina's study, utilizing a token-based approach to measure word order variability. However, it goes further by investigating the previously unexplored interaction between word order variability and agreement. As noted earlier, the underlying rationale is the assumption that both flexible word order and agreement-marking serve to manage information structure (section 1.3). This overlapping function leads to the prediction of a unidirectional implication between the two strategies (section 1.4). Ultimately, to conduct such a typological investigation using a token-based approach, a multilingual and annotated corpus is essential (section 1.5).

1.1 Type-based approaches

Most of the research regarding the interaction of different argument-marking strategies has used a type-based approach to word order. The word order types most commonly investigated involve the various positions of the verb in relation to its subject and object. The following three positions of the verb (V) are possible: i) the verb preceding its arguments (V-initial), ii) the verb positioned between its arguments (V-medial), and iii) the verb following its arguments (V-final). Hawkins (2002) hypothesizes from a sentence processing perspective that if the verb is in the last position and shows agreement with one or both arguments, the addressee must wait until the verb is processed to disambiguate the arguments based on agreement. Consequently, it takes more time to distinguish the subject from the object, making case marking desirable to differentiate the two arguments early on. Conversely, if the verb appears in the first position (V-initial), verbal agreement is preferable as it tells the listener what to expect. Hawkins (2002, p.100) illustrates this with the following examples:

(1) Rich Verb Agreement preferred in Verb-early Languages
I.e. V-initial > agr SVO > agr SOV (> agr = 'has more agreement than')

(2) Rich Case Marking preferred in NP-early Languages
I.e. SOV > cm SVO > cm V-initial (> cm = 'has more case marking than')

Dryer (2002) takes a slightly different approach and hypothesizes that if both the subject and the object appear on the same side of the verb (V-initial or V-final), it is not possible to tell their function apart if one of them is dropped. Dropping either the subject or the object often happens in natural language and therefore, marking the arguments with either flagging or indexing would be useful. Dryer tests this hypothesis using his own typological database, which includes a sample of 502 languages and is later incorporated into the online WALS (Dryer & Haspelmath, 2013). The main findings of his study are that V-medial order generally obviates the need for argument marking, while both V-initial and V-final word orders often combine with case marking and agreement. These findings are in line with his hypothesis.

Using a similar approach, Siewierska & Bakker (2008) looked at the interaction between word order, case marking and agreement in a sample of 417 languages. They similarly find that the V-medial order reduces the necessity for case marking, whereas the V-final word order is more likely to utilize case marking. Additionally, they find that agreement is predominantly associated with V-initial orders, but is also common in V-final orders, often in combination with case marking. A possible explanation for this is likely due to the different primary functions of case marking and agreement, allowing them to co-occur. While case marking is primarily used for argument disambiguation, agreement is more closely related to managing information structure, which will be further elaborated in section 1.3.

1.2 Token-based approach with UD

The problem of the type-based approach is that the variability of the word order types within a language is not captured. Therefore, Levshina et al. (2023) advocate for a token-based approach with a gradient measure of word order variability. However, token-based approaches are still rare. One reason for this is that they require large multilingual and annotated corpora, as well as software to process and analyze the data statistically.

Levshina (2019) is the first study to look at the correlation between word order variability and case marking using the Universal Dependency (UD) corpora. UD is a standardized framework with guidelines for annotating written text corpora and as of 2019 included annotated corpora for 60 languages. Its main annotations are head and dependent elements, part-of-speech tags, morphological features (e.g., number, gender, case) and syntactic features (e.g., mood, aspect, voice). Levshina (2019) computes entropy scores in order to measure the variability of different word order patterns. Entropy is a measure of variability and has scores ranging from 0 to 1.[3] The proportion of identical S and O forms measures the amount of case marking in a language: The higher the confusability between the subject and object, the less case marking. To investigate the interaction between word order variability and case marking, Levshina maps all 60 languages to their Subject-Object (S-O) word order entropy on one axis and their proportion of identical S and O forms per lemma on the other axis. The results show a unidirectional implicational universal between the two argument-marking strategies: High confusability (little case marking) implies low entropy (word order fixation), but low confusability does not necessarily lead to high word order entropy. Levshina explains this outcome by pointing to the communicative goal of minimizing ambiguity. When the subject and object are highly confusable, a fixed word order is used to resolve the ambiguity.

1.3 The role of information structure management

If each of the three argument-marking strategies discussed so far were exclusively being used for argument disambiguation, the different strategies would be expected to mutually exclude each other. However, as Dryer (2002) and Siewierska & Bakker (2008) show in their research, most languages combine several strategies. This suggests that other functions, besides argument disambiguation, must be involved.

According to Lestrade (2015), the "basic function of word order (…) is the sequencing of information to reflect communicative intentions and optimize processing". This "sequencing of information" is also known under the term 'Information Structure Management (ISM)' and is only possible if the word order is somewhat flexible. Levshina (2019) provides an example for this: In English you can say "*In the middle of the room stood* ***a table",*** but also "**The table** stood in the middle of the room". In the first sentence the nominal subject "*a table*" unexpectedly conveys new information and therefore shows a non-canonical (VS) order. In the second sentence, the subject "*the table*" refers to already known information and therefore stays in its typical position before the verb. This is an example of how the flexibility of word order can manage information structure in discourse. Interestingly, Levshina observes that the position of pronominal subjects relative to verbs varies less within a language compared to the position of nominal subjects. She hypothesizes that the decrease in intra-linguistic variability of pronominal subjects might be caused by their lacking ability to

[3] Section 2.4 explains word order entropy and how it is used in this paper in more detail.

introduce new information. To address this difference, pronominal and nominal word order patterns will be treated separately in this paper.

Similar to word order, verb agreement is known for organizing information by keeping track of salient discourse referents through their index of features (Siewierska & Bakker, 2008). In fact, according to Lestrade (2015), agreement "is hardly reliable for differentiating arguments", because arguments often share the relevant agreement features, such as animacy, person or gender. Therefore, in his opinion, the primary function of agreement marking is to "identify highly accessible referents in discourse". Furthermore, Simpson (2012) thoroughly discusses the relationship of agreement to Silverstein's influential hierarchy (1976), which he argues is essentially a hierarchy of information structure resources. Originally, the hierarchy was developed by Silverstein to capture case and agreement marking strategies. The hierarchy shows that the 1st/2nd person features rank higher than 3rd person, pronouns rank higher than nouns and animates rank higher than inanimates. The scale reflects the expected features that a subject (associated with features high on scale) and an object (associated with features low on scale) usually have. If the features of a subject or an object unexpectedly deviate from the scale, the argument is claimed to be marked by case marking. This aligns with Haspelmath's (2021) frequency-based efficiency explanation. However, already Bickel (2008) observed that the referential hierarchy might work differently for case marking than for agreement marking. Simpson (2012) explains the reason for this. He argues that the referential hierarchy is essentially a hierarchy of information structure tools with agreement being an element of it. For example, the typical function of pronouns and proper nouns is to mark given information and continuing topic. In contrast, the typical function of nouns is to mark new information and focus. Similar to pronouns, the index marker on the verb can express continuing topic. Therefore, Simpson concludes that agreement constitutes one of the components of information structure resources available to languages, alongside pronouns, nouns, proper names, clitics, demonstratives, word order, prosody, and non-verbal resources such as gestures and eye-gaze.

1.4 Prediction

Given that both word order and verb agreement serve to organize information structure in discourse, it is not economically efficient to use two resources for the same purpose. Thus, it is unlikely that flexible word order and extensive verb agreement would coexist. Therefore, we expect to see an implicational relationship between word order entropy and agreement rate: the higher the amount of agreement in a language, the lower its word order entropy, indicating a more rigid word order.

1.5 Corpus-based typology

In order to conduct a crosslinguistic typological study involving word order entropy, large multilingual corpora are needed. In contrast to grammar-based typology, corpus-based typology “focuses on variation within languages and seeks to determine structural probabilities, which are then compared across languages” (Schnell & Schiborr, 2022). So far, most corpus-based research on word order typology has used the UD framework (cf. Levshina, 2019), due to its large number of available languages (167 languages as of November 2024). However, the UD corpora consist primarily of written texts. This is problematic because written texts generally reflect the so called “LOL” languages (“Literate, Official, with Lots of users”, Dahl, 2015) rather than naturalistic, spoken language. Consequently, the UD corpora further exhibit a bias towards the Indo-European and Eurasian languages (Levshina, 2019). Therefore, throughout the past decade Geoffrey Haig & Stefan Schnell (2023) developed multi-CAST (Multilingual Corpus of Annotated Spoken Texts) as an initiative to “counteract the existing bias in corpus-based typology towards written forms of well-researched and generally standardized languages” (Haig et al., 2022). The multi-CAST corpus provides the data for the present study, which aims to analyze the interaction of word order variability and verb agreement.

2. Methodology

2.1 Multi-CAST corpus design

Multi-CAST uses data that was collected in collaboration with several linguists who obtained expertise in their individual languages during multiple language documentation and fieldwork projects. All data are freely accessible online under a Creative Commons licence (CC-BY 4.0). The current version from November 2023 contains data from 18 languages (see Table 1). Each language subcorpus contains at least 1000 clause units and consists of several monological texts, ranging from two to twenty-nine per language. This is still a relatively small amount of data, mainly due to the “extremely labor-intensive” (Haig et al, 2022) nature of processing and annotating naturalistic spoken data. The small sample of languages presents a bias towards Western Asia and the Pacific region, the geographic foci of the two editors Geoffrey Haig and Stefan Schnell. However, additional languages are constantly integrated into the corpus as they become available.

Table 1: Overview of the Multi-CAST corpora (version 2311)

Corpus	Genus	Citation	Texts	Clause Units

Arta	Malayo-Polynesian	Kimoto 2019	11	1030
Bora	Boran	Seifart & Hong 2022	2	1241
Cypriot Greek	Greek	Hadjidas & Vollmer 2015	3	1070
English	Germanic	Schiborr 2015	5	5649
Jinghpaw	Tibeto-Burman	Kurabe 2021	11	1276
Kalamang	West Bomberai	Visser 2021	6	1051
Mandarin	Sinitic	Vollmer, 2020	3	1194
Matukar Panau	Oceanic	Barth et al. 2023	8	1423
Nafsan	Oceanic	Thieberger & Brickel 2019	9	1012
Northern Kurdish	Iranian	Haig et al. 2019	3	1841
Persian	Iranian	Adibifar 2016	29	1418
Sanzhi Dargwa	Dargwic	Forker & Schiborr 2019	8	1066
Sumbawa	Malayo-Polynesian	Shiohara 2022	5	1084
Tabasaran	Lezgic	Bogomolova et al. 2021	5	1383
Teop	Oceanic	Mosel & Schnell 2015	4	1303
Tondano	Malayo-Polynesian	Brickell 2016	8	1085
Tulil	Taulil-Butam	Meng 2019	6	1264
Vera'a	Oceanic	Schnell 2015	10	3608
Totals			136	28,998

2.2 Multi-CAST annotation process

The multi-CAST corpora receive gold standard labels through a collaborative annotation process involving both the individual corpus contributors and the core multi-CAST team (Geoffrey Haig, Stefan Schnell, and Nils Schiborr). Instead of relying on a single annotator working independently, the annotations are created through a guided process in which the primary contributor manually annotates a small sample of data, followed by detailed feedback and revisions from the MC team. This process is repeated until consensus is reached. Throughout this collaboration, decisions are documented in annotation notes accompanying each corpus, co-authored by the contributor and the MC team. Additionally, annotations are accompanied by transcriptions, morphological glossing[4] and English translation, ensuring that they are accessible and interpretable to all users. Overall, the annotations follow the GRAID ('Grammatical Relations and Animacy in Discourse', Haig & Schnell, 2014) scheme. The annotation tier of the GRAID scheme reflects the syntactic functions of referential expressions and assigns the values of animacy and person. Furthermore, GRAID also distinguishes predicates, marks clause boundaries and captures clausal operators, such as negation or subordination. Another distinctive feature of the GRAID scheme is that its universal annotation tags can be used to detect inflectional marking directly, allowing for a more precise examination of word inflection. This sets it apart from previous approaches based on the UD annotation scheme, such as Levshina's 2019 study (cf. section 1.2), which rely only on surface-level differences in word forms to identify inflection. Ultimately, according to Haig et al. (2022), the overall function of GRAID is to provide the annotator with a set of tags which should primarily target referential expressions that introduce and track discourse referents. This discourse-oriented focus of the annotations makes the multi-CAST corpus a perfect dataset for the present study which aims to find out the roles of word order variability and verb agreement in relation to information management in discourse.

2.3 Word Order patterns

Sentences consist of multiple constituents and can therefore have many different word orders. Levshina (2019) for instance distinguishes 24 different word order patterns in her research on token-based typology and word order entropy. Due to the smaller scale of the present study and the focus on word order in relation to verb agreement, many of the word order patterns used by Levshina were excluded. Consequently, word order variability is only measured with regard to the core arguments of the verb. In the multi-CAST corpus, the core arguments are

[4] Morphological glosses follow the Leipzig Glossing Rules (Comrie et al. 2008).

annotated following Andrews's approach (2007), distinguishing intransitive subjects (S, annotated ⟨:s⟩), transitive subjects (A, ⟨:a⟩), and direct objects (P, ⟨:p⟩). As previously noted, Levshina (2019) identifies significant differences in the entropy values between nominal and pronominal objects and subjects. Therefore, nominal and pronominal word orders are addressed separately in this paper (see Table 2). Note that unexpressed arguments, annotated as zero arguments, as well as non-canonical subjects (labeled as 'ncs') and oblique objects (e.g., locatives labeled as 'l' or goals labeled as 'g'), are not considered in this analysis.

Table 2: Word order patterns examined in the present paper

Type	Word Order Pair	Label for Word Order Pattern
Nominals and their verb[5] (V)	nomS + V and V + nomS	nomS_V
	nomA + V and V + nomA	nomA_V
	nomP + V and V + nomP	nomP_V
Pronominals and their verb (V)	pronS + V and V + pronS	pronS_V
	pronA + V and V + pronA	pronA_V
	pronP + V and V + pronP	pronP_V

2.4 *Word order entropy*

In order to measure the variability of the different word order patterns, I adopt the Shannon entropy measurement, used by Levshina (2019). The measurement has scores ranging from 0-1, where 1 means that a word order pair has reached the highest score of variability. In other words, its frequency proportion lies at 50% (e.g., 50% V-S and 50% S-V). Therefore, the first step to measure word order variability is to extract the frequencies of all word order pairs in every text for each language with the help of an R script.[6] To ensure the accuracy and comparability of the results across the 18 languages, some contextual restrictions are applied. Specifically, arguments and their predicates are only counted when they appear in the main clause. Furthermore, questions and exclamatory sentences are disregarded. Then, after the extraction of the frequencies, the proportions of the word order pairs are computed for each language. Finally, the entropy of a word order pair is calculated with the proportion values and the Shannon entropy formula (Shannon, 1948):

(3) $H(X) = -\sum_{i=1}^{2} P(x_i) \log_2 P(x_i)$

[5] The term verb (V) is defined as all tokens that are annotated as verbal predicates in the GRAID tier of the multi-CAST corpus.

[6] You can access a repository with the R scripts here: https://github.com/anni22/BA-Thesis-Linguistics

In the formula, X represents the two possible word orders of a word order pair (e.g., S-V and V-S) and $P(x_i)$ the probability of one of the orders. This probability equals the proportion computed of the two frequencies of a specific word order pair in a language. The relationship between probability and entropy is non-linear. Therefore, already a small amount of word order variation results in relatively high values of entropy. For example, if the proportion of nomS + V in a language is 0.9 (90%) and the proportion V + nomS is 0.1 (10%), then the entropy of this word order pattern would be 0.47 according to the formula in (3).

2.5 Syntactic agreement

Siewierska (2004, p.126) and Hengeveld (2012) distinguish two types of agreement markers which fall under the term indexing. The first type are syntactic agreement markers which are "the result of a mere copying mechanism and do not contribute to the meaning of the sentence" (Hengeveld, 2012). For the second type of agreement markers, Siewierska uses the term anaphoric agreement markers as they lack a controller[7] in their clause with which they agree and are rather referential in nature by themselves. However, Hengeveld criticizes the use of this term because "in the absence of other elements in the clause there is no 'agreement' in the strict sense involved". Therefore, in the following, I will adhere to Hengeveld's definition of agreement which is restricted to purely syntactic agreement. Consequently, only agreement markers that agree with an overt noun phrase or free pronoun are looked at. This approach may lead to the same agreement marker being counted in one clause, but not in another when its argument is dropped. Such markers are known as ambiguous agreement markers (Siewierska, 2004, p.126), as they exhibit characteristics of both syntactic and referential agreement markers. Further elaboration on this topic is provided in the discussion section (4.3).

2.5.1 Measuring agreement rate

To get the amount of syntactic agreement for each language, the agreement rate is measured. The agreement rate is computed by the total number of agreement markers divided by the total number of possible agreement slots that the language has. To extract the total number of syntactic agreement markers for each language, an R script was developed. First, the algorithm of the script excludes all predicates[8] that do not contain any noun phrases or free pronouns in their

[7] Siewierska uses the term 'controller' which refers to either a noun phrase or a free pronoun with which the verb agrees.
[8] Haig & Schnell (2014) define the term 'predicate' as all predicative expressions including non-canonical verb forms, overt copular verbs and suffixal and clitic auxiliaries.

clause. Additionally, predicates are classified as either transitive or intransitive based on the annotation of their argument(s) (S or A and/or P). Transitive predicates receive two possible agreement slots (A and P arguments) and intransitive predicates only receive one possible agreement slot (S argument). If the algorithm then detects an agreement marker on the verb, it is only counted if the clause of the predicate also contains the argument with which it agrees. In this way, it is made sure that only syntactic agreement markers and no referential marking is counted. The agreement features being measured are based on the available agreement annotations of the GRAID and the Gloss tier. The GRAID tier only shows 1st and 2nd person agreement and the 3rd person in the form of animacy agreement (human, anthropomorphic and non-human). Therefore, only person[9] agreement is taken into account.

2.5.2 Algorithm procedure

The algorithm may detect one agreement marker for an intransitive predicate, and up to two agreement markers for transitive predicates. To do so, three steps on different tier levels and tokens are employed. The first step of the algorithm is to analyze in the GRAID tier all affixes and/or clitics[10] to the right and to the left of the predicate and to check if any of them mark person agreement. For instance, in example 4 from Matukar, the algorithm detects two agreement markers during this step.[11] However, using the GRAID tier is not enough to detect all agreement in multi-CAST. According to the GRAID manual, "the overarching principle behind annotating person markers in GRAID is that we annotate those positions which permit variation" (Haig & Schnell, 2014). This means that obligatory person markers are not annotated in GRAID as their presence can be inferred via a general grammatical rule. Therefore, in order to detect obligatory agreement markers, the algorithm needs to check the Gloss tier. Consequently, the second step involves examining the specific tokens in the Gloss tier that were detected to be annotated as separate affixes/clitics in the initial step, but were not recognized as agreement markers. It is during this stage that agreement markers like those found in Nafsan (example 5) are identified. The last step of the algorithm is to check the Gloss tier of the predicate token itself. At this stage, affixes whose morpheme is not annotated as a separate token in the Word tier, are recognized (e.g., example 6 from Northern Kurdish). Additionally, person agreement markers that are annotated on the root of the verb in the Gloss tier, but not

[9] Animacy agreement will be counted as the 3rd person agreement marker.
[10] No differentiation between affixal and clitic markers is being made due to their distinction being rather scalar than discrete.
[11] In the examples, tokens that carry agreement markers are marked in yellow and the predicates are marked in blue. Specific annotation symbols that the algorithm uses to detect the agreement marker, are marked in red.

delimited as an affix in the Word tier (e.g., example 7 from Persian[12]), are detected. If at any stage, the maximum number of agreement markers for a predicate has been counted, the remaining steps are skipped and the algorithm jumps to the next predicate in the language. Ultimately, all person agreement markers of the predicates (maximum 2 per predicate) are summed up to get the total number of agreement markers for each language.

(4) Hudungan main Yali ditekaman.

Word:	hudungan	main	Yali	di-	te-k-ama	-n
Gloss:	every	TOP	Yali	3PL.A-	see-NOM-APS	-3SG
GRAID:	#np.h:a	rn	np.h:p	-pro_h_a	v:pred	-pro_h_p

Translation: 'All these they saw Yali.'

[mc_matukar_yali_0159]

(5) …ntwam itnoli.

Word:	ntwam	i=	tnol-I	ZERO
Gloss:	devil	3S.RS=	swallow-TS	ZERO
GRAID:	##np.d:a	=lv	v:pred	0:p

Translation: '…the devil swallowed it.'

[mc_nafsan_kori_0028]

(6) Ewana fal davêjin, …

Word:	ewana	fal	davêj-in
Gloss:	3PL	fortune	IND.throw.PRS-3PL
GRAID:	## pro.h:a	np:p	v:pred

Translation: 'They tell fortunes (by throwing special dice), …'

[mc_nkurd_muserz01_0008]

(7) …in se tā pesarbače miāynd…

Word:	in	se	tā	pesarbače	miāyand
Gloss:	this	three	piece	little.boy	come.PRS.IND.3PL
GRAID:	#ln_dem	ln_qu	ln_class	np.h:s	v:pred

Translation: '…these three boys come…'

[mc_persian_g1-f-01_0012]

---

[12] Note that in this particular example, the segmentation of the annotation is incorrect because '*-and*' functions as a 3PL suffix and should therefore be annotated as a separate token (following a personal communication with Pegah Faghiri). Nonetheless, the algorithm identifies the agreement marker correctly.

3. Results and analysis

3.1 Interaction between word order entropy and syntactic agreement

As outlined in the methodology, variability in word order is assessed using the overall word order entropy, calculated as the average of the individual entropies of the six word order pairs. Across the 18 languages studied, all individual word order entropies exhibit positive correlations.[13] This indicates that the averaged entropy value for each language represents the variability of each of the word orders within that language well. The agreement rate measures the total number of agreement markers divided by the total possible agreement slots. Consequently, languages with higher agreement rates exhibit greater overall agreement. Figure 1 illustrates the resulting word order entropy and agreement rate values for each of the 18 languages[14].

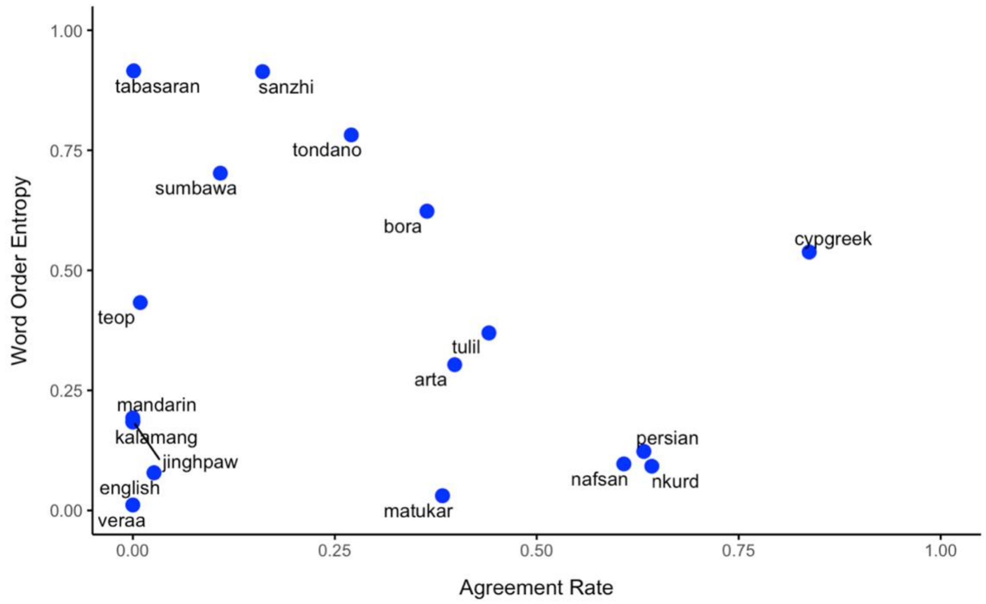

Figure 1:

Vertical axis: Overall word order entropy for each language in the multi-CAST corpus
Horizontal axis: Agreement rate measured as the proportion of actual syntactic agreement markers present compared to the possible syntactic agreement slots available for each language

[13] See Appendix B for the correlogram of the word order patterns.
[14] See Appendix A for the exact overall and individual entropy values. Appendix C shows the exact agreement values, amount of possible agreement slots and individual frequencies for each language.

In Figure 1, languages with low word order entropy are distributed across most of the x axis. This means that among these languages, some have a low agreement rate (e.g., English and Vera'a), while others have a moderate or high agreement rate (e.g., Nkurd, Persian). In contrast, languages with high word order entropy exhibit only low to moderate agreement values (e.g., Tabasaran, Sanzhi). Notably, no languages have both high word order entropy and a high agreement rate, as indicated by the empty space in the top right corner of the figure. Though, it is important to consider that the agreement rate only reflects person agreement, excluding other types of agreement markers which could result in higher overall agreement rates in some languages (see Discussion 4.3). Nevertheless, the results align with my prediction that languages with higher agreement rates tend to have lower word order entropy.

Cypriot Greek scores the highest agreement rate with a value of 0.8373. In addition, it has a moderately high word order entropy of 0.5383. This gives the language an outlier position. An outlier analysis using the Local Outlier Factor (LOF) method with a threshold value of 1.5 validates the visual finding.[15] A possible reason for Cypriot Greek to take this outlier position might be due to its strict subject agreement being of a purely grammatical nature. A further investigation is done in the discussion section. However, to ensure that the analysis accurately reflects the general trends and patterns present in the other languages, Cypriot Greek is excluded from the subsequent analysis.

3.2 Linear mixed model

A linear mixed-effects model[16] with genera as random intercepts reveals a regression line with a negative slope between verbal agreement and word order entropy in Figure 2.

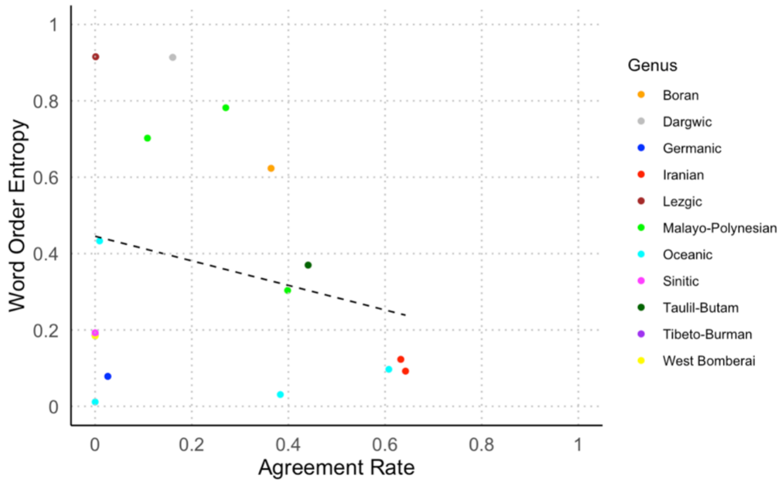

Figure 2: Linear mixed model with genera as random effects and negative regression line

[15] See Appendix D for the R code and outcome of the outlier analysis.
[16] The complete output of the mixed-effects model can be found in Appendix E.

The negative coefficient for syntactic agreement (estimate = –0.322, p = .272) suggests that as the amount of agreement increases, the word order entropy decreases. While this relationship is not statistically significant, its general direction is in line with my theoretical prediction. The high standard error (0.282) of the estimate further indicates considerable uncertainty, suggesting that the data does not strongly support a consistent negative relationship between agreement values and word order entropy.

Importantly, I used a mixed-effect model to account for the potential similarities among languages from the same genus, which may share historical or areal influences. Including genera as random intercepts allows the model to adjust for group-level variance without attributing it incorrectly to the fixed effect of agreement. A comparison with a fixed-effect model[17] (estimate = –0.2874, p = .373, AIC = 16.86) shows that the mixed-effect model has a moderately lower AIC (16.32, Δ0.54), suggesting a slightly better overall model fit, though the change in AIC of 0.54 does not indicate a significant improvement.

The Oceanic and Malayo-Polynesian languages illustrate the value of including random effects: while all Oceanic languages (blue data points) lie below the regression line, most of the Malayo-Polynesian languages (green data points) lie above it. However, when considering agreement values, they display a wide distribution. This shows that while the random effects of genera account for some variability in word order entropy, they do not show a clear, significant pattern when we look at the agreement rate.

4. Discussion

This study explored the relationship between word order entropy and syntactic agreement across languages using data from the multi-CAST corpus. The findings demonstrate a unidirectional implication, supporting the prediction that languages with high agreement rates tend to have low word order entropies. This suggests an inverse relationship between the two argument-marking strategies which supports their function of managing information structure. However, due to the lack of statistical significance, no definite conclusions can be drawn. A post hoc power analysis reveals that the linear mixed model used to analyze the data, has a power of 19%.[18] This implies a limited 19% probability of detecting a statistically significant effect between the variables which is well below the commonly desired power threshold of 80%. Therefore, the observed high p-value might be due to insufficient power rather than the absence of an effect. Thus,

[17] The complete output of the fixed-effects model can be found in Appendix F.
[18] See Appendix G for the output of the post hoc power analysis of the linear mixed model in Figure 2.

adding more languages to the multi-CAST corpus is crucial for improving the power of the model.

4.1 Word order results

Levshina (2019) finds that many languages that have high word order entropies and rich case marking are synthetic. This is expected, as these languages require numerous morphological markers to facilitate extensive case marking used for argument disambiguation. However, she also notes that some Asian languages (e.g., Japanese) maintain rigid word orders despite their synthetic nature. Therefore, Levshina (2019) hypothesizes that “(s)yntheticity, which is usually accompanied by abundance of grammatical markers, seems to be a necessary, but not sufficient condition for high entropy”. The word order entropy results of the present study (Appendix A) support this hypothesis. Highly synthetic languages such as Tabasaran and Sanzhi tend to exhibit higher entropy scores compared to analytic languages like Mandarin and English. However, exceptions exist, such as the Iranian languages Persian and Northern Kurdish which exhibit low word order entropies despite their synthetic characteristics. A likely explanation for these exceptions is their high agreement rates, leading to many morphological markers, which contribute to the languages' synthetic nature. Therefore, it appears that when the need to manage information structure is involved, languages display synthetic characteristics even with a low word order entropy.

4.2 Agreement results

When looking at the individual frequencies of the agreement results in Appendix C, it is evident that the languages with the highest agreement rates (Persian, Northern Kurdish, Cypriot Greek) are primarily determined by agreement markers counted on the predicate token itself in the Gloss tier (step 3 of algorithm procedure). The algorithm counts few affixal agreement markers from the GRAID tier for these languages. To investigate this phenomenon further, a closer look has been taken into the agreement systems of Persian and Cypriot Greek. Persian has very strict subject agreement. A strict grammatical rule requires the verb to inherently contain a person marker, agreeing with the subject. This results in the high number of non-affixal agreement markers counted in the Gloss tier. Similarly, Cypriot Greek also shows strict subject agreement. Interestingly, while in all other languages, affixes marked in the Gloss tier are also marked in the Word tier, the annotators of Cypriot Greek do not annotate the person affixes as separate morphemes in the Word tier. This indicates how strongly the person affix has fused with the verb in Cypriot Greek, explaining its high agreement rate. However, it is surprising that, in addition to its high agreement rate, the language also has a relatively high word order entropy. One explanation might be that agreement systems that are purely derived from a grammatical rule are not able to carry the function of information structure management. Therefore, word order and agreement in Cypriot Greek, as well as in Persian, may not have overlapping functions. This hypothesis would explain the outlier position of Cypriot Greek.

4.3 Limitations and future research

Besides the limitation of the current sample size, corpus-based typology generally presents additional limitations. For example, the corpus data make it difficult to distinguish between syntactic and ambiguous agreement markers. Ambiguous agreement markers, as defined by Siewierska (2004, p.126), may sometimes appear with a noun phrase and sometimes without. Therefore, they are somewhat in between syntactic and referential agreement. The algorithm of the present study counts all agreement markers that appear with a noun phrase. Thus, some ambiguous agreement markers are by default counted as syntactic agreement markers. This binary classification (syntactic vs non-syntactic agreement marker) fails to capture the nuances of ambiguous markers that are visible in grammar-based typology. Therefore, future research should aim to develop methods for recognizing and categorizing ambiguous agreement markers within corpora to provide a more accurate analysis of syntactic agreement.

Additionally, the current study ignores anaphoric agreement markers (defined as referential marking by Hengeveld (2012)). Therefore, future research could expand the scope to test word order entropy for a possible interaction with referential marking. To thoroughly investigate referential agreement, it is necessary to include additional agreement types beyond verbal agreement, because verbal agreement in predicates is 77% grammatical and only 23% anaphoric (Siewierska, 2004, p.127). According to Siewierska, possessed nouns and adpositions are typologically more prone to exhibit anaphoric agreement than verbal predicates. Thus, when investigating referential agreement marking core arguments, it would be important to examine not only verbal predicates but also possessed nouns and adpositions.

Moreover, the present study only considered person agreement, leaving out other types of agreement, such as number, gender, animacy or definiteness. This is due to current limitations within the annotations of the multi-CAST corpus. The GRAID guidelines, which primarily focus on person and some animacy agreement, need to be revised to include these other agreement types. Expanding the GRAID guidelines would allow for a more detailed and inclusive analysis of agreement phenomena, providing insights into how different types of agreement interact with word order entropy.

Another factor to consider is the relationship between agreement and case marking. Lestrade (2015), as well as Siewierska & Bakker (2008), find in their research that case marking and agreement often occur together. Therefore, Lestrade hypothesizes that there might be a correlation between case marking and agreement, which still needs to be investigated further. This suggests that case marking might interplay with agreement systems, potentially influencing their relationship with word order. In any case, it would be interesting to examine the interaction between case marking and agreement in future research.

Ultimately, intonation, as a strategy for managing information structure, could significantly impact the relationship between word order and agreement. Therefore, exploring a possible interaction between intonation, word order and agreement in spoken data could be beneficial.

5. Conclusion

In summary, the results provide preliminary support for the assumption that flexible word order and verbal agreement are functionally overlapping mechanisms for managing information structure in discourse. The lack of statistical significance for the effect may be due to the limited sample size of 17 languages, suggesting that further research with more languages is necessary to confirm these findings. Nonetheless, the overall negative regression line observed between word order flexibility and syntactic agreement aligns with the theoretical prediction of an implication between these two variables. The word order entropy results support Levshina's (2019) hypothesis that syntheticity is a necessary condition for high word order entropy. However, when managing information structure through agreement, languages may display synthetic characteristics despite having low word order entropy. In terms of syntactic agreement, languages such as Persian and Cypriot Greek, which exhibit high agreement rates, were found to have strict grammatical rules enforcing subject agreement. This potentially explains why Cypriot Greek does not show the expected trade-off between a high agreement rate and flexible word order, as seen in other languages.

To conclude, the study provides the first token-based approach to investigate the interaction between word order entropy and verbal agreement. The findings suggest that the relationship between word order entropy and syntactic agreement may be more complex than initially hypothesized, and further investigation is needed to fully understand the interaction between the two argument-marking strategies. Future studies should expand their scope to include a larger typological sample, referential marking, different types of agreement, case marking, and intonation. This will help provide a better understanding of the complex relationship between word order entropy and agreement.

6. References

Adibifar, S. 2016. Multi-CAST Persian. In G. Haig & S. Schnell (eds.), Multi-CAST: Multilingual Corpus of Annotated Spoken Texts (Version 2311). Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/#persian

Andrews, A. 2007. The major functions of the noun phrase. In T. Shopen (ed.), *Language Typology and Syntactic Description, Volume 1: Clause Structure*. Cambridge: Cambridge University Press, 132–223.

Barth, D., K. Davey & M. Matheas. 2023. Multi-CAST Matukar Panau. In G. Haig & S. Schnell (eds.), Multi-CAST: Multilingual Corpus of Annotated Spoken Texts (Version 2311). Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/#matukar

Bogomolova, N., D. Ganenkov & N. N. Schiborr. 2021. Multi-CAST Tabasaran. In G. Haig & S. Schnell (eds.), Multi-CAST: Multilingual Corpus of Annotated Spoken Texts (Version 2311). Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/#tabasaran

Bickel, B. 2008. On the scope of the referential hierarchy in the typology of grammatical relations. In G. G. Corbett & M. Noonan (eds.), *Case and Grammatical Relations: Studies in Honor of Bernard Comrie*. Amsterdam: John Benjamins, 191–210.

Brickell, T. 2016. Multi-CAST Tondano. In G. Haig & S. Schnell (eds.), Multi-CAST: Multilingual Corpus of Annotated Spoken Texts (Version 2311). Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/#tondano

Comrie, B., M. Haspelmath & B. Bickel. 2008. The Leipzig Glossing Rules: Conventions for Interlinear Morpheme-by-Morpheme Glosses. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Dahl, Ö. 2015. How WEIRD are WALS languages? Paper presented at the Closing Conference of the Department of Linguistics at the Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany. https://www.eva.mpg.de/fileadmin/content_files/linguistics/conferences/2015-diversity-linguistics/Dahl_slides.pdf

Dryer, M. S. & M. Haspelmath (eds.). 2013. WALS/Online. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from http://wals.info/

Dryer, M. S. 2002. Case distinctions, rich verb agreement, and word order type. *Theoretical Linguistics 28*: 151–157.

Forker, D. & N. N. Schiborr. 2019. Multi-CAST Sanzhi Dargwa. In G. Haig & S. Schnell (eds.), Multi-CAST: Multilingual Corpus of Annotated Spoken Texts (Version 2311). Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/#sanzhi

Hadjidas, H. & M. Vollmer. 2015. Multi-CAST Cypriot Greek. In G. Haig & S. Schnell (eds.), Multi-CAST: Multilingual Corpus of Annotated Spoken Texts (Version 2311). Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/#cypgreek

Haig, G. & S. Schnell. 2014. Annotations using GRAID (Grammatical Relations and Animacy in Discourse): Introduction and guidelines for annotators. Version 7.0. https://multicast.aspra.uni-bamberg.de/

Haig, G. & S. Schnell. 2015. Multi-CAST: Multilingual Corpus of Annotated Spoken Texts. Bamberg: University of Bamberg. https://multicast.aspra.uni-bamberg.de/

Haig, G. & S. Schnell (eds.). 2023. Multi-CAST: Multilingual Corpus of Annotated Spoken Texts (Version 2311). Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/

Haig, G., S. Schnell & F. Seifart (eds.). 2022. *Doing Corpus-Based Typology With Spoken Language Corpora: State of the Art.* Language Documentation & Conservation Special Publication 25. Honolulu: University of Hawaiʻi Press. https://doi.org/10.5167/uzh-217096

Haig, G., S. Schnell & N. N. Schiborr. 2021. Universals of reference in discourse and grammar: Evidence from the Multi-CAST collection of spoken corpora. In G. Haig, S. Schnell & F. Seifart (eds.), *Doing Corpus-Based Typology with Spoken Language Corpora*. Language Documentation & Conservation Special Publication 25. Honolulu: University of Hawai'i Press.

Haig, G., M. Vollmer & H. Thiele. 2019. Multi-CAST Northern Kurdish. In G. Haig & S. Schnell (eds.), Multi-CAST: Multilingual Corpus of Annotated Spoken Texts (Version 2311). Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/#nkurd

Haspelmath, M. 2021. Role-reference associations and the explanation of argument coding splits. *Linguistics* 59(1): 123–174.

Hawkins, S. 2002. Symmetries and asymmetries: Their grammar, typology and parsing. *Theoretical Linguistics* 28: 95–149.

Hengeveld, K. 2012. Referential markers and agreement markers in Functional Discourse Grammar. *Language Sciences 34*(4): 468–479.

Holton, D., P. Mackridge & I. Philippaki-Warburton. 2016. *Greek: An Essential Grammar of the Modern Language*. Routledge.

Kimoto, Y. 2019. Multi-CAST Arta. In G. Haig & S. Schnell (eds.), Multi-CAST: Multilingual Corpus of Annotated Spoken Texts (Version 2311). Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/#arta

Kurabe, K. 2021. Multi-CAST Jinghpaw. In G. Haig & S. Schnell (eds.), Multi-CAST: Multilingual Corpus of Annotated Spoken Texts (Version 2311). Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/#jinghpaw

Lestrade, S. 2015. The interaction of argument-marking strategies. In S. Lestrade, P. J. F. de Swart & L. Hogeweg (eds.), *Addenda: Artikelen voor Ad Foolen*, 251–256. https://hdl.handle.net/2066/142373

Levshina, N. et al. 2023. Why we need a gradient approach to word order. *Linguistics* 61(4): 825–883. https://doi.org/10.1515/ling-2021-0098

Levshina, N. 2019. Token-based typology and word order entropy: A study based on Universal Dependencies. *Linguistic Typology* 23(3): 533–572. https://doi.org/10.1515/lingty-2019-0025

Meng, C. 2019. Multi-CAST Tulil. In G. Haig & S. Schnell (eds.), Multi-CAST: Multilingual Corpus of Annotated Spoken Texts (Version 2311). Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/#tulil

Mosel, U. & S. Schnell. 2015. Multi-CAST Teop. In G. Haig & S. Schnell (eds.), Multi-CAST: Multilingual Corpus of Annotated Spoken Texts (Version 2311). Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/#teop

Schiborr, N. N. 2015. Multi-CAST English. In G. Haig & S. Schnell (eds.), Multi-CAST: Multilingual Corpus of Annotated Spoken Texts (Version 2311). Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/#english

Schnell, S. & N. N. Schiborr. 2022. Cross-linguistic corpus studies and linguistic typology. *Annual Review of Linguistics* 8: 171–191. https://doi.org/10.1146/annurev-linguistics-031120-104629

Schnell, S. 2015. Multi-CAST Vera'a. In G. Haig & S. Schnell (eds.), Multi-CAST: Multilingual Corpus of Annotated Spoken Texts (Version 2311). Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/#veraa

Seifart, F. & T. Hong. 2022. Multi-CAST Bora. In G. Haig & S. Schnell (eds.), Multi-CAST: Multilingual Corpus of Annotated Spoken Texts (Version 2211). Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/#bora

Shannon, C. E. 1948. A mathematical theory of communication. *Bell System Technical Journal* 27(3): 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Shiohara, A. 2022. Multi-CAST Sumbawa. In G. Haig & S. Schnell (eds.), Multi-CAST: Multilingual Corpus of Annotated Spoken Texts (Version 2211). Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/#sumbawa

Siewierska, A., & D. Bakker. 2008. Case and alternative strategies: Word order and agreement marking. In A. L. Malchukov & A. Spencer (eds.). 2012. *The Oxford Handbook of Case*. Oxford: Oxford University Press. Online edition, Oxford Academic. https://doi.org/10.1093/oxfordhb/9780199206476.013.0020

Siewierska, A. 2004. *Person*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511812729

Silverstein, M. 1976. Hierarchy of features and ergativity. In R. M. W. Dixon (ed.), *Grammatical Categories in Australian Languages*. Canberra: Australian National University, 112–171.

Simpson, J. 2012. Information structure, variation and the referential hierarchy. In F. Seifart, G. Haig, N. P. Himmelmann, D. Jung, A. Margetts & P. Trilsbeek (eds.), *Potentials of Language Documentation: Methods, Analyses, and Utilization*. University of Hawaii Press, 73–82.

Thieberger, N. & T. Brickell. 2019. Multi-CAST Nafsan. In G. Haig & S. Schnell (eds.), Multi-CAST: Multilingual Corpus of Annotated Spoken Texts (Version 2311). Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/#nafsan

Visser, E. 2021. Multi-CAST Kalamang. In G. Haig & S. Schnell (eds.), Multi-CAST: Multilingual Corpus of Annotated Spoken Texts (Version 2311). Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/#kalamang

Vollmer, M. 2020. Multi-CAST Mandarin. In G. Haig & S. Schnell (eds.), Multi-CAST: Multilingual Corpus of Annotated Spoken Texts (Version 2311). Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/#mandarin

7. Appendices

Appendix A: Results of Word Order Scripts

Table 3: Overall entropy values and the individual word order entropies for each language.

		Individual Entropy values					
Language	Overall Entropy (individual entropies averaged)	nomS_V	nomA_V	nomP_V	pronS_V	pronA_V	pronP_V
Veraa	0.0114	0	0	0	0	0.0686	0
Matukar	0.0308	0	0	0.1847	0	0	0
English	0.0784	0.1166	0.0731	0.1096	0.0485	0.1009	0.022
Nkurd	0.0922	0.1292	0	0.2727	0	0	0.1511
Nafsan	0.0969	0.5813	0	0	0	0	0
Persian	0.1230	0.0905	0	0.1437	0	0	0.5033
Kalamang	0.1840	0.0789	0.2285	0.4855	0	0.1392	0.172
Jinghpaw	0.1861	0	0.1872	0.4418	0.1033	0.3843	0
Mandarin	0.1926	0.5687	0	0.2056	0	0	0.3809
Arta	0.3036	0.3534	0.2164	0.2879	0	0	0.9641
Tulil	0.3698	0.795	0	0.0705	0.9995	0.3534	0
Teop	0.4330	0.5472	0.5127	0.1936	0.1556	0.5917	0.5976
Cypgreek	0.5383	0.2975	0.9321	0.3877	0.7219	0.8905	0
Bora	0.6232	0.6122	0.6292	0.8356	0.3451	0.4328	0.8841
Sumbawa	0.7023	0.8882	0.9569	0.3893	0.9975	0.9819	0
Tondano	0.7819	0.971	0.8865	0.5262	0.5436	0.7642	1
Sanzhi	0.9138	0.976	0.8366	0.7928	0.9321	0.9812	0.9641
Tabasaran	0.9154	0.7548	0.9341	0.9507	0.9975	0.9367	0.9183

Appendix B: Correlation between Word Orders

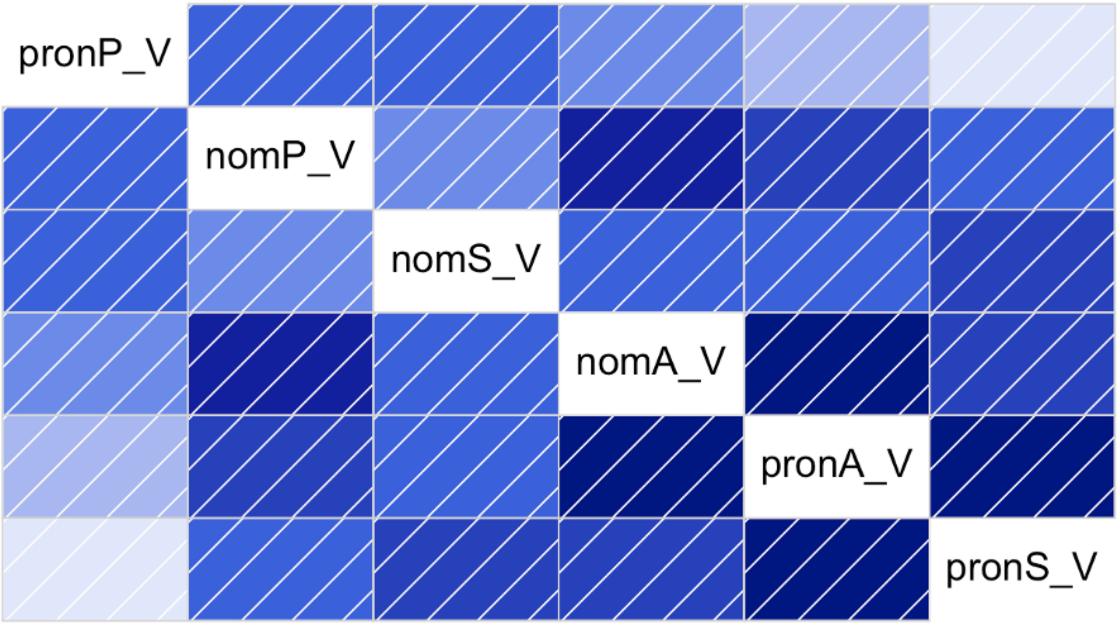

Figure 3: A correlogram, using the individual entropy values from table 3 as data points, displays the Spearman rank-order correlation between the 6 word orders. Each blue rectangle indicates a positive correlation between the two word orders in its vertical and horizontal line, respectively. The color intensity represents correlation strength.

Appendix C: Results of Agreement Script

Table 4: For each language, the exact agreement values, amount of possible agreement slots and individual frequencies are listed.

				Individual Frequencies		
Languages	Agree-ment Rate	Possible Agree-ment Slots	Total Number of Agree-ment Markers	Agreement Markers counted during step 1 of Algorithm Procedure	Agreement Markers counted during step 2 of Algorithm Procedure	Agreement Markers counted during step 3 of Algorithm Procedure
Mandarin	0	745	0	0	0	0
Kalamang	0	858	0	0	0	0
Jinghpaw	0	0	0	0	0	0
Veraa	0	3070	0	0	0	0
Tabasaran	0.0011	917	1	0	0	1
Teop	0.0092	976	9	9	0	0
English	0.0262	5389	141	10	0	131
Sumbawa	0.1083	628	68	65	3	0
Sanzhi	0.1606	467	75	0	0	75
Tondano	0.2704	429	116	109	7	0
Bora	0.3641	541	197	176	20	1
Matukar	0.3833	694	266	119	147	0
Arta	0.3984	374	149	149	0	0
Tulil	0.4408	524	331	4	0	227
Nafsan	0.6078	487	296	8	288	0
Persian	0.6325	762	482	1	0	481
Nkurd	0.6424	1046	672	10	0	662
Cypgreek	0.8373	381	319	41	0	278

Appendix D: Local Outlier Factor (LOF) Analysis

<u>R code:</u>

```
library(DMwR2)
lof_scores <- lofactor(data[, c("Agreement", "WordOrderEntropy")], k = 4)
lof_outliers <- data[lof_scores > 1.5, ]
print(lof_outliers)
```

→ output: Cypgreek

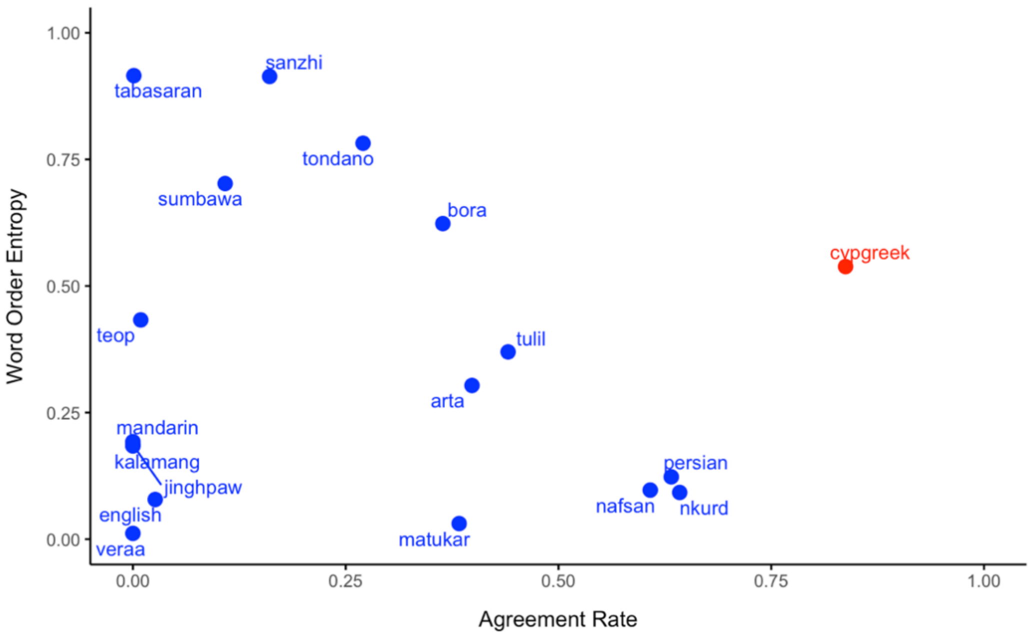

Figure 4: The data point cypgreek (Cypriot Greeek) in its isolated position.

Appendix E: R Output of linear mixed-effect Model

```
Linear mixed model fit by REML ['lmerMod']
Formula: wo ~ agr + (1 | Genus)
   Data: data

REML criterion at convergence: 8.3

Scaled residuals:
    Min      1Q   Median     3Q     Max
-1.1803 -0.4813 -0.0972  0.5266  1.1370

Random effects:
 Groups   Name        Variance Std.Dev.
 Genus    (Intercept) 0.0684   0.262
 Residual             0.0395   0.199
Number of obs: 17, groups:  Genus, 11

Fixed effects:
            Estimate Std. Error t value
(Intercept)    0.445      0.112    3.99
agr           -0.322      0.282   -1.14

Correlation of Fixed Effects:
    (Intr)
agr -0.521
```

Calculating p-value:

```
t_value <- -1.14
degrees_of_freedom <- 17 - 2
p_value <- 2 * pt(abs(t_value), df = degrees_of_freedom, lower.tail = FALSE)
```

→ p-value: 0.272

Appendix F: Fixed-Effect Model

Model fit R code:

```
model_glmmTMB <- glmmTMB(wo ~ agr, data = data, REML = TRUE)
```

R Output:

```
Family: gaussian  ( identity )
Formula:          wo ~ agr
Data: data

     AIC       BIC    logLik -2*log(L)  df.resid
    16.9      19.4      -5.4      10.9        14

Dispersion estimate for gaussian family (sigma^2):  0.1

Conditional model:
               Estimate     Std. Error    z value     Pr(>|z|)
(Intercept)    0.4236        0.1086       3.901       9.56e-05 ***
agr            -0.2874       0.3226       -0.891      0.373
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Appendix G: Post Hoc Power Analysis

Power for predictor 'agr', (95% confidence interval):
18.00% (11.03, 26.95)

Test: Kenward Roger (package pbkrtest)
Effect size for agr is -0.32

Based on 100 simulations, (0 warnings, 0 errors)
alpha = 0.05, nrow = 17