Case inflection and the functional indeterminacy of nouns: A cross-linguistic analysis

Nicholas A. Lester (nlester@umail.ucsb.edu)
Sandra Auderset (auderset@umail.ucsb.edu)
Phillip G. Rogers (phillip@umail.ucsb.edu)
Department of Linguistics
University of California, Santa Barbara, South Hall 3432
Santa Barbara, CA 93106 USA

Abstract

Prior research shows that languages balance syntactic complexity against morphological complexity. We explore this relationship using a new measure of syntactic complexity, *functional indeterminacy*, which measures the aggregate uncertainty of mapping from lexical items to syntactic function. We predict that greater functional indeterminacy for nouns will correlate with languages having case systems, and for those with case systems, increased number of cases. We operationalize indeterminacy as the simple and normalized conditional entropies of the summed frequency distributions of nouns across syntactic dependencies. We compute these measures for 44 languages. We then correlate the measures with presence and number of cases in two regression analyses, controlling for genetic affiliation between languages. Results show that as the functional indeterminacy of nouns increases, languages are more likely to have case systems, and if so, to have more cases. These data provide new support for the functionally motivated relationship between morphological and syntactic complexity.

Keywords: syntax-morphology trade-off; case marking; cross-linguistic variation; dependency syntax; entropy

Introduction

Languages are structured at multiple levels of representation. In some cases, these layers of organization overlap in the information they express. For example, information about the syntactic functions of words may be encoded by word order and/or morphology (among other things; e.g., adpositions, clitics, and so on). A long-standing theory in linguistics posits that where such overlap occurs, languages will balance the expressiveness of one layer of representation against the other. Some languages develop rich systems of inflectional morphology while leaving word order relatively free (e.g., Latin, Russian, Hungarian etc.). Other languages have shallow morphology but rigid constraints on word order (e.g., English, Vietnamese, Indonesian, etc.).

These global trade-offs in complexity between morphology and syntax have been well documented. Sinnemäki (2008) defines complexity in terms of "functional load." He defines functional load relative to four categorical levels based on the strategies that languages employ to disambiguate the syntactic roles of core arguments. He codes a sample of 50 languages for these levels of functional load in word order and morphological marking and finds significant inverse correlations between the two. Koplenig, Meyer, Wolter, & Müller-Spitzer (2017) provide a more direct, text-based measure of complexity for syntax and morphology in a sample of over 1,000 languages. They operationalize complexity in morphology and word order as the increase in entropy that results from randomizing continuous texts word-internally and word-externally, respectively. They likewise find strong negative correlations between the two measures. Other evidence comes from experimental studies. For example, Fedzechkina, Newport, & Jaeger (2017) used an artificial grammar paradigm in which they manipulated the freedom of word order while including optional case inflections on nouns. They found that learners of fixed word order languages relied less on case marking than learners of free word order languages.

These findings are often interpreted as evidence for competition between two forces: maximization of clarity and minimization of effort. The balancing act between clarity and effort reflects a tendency for languages to maximize communicative efficiency (see Jaeger, 2010; Levy, 2008). Languages with rich inflectional morphology and strict word order double down on effort to maximize clarity (e.g., Icelandic). Conversely, languages with no morphology and free word order minimize effort but risk unlimited ambiguity (e.g., Riau Indonesian; Gil, 1994). Most languages fall somewhere between these two extremes, committing to some effort in the morphology or word order, but permitting some ambiguity. The primary goal of the present study is to explore whether the ambiguity of syntactic function – that aspect of linguistic representation that both word order and morphology aim to measure – correlates with the accretion of complex morphology.

Prior work on these issues focuses on word order and inflectional morphology. The language of 'trade-off' assumes that word order and morphology are overt manifestations of the same underlying informational signal. We refer to this underlying signal as *syntactic function*. Prior research has therefore approximated syntactic function by measuring properties of the mechanisms by which it is conveyed. Importantly, the association of syntactic function to the means of its encoding have not been consistent. In some studies, syntactic function has been unambiguously linked to a morphological form or position within the clause, but

for only a highly limited subset of the overall syntactic-functional space (Fedzechkina, et al., 2017; Sinnemäki, 2008). This scenario is unlikely to arise in natural languages, introducing the question of scalability. Other studies measure word order and morphology without considering syntactic function at all (Koplenig et al., 2017).

To avoid this issue, we measure syntactic function directly, independently of both word order and morphology. We focus on nouns, whose syntactic function may differ based on word order and inflectional morphology in the form of case marking. Specifically, we operationalize the ambiguity within syntactic function as the predictability with which a given noun is mapped into any of the syntactic functions available to its class. We refer to this dimension as *functional indeterminacy*. On analogy to prior findings, we expect languages with greater functional indeterminacy (i.e., less predictable mappings between nouns and syntactic relations) to develop more robust systems of case marking. We test this hypothesis using a sample of 44 languages from seven language families and 17 genera (sub-families).

Case inflection and syntactic function

Case inflection is a form of dependent-marking whereby the syntactic function of a noun is marked by a morphological change to the stem, typically by an affix. For example, the Hungarian stem *hegy-* 'mountain' may be suffixed with *–et* to create *hegyet*, which reflects its syntactic status as direct object. Case-marking is thus a form of local syntactic disambiguation. Languages that exhibit case marking differ in the number of cases they distinguish. Languages with more cases provide a more powerful system for disambiguating the syntactic function of nouns in context. Following Ackerman and Malouf (2013), we refer to this measure as *enumerative complexity*.

Other studies have attempted to measure the complexity of inflectional paradigms like case using frequency distributions. For example, Moscoso del Prado Martín, Kostić, and Baayen (2004) show that the uncertainty of the frequency distributions of nouns across their various inflected variants influences how quickly they are processed in the visual lexical decision paradigm. These measures have the advantage of accounting for the operative complexity of the case system. For example, a language may have 7 cases, but speakers may only produce instances of 5 of those cases with any regularity.

Regarding syntax, we take as our level of analysis the syntactic relations in which nouns are observed. This approach complements prior work on word order and inflectional morphology. In particular, neither word order nor inflectional morphology, nor even their combination fully disambiguates the syntactic function of nouns in context. For example, German syntax allows nouns to occupy pre- and post-verbal positions, irrespective of the type of relation (e.g., *Die Frau begrüßt das Mädchen / Das Mädchen begrüßt die Frau*[1] 'The woman greets the girl.'). In addition, German verbs may require specific case inflections for their argument nouns, irrespective of the more general syntactic functions of those inflections (e.g., *antworten* 'to answer' requires dative case for direct objects rather than the more productive accusative case). Therefore, to the extent that a given language exhibits these kinds of ambiguities, the system of syntactic functions must carry some information independent of word order and/or morphology.

Having defined the syntactic space, we must address how nouns are distributed within that space. Speakers tend not to combine all nouns in equal proportion to the set of available syntactic relations. They may favor some relations and eschew others given the communicative demands of naturally occurring discourse. These distributional asymmetries reduce the problem space, rendering some part of the ostensible complexity moot. We refer to the residual complexity carried by the syntactic distribution of nouns as functional indeterminacy. Crucially, functional indeterminacy may differ cross-linguistically, and part of this variability may be functionally determined relative to other components of the grammar, such as inflectional morphology (Kostić et al., 2003).

Based on these considerations, we formulate the following two hypotheses ($H_{1\text{-}2}$):

H_1: As the average functional indeterminacy of nouns increases, languages will be more likely to have a case system than not.

H_2: As the average functional indeterminacy of nouns increases, the enumerative complexity of the case system also increases.

H_2 assumes that when nouns engage in a diverse array of syntactic relationships on average (high indeterminacy), the syntactic function of nouns are generally more difficult to recover, all else being equal. To counter the functional indeterminacy, languages develop local, explicit cues to disambiguate syntactic function. Notice that this prediction is functionally motivated in that it maximizes communicative efficiency: speakers compensate for functional ambiguity by providing explicit, local cues.

Data and Methods

Estimating the functional indeterminacy of nouns requires parsed corpora. Because we are interested in comparing the relationship between indeterminacy and case inflection across languages, the ideal corpora should be parsed according to comparable standards. We therefore select the Universal Dependency Treebanks (UDT v2.6), which contains treebanks for 50 morphosyntactically and genetically diverse languages. All corpora in UDT have been parsed according to a central set of standards, with some variation allowed for language-specific categories.

[1]The second formulation, in which the direct object precedes the verb, is discourse-pragmatically constrained, but not syntactically illicit.

Where possible, we extract the number of case-inflectional categories from the World Atlas of Languages (WALS) database (Iggeson, 2013). For the remaining languages, this information was extracted from Ethnologue (Simons & Fennig, 2017). We further extracted genetic affiliations (at two levels, *genus* and *family*) from Glottolog (Hammarström, Bank, Forkel, & Haspelmath, 2017). We cross-checked the numbers of cases against reference grammars. In cases of disagreement, we selected the most conservative (lowest) number of cases per language. We only counted cases that surface on lexical nouns, and so ignore case distinctions made only for pronouns. For example, WALS states that English has two cases, presumably referring to the subject/object case distinction for pronouns (*he/him*). However, lexical nouns do not reflect this contrast (*dog/dog*); therefore, we treat English as having zero case-inflectional categories. Following WALS, we also only counted cases that are productive at least at the level of declension class (where applicable). For example, German has four cases, but only genitive (for masculine and neuter) and dative require distinctive forms for the noun stems (with a third unmarked form for nominatives and accusatives). Also following WALS, we treated instances of syncretism across cases within a language as single cases. For example, Croatian has six cases, but the dative and locative cases have the same form in all declension classes. Where syncretism differed across declension classes, we took the number of cases that are distinguished in at least one declension class. Finally, the genitive case sometimes surfaces as a phrasal clitic. We counted the genitive as an inflectional category only if the genitive morpheme must attach to the noun stem, and not if it attaches to the edge of the full NP. For example, English is not considered to have an inflectional genitive because the morphological genitive *'s* attaches to the end of the NP (e.g., *the dog with the brown spot's bowl*).

Table 1 lists the languages in our final sample, along with the sample size, number of case-inflectional categories, and genetic affiliations.

Table 1: Languages in the sample[2]

Language	Sample size	Cases	Genus	Family
A. Greek	414K	5	Greek	IE
Arabic	1.042M	3	Semitic	Afroasiatic
Basque	121K	12	Basque	Isolate
Belarusian	8K	6	Slavic	IE
Bulgarian	156K	0	Slavic	IE
Catalan	531K	0	Italic	IE
Coptic	11K	0	Egyptian	Afroasiatic
Croatian	197K	5	Slavic	IE
Czech	2.222M	7	Slavic	IE
Danish	100K	0	Germanic	IE
Dutch	310K	0	Germanic	IE
English	496K	0	Germanic	IE
Estonian	106K	14	Finnic	Uralic
Finnish	377K	15	Finnic	Uralic
French	1.099M	0	Italic	IE
Galician	164K	0	Italic	IE
German	313K	3	Germanic	IE
Gothic	55K	5	Germanic	IE
M. Greek	63K	4	Greek	IE
Hebrew	161K	0	Semitic	Afroasiatic
Hindi	375K	2	Indic	IE
Hungarian	42K	17	Ugric	Uralic
Irish	23K	2	Celtic	IE
Italian	436K	0	Italic	IE
Japanese	402K	0	Japanese	Japonic
Latin	491K	6	Italic	IE
Latvian	90K	5	Baltic	IE
Lithuanian	5K	7	Baltic	IE
Norwegian	625K	0	Germanic	IE
OCS	57K	7	Slavic	IE
Persian	152K	2	Iranian	IE
Polish	83K	7	Slavic	IE
Portuguese	570K	0	Italic	IE
Romanian	239K	2	Italic	IE
Russian	99K	6	Slavic	IE
Slovak	106K	6	Slavic	IE
Slovenian	170K	6	Slavic	IE
Spanish	1.004M	0	Italic	IE
Swedish	195K	2	Germanic	IE
Tamil	9K	8	S. Dravidian	Dravidian
Turkish	74K	6	Turkic	Turkic
Ukrainian	100K	7	Slavic	IE
Urdu	138K	2	Indic	IE
Vietnamese	43K	0	Viet-Muong	Austroasiatic

[2]A. Greek = Ancient Greek; M. Greek = Modern Greek; OCS = Old Church Slavonic; S. Dravidian = Southern Dravidian; IE = Indo-European; K = thousand; M = million.

Measures

The UDT parses follow the basic structure of dependency grammar (DG). In this formalism, syntactic structure is expressed in the form of acyclic graphs whose nodes are words and whose edges are typed functional relations (this differs from the more familiar phrase-structure trees, which

are built of abstract phrasal nodes that bind groups of terminal nodes into constituents via label-less arcs). Dependency graphs are hierarchically organized such that each word is dominated, or *headed*, by exactly one other word, though each word may head indefinitely many dependents, or *modifiers*. Again, each of these head or modifier relations is directly labeled for syntactic function. For example, in the phrase *the aged cheese*, the word *the* is headed by *cheese* via the DET (determiner) relation, while the word *aged* is headed by *cheese* via the AMOD (adjectival modification) relation. *The* and *aged* are thus modifiers of the head *cheese*, while *det* and *amod* label their respective syntactic functions.

We base our measure of functional indeterminacy on the frequency distribution of nouns across the set of typed syntactic relations that occur within each language. To simplify, we ignore whether the target nouns serve as head or modifier in the relations and the word order of the head relative to the modifier. To avoid a morphological confound, we compute these measures over lemmas (i.e., stemmed forms). Using information theory, the functional indeterminacy of a noun can be understood as the entropy of its frequency distribution across the set of dependency relation types, where the entropy is defined as in Eq. 1.

$$H(D) = -\sum_{d \in D} p(d) \log p(d) \quad (1)$$

In Eq. 1, $p(d)$[3] reflects the probability with which a noun occurs in a given dependency relation d from the set of all dependencies D. $H(D)$ is highest when a noun is distributed evenly across all dependency relations d in D. It approaches zero as a noun tends to occur only in a single dependency.

Each of these dependency relations is bound up with a non-target word that co-instantiates the relation (e.g., *aged* is bound to the target *cheese* within the *amod* relation). Lexical co-distributions are known to tap into semantics (Bullinaria & Levy, 2012). If the syntactic relation is fully or partially redundant given the words that instantiate it, then the entropy is ambiguous between semantic and syntactic information. We therefore require some means of removing the lexical information to arrive at a more thoroughly syntactic distribution. Otherwise, the interpretability of the measure is hindered. We handle this by appealing to the information-theoretic notion of *conditional entropy*. Instead of taking $p(d)$ directly, conditional entropy requires that we take two distributions. First, we take the joint distribution of the noun across each combination of dependency relation d and non-target word w in the set of words W, or $p(d,w)$. Then, we take the distribution across the non-target words, irrespective of dependency relation $p(w)$. We take the entropy of each distribution independently and subtract the lexical entropy $H(W)$ from the joint entropy $H(D,W)$. This relationship is formalized in Eqs. 2-4. Conditional entropy allows us to remove the information carried by non-target words from the information jointly carried by non-targets and their associated dependencies. This leaves us with the information carried by the dependencies given (i.e., without) the information carried by the non-targets.

$$H(D \mid W) = H(D, W) - H(W) \quad (2)$$

$$H(D, W) = H(D) + H(W) - I(D; W) \quad (3)$$

$$I(D; W) = \sum_{d \in D} \sum_{w \in W} p(d, w) \log p \frac{(d, w)}{p(d)\, p(w)} \quad (4)$$

Estimates of these entropies are subject to an underestimation bias (Miller, 1955). We therefore correct the entropies using the estimator introduced by Chao, Wang, and Jost (2013), which is based on the species accumulation curve. This estimator is thought to be the least biased by sample size, making it ideal for handling Zipf-distributed lexical frequency distributions (Moscoso del Prado Martín, 2016)

The conditional entropy for a given language measures how much functional indeterminacy speakers actually build into their utterances. Another dimension of use is whether speakers are as syntactically indeterminate as they could be given the structure of the language. To account for this dimension, we also compute the normalized conditional entropy. Normalized conditional entropy is defined as the conditional entropy $H(d \mid w)$ divided by the maximum entropy $H_{max}(d)$, where $H_{max}(d)$ is defined as the logarithm of the number of possible dependencies in d. The resulting value is a proportion representing how much of the possible indeterminacy the speakers of a language actually exploit.

To capture the behavior of the system as a whole, we first sum the distributions of all nouns within the syntactic space. We then compute the conditional entropy and normalized conditional entropy over the summed vectors. We do this for each language in the sample.

Results

First, we test H_1 using a generalized linear mixed-effect model (GLMM) to predict whether a language has a case system (+case) or not (-case). While conceptually distinct, conditional entropy and normalized conditional entropy are highly correlated (ρ = .96, p < .001). This correlation can cause problems of interpretation if both measures are included within the GLMM. To handle this problem, we perform an Independent Component Analysis (ICA) to decorrelate the measures. Because the magnitude of loadings within the ICA is sensitive to scale, we rescale the measures by taking the z-scores of each. We use ICA to derive a single component which loads strongly and positively for both variables. Positive scores for the component represent increasing conditional entropy and proportion of maximum conditional entropy. This means

[3] The true probability is conditioned on the target word t: $p(d \mid t$ = target). We simplify the equations by removing the conditioning expression.

that the component captures information shared by both conditional entropy and maximization of conditional entropy. A visual examination of the ICA scores revealed two outliers, Vietnamese (-case) and Turkish (+case). We removed these from the dataset before proceeding. We fit the GLMM with ICA scores as independent variables and random intercepts of genus nested into family (to control for genetic factors; see Jaeger, Graff, Croft, & Pontillo, 2011). The model with the ICA scores significantly outperformed the null model (χ^2_Δ = 4.99; p = .02). As the average functional indeterminacy of nouns increases for a language, so does the likelihood that the language will have a case system (β = 1.23, SE = 0.62, p = .05).

Next, to test H_2, we perform a linear mixed effect regression with number of case-inflectional categories for only those languages that have cases. The rest of the model was identical to the GLMM reported above. A Box-Cox power analysis (Box & Cox, 1964) revealed that we should take the logarithm of the number of cases to better approximate normality. We therefore substitute the log-transformed number of cases as dependent variable. This step is necessary in order to satisfy the assumptions of the model. The model with the ICA scores significantly outperformed the null model (χ^2_Δ = 3.87; p = .05). As functional indeterminacy increases, so does the number of inflectional categories (β = 0.14, SE = .06; F(1, 16.26) = 6.41, p = .02). This effect is plotted in Figure 1.

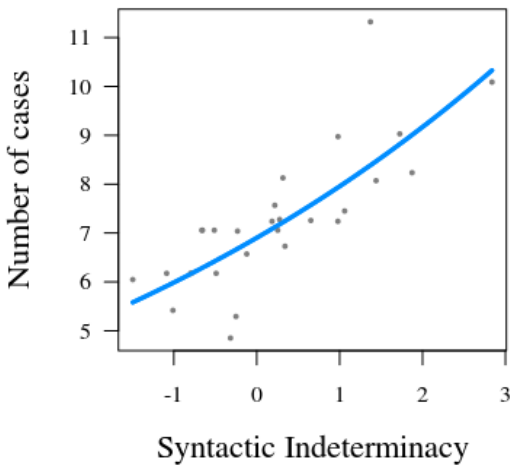

Figure 1: Effect of functional indeterminacy (ICA scores) on the number of cases in languages with case systems. More positive values on the x-axis reflect increasing conditional entropy and proportion of maximum conditional entropy. Points represent partial residuals per language.

Discussion

Many recent experimental and observational studies have demonstrated a number of trade-offs in complexity among the various tiers of linguistic representation (see the papers in Miestamo, Sinnemäki, & Karlsson, 2008). A notable finding from this literature concerns the relationship between syntax and morphology: across languages, morphological complexity is inversely proportional to syntactic complexity. One explanation for this relationship is that languages are organized to minimize both ambiguity and effort to promote communicative efficiency (e.g., Jaeger, 2010). Most of this research has compared word order to the morphological structure of words. However, both of these phenomena are overt instantiations of the same underlying principle, what we refer to as *syntactic function*. Using information theory, we developed a novel way to measure the complexity of syntactic function for nouns across languages, which we refer to as functional indeterminacy. Based on the logic of communicative efficiency, we proposed two new hypotheses relating the functional indeterminacy to the prevalence of case marking across 44 languages. We tested these hypotheses in two regression analyses.

The results of the analyses confirm our hypotheses. First, case systems are more common for languages that have more complex functional-syntactic distributions for nouns on average. Second, for languages with case systems, the number of cases increases with the indeterminacy of noun syntax. Importantly, these relationships were predicted based on a principle of communicative efficiency. If we can be less sure of the syntactic functions of nouns generally, we can increase our certainty by changing their form, thereby making a successful parse on the part of our listeners more likely. The fact that we demonstrate these relationships across a broad number of languages supports the more general argument that typological features, such as the presence or absence of case marking, are functionally motivated (e.g., Fedzechkina et al., 2017).

Future Directions

First, our measure of case richness ignores both phonological and distributional factors, both of which contribute to the complexity of the case system. This issue is compounded by the potential differences that may occur between declension classes within any given language. Future research should account for these factors, most likely through multiple competing measures. For example, one could take the integrative complexity, which is defined relative to the structural options within languages, as well as the distributional variability of nouns across their inflectional categories. Additionally, distribution-based measures would have to be conditioned on declension class to account for paradigm-specific behaviors (e.g., Milin et al., 2009). A crucial problem concerns how to integrate complexity estimates across declension classes.

Another outstanding issue concerns the relationship between word order and syntactic function. We have argued that in natural languages, word order is not an unambiguous cue of syntactic function, even in languages with highly inflexible word order. However, this is in fact an empirical question. It is possible that the majority of languages with rigid word order have near one-to-one correspondences between position and syntactic function. This problem may be addressed using the dependency treebanks used here,

which also contain information about word order. However, word order is partially dependent on hierarchical status, that is, status as head or modifier. At the very least, one would need to measure the information carried by word order as conditioned on the hierarchical status of the target word. Additionally, one would need to strip away the information carried by the non-target lexemes, for the same reasons that we describe in Measures. The optimal formalization for these measures is unclear. A serious problem concerns how to achieve accurate estimates. These distributions are multiply conditioned and hence require large corpora to achieve a reasonable number of observations per cell – exponentially larger than the majority of samples considered here.

Future research should also attempt to increase the typological diversity of the sample. As Table 1 illustrates, our sample is comprised mostly of languages from the Indo-European family. A glaring omission is that we have no indigenous languages of the Americas, many of which differ substantially in their inflectional potential from European languages (e.g., Mithun, 1999). Other omissions include the languages of sub-Saharan Africa, Australia, and Oceania.

Although we have demonstrated a correlation, we expect this relationship to be established over time. In particular, time-series analysis of reasonably sized diachronic corpora should allow us to determine the direction of causality between the development of complexity in case marking and functional indeterminacy. Does case-inflectional complexity drive increases in functional indeterminacy, or vice versa? Evidence from Icelandic suggests that changes in morphological structure drive changes in syntactic structure (Moscoso del Prado Martín, 2014). This finding predicts that the behavior of the syntactic system bends to fit the morphological structure of the language. We await the development of large-scale, longitudinal corpora, taken from diverse languages, which would allow us to test this prediction.

References

Ackerman, F., & Malouf, R. (2013). Morphological organization: The low conditional entropy conjecture. Language. *89*, 429-464.

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society*, *Series B*. *26*, 211–252.

Bullinaria, J. A., & Levy, J. P. (2012). Extracting Semantic Representations from Word Co-occurrence Statistics: Stop-lists, Stemming and SVD. *Behavior Research Methods*, *44*, 890-907.

Chao, A., Wang, Y. T., & Jost, L. (2013). Entropy and the species accumulation curve: a novel estimator of entropy via discovery rates of new species. *Methods in Ecology and Evolution*, *4*, 1091–1110.

Fedzechkina, M., Newport, E. L., & Jaeger, T. F. (2017).Balancing effort and information transmission during language acquisition: Evidence from word order and case marking. *Cognitive Science*, *41*, 416-446.

Gil, D. (1994).The structure of Riau Indonesian. *Nordic Journal of Linguistics*, *17*, 179-200.

Hammarström, H., Bank, S., Forkel, R., & Haspelmath, M. (2017). *Glottolog 3.1*. Jena: Max Planck Institute for the Science of Human History.

Iggesen, O. A. (2013). Number of Cases. In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, *61*, 23-62.

Jaeger, T. F., Graff, P., Croft, W., & Pontillo, D. (2011). Mixed effect models for genetic and areal dependencies in linguistic typology

Koplenig, A., Meyer, P., Wolfer, S., & Müller-Spitzer, C. (2017). The statistical trade-off between word order and word structure: Large-scale evidence for the principle of least effort. *PloS ONE*, *12*, 1-25.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*, 1126-1177.

Miestamo, M., Sinnemäki, K., & Karlsson, F. (Eds.) (2008). *Language complexity: Typology, contact, change*. Amsterdam: John Benjamins.

Milin P, Filipović Đurđević, D., & Moscoso del Prado Martín, F. (2009). The simultaneous effects of inflectional paradigms and classes on lexical recognition: Evidence from Serbian. *Journal of Memory and Language*, *60*, 50-64.

Miller, G. (1955). Note on the Bias of Information Estimates. In H. Quastler (Ed.), *Information Theory in Psychology: Problems and Methods* (pp. 95–100). Glencoe: Free Press.

Mithun, M. (1999). *The languages of native North America*. Cambridge: Cambridge University Press.

Moscoso del Prado Martín, F. (2014). Grammatical change begins within the word: Causal modeling of the co-evolution of Icelandic morphology and syntax. In P. Bello, M. Guarini, M. McShane, and B. Scassellati (Eds.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 2657–2662). Austin: Cognitive Science Society.

Moscoso del Prado Martín, F. (2016). Vocabulary, grammar, sex, and aging. *Cognitive Science*, *41*, 950-975.

Moscoso del Prado Martín, F., Kostić, A., Baayen, R. H. (2004). Putting the bits together: An information theoretical perspective on morphological processing. *Cognition*, *94*, 1-18.

Simons, G. F., & Fennig, C. D. (2017). *Ethnologue: Languages of the world, Twentieth Edition*. Dallas: SIL International.

Sinnemäki, K. (2008). Complexity trade-offs in core argument marking. In M. Miestamo, K. Sinnemäki, & F. Karlsson (Eds.), *Language complexity: Typology, contact, change* (pp. 67–88). Amsterdam: John Benjamins.