



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2023

**The Entropy of Morphological Systems in Natural Languages Is Modulated by
Functional and Semantic Properties**

Franzon, Francesca ; Zanini, Chiara

DOI: <https://doi.org/10.1080/09296174.2022.2063501>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-220374>

Journal Article

Accepted Version

Originally published at:

Franzon, Francesca; Zanini, Chiara (2023). The Entropy of Morphological Systems in Natural Languages Is Modulated by Functional and Semantic Properties. *Journal of Quantitative Linguistics*, 30(1):42-66.

DOI: <https://doi.org/10.1080/09296174.2022.2063501>

The entropy of morphological systems in natural languages is modulated by functional and semantic properties.

Francesca Franzon
Neuroscience Area
International School for Advanced Studies
ffranzon@sissa.it
orcid: 0000-0003-0503-2792

Chiara Zanini
Romanisches Seminar
Universität Zürich
chiara.zanini@uzh.ch
orcid: 0000-0002-8786-4973

Abstract. In most natural languages, grammatical gender and number features encode semantic attributes concerning the animacy, sex, and numerosity of the referents. Despite the likely advantage of promptly communicating about such salient attributes, inflectional systems rarely display consistently bijective correspondences between the semantic attributes and the grammatical feature values. In a study on Italian, we explored how this apparently noisy encoding depends on a trade-off between the semantic and the functional aspects of grammatical features. By means of entropy metrics, we assessed the primarily functional purpose of gender and number features in the lexicon, observing a distribution of nouns that can optimally serve agreement-based parsing and prediction of words in sentences. A novel context entropy measure, introduced in this study to assess meaning specificity, revealed a semantic underspecification in masculine and singular nouns denoting animate referents. We argue that underspecification is the hallmark of the particular type of information compression occurring in inflectional systems. In binary inflectional systems, one value specifically encodes a semantic attribute, while the other value does not encode any semantic information, and surfaces as a default for functional purposes; the encoding of two antithetic semantic attributes is less likely. By providing an information-theoretical account of the role of grammatical features based on large-scale data, we set the basis for a scientifically informed pursue of language inclusiveness.

Keywords. Maximum entropy, inflectional morphology, prediction, agreement sustained discriminability, grammatical gender, grammatical number, cognitive salience, efficiency in human languages.

1 Introduction

In natural languages, inflectional features such as gender and number express grammatical contrasts; for example, in English, nouns can be specified for two different inflectional number values, namely singular (SG, e.g., *cat*) and plural (PL, e.g., *cats*). Inflectional values are key for the functional operations that link words in sentences, such as tracing subject-verb agreement relations (e.g., *that cat is black vs. those cats are black*). Moreover, inflectional values can encode semantic attributes of the referents, such as numerosity, sex or animacy. The semantic attributes that can be encoded into inflectional features are a very limited set, if compared to the diverse information that can be encoded in derivational morphology, not to mention lexical items. This set is also surprisingly consistent across languages. At least to our knowledge, while no language has dedicated inflectional means to convey information about, for example, colours or odours, grammatical devices to express the numerosity of the referents are found quasi-universally (WALS, Chapter 33A, Dryer, 2013; Dryer and Haspelmath, 2013). Similarly, semantic attributes related to animacy, such as humanness and biological sex, are grammatically encoded at least in one part of the lexicon of any language that marks grammatical gender (Corbett, 1991; Fraser and Corbett, 2000). Besides numerosity and animacy, inflection encodes references to time and durations (that surfaces in verbal tense and aspect), and space (that surfaces in deixis).

Recent studies have highlighted the saliency of such semantic attributes and their close relation to the environmental information that can be processed by the core knowledge systems (Franzon and Zanini, 2019; Franzon et al., 2019, 2020; Golston, 2018; Strickland, 2017). The core knowledge systems are a toolkit of cognitive abilities, phylogenetically ancient and available soon after birth, that enable humans to represent the most salient aspects of the environment, such as the presence of inanimate and animate physical objects, places in the spatial layout and their geometric relationships, and numerosities (Carey, 2004; Spelke, 2000; Vallortigara et al., 2010). Grammatical encoding would enable the possibility to promptly communicate to conspecifics the semantic attributes processed by core knowledge, following the salience of this information from the cognitive and the referential point of view.

Nevertheless, grammatical features cannot be interpreted as unambiguously related to the actual dimensions on which they are based. The grammatical feature of gender is illustrative in this regard. A strict semantic system—in which the gender of a noun can be guessed by the meaning of that noun and vice versa—is surprisingly uncommon from a typological perspective (Corbett, 1991). Many more languages, instead, have predominantly formal systems in which the grammatical gender of the great majority of the nouns is not assigned on a semantic basis (Corbett, 1991). Likewise, number systems do not encode all the numerical information that can be present in the reference, nor do they encode all of the information potentially perceivable by speakers. The core knowledge systems dealing with numerosities and quantities handle distinct numerical representations up to four units and can distinguish different magnitudes (Dehaene, 2011). Nevertheless, most number systems observed cross-linguistically only comprise singular and plural; the rest of the semantic information is seemingly compressed into these two values.

Concerning the broader issue of the relation between language and cognition, one may wonder whether an inflectional system organized around a binary opposition of grammatical values (e.g., singular (SG) vs. plural (PL), and masculine (M) vs. feminine (F)) is the signpost of a binary conceptualization of the world. Notably, in a system that involves the encoding of semantic attributes, some feature values do not occur unambiguously related to a semantic attribute of the reference. For instance, a number value of singular can be used to denote no numerosity, by denying countability as in mass references (e.g., *some milk*), or by referring to a kind (e.g., the singular form *lion* in the sentence *the lion is the king of the savannah* does not imply that there is just one lion)¹. Similarly, in some Indo-European languages (such as Italian) some lexemes can be inflected both in the feminine (*gatte* 'cat(F).PL') or in the masculine (*gatti* 'cat(M).PL'). While the value of feminine can unambiguously refer to female individuals, the value of masculine can denote also non-males. Hence, this ambiguity implies a loss in the encoding/decoding of the referential information. Yet, these phenomena are regularly present in natural languages, and data from language acquisition have shown that their comprehension is enabled since infancy. In a comprehension task involving human referents and nouns marking the semantic feature of sex in gender values, twelve-month-old infants preferably link feminine nouns to female referents, and masculine nouns to both female and male referents. (Ferry et al., 2020). The presence of non-biunivocal morphological encodings in the inflection of natural languages is pervasive across and within languages. Are these patterns better described as noise in otherwise preferably bi-univocal systems, or do they respond to some structural need?

We observe that, despite its cognitive relevance, the encoding of information about semantic attributes in grammatical features is not maximized. Since inflectional systems display non-random similarities across diverse languages, it can be hypothesized that the encoded information follows some common patterns of compression, related to an optimization process in the encoding/transmission of the signals, or in the learning of the code by speakers (Franzon et al., 2020). Efficient communication codes balance between the total amount of encoded information and the costs related to its processing (Ferrer-i Cancho et al., 2020; Piantadosi et al., 2011; Shannon, 1948; Zipf, 1935) and inflectional systems are unlikely an exception. Crucially, it has been reported that inflectional systems attested in natural languages are easier to acquire than unattested combinations of inflectional values (Kirby et al., 2007; Malouf et al., 2015).

In this light, inconsistency in morphological systems may not be seen as a bizarrely frequent idiosyncrasy, since compression can be explained as an optimization property of codes in general, and related to the processing of the linguistic signals in particular. If so, in addition to the semantic side, it is crucial to consider another facet of inflection, whose relevance is central to the construction of the linguistic signal. Gender and number are in fact morphosyntactic features, since their morphological values contribute to trace syntactic relations between words, by sustaining agreement dependencies (Corbett, 2012; Stump, 2015). Agreement is a core property of natural languages, that can be described as the «systematic covariance between a semantic or formal property of one element and a formal property of another» (Steele, 1978, 610) (e.g., *that cat is black*, but not **those cat are black*).

¹Languages differ in the way they encode kind references; in English, the generic interpretation of a kind can also be encoded in plurals (Lazaridou-Chatzigoga et al., 2017).

One of the reasons making agreement so intriguing is that it involves “displaced” information, namely one element (the controller; e.g., *cat* in the example above) determines the form of other elements (the targets; e.g., *that* and *is* in the example above). Since linguistic signals unfold as linear sequences through time (or space), morphosyntactic features can anticipate some information about what words will come next; for instance, hearing a determiner in the singular can signal the presence of a forthcoming related noun in the singular.

A large body of psycho- and neuro-linguistics literature has indeed shown that this displacement of information would favour word prediction during sentence processing by restricting the number of alternatives in the possible upcoming words, thus reducing cognitive effort². Here, we do not go through the many nuances the term prediction implies (e.g., prediction, expectation, anticipation; for a review: Kuperberg and Jaeger, 2016). Though, we observe that anticipatory processing is found in many cognitive domains, and the grammars of human languages unlikely represent an exception to this, as reported in many eye-tracking studies (Altmann and Kamide, 2007; Dutch: Huettig and Janse, 2016; French: Dahan et al., 2000; Spanish: Grüter et al., 2012; Lew-Williams and Fernald, 2007; Russian: Sekerina and Trueswell, 2011; German: Hopp, 2016; for a review: Huettig et al., 2011). Similarly, event-related potential (ERP) studies also observed neural activity modulation prior to the onset of predicted vs. unpredicted words (Wicha et al., 2004). Analogous results have been reported in a spoken language comprehension study involving Dutch participants (Van Petten and Luka, 2012) and neural network simulations (Lakretz et al., 2020).

While psycholinguistic research has highlighted the relevance of inflectional morphology for prediction processes, only few studies have examined whether the structure of languages is suitable to systematically sustain them. Recently, Dye et al. (2017) noted that, in German, grouping beverages like *gin*, *coffee*, *juice* (gender: masculine) on the one side, and *water*, *beer* on the other (gender: neuter) seems pointless from a merely taxonomic viewpoint; nevertheless, once the frequency of occurrence of the nouns is considered, a breakdown by gender efficiently decreases the number of competitors following words that have introduced the semantic field of drinks. More generally, inflectional systems seem indeed to be optimised «for discriminating between alternatives of differing probabilities» (Dye et al., 2017, 228). These observations open a different approach to the questions why inflectional systems are not consistently semantic, and at the same time highlight the relevance of functional aspects for language processing.

The behaviour of linguistic signals, as observed in language corpora, discloses how grammatical aspects of languages respond to the integration of semantic and functional needs, and how they relate to cognitive processes that deal with discrimination and prediction. Why is a one-to-one mapping between a semantic attribute and a morphological value not always recognizable? A first possibility is that a maximally informative, one-to-one encoding is optimal and any instance of non bi-univocal encoding is noise. A certain amount of noise would be tolerated in the system, although dispreferred. According to this “full specification” view, in principle, all inflectional values would be interpreted as

²Some scholars have argued that prediction would instead be a waste of cognitive resources given the large number of nouns that may be suitable for occurring in any context (for a discussion cf. Van Petten and Luka, 2012). However, in the perspective of information theory (since Shannon, 1948), prediction does appear an efficient tool given noise occurring during communication (cf. Kleinschmidt and Jaeger, 2015).

linked to a semantic attribute of the referent. A second possibility is that a compressed encoding is instead preferred in grammatical systems; the related loss of information reflects a structural property of language, derived by a trade-off between the costs and benefits in encoding. According to this "underspecification" account, at least one of the grammatical values in a system would not encode straightforwardly one and only one semantic attribute.

2 The study

This study assesses to what extent the distribution of words in morphological systems respond to the pressures of functional properties, and whether the encoding of the semantic attributes, when present, is fully specified or underspecified. As a case study, we measured the distributions of nouns across inflectional feature values in Italian. The study focuses on nouns, as this part of speech often serves as a controller in agreement (see §1). In what follows, we briefly describe the Italian noun inflectional system (§2.1) and introduce the operational definitions of entropy and agreement-sustained discriminability used for calculation in this work (§2.2). Then, we present the results relating to three different data sets. We first examined the distribution of nouns in the lexicon (§2.3). Second, we selected a subset of animate nouns and explored how the semantic attributes of sex and animacy modulate the distributional space of this system (§2.4). Eventually, for the first time, we measured the semantic specificity of nouns based on the entropy of their contexts of occurrence, to evaluate the impact of a the possible underspecification of the inflectional values on the system (§2.5). The results are discussed in §3; §4 concludes the article.

2.1 Gender and number in Italian

In Italian, nouns occur mandatorily inflected for gender and number, whose values are respectively singular vs. plural, and feminine vs. masculine. Thus, overall, the Italian noun inflectional system comprises four possible combinations of values (i.e. masculine singular, masculine plural, feminine singular, and feminine plural). Number usually encodes the numerosity of the referent, but singulars can also denote mass references (e.g., *poco latte* 'some.M.SG milk(M).SG'; for more details see Franzon et al., 2021; Franzon and Zanini, 2019; Zanini et al., 2017). Gender does not encode semantic attributes in most of the Italian lexicon, since in nouns denoting inanimate entities it merely surfaces as a functional feature (e.g., *sedia* 'chair(F).SG'; *divano* 'couch(M).SG'). However, it can mark the sex of the referent in some animate nouns.

Animacy *per se* is not sufficient to licence the encoding of the sex of the referent (for more details see, among others, Thornton, 2003; Zanini et al., 2020). Some masculine animate nouns do not evidently denote male entities (*struzzo* 'ostrich' is formally masculine, but can refer to both male and female specimens of ostrich; a feminine form **struzza* is unattested). Similarly, some feminine animate nouns do not denote only female entities (*pantera* 'panther' is formally feminine, but can refer to both male and female specimens of panther; a masculine form **pantero* is unattested). Nevertheless, animacy is necessary to licence sex-related oppositions such as *gatto* 'cat(M).SG', *gatta* 'female

cat(F).SG', *gatti* 'cats(M).PL', *gatte* 'female cats(F).PL'³. In this type of nouns, feminines denote only individuals specifying that they are females, whereas masculine singulars (e.g., *gatto*) can be used to denote individuals without strictly specifying their sex, and masculine plurals (e.g., *gatti*) can denote a group comprising also non-male individuals; e.g., in the case there is only one male individual in a group, the assigned value is masculine (Corbett, 1991).

The use of the feature values in the latter kind of nouns could be conceived either as a "full specification" or as an "underspecification", as described in the end of the introduction. Following a "full specification" hypothesis, the gender system is based on an opposition between two antithetic semantic attributes [+female] vs. [+male]. These semantic attributes would respectively be encoded one-to-one in the morphological feminine and masculine values. The overextended use of the gender value of masculine could be interpreted as a dominance of the attribute [+male] over the attribute [+female]. As a consequence, a masculine value would be always interpreted as denoting at least a [+male] individual. Thus, the dominance of this semantic attribute would override the reference to non-males, such as female or non-binary individuals, and individuals whose sex is not known or pertinent. Conversely, according to the underspecification hypothesis, the system is built around only one specified feature, namely the [\pm female] attribute. In this case, the feminine value is used to denote individuals for which it is pertinent to unambiguously specify they are female; the masculine value is used for all the other cases and does not necessarily entail an interpretation of [+male]. The semantic interpretation of the grammatical masculine would be [-female], namely "not specifically female". The masculine feature value would thus host the reference to specifically male individuals, as well as individuals whose sex is not known, not binary, or not pertinent to express, which could not be comprised in the feminine value exclusively encoding [+female].

Similarly, either of these two hypotheses could describe the number system. A full specification account can formalize singular and plural as the morphological realization of two complementary attributes [+one] vs. [+different from one]; an underspecification would describe these values as [\pm different from one] referring to either a numerosity specifically different from one (plural) or to a reference possibly comprising also numerosities of one and to uncountability (singular). The distribution of nouns is a privileged testbed to assess the compliance to functional aspects in inflection, and to assess which of these two accounts is more suitable to describe the relation between a semantic attribute and a grammatical feature in gender and number systems. In what follows we describe the methodology used to collect the measures.

2.2 General methods

In order to efficiently sustain prediction and discrimination, the lexicon should be distributed so that ambiguity is reduced, and competing forms are cut off from the possible upcoming alternatives (Ramscar et al., 2010). Inflectional systems should favour this purpose. A system with two equiprobable values in noun inflection could potentially cut off one half of competitors from the expected word form

³For nouns denoting animate non-human referents, the oppositions likely denote the most frequent configurations of sex. For referents in higher positions in the animacy scales, including humans, attributes pertaining to the social and cultural construction of gender can be encoded as well. Notably, the latter varies across social and cultural groups and likely comprises more than two semantic attributes.

in the average utterance of a noun phrase. Evaluating the Shannon entropy of such a system following Eq.1 ,

$$H = - \sum_x p(x) \log_2 p(x) \quad (1)$$

we find that $H = 1$, corresponding to the maximum for a two-value system. If, in a two-value system, one of the two values occurs with higher probability, the entropy decreases. For instance, if one of the two values has probability 0.9, $H = 0.47$. Defining the self-information of the value x , also known as surprisal ($s(x)$), as in Eq.2

$$s(x) = -\log_2 p(x) \quad (2)$$

we see that it is higher for the less probable value (Shannon, 1948).

The uncertainty reduction will be favoured in the occurrences of the less probable value, but the system overall would not be as discriminative; in 90% of the utterances, only the 10% of the competitors would be cut off from the prediction. Thus, we define the the potential agreement-sustained discriminability ($AggrD$) of the system as the proportion of lexical competitors that are cut off if compared to a system that allows no prediction, as in Eq.3.

$$AggrD = 1 - \frac{1}{2^H} \quad (3)$$

Therefore, the highest discriminability of a morphological system of a language corresponds to the maximum entropy allowed by the number of values available in its grammar. A 2-bit system of four equiprobable values, will be effective in cutting off the 75% of competitor forms in the average utterance. Such a system was used as a reference for this study, as we calculated the entropy of the Italian noun inflectional system based on the distribution of nouns across the four possible combinations of the inflectional values (see §2.1).

In particular, we considered both type frequency (every different type of word form in the database, regardless of its total number of occurrences; e.g., even if the plural form *gatti* 'cats' occurs n times, it is counted only once) and token frequency (the total number of occurrences of a word form in the database; e.g., the plural form *gatti* 'cats' occurs n times). Type frequency is expected to be representative of the potential discriminative properties of language, whereas token frequency is more representative of the state of the language as actually used by speakers. We counted the number of masculine singular, masculine plural, feminine singular, and feminine plural noun types occurring in the lexicon, as listed in the Morph-it! database (Zanchetta and Baroni, 2005). For each noun type, token frequency was collected from the It-Wac corpus (Baroni et al., 2009). The details of the resources are described in §2.3.1. Statistical analysis was performed using R v. 4.0.3; (R Core Team, 2021), packages *kolmim* (v.1.0, Carvalho, 2015) and *LaplacesDemon* (v.16.1.4, Statisticat and LLC., 2020). The dataset used in the study and the reproducible code are available at: https://github.com/franfranz/Morphological_Systems_Entropy.

A vector of 10,000 observations of a uniform probability distribution over four discrete values was

generated to be used as a reference. The reference values are reported in Table 1 - Discrete uniform distribution. We compared the maximum entropy measured in the reference distribution ($H = 2.000$) to the entropy of the observed systems and calculated their Kullback-Leibler divergence (D_{KL}) as a measure of their difference. Given two discrete probability distributions $p(x)$ and $q(x)$, the Kullback-Leibler divergence is defined as:

$$D_{KL} = \sum_x p(x) \log_2 \left(\frac{q(x)}{p(x)} \right) \quad (4)$$

We consider $q(x)$ as the reference and $p(x)$ as the observed probability distributions.

In what follows, we explore to what extent the Italian noun inflectional system complies to functional pressures. A close-to-maximum measured entropy would indicate that the noun inflectional system of Italian is suitable to efficiently support agreement-sustained discrimination and prediction processes. Then, we investigate the distribution of a sample of nouns whose morphological features of gender and number can encode semantic attributes. If the semantic encoding alters the system, we expect to observe differences in the distribution of nouns across the feature values in this sample, compared to the whole lexicon. We also measure the semantic specificity of nouns by measuring the diversity of their contexts of occurrence. We explore to what extent these distributions comply with the possibility that the semantic features are fully specified, or rather underspecified.

2.3 Entropy of Italian nominal inflectional system

2.3.1 Data set: all nouns

A data set including the largest available sample of Italian nouns was obtained by merging two freely available resources. Noun types were collected from Morph-It!, a morphologically annotated lexicon containing approximately 500,000 word forms (Zanchetta and Baroni, 2005). In Morph-it!, the word-form tagging includes the level of the inflectional values, so word forms are also annotated for grammatical gender and for number. Token frequencies of the noun forms listed in Morph-it! were obtained from ItWaC, a 1.9 billion token corpus of texts collected from websites in Italian (Baroni et al., 2009), in which word forms are tagged for part of speech and for lemma. Only the nouns present in both resources were included in our data set. Some of the noun forms can also be homographs to other parts of speech (such as *apparecchio*, that can be both a verb ‘I prepare’ and a noun ‘device’). Here, we considered only the frequency of the words tagged as nouns. In addition, we excluded also homograph forms within the class of nouns that could be ambiguously linked to more than one combination of feature values (e.g., *cameriere* ‘waiter(M).SG’ or ‘waitresses(F).PL’; *fine* ‘aim(M).SG’ or ‘end(F).SG’), obtaining a total of 296,217,261 tokens of 29,126 noun types.

2.3.2 Results: all nouns

The collected measures for all the lexicon are reported in Table 1 - All nouns. The probability distributions of types and tokens across the feature values are plotted in Figure 1. The dashed line in the figure represents the reference distribution. Noun types are quite evenly distributed across

the features, entailing a near maximum entropy for a system entailing four possible combinations of values ($H = 1.976, D_{KL} = 0.016$). When considering token frequency, a slight decrease in entropy ($H = 1.864, D_{KL} = 0.093$) is measured with respect to the one potentially provided by types, mainly related to the higher token frequency of singular forms compared to plural forms.

2.4 Semantically interpreted subsystems

2.4.1 Data set: animate nouns

We selected a subset of animate nouns whose forms are phonologically and orthographically transparent with respect to the feature values they realize. As a result, in this subset, each stem (e.g., *gatt*-‘cat-uninflected’) surfaces in four distinct forms, one for each possible combination of feature values (e.g., *gatta* ‘cat(F).SG’, *gatto* ‘cat(M).SG’, *gatte* ‘cat(F).PL’, *gatti* ‘cat(M).PL’). We selected 90 stems, for a total of 360 types and 4,404,872 tokens.

2.4.2 Results: animate nouns

The distribution of the animate nouns is reported in Table 1 – “Animate nouns”, and plotted in Figure 1. The selection criteria for this sample forcibly lead to a maximum entropy in the distribution of types, being the noun forms uniformly distributed across the four combination of feature values ($H = 2, D_{KL} = 0$). Noteworthy, when token frequency is computed, the entropy decreases ($H = 1.502, D_{KL} = 0.342$), due to the increased token frequency of (singular and plural) masculine forms, if compared to the sample comprising all nouns.

A Wilcoxon rank-sum test was conducted as a pairwise comparison of token frequency within each of the combination of feature values, between the two samples. No significant differences were found for feminine plural forms ($W = 294828, p = 0.069$). Instead, differences in the medians (reported in Table 2) were found significant in feminine singulars, ($W = 320921, p = 0.003$), in masculine plurals ($W = 17306, p < 0.001$), and in masculine singulars ($W = 236693, p < 0.001$), showing that in the animate sample feminine singular forms occur at lower token frequencies whereas masculine forms occur at higher token frequencies (as reported in Table 2 and shown in Figure 2).

2.5 Context entropy

The increased proportion of masculine tokens in the animate sample could be interpreted as a larger presence of [+male] referents, if the full specification account is assumed. Else, it could be the hallmark of underspecification in a system build around the [\pm female] attribute if the underspecification account is assumed (see §2.1). This possibility will give reason to their increased frequency, which would in turn correspond to less self-information of the masculine. Does this correspond also to a less specific interpretation of nouns inflected in the masculine? The higher token frequency of occurrence is not sufficient *per se* in upholding that the masculine value is not linked to a specific interpretation of [+male]. A measure of specificity should also consider the increased diversity of the contexts in which the nouns occur. The basic assumption is that words with a less specific meaning can occur in diverse

contexts, namely they co-occur with many different words, whereas words with a more specific meaning would occur only in few contexts, namely they can co-occur with few diverse words. Research in lexical semantics has indeed highlighted the role of contexts in defining the semantic specificity. Several measures based on corpora have been used to define the extent of words specificity by counting or modeling their possibility to occur in diverse contexts. Context-based methods range from a simple count of number of texts in which a word occurs (Adelman et al., 2006) to more refined and computationally complex diversity indexes based on Latent Semantic Analysis spaces for large corpora as in the case of Semantic Diversity (SemD; Hoffman et al., 2013) or Semantic Distinctiveness (Johns et al., 2016; Jones et al., 2012), and word-centered information-theoretical measures as the Contextual Distinctiveness (McDonald and Shillcock, 2001). Semantic spaces drawn on contextualized word-embeddings (Mikolov et al., 2013; Peters et al., 2018) and more recent models deliver outstanding performances in catching fine grained semantic relations, including hyponymy and polysemy of words (Devlin et al., 2018; Wiedemann et al., 2019).

Since we aim to assess differences in specificity observable at a macroscopic level, we collected a coarse-grained measure that can be easily calculated for any word form in any non-tokenized corpus with a low computational cost: context entropy. We aimed to gain a deeper understanding of the relation between grammatical features and the semantic specificity of inflected forms that realize these features. Two inflected forms of the same lexeme share most of their meaning, except for the semantic attribute encoded by the grammatical feature itself. However, the inflected form with a less specific meaning will co-occur with many diverse word types, whereas the inflected form with a more specific meaning will co-occur with few, less diverse, word types. Context entropy assesses the semantic specificity based on the weighted diversity of lexical types surrounding the target word and is, at least in principle, independent from the total number of times a word occurs.

2.5.1 Methods and data sets: context entropy

In order to obtain the context entropy of a noun, we extracted the word forms occurring in a 20-word-window around each occurrence of the target form (10 words preceding and 10 words following), collected from the ItWaC corpus. All the words were lowercased; punctuation, digits, and function words were removed, leaving only lexical words tagged as nouns, verbs, adverbs, and adjectives. All of the words occurring in the window surrounding each target form were concatenated into a single vector, corresponding to the context of that form. We calculated the entropy on each of the vectors following Eq.1. Eventually, we compared the context entropy across the four combination of feature values in the subset of the animate nouns (see §2.4.1) and in a control subset of nouns selected from the data set containing all nouns (see §2.3.1). In particular, a sample of singular nouns (47 masculine and 47 feminine) was randomly selected; each of these singular forms was matched with its plural form, based on the lemma information available in the Morph-it! list. Similarly, a sample of plural nouns (47 masculine and 47 feminine) was selected and each plural form matched with its singular counterpart. Identical nouns incidentally appearing in both samples were discarded. The resulting sample included 364 noun types. Type and token frequency distributions of the sample are reported in Table 1 - Control Sample.

2.5.2 Results: context entropy

A summary of the context entropy in the animate sample and in the control sample is reported in Table 3; the probability density functions for the distributions of the context entropy are plotted in Figure 3. Due to the divergence of samples from normality, Kolmogorov-Smirnov tests were run to pairwise compare the entropy of each feature between the two samples. No significant differences in context entropy were found in feminine plural forms ($D = 0.187, p = 0.08$) whereas, regarding the nouns of the animate sample, the context entropy was found significantly lower in feminine singular forms, and higher in (singular and plural) masculine forms (feminine singular: $D = 0.245, p = 0.008$; masculine plural: $D = 0.431, p < 0.001$; masculine singular: $D = 0.333, p < 0.001$).

3 Discussion

In this study, we explored the impact of formal properties and semantic attributes on the distribution of nouns across the grammatical features of gender (masculine and feminine) and number (singular and plural) in the Italian inflectional system. By means of entropy metrics, the study aimed to assess whether this system can sustain prediction processes reducing the number of potentially upcoming words through agreement-sustained discriminability. Furthermore, since gender and number can also encode semantic attributes related to biologically relevant aspects of the referential environment (see §1), we measured the impact of their semantic interpretability on the system. Notably, in the literature, topics concerning inflection have mostly been addressed on the paradigmatic side (in the sense of De Saussure, 1922). This is not surprising, since the nature of inflection unfolds primarily along the paradigmatic axis, where a paradigm of a word lists the number of the forms that word can take based on the available feature values (e.g., Stump and Finkel, 2013). This roughly corresponds to our analysis in terms of types. Yet, the patterns displaying on the syntagmatic axis cannot be disregarded. In our study, the analysis of the noun form distribution in connection with other words (i.e. the analysis of tokens in context) have revealed the more complex story we will discuss in what follows.

3.1 Agreement-based discriminability

We compared the entropy of the inflectional system to the maximum entropy of a uniform discrete distribution over four discrete combinations of feature values (i.e. masculine singular, masculine plural, feminine singular, and feminine plural; see §2.3). The close-to-maximum entropy of type distribution in the sample comprising all nouns suggests that Italian lexicon is potentially arranged in a way that allows to systematically reduce uncertainty through a high agreement-sustained discriminability. While type frequency may reflect a potential discriminative power of the language, token frequency is a closer estimation of the concrete realization of such potential. Interestingly, when token frequency is considered, the entropy of the noun inflectional system and its agreement-sustained discriminability decreases. This loss in the self-information of the singular, related to the increased proportion of singular forms with respect to plural forms, can possibly be the signpost of a semantic underspecification effect. Possibly, the singular is more likely than the plural to be used as a default value to underspecify

the referential numerosity while still being available for agreement purposes (see §1, §2.1). In particular, singular is the default value for underdetermine both referential numerosity and countability, as in the case of mass expressions, whereas the value of plural is never used to underdetermine countability. The increased self-information observed for the plural forms in this study corresponds to the higher semantic specificity of plurals compared to singulars, reported in experimental literature (Arcara et al., 2019; Franzon et al., 2016, 2021). The underspecification is further confirmed by the distributions observed in the subset of animate nouns, as discussed in the next session.

3.2 Semantic underspecification

A second experiment was designed to investigate more deeply the role of semantics in a subset of animate nouns with semantically interpretable grammatical gender and number (§2.4). In the selected subset, each of the combinations of feature values entailed the same number of types; thus, its maximum entropy should allow to disambiguate animate male referents from animate female referents and animate singulars from animate plurals. However, in terms of tokens, the entropy of the system drops due to the increased proportion of masculine forms, both in the singular and in the plural. Two alternate explanations can account for these results. A full specification hypothesis would assume that the increased proportion of masculine forms is a consequence of a wider presence of males in the reference denoting males or mixed groups. Conversely, the underspecification hypothesis would assume that the increased proportion of masculine forms is a consequence of the fact that the value of masculine occurs whenever it is not pertinent to denote the reference as comprising specifically and only [+female] individuals. These cases are more frequent not only because they would include reference to males—which can be indeed more frequent in the considered referents—but also because they would denote referents whose sex is unknown or not relevant to the communication. The meaning of the lexical animate nouns inflected in the masculine would be thus unspecified with respect to the sex of the referents. For example, the masculine noun *gatto* 'cat' would mean "small domestic feline" without implying "male small domestic feline". In Italian, a few nouns do encode the attribute [+male] at the lexical level (e.g., *padre* 'father', *toro* 'bull', *verro* 'male pig'), but it is unlikely that [+male] is ever encoded in inflectional morphology.

To further investigate the possibility that the feature values are fully specified or underspecified with respect to the referential sex and numerosity, we collected the context entropy of nouns in a third experiment (§2.5). This measure is in principle independent from the frequency of occurrence of nouns and captures the fact that nouns with a specific meaning can occur in less diverse contexts (low context entropy) when compared to nouns with more general meaning, that can instead occur in more diverse contexts (high context entropy). We measured the context entropy of forms inflected in each of the combinations of feature values in the sample of animate nouns and in a control sample. Whereas a full specification account predicts no differences in the context entropy between the two samples, an increased context entropy would be the hallmark of semantic underspecification of a feature value.

We found that in masculine nouns context entropy is significantly higher in the animate sample compared to the control sample. Hence, the context entropy measures suggest that the grammatical gender opposition feminine vs. masculine in Italian does not encode a [+female] vs. [+male] semantic opposition as supposed in a full specification hypothesis. Rather, it is more economically interpreted following an underspecification account, in which [\pm female] is the only specified attribute.

3.3 The primacy of functional aspects

Taken together, these results suggest a primacy of the functional properties over the semantic attributes in inflectional morphology as the inflected noun forms are mainly shaped to sustain functional purposes such as agreement (see also Pescuma et al., 2021). While agreement in Italian is always mandatory in sentence construction, referential information about numerosity, sex, and animacy is not always available or relevant to communicate, although it is salient from a cognitive perspective and, paradoxically, the basis of the inflectional systems. As a consequence, default values surface whenever agreement is required without being linked to semantic attributes. The pervasiveness of masculine and singular as default values is evident also in other inflected parts of speech. For example, under certain conditions, in Italian, verbs in the past participle must agree in gender and number with the controller (e.g. *queste gatte sono state sfamate* ‘these(F).PL cats(F).PL have been fed-F.PL’). In all other cases, a default agreement takes place and the past participle form surfaces in the masculine singular (e.g., *queste gatte hanno dormito tutto il giorno* ‘these(F).PL cats(F).PL have slept-M.SG all day long’). This value does not encode any semantic attribute. Indeed, the past participle surfaces in masculine singular also when the subject is not linked to any referent as in the case of weather verbs (e.g. *ha piovuto* ‘it has rained-M.SG’).

Our results clearly show that even in cases of animate nouns that allow the encoding of semantic attributes in grammatical features, some feature values (i.e. masculine and singular) surface more frequently than others and are found in more diverse contexts. Likely, this is not or not only a consequence of an alleged disproportion in the referential world (between male and female entities or between singular and plural entities); instead it is due to the semantic underspecification of masculine and singular values, which surface as the default values in the Italian inflectional system⁴. The fact that the grammatical gender and number in Italian noun inflection comprise two values each does not entail that the semantic attributes encoded in this system are *per se* conceived or perceived as binary. In the case of number, input information processed by numerical cognition is quite rich and, certainly, not limited to the recognition of singularity and plurality (see §1). The information encoded into the gender system likely results from the compression of an even wider and blurred set of attributes related to the animacy hierarchy and to cultural features in general (Corbett, 1991; Dahl, 2011; Smith-Stark, 1974). The ways and the reasons driving the compression of semantic information in the observed directions are to be further explored experimentally. Yet, a systematic one-to-one encoding of semantic attributes into grammatical features, which is more demanding at least in terms of memory costs, is

⁴It is likely that, due to historical and societal configurations, human male referents in public roles are a majority; this may be reflected in the higher frequencies of occurrences of masculine forms. Even this case, though, can be economically described within an underspecification account. Indeed, the male referents cannot be denoted as specifically [+female], therefore the available grammatical value is masculine.

dispreferred in natural languages and in communication codes in general. The possibility to talk about a thing without knowing its physics defines one of the key features of human languages. In the words of Borges on Wilkins' essay, «[...] there is no classification of the universe that is not arbitrary and conjectural. The reason is very simple: we don't know what the universe is» (Borges, 1952). There is some degree of arbitrariness in the attribute encoded in each language; nonetheless, the possibility to express any reference is warranted by the compression of information into underspecified values. A similar phenomenon is observable at the lexical level; for example, the noun 'cat' can encode reference to Persian, Siamese, Maine Coon cats as well as to any cat of unknown or non pertinent breed. Our data suggests that a similar relation occurs between inflected forms in grammatical gender and number. The information-theoretic analysis points to the fact that the grammatical gender opposition is economically interpreted as [+female] vs. [not specifically female] from a semantic point of view. A similar pattern is observed in neural network models trained to solve tasks involving grammatical gender and number agreement or coreference resolution. The strategy developed by the models to solve these tasks highlights that a compressed encoding emerges spontaneously in communication games and in the interpretation of gender and number values as specific/generic in neural networks (Jumelet et al., 2019; Lakretz et al., 2020). More in general, this seems to be a property of all discrete communicative systems dealing with semantic categorization, as suggested by studies on language emergence in artificial neural networks (see, for example, Dessì et al., 2019 and Chaabouni et al., 2021 on color-naming systems).

Compression is a general property of codes; underspecification is its particular realization in inflectional systems. Due to underspecification, one of the values in the system is not linked to any semantic attribute and is used as a default for functional purposes, like agreement. Hence, a binary inflectional system does not entail a conceptualization of the referential world along two alleged complementary attributes; rather, in a system comprising two values, one value encodes the specific reference to one attribute, and the other value denotes no attribute. Our measurements suggest that the binary gender system of Italian can be described as "common" vs. "specifically female". Similarly, the binary number system of Italian is more economically described as "common" vs. "specifically plural". The common gender encodes any possible referent as not specifically female, just as the common number could encode referents independently of their numerosity, including the cases of uncountability. Hence, underspecification allows the intrinsic possibility for natural languages to denote any possible referent.

Possibly, the labels "masculine" and "feminine" favored the misleading interpretation of an alleged semantic opposition [+male] vs. [+female] as foundational attributes of the grammatical gender system of Italian. This assumption does not economically describe the data. Rather, it reflects a strong cultural bias present in society, which surfaces in the interpretation of grammatical facts. This long-standing bias derives from the assumption that the possible animate referents are either males or females, and that the semantic attributes underlying grammatical gender values have to adhere to this split. Notably, this type of categorization cannot completely cover the semantic space of genders, and it also finds no correspondence in the actual distributions measured in language. We think that the debate about language inclusiveness can benefit from these measurements. It is desirable that

human society puts all their effort to remove any obstacle to an inclusive society. A data-driven approach helps to overcome the cultural biases that influence the conception of grammatical structures as complicit in promoting non-inclusive societies. Is rethinking grammatical structures a sufficient or at least a necessary step towards inclusiveness? Whatever the case is, measurements are essential for an unbiased interpretation of language functioning. This is crucial to perform an evaluation of the current state of language and to assess language change in response to policies of language inclusiveness.

4 Conclusions

Grammatical features play a crucial role in language, by providing functional cues relevant to discrimination and prediction (e.g., gender and number values), and by encoding semantic attributes of the environment that are salient from a cognitive perspective (e.g., the sex and the numerosity of the referents). We explored the interaction of these factors under an information-theoretic perspective to measure how the structure of inflectional systems responds efficiently to general principles of codes, such as the necessity to reduce ambiguity in the signal as well as in the reference.

For this purpose, we analyzed the Italian noun inflectional system as a case study by means of entropy metrics, including the novel context entropy. Our results clearly show the primacy of functional properties related to the inflectional values. In animate nouns whose inflectional values can encode semantic attributes, some features values are used more than others and found in more diverse contexts. This is the hallmark of underspecification, namely the type of compression taking place in inflectional systems, which enforces at least one of the values to surface as a default for functional purposes. In Italian underspecification involves masculine and singular values, which are not interpretable as [+male] and [+one], respectively. The observed distributions highlight that the attributes underlying the gender and number systems are more economically formalized as common vs. specifically female, and common vs. specifically plural. Indeed, a binary system denotes the presence or absence of one specific attribute, since underspecification of the semantic attributes unlocks the intrinsic possibility for natural languages to denote any possible referent. In this sense, a binary system does not entail a conceptualization of the referential world along two alleged complementary attributes.

We wish that this study can encourage further typological research, to assess universal trends and particular behaviors of inflectional systems in natural languages. It would be interesting to verify whether, crosslinguistically, the semantic specificity of certain feature values is always related to low context entropy values. The employed measures could also be usefully exploited to assess diachronic dynamics in language change, including possible evolution in response to the application of policies for an inclusive use of language involving changes in grammatical gender.

Disclosure Statement

We report no potential conflict of interest.

Acknowledgements

We wish to thank Claudia Artiaco, Davide Crepaldi and Alessandro Treves for their insightful comments and fruitful discussion.

References

- Adelman, J. S., Brown, G. D., and Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological science*, 17(9):814–823.
- Altmann, G. T. and Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of memory and language*, 57(4):502–518.
- Arcara, G., Franzon, F., Gastaldon, S., Brotto, S., Semenza, C., Peressotti, F., and Zanini, C. (2019). One can be some but some cannot be one: Erp correlates of numerosity incongruence are different for singular and plural. *Cortex*, 116:104–121.
- Baroni, M., Bernardini, S., Ferraresi, A., and Zanchetta, E. (2009). The wacky wide web: a collection of very large linguistically processed web-crawled corpora. *Language resources and evaluation*, 43(3):209–226.
- Borges, J. L. (1952). *El idioma analítico de John Wilkins*. Emecé Editorial Buenos Aires.
- Carey, S. (2004). Bootstrapping & the origin of concepts. *Daedalus*, 133(1):59–68.
- Carvalho, L. (2015). An improved evaluation of kolmogorov s distribution. *Journal of Statistical Software*, 65(1):1–8.
- Chaabouni, R., Kharitonov, E., Dupoux, E., and Baroni, M. (2021). Communicating artificial neural networks develop efficient color-naming systems. *Proceedings of the National Academy of Sciences*, 118(12).
- Corbett, G. (1991). *Gender*. Cambridge: Cambridge University Press.
- Corbett, G. (2012). *Features*. Cambridge: Cambridge University Press.
- Dahan, D., Swingle, D., Tanenhaus, M. K., and Magnuson, J. S. (2000). Linguistic gender and spoken-word recognition in french. *Journal of memory and Language*, 42(4):465–480.
- Dahl, Ö. (2011). Animacy and the notion of semantic gender. In *Gender in grammar and cognition*, pages 99–116. De Gruyter Mouton.
- De Saussure, F. (1922). *Cours de linguistique générale*. Otto Harrassowitz Verlag.
- Dehaene, S. (2011). *The number sense: How the mind creates mathematics*. OUP USA.

- Dessi, R., Bouchacourt, D., Crepaldi, D., Baroni, M., et al. (2019). Focus on what's informative and ignore what's not: Communication strategies in a referential game. In *3rd Conference on Neural Information Processing Systems (NeurIPS 2019)*, Vancouver, Canada. Neural Info Process Sys.
- Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Dryer, M. S. (2013). Coding of nominal plurality. In Dryer, M. S. and Haspelmath, M., editors, *The World Atlas of Language Structures Online*. Max Planck Institute for Evolutionary Anthropology, Leipzig.
- Dryer, M. S. and Haspelmath, M., editors (2013). *WALS Online*. Max Planck Institute for Evolutionary Anthropology, Leipzig.
- Dye, M., Milin, P., Futrell, R., and Ramscar, M. (2017). A functional theory of gender paradigms. In *Perspectives on morphological organization*, pages 212–239. Brill.
- Ferrer-i Cancho, R., Bentz, C., and Seguin, C. (2020). Optimal coding and the origins of zipfian laws. *Journal of Quantitative Linguistics*, pages 1–30.
- Ferry, A., Nespor, M., and Mehler, J. (2020). Twelve to 24-month-olds can understand the meaning of morphological regularities in their language. *Developmental psychology*, 56(1):40.
- Franzon, F., Arcara, G., and Zanini, C. (2016). Lexical categories or frequency effects? a feedback from quantitative methods applied to psycholinguistic models in two studies on italian. In *CLiC-it/EVALITA*.
- Franzon, F., Arcara, G., and Zanini, C. (2021). The role of the context and cognition in countability. a psycholinguistic account of lexical distributions. In Pelletier, F. J., Kiss, T., and Husić, H., editors, *The Semantics of Count-Mass Distinction*. Cambridge University Press.
- Franzon, F. and Zanini, C. (2019). Different degrees of abstraction from visual cues in processing concrete nouns. In Bolognesi, M. and Steen, G. J., editors, *Perspectives on Abstract Concepts: Cognition, language and communication*, volume 65, pages 167–184. John Benjamins Publishing Company.
- Franzon, F., Zanini, C., and Rugani, R. (2019). Do non-verbal number systems shape grammar? numerical cognition and number morphology compared. *Mind & Language*, 34(1):37–58.
- Franzon, F., Zanini, C., and Rugani, R. (2020). Cognitive and communicative pressures in the emergence of grammatical structure: A closer look at whether number sense is encoded in privileged ways. *Cognitive Neuropsychology*, 37(5-6):355–358.
- Fraser, N. and Corbett, G. G. (2000). Gender assignment: a typology and a model. In *Systems of Nominal Classification (Language, Culture and Cognition 4)*, pages 293–325. Cambridge University Press.

- Golston, C. (2018). Phi-features in animal cognition. *Biolinguistics*, 12:055–098.
- Grüter, T., Lew-Williams, C., and Fernald, A. (2012). Grammatical gender in L2: A production or a real-time processing problem? *Second Language Research*, 28(2):191–215.
- Hoffman, P., Ralph, M. A. L., and Rogers, T. T. (2013). Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words. *Behavior research methods*, 45(3):718–730.
- Hopp, H. (2016). Learning (not) to predict: Grammatical gender processing in second language acquisition. *Second Language Research*, 32(2):277–307.
- Huettig, F. and Janse, E. (2016). Individual differences in working memory and processing speed predict anticipatory spoken language processing in the visual world. *Language, Cognition and Neuroscience*, 31(1):80–93.
- Huettig, F., Rommers, J., and Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta psychologica*, 137(2):151–171.
- Johns, B. T., Dye, M., and Jones, M. N. (2016). The influence of contextual diversity on word learning. *Psychonomic bulletin & review*, 23(4):1214–1220.
- Jones, M. N., Johns, B. T., and Recchia, G. (2012). The role of semantic diversity in lexical organization. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 66(2):115.
- Jumelet, J., Zuidema, W., and Hupkes, D. (2019). Analysing neural language models: Contextual decomposition reveals default reasoning in number and gender assignment. In *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*, pages 1–11.
- Kirby, S., Dowman, M., and Griffiths, T. L. (2007). Innateness and culture in the evolution of language. *Proceedings of the National Academy of Sciences*, 104(12):5241–5245.
- Kleinschmidt, D. F. and Jaeger, T. F. (2015). Robust speech perception: recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological review*, 122(2):148.
- Kuperberg, G. R. and Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, cognition and neuroscience*, 31(1):32–59.
- Lakretz, Y., Dehaene, S., and King, J.-R. (2020). What limits our capacity to process nested long-range dependencies in sentence comprehension? *Entropy*, 22(4):446.
- Lazaridou-Chatzigoga, D., Stockall, L., and Katsos, N. (2017). A new look at the ‘generic overgeneralisation’ effect. *Inquiry*, pages 1–27.
- Lew-Williams, C. and Fernald, A. (2007). Young children learning spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science*, 18(3):193–198.

- Malouf, R., Ackerman, F., and Seyfarth, S. (2015). Explaining the number hierarchy.
- McDonald, S. A. and Shillcock, R. C. (2001). Rethinking the word frequency effect: The neglected role of distributional information in lexical processing. *Language and Speech*, 44(3):295–322.
- Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.
- Pescuma, V. N., Zanini, C., Crepaldi, D., and Franzon, F. (2021). Form and function: A study on the distribution of the inflectional endings in italian nouns and adjectives. *Frontiers in Psychology*, 12:4422.
- Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., and Zettlemoyer, L. (2018). Deep contextualized word representations. *arXiv preprint arXiv:1802.05365*.
- Piantadosi, S. T., Tily, H., and Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences*, 108(9):3526–3529.
- R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Ramscar, M., Yarlett, D., Dye, M., Denny, K., and Thorpe, K. (2010). The effects of feature-label-order and their implications for symbolic learning. *Cognitive science*, 34(6):909–957.
- Sekerina, I. A. and Trueswell, J. C. (2011). Processing of contrastiveness by heritage russian bilinguals. *Bilingualism: Language and cognition*, 14(3):280–300.
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423.
- Smith-Stark, T. C. (1974). The plurality split. In *Chicago Linguistic Society*, volume 10, pages 657–672.
- Spelke, E. S. (2000). Core knowledge. *American psychologist*, 55(11):1233.
- Statisticat and LLC. (2020). *LaplacesDemon: Complete Environment for Bayesian Inference*. R package version 16.1.4.
- Steele, S. (1978). Word order variation. a typological study. In *Universals of human language*, volume 4: Syntax. Stanford University Press.
- Strickland, B. (2017). Language reflects “core” cognition: A new theory about the origin of cross-linguistic regularities. *Cognitive science*, 41(1):70–101.
- Stump, G. (2015). *Inflectional paradigms: Content and form at the syntax-morphology interface*, volume 149. Cambridge University Press.
- Stump, G. and Finkel, R. A. (2013). *Morphological typology: From word to paradigm*, volume 138. Cambridge University Press.

- Thornton, A. M. (2003). L'assegnazione del genere in italiano. In *Actas del XXIII Congreso Internacional de Lingüística y Filología Románica*, volume 1, pages 467–481. Niemeyer Tübingen.
- Vallortigara, G., Chiandetti, C., Rugani, R., Sovrano, V. A., and Regolin, L. (2010). Animal cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(6):882–893.
- Van Petten, C. and Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and erp components. *International Journal of Psychophysiology*, 83(2):176–190.
- Wicha, N. Y., Moreno, E. M., and Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in spanish sentence reading. *Journal of cognitive neuroscience*, 16(7):1272–1288.
- Wiedemann, G., Remus, S., Chawla, A., and Biemann, C. (2019). Does bert make any sense? interpretable word sense disambiguation with contextualized embeddings. *arXiv preprint arXiv:1909.10430*.
- Zanchetta, E. and Baroni, M. (2005). Morph-it! a free corpus-based morphological resource for the italian language. *Corpus Linguistics*, 1(1):2005.
- Zanini, C., Benavides-Varela, S., Lorusso, R., and Franzon, F. (2017). Mass is more: The conceiving of (un) countability and its encoding into language in 5-year-old-children. *Psychonomic Bulletin & Review*, 24(4):1330–1340.
- Zanini, C., Rugani, R., Giomo, D., Peressotti, F., and Franzon, F. (2020). Effects of animacy on the processing of morphological number: a cognitive inheritance? *Word Structure*, 13(1):22–44.
- Zipf, G. K. (1935). *The psycho-biology of language: An introduction to dynamic philology*, volume 21. Psychology Press.

	M SG	M PL	F SG	F PL	H (bits)	D_{KL}
Discrete uniform distribution (4 values)	2505 <i>0.250</i>	2487 <i>0.249</i>	2561 <i>0.256</i>	2447 <i>0.245</i>	2	0
All nouns – Type	8631 <i>0.296</i>	8565 <i>0.294</i>	6035 <i>0.207</i>	5895 <i>0.202</i>	1.976	0.016
All nouns — Token	110332997 <i>0.372</i>	50603706 <i>0.171</i>	99629938 <i>0.336</i>	35650620 <i>0.120</i>	1.863	0.093
Animate nouns – Type	90 <i>0.250</i>	90 <i>0.250</i>	90 <i>0.250</i>	90 <i>0.250</i>	2	0
Animate nouns – Token	2234461 <i>0.507</i>	1627482 <i>0.369</i>	400700 <i>0.091</i>	142229 <i>0.032</i>	1.502	0.342
Control sample — Type	90 <i>0.247</i>	89 <i>0.245</i>	93 <i>0.255</i>	92 <i>0.253</i>	2	-
Control sample — Token	2100306 <i>0.395</i>	902969 <i>0.170</i>	1607854 <i>0.302</i>	705537 <i>0.133</i>	1.872	-

Table 1: Distribution of nouns across inflectional feature values in the considered sets of nouns (count, *probability*), as compared with a reference distribution.

	Token Frequency (log) - Median, SD			
	M SG	M PL	F SG	F PL
All nouns	6.742 <i>2.097</i>	5.981 <i>2.159</i>	7.402 <i>2.076</i>	6.019 <i>2.286</i>
Animate nouns	8.263 <i>1.822</i>	8.436 <i>2.104</i>	6.737 <i>1.493</i>	5.611 <i>1.632</i>
Control sample	6.826 <i>2.223</i>	6.267 <i>2.470</i>	7.454 <i>1.857</i>	6.020 <i>2.242</i>

Table 2: Token frequency of nouns in the considered set.

	Context entropy (bits) - Median, SD			
	M SG	M PL	F SG	F PL
Control sample	10.157 <i>1.388</i>	9.664 <i>2.065</i>	10.538 <i>1.020</i>	9.686 <i>1.732</i>
Animate nouns	10.918 <i>0.916</i>	10.940 <i>1.081</i>	10.109 <i>0.939</i>	9.098 <i>1.454</i>

Table 3: Context entropy in animate nouns and control sample nouns.

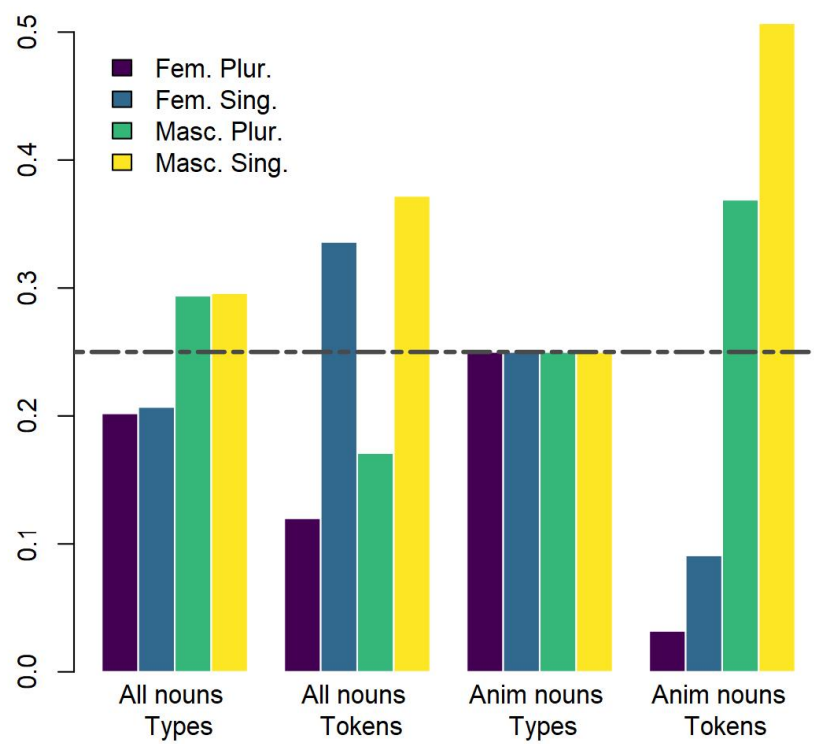
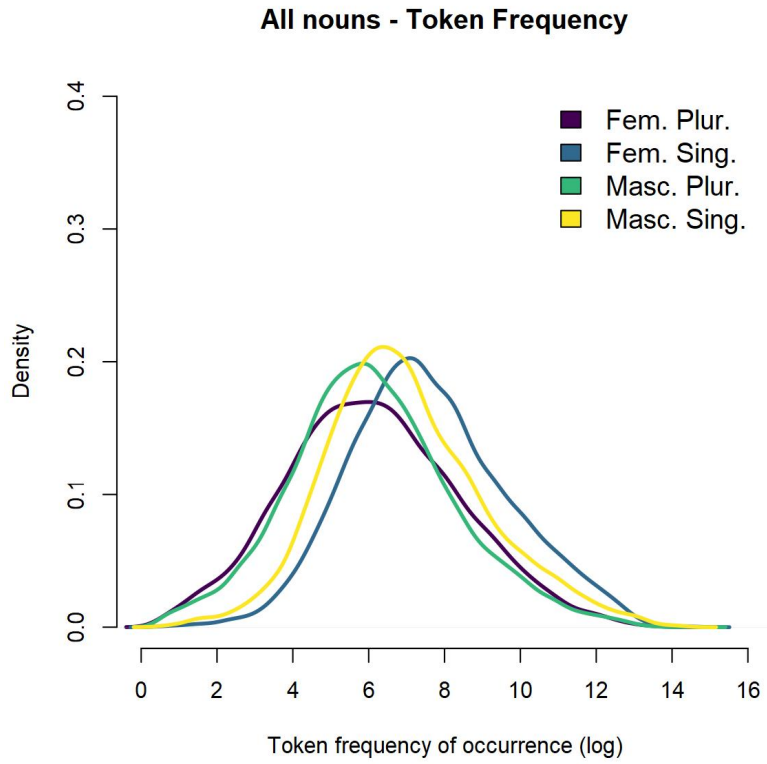
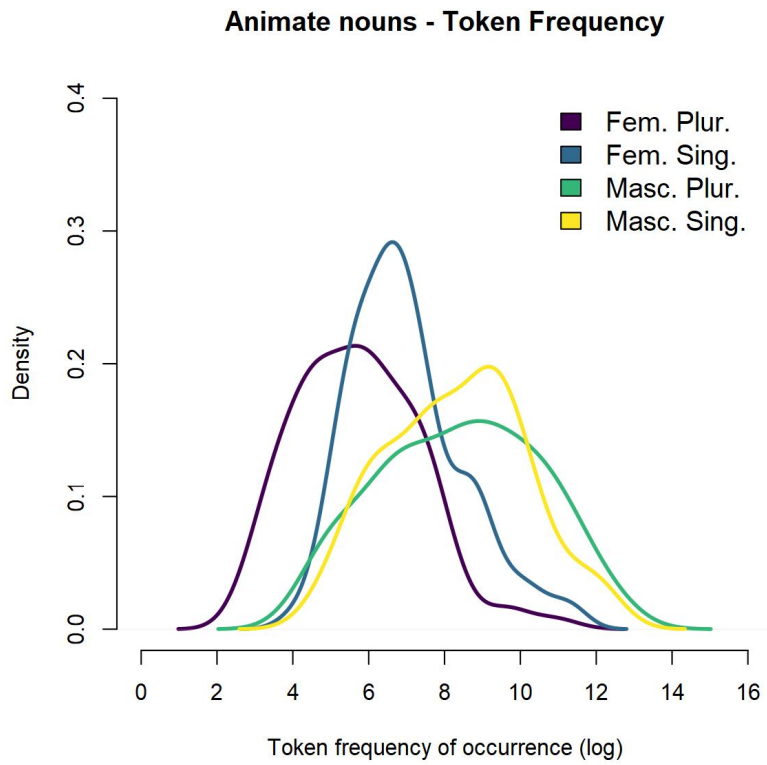


Figure 1: Distribution of nouns across the inflectional features. The dashed line represents the reference distribution.

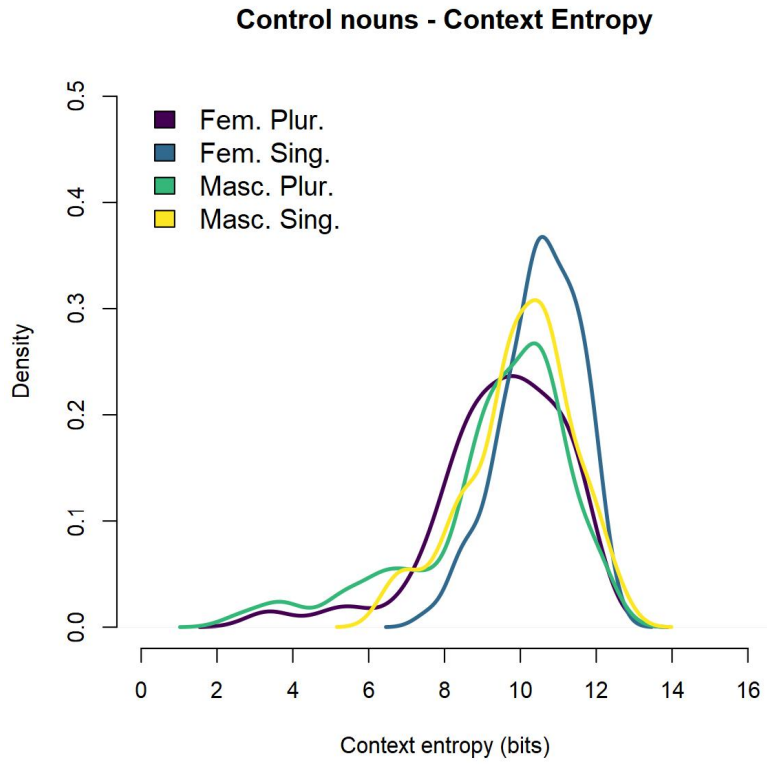


(a) Probability density function of token frequency — All nouns

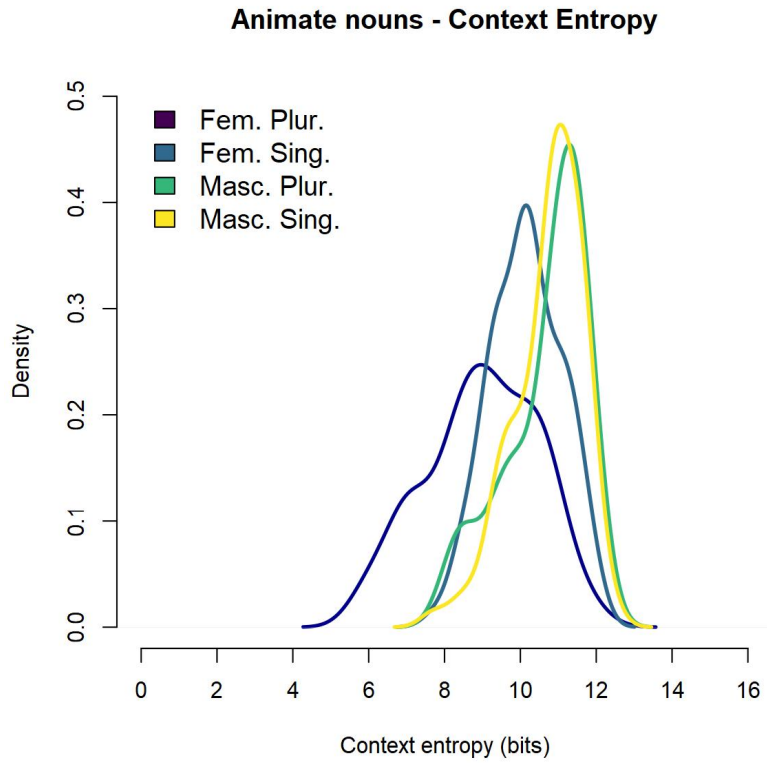


(b) Probability density function of token frequency — Animate nouns

Figure 2: Probability density function of token frequency across inflectional feature values.



(a) Probability density function of Context Entropy – Control Sample



(b) Distribution of tokens across inflectional features — All nouns

Figure 3: Probability density function of Context Entropy across inflectional feature values.