Implicational generalizations in morphological syncretism: the role of communicative biases[1]

BENJAMIN STORME

Université de Lausanne

Cross-linguistic generalizations about grammatical contexts favoring syncretism often have an implicational form. This paper shows that this is expected if (i) morphological paradigms are required to be both as small and as unambiguous as possible, (ii) languages may prioritize these requirements differently, and (iii) probability distributions for grammatical features interacting in syncretic patterns are fixed across languages. More specifically, this approach predicts that grammatical contexts that are less probable or more informative about a target grammatical feature T should favor syncretism of T cross-linguistically. The paper provides evidence for these predictions based on four detailed case studies involving well-known patterns of contextual syncretism (gender syncretism based on number, gender syncretism based on person, aspect syncretism based on tense, and case syncretism based on animacy).

Keywords: morphology, syncretism, typology, communicative efficiency, Harmonic Grammar, gender, number, person, case, aspect, tense

1. Introduction

Across languages, morphology often fails to mark grammatically relevant distinctions in some contexts. For instance, Danish marks the gender of third person pronouns in the singular, as shown in (1a), but fails to do so in the plural, as shown in (1b). In the plural, a single form is used, whether the referent is a group containing only female individuals, male individuals or both.

[1] I thank Edward Flemming for his invaluable feedback on an earlier version of this work. I also thank audience members at AIMM3 in Amherst, MA in 2015, at LAGB in London in 2015, and at the Linglunch Paris Diderot in 2018. Finally, I thank the anonymous reviewers for the *Journal of Linguistics* and for *Language* for helping me improve the quality of this paper.

(1) Gender syncretism in Danish pronouns

(a) hun ‘she’, han ‘he’

(b) de ‘they’

Cases where morphological paradigms are less rich than what the free combination of grammatical features available in a language would lead us to expect are known as cases of syncretism (e.g. Greenberg 1966: 27; Baerman et al. 2005). For instance, gender is syncretic in plural pronouns but not in singular pronouns in Danish. Interestingly, the range of attested syncretic patterns is more restricted than what is logically possible: across languages, some grammatical contexts favor the syncretic expression of other grammatical features and cross-linguistic generalizations about patterns of syncretism often have an implicational form (Greenberg 1963, 1966). For instance, gender is often syncretic in plural pronouns but less so in singular pronouns and, cross-linguistically, the presence of gender distinctions in the plural asymmetrically entails the presence of gender distinctions in the singular (cf. Greenberg’s 1963 Universal 45).

What is the source of syncretism? Why do some grammatical contexts favor syncretism of other grammatical features and why do generalizations in this domain have an implicational form? Since Greenberg’s (1963) seminal work, several concurrent analyses have been proposed to answer these questions, ranging from frequency-based explanations (e.g. Greenberg 1966, Croft 1990, Jäger 2007, Haspelmath & Sims 2010) to structural explanations based on Universal Grammar (e.g. Harley & Ritter 2002). This paper adopts a specific version of the frequency-based approach, where morphological patterns are shaped by communicative biases towards accurate message transmission and low resource cost (e.g. Martinet 1962, Jäger 2007, Piantadosi et al. 2012, Gibson et al. 2019). Under this view, syncretism arises when the resource cost of expressing a grammatical distinction morphologically in a context is not compensated by a large enough gain in decoding accuracy. Language variation results from different ways of resolving the conflict between minimizing resource cost on the speaker’s part (which favors small morphological paradigms) and maximizing decoding

accuracy on the listener's part (which favors morphological paradigms that are as large as allowed by the free combination of available grammatical features). The interaction of these two conflicting goals takes place in a synchronic model of the speaker's morphological productions using weighted constraints (Smolensky & Legendre 2006). These constraint-based models are widely used in phonology to derive implicational generalizations in sound patterns.

Section 2 shows how this approach together with the assumption that grammatical features are drawn from the same probability distributions across languages predicts that cross-linguistic generalizations about morphological syncretism should have an implicational form. More specifically, this approach predicts that grammatical contexts that are less probable or more informative about a target grammatical feature *T* should favor syncretism of *T* cross-linguistically. Section 3 shows how the theory derives the typology of syncretism in three well-known case studies: gender syncretism based on number, gender syncretism based on person, and aspect syncretism based on tense. Section 4 extends the model introduced in Section 2 to deal with more complex cases where the cost of syncretism is not independent from context to context. The extended model is applied successfully to another well-known case study: case syncretism based on animacy.

The four generalizations addressed in Sections 3 and 4 are, to the author's knowledge, among the most discussed implicational generalizations in morphological syncretism. They therefore constitute a plausible set of cases against which models of syncretism should be evaluated. Section 5 discusses alternative analyses that do not rely on synchronic (possibly implicit) communicative biases and shows that they all fail on at least one of the four case studies. In particular, the current approach is compared with the closely related frequency-based approach where morphological asymmetries arise diachronically in the transmission from one generation to the next, but crucially without any specific bias towards communicatively efficient patterns from learners or speakers (e.g. Haspelmath & Sims 2010: chapter 12). In this approach, syncretism arises when one of the values

of a grammatical feature is rare in a grammatical context in speakers' productions and therefore the corresponding morph fails to be correctly learned by learners of the language (Haspelmath & Sims 2010: 272-273). The paper shows that the two approaches make different predictions when the recoverability of a grammatical feature is high in a given grammatical context but the values of the relevant feature all occur frequently in this context: the synchronic model predicts that syncretism should be likely in this case, due to high recoverability, whereas the diachronic model predicts that syncretism should be unlikely, due to high frequency. The case study on gender syncretism based on person discussed in Section 3 will be shown to support the model assuming synchronic communicative biases on the part of learners or speakers.

The idea that communicative efficiency plays a role in shaping linguistic patterns across languages, and in particular morphological patterns, is not new (e.g. Martinet 1962, Hawkins 2004, Jäger 2007, Piantadosi et al. 2012, Gibson et al. 2019, Haspelmath 2021). In particular, a growing body of experimental evidence has accumulated that speakers make communicatively efficient choices when producing morphology (e.g. Kurumada & Jaeger 2005) and that learners may reshape their morphological input in order to improve its efficiency (e.g. Fedzechkina et al. 2012). The specific contribution this paper makes to this research theme is threefold. First, it focuses specifically on the question of whether a feature is expressed syncretically or not in a given context (i.e. whether the different values of that feature are expressed ambiguously or not). This question differs from the question that is more often addressed in the literature on morphological marking across languages (e.g. Haspelmath 2021), namely the question of how a given feature value is expressed morphologically (through overt or zero marking). Second, the paper includes a comparison of different approaches to syncretism: all four approaches considered here are shown to make different predictions on the four case studies in the paper, a fact that has not been noted before, to the author's knowledge. Finally, the analysis is couched in a grammatical model familiar to linguists, and in particular to phonologists,

namely a model with weighted constraints. Adopting a common framework to account for phonology and morphology is a step towards a better understanding of the similarity between morphological syncretism and phonological neutralization (Greenberg 1966: 29; Martinet 1968), as will be discussed in Section 6.

2. MODEL

This section proposes a model of the syntax-morphology interface where the mapping from grammatical features to morphs is regulated by two constraints: a constraint that aims to minimize misinterpretation of the morphology on the part of the hearer and a constraint that aims to minimize the size of morphological paradigms. It is not possible to satisfy completely both constraints as they correspond to contradictory demands: minimizing ambiguity is only possible at the cost of making the morphology more complex. Language variation comes from different ways of resolving the conflict between universal but contradictory demands.

The model focuses on the simplest case where a binary grammatical feature T with two exhaustive and mutually exclusive feature values t_1 and t_2 (e.g. sex-based gender, with feature values 'feminine' and 'masculine') is targeted by syncretism in a grammatical context c_i (belonging to a grammatical feature C; e.g. 'plural' belonging to the feature 'number'). There are cases where non-binary grammatical features are subject to syncretism. For instance, Latin does not distinguish masculine, feminine, and neuter nouns in the dative and ablative plural (e.g. Latin *-is* is used to mark the ablative plural of masculine, feminine, and neuter nouns, as in *domin-is* 'master'-MASC.ABL.PL, *ros-is* 'rose'-FEM.ABL.PL, and *uerb-is* 'word'-NEUT.ABL.PL). In this case, syncretism therefore targets a ternary feature. However these cases are relatively marginal in the literature on implicational generalizations: in particular, the four cases studies treated in this paper all involve binary grammatical features. As a consequence, cases of syncretism targeting non-binary features will be left aside here. There are also cases of syncretism that do not involve a single grammatical feature but

combinations of grammatical features. For instance, in Latin, the same morph *-ae* conveys the genitive singular and the nominative plural of feminine nouns (e.g. *famili-ae* 'family'-GEN.SG/NOM.PL). The two combinations of feature values (genitive singular vs. nominative plural) do not form a minimal pair: both case and number vary. These patterns are often treated as involving accidental or arbitrary homophony (see Baerman et al. 2005 on the distinction between accidental and systematic syncretism). Accidental syncretism is not prominent in the literature on implicational generalizations and therefore will be left aside as well.

Technically, not expressing a grammatically relevant distinction morphologically in a particular grammatical context incurs a cost, called the 'ambiguity cost' of a paradigm (see Section 2.1). Expressing this grammatical distinction morphologically also incurs a cost, called the 'size cost' of a paradigm (see Section 2.2). The size of a paradigm (i.e. the number of distinct cells in this paradigm) is used as a proxy for resource cost: smaller paradigms should be easier to store, process and produce. Morphological paradigms are evaluated with respect to a weighted sum of their ambiguity and size costs (see Section 2.3), as in Harmonic Grammar (HG; Smolensky & Legendre 2006). These costs contain (i) a language-universal component, namely the definitions of the costs, and (ii) a language-specific component, represented by language-specific, positive weights associated with each constraint. Different choices of weights will result in different trade-offs between clarity and morphological complexity. Section 2.4 describes how this model predicts the existence of implicational generalizations in patterns of morphological syncretism.

2.1. Ambiguity cost

Formally, the ambiguity cost of mapping values t_1 and t_2 of a binary grammatical feature T to the same morph m in a grammatical context c_i is assumed to be proportional to the probability that the hearer misinterprets m in c_i. The hearer misinterprets m when they interpret a speaker uttering m and intending t_1 as intending t_2 or the other way around. Note that the model implies that the hearer

always assumes that the speaker meant either t_1 or t_2 when using a syncretic form to convey T. In other words, the model assumes that syncretic morphology is semantically ambiguous (m means t_1 or means t_2) rather than semantically underspecified (m denotes the disjunction of the two feature values t_1 and t_2). This assumption will be further discussed at the end of this section.

The probability of error is noted as $P(error_m|c_i)$: it corresponds to the conditional probability that the hearer misinterprets m given that the grammatical context is c_i. Sections 2 and 3 focus on cases where this probability can be assumed to be independent of whether T is syncretic in another context c_j. Section 4 will extend the model to cases where the independence assumption does not hold.

The hearer is assumed to use the strategy that minimizes the probability that they will make an error: they default to the most likely feature value among t_1 and t_2 in c_i in case of syncretism. Because t_1 and t_2 are the only values available for the grammatical feature T (because T is binary by assumption) and cannot co-occur in c_i (because t_1 and t_2 are mutually exclusive by assumption), the two conditional probabilities $P(t_1|c_i)$ and $P(t_2|c_i)$ sum to one. As a consequence, the probability of error in c_i will be equal to the probability of the least likely feature value in c_i (Equation 1).

$$P(error_m|c_i) = \min\{P(t_1|c_i), P(t_2|c_i)\} \quad (1)$$

This approach predicts that, in case of syncretism, the probability of error will be smaller in contexts where the two feature values are more imbalanced. For instance, in a context c_i where the two feature values have imbalanced probabilities, e.g. $P(t_1|c_i) = 0.9$ and $P(t_2|c_i) = 0.1$, the probability of incorrectly identifying the feature value intended by the speaker is rather small (here it is equal to 0.1). In a context c_j where the probabilities are closer to 0.5, e.g. $P(t_1|c_j) = 0.6$ and $P(t_2|c_j) = 0.4$, the probability of error is larger (here it is equal to 0.4). As one of the two feature values becomes less likely relative to the other

one, the probability of error in case of syncretism decreases.[1]

To obtain the contribution of syncretic morphology m in context c_i to the overall probability of error, the conditional probability $P(error_m|c_i)$ is multiplied by the probability of context c_i, $P(c_i)$. This captures the following intuition: the more frequent the context conditioning the syncretic expression of a grammatical distinction, the larger the number of errors on the part of the listener.

The ambiguity cost of mapping t_1 and t_2 to two different morphs m_1 and m_2 is assumed to be equal to zero. If the two feature values are distinguished morphologically (and assuming that the likelihood of misidentifying two phonetically distinct morphs is negligible), a hearer can deterministically recover the meaning intended by the speaker.

To account for the fact that languages show different degrees of syncretism, languages are further assumed to vary in the importance attributed to minimizing misinterpretation. This is implemented by multiplying the ambiguity cost defined above by a language-specific, positive weight w_A, where A stands for ambiguity. The weight on the ambiguity cost is also assumed to be specific to a particular grammatical feature: a language therefore has potentially different weights for each grammatical feature (but this weight is the same for a given feature across contexts). This allows a language to adopt different morphological strategies for different grammatical features.

In sum, the ambiguity cost for the syncretic and non-syncretic expressions of two feature values t_1 and t_2 in a context c_i and a language with ambiguity weight w_A for that feature can be written as in Equation 2.

---

[1] As noted by a reviewer, Shannon's entropy could be used instead of the probability of error. The predictions would be very similar. The entropy of a binary random variable increases as the probability of success gets closer to one half (MacKay 2003: 2). In other words, the uncertainty about whether the syncretic morph means t_1 or t_2 increases as the probabilities get closer to one half.

$$\begin{cases} cost_A = 0 & \text{if } t_1 \text{ and } t_2 \text{ are not syncretic in } c_i \\ cost_A = \min\{P(t_1|c_i), P(t_2|c_i)\}P(c_i)w_A & \text{if } t_1 \text{ and } t_2 \text{ are syncretic in } c_i \end{cases} \quad (2)$$

As pointed out at the beginning of this section, the model assumes that a syncretic morph is semantically ambiguous rather than semantically underspecified. However, the model could probably be restated in terms of underspecification. In particular, syncretism should lead to more uncertainty about the speaker's message under the syncretism-as-underspecification view as well. For instance, if a speaker uses a morph conveying an underspecified predicate including both animate and inanimate referents in its denotation to refer to an animate referent, the hearer will be more likely to make an incorrect identification of the intended referent than if the speaker uses a morph strictly denoting the set of animate referents. Indeed, the set of animates is a strict subset of the set of animates and inanimates. The syncretism-as-ambiguity view was adopted here mainly for practical purposes. It is easier to represent the choice between two alternative features (for instance animate and inanimate) than between multiple referents (in the case of pronominal gender for instance). Also, under the ambiguity view, one can easily obtain information on the frequencies of features using corpus data and then infer the ambiguity cost of syncretism using Equation 2.

2.2. Size cost

For the same underlying feature system, a paradigm with syncretism includes a smaller number of distinct word-forms than a morphological paradigm without syncretism. Syncretism therefore allows for a reduction of the size of the mental lexicon. For instance, the paradigm of English animate third person pronouns features gender syncretism in the plural and contains three distinct word-forms (*he*, *she*, *they*). The corresponding paradigm in French maintains a gender distinction across both numbers but does so at the cost of increasing the size of the paradigm: the French paradigm contains four instead of three distinct word-forms

(*il*, *elle*, *ils*, *elles*).[2]

Several benefits follow from having a smaller lexicon in general. A smaller lexicon provides an obvious advantage in terms of storage. As Croft (1990: 254) puts it, 'minimizing the number of distinct linguistic forms that must be acquired and retained presumably minimizes the load on memory.' A smaller lexicon also provides benefits in terms of production and processing cost. Indeed, if there are fewer words in a lexicon, these words are allowed to be shorter, more frequent, and more probable phonotactically, making production and processing more efficient (e.g. Jaeger & Tily 2011, Piantadosi et al. 2012).[3] For practical purposes, the size of the paradigm will be used as a proxy for storage, production, and processing costs in the remainder of the paper.

The size cost of having two distinct morphs to express two feature values is assumed to be equal to one for any context. The size cost of having a single morph to express the two feature values is assumed to be equal to zero for any context. The probability of the context is not assumed to play a role in this cost for the following reason: although contexts that are more frequent should increase the overall processing cost of a morphological distinction as compared to less frequent contexts (in the same way as ambiguity in more frequent contexts

---

[2] As noted by a reviewer, word-forms in a paradigm are here assumed to be stored holistically. Indeed syncretism would not necessarily minimize the size of the mental lexicon if the mental lexicon only consisted of morphemes: for instance, in terms of morphemes, the non-syncretic paradigm of French pronouns contains as many units (*il*, *elle*, *-s*) as the partially syncretic paradigm of English pronouns (*he*, *she*, *they*). Evidence for the storage of morphologically complex word-forms comes from studies showing that the frequency of a complex word is predictive for processing latencies, independently of the frequencies of its constituents. Crucially, the word frequency effect was found not only for irregular complex words but also for highly regular complex words such as English plurals in *-s* (see Baayen 2007 for a review). The view that word-forms in a paradigm are stored does not exclude the possibility that individual morphemes are also stored. For instance, in dual access models of morphology, complex words can be accessed both directly or through morpheme decomposition.

[3] A reviewer agues that syncretism should reduce production cost when it involves zero marking but not when it involves using the same overt marking across multiple features (e.g. Latin *uerb-um* 'word'-NOM/ACC.NEUT.SG, where the same overt marker *-um* is used across nominative and accusative cases). However, regardless of the type of marking, syncretism results in a smaller number of word-forms and this fact alone should allow in principle for a number of benefits such as shorter word-forms, more probable phonotactics, etc. Also, a single overt marker that does not compete with any other marker to express a grammatical feature can be subject to reduction processes without any risk of confusion. This is not the case when two markers compete: reduction processes can be blocked in order to maintain the two markers sufficiently distinct acoustically, as attested in patterns of paradigmatic contrast (Kenstowicz 2005).

resulted in a larger ambiguity cost), more frequent forms are also easier to process, according to the well-known word frequency effect (Brysbaert et al. 2018). A simple way to accommodate these contradictory effects is to assume that they cancel each other.

Languages are further assumed to vary in the importance they attribute to minimizing the size of morphological paradigms. This is implemented by multiplying the size cost in syncretic and non-syncretic contexts by a language-specific, positive weight w_S, where S stands for size. As for the ambiguity cost, the weight on the size cost is assumed to be specific to a particular grammatical feature and constant for all contexts in which this feature occurs.

In sum, the size cost for syncretic and non-syncretic expressions of two feature values t_1 and t_2 in a context c_i and a language with weight w_S for that particular feature can be written as in Equation 3.

$$\begin{cases} cost_S = w_S & \text{if } t_1 \text{ and } t_2 \text{ are not syncretic in } c_i \\ cost_S = 0 & \text{if } t_1 \text{ and } t_2 \text{ are syncretic in } c_i \end{cases} \tag{3}$$

2.3. Balancing ambiguity and size

The preference for syncretism or no syncretism depends on which of the ambiguity cost or size cost is larger. It is equally good to have syncretism or no syncretism in a given context if the two costs as defined in Equations 2 and 3 are equal, that is if Equation 4 holds.

$$w_S = \min\{P(t_1|c_i), P(t_2|c_i)\}P(c_i)w_A \tag{4}$$

Assuming trivially that $P(c_i)$ is non-null, Equation 4 is equivalent to Equation 5.

$$\min\{P(t_1|c_i), P(t_2|c_i)\} = \frac{w_S}{w_A} \times \frac{1}{P(c_i)} \tag{5}$$

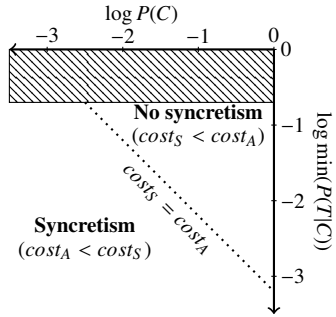


Figure 1: Morphological expression (syncretism vs. no syncretism) of a target grammatical (binary) feature T depending on the grammatical context C where it occurs (with arbitrarily chosen weights for the ambiguity and size costs).

Assuming trivially that all probabilities are non-null, it is possible take the natural logarithm of both sides of Equation 5 to obtain Equation 6. The logarithmic transformation makes it possible to express $\log \min\{P(t_1|c_i), P(t_2|c_i)\}$ as a linear function of $\log P(c_i)$.

$$\log(\min\{P(t_1|c_i), P(t_2|c_i)\}) = -\log P(c_i) + (\log w_S - \log w_A) \tag{6}$$

Equation 6 is the equation of a line with slope -1 and intercept $\log w_S - \log w_A$. This line is graphically represented in Figure 1, for arbitrarily chosen weights. $P(c_i)$ takes values between 0 and 1. Therefore $\log P(c_i)$ takes values between $-\infty$ and 0. The probability $\min\{P(t_1|c_i), P(t_2|c_i)\}$ takes values between 0 and 0.5. It cannot be larger than 0.5 because it is defined as the smaller of two probabilities summing to one. Therefore $\log(\min\{P(t_1|c_i), P(t_2|c_i)\})$ takes values between $-\infty$ and $\log(0.5) \approx -0.7$. Accordingly, the area above $y = -0.7$ is grayed out in Figure 1.

The dotted line depicted in Figure 1 corresponds to values of $P(c_i)$ and $\min\{P(t_1|c_i), P(t_2|c_i)\}$ for which the syncretic and non-syncretic expressions of t_1 and t_2 in context c_i have the same cost. The region under the line corresponds to contexts where the cost of adding a morphological distinction outweighs the benefit in terms of disambiguation, namely where syncretism is enforced. The

region above the line corresponds to contexts where the interpretative benefit of adding a morphological distinction outweighs the cost in terms of size, namely where syncretism is banned.

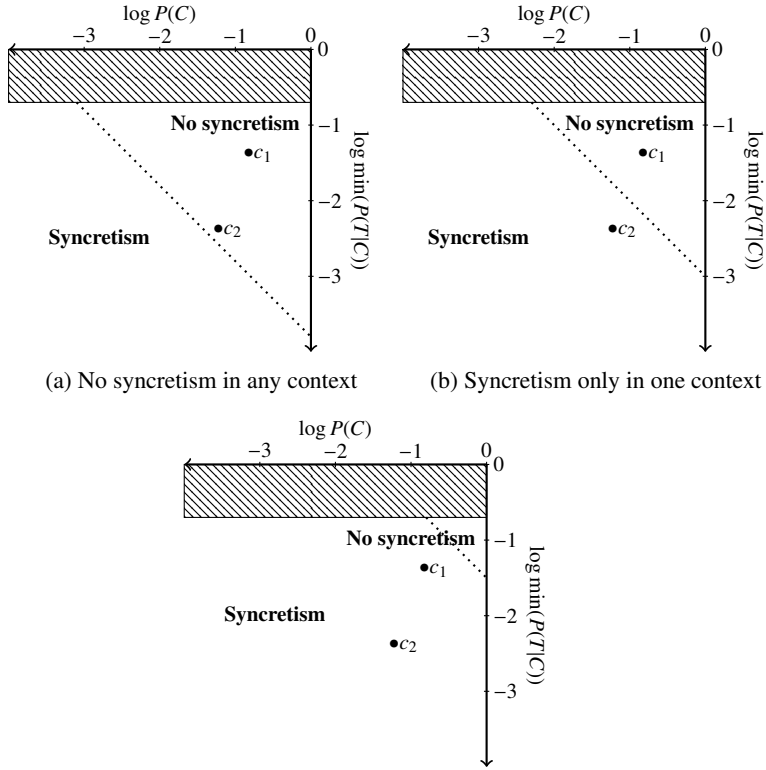
The intercept varies as a function of w_A and w_S: the larger w_A is with respect to w_S, the smaller the intercept $\log w_S - \log w_A$ and the smaller the set of contexts allowing for syncretism of T. This captures the desired effect: as it becomes more important to increase decoding accuracy, morphological distinctions are made in a larger set of contexts (and therefore morphological paradigms get larger). However, the slope does not vary as a function of the weights and is therefore constant and equal to -1 across languages.

2.4. Deriving implicational generalizations

Analyses using frequency asymmetries to explain typological generalizations about linguistic patterns hypothesize that these frequency asymmetries are universal (Greenberg 1966, Croft 1990, Jäger 2007). This paper follows this tradition. In the specific model discussed here, the hypothesis of universal frequency asymmetries is formalized by positing that, for all grammatical features T and C, the joint probability distribution of the two features, $P(T, C)$, is fixed across languages. This hypothesis will remain an assumption throughout this work.

With this assumption in place, the main locus of typological variation predicted by the model lies in the weights w_A and w_S. Because the slope of the line separating syncretic and non-syncretic contexts in the probability space is constant across languages (see Section 2.4), the following prediction is made: the presence of syncretism in some contexts asymmetrically entails the presence of syncretism in other contexts.

This is illustrated in Figure 2, for two grammatical contexts c_1 and c_2. Among the four logically possible ways of expressing the distinction between t_1 and t_2 in c_1 and c_2, only three are predicted to be attested: the pattern without syncretism in any of the two contexts (Figure 2a), the pattern with syncretism only in c_2 (Figure 2b) and the pattern with syncretism across the two contexts (Figure 2c).



(a) No syncretism in any context

(b) Syncretism only in one context

(c) Syncretism in both contexts

Figure 2: Deriving an implicational generalization in morphological syncretism.

The pattern with syncretism only in c_1 is predicted to be impossible: there is no line with slope equal to -1 that is above c_1 and under c_2. In other words, the following implicational generalization is derived: syncretism in context c_1 entails syncretism in context c_2 (or equivalently, a morphological distinction in c_2 entails a morphological distinction in c_1).

The model derives implicational generalizations in the typology of morphological syncretism and predicts furthermore that contexts that lead to fewer identification errors of T in case of syncretism should favor syncretism of T. Indeed, if a morphological distinction can be made only in one of two contexts, the model predicts that it will be in the context where the probability of error is

the smaller. Indeed, the size cost of a morphological distinction is not context-dependent (see Section 2.2 for justification). Therefore, the choice to neutralize a morphological distinction in a context c_i or a context c_j only depends on the ambiguity cost of syncretic morphology in the two contexts: whichever context corresponds to the larger ambiguity cost for the syncretic expression of T should be more likely to distinguish t_1 and t_2 morphologically. Because the weight on the ambiguity cost relative to T is assumed to be the same for the two contexts in a given language, which context will favor syncretism will ultimately depend on the relationship between the probabilities of misidentifying T in the two contexts. Context c_i will favor syncretism of T and context c_j will favor the non-syncretic expression of T in the language if the probability of misidentifying T is smaller in c_i than in c_j, that is if the following inequality holds:

$$\min\{P(t_1|c_i), P(t_2|c_i)\}P(c_i) < \min\{P(t_1|c_j), P(t_2|c_j)\}P(c_j) \tag{7}$$

or equivalently:

$$\min\{P(t_1, c_i), P(t_2, c_i)\} < \min\{P(t_1, c_j), P(t_2, c_j)\}) \tag{8}$$

As shown in Equation 7, two probabilities contribute to the overall probabilities of error shown in Equation 8: the probability of the grammatical context, $P(C)$, and the conditional probability distribution of the target grammatical feature given that context, $P(T|C)$. The two paragraphs below explain how these two probabilities should affect the morphological expression of T, pointing to places in the literature where the role of these probabilities has been discussed in connection with morphological syncretism.

Probability of the context. Everything else being equal, syncretism in a context with a larger probability should entail syncretism in a context with a smaller probability (see the x-axis in Figure 2). The reason is that, everything else being equal, syncretism in a context that is more probable will result in more errors on the part of the listener than syncretism in a context that is less probable. If

a speaker is willing to neutralize a morphological distinction in a context where it is more helpful to the listener, then they should also be willing to neutralize it in a context where it is less helpful to the listener. The hypothesis that a grammatical context that is less probable is more likely to favor syncretism of another grammatical feature is well known in the literature (e.g. Croft 1990: 72, 158; Haspelmath & Sims 2010: chapter 12) and dates back at least to Greenberg (1966).

Conditional probability of the target grammatical feature. Everything else being equal, syncretism in a context c_1 where the probability distribution $P(T|c_1)$ is less skewed towards one of the two feature values should entail syncretism in a context c_2 where $P(T|c_2)$ is more skewed (see the y-axis in Figure 2; $\min P(T|c_1)$ is closer to 0.5 than $\min P(T|c_2)$, and therefore $\log \min P(T|c_1)$ is closer to -0.7 than $\log \min P(T|c_2)$). The reason is that defaulting to the most likely interpretation should result in more identification errors on the hearer's part in case the conditional probability distribution is less skewed (e.g. in c_1 than in c_2). If a speaker is willing to neutralize a morphological distinction in a context where this distinction would be more helpful to the listener (e.g. c_1 in the example in Figure 2), they should also be willing to neutralize it in a context where it would be less helpful to the listener (e.g. c_2 in the example in Figure 2). This prediction can be restated in information-theoretic terms: if there is syncretism in a context that is less informative about a grammatical feature T (e.g. c_1 in the example in Figure 2), there should be syncretism in a context that is more informative about T (e.g. c_2 in the example in Figure 2).

The latter prediction does not seem to have received as much emphasis in the literature on syncretism as the former one. For instance, Haspelmath and Sims' (2010: chapter 12) morphology textbook only highlights the probability of the context. In the present approach, the conditional probability of the target grammatical feature is crucial because it reflects how informative the context is about that feature. The relevance of conditional probability to linguistic patterns has been emphasized more generally in information-theoretic accounts of lexical

ambiguity (Piantadosi et al. 2012).

3. CASE STUDIES

The approach presented in Section 2 makes specific predictions about contexts that should favor morphological syncretism typologically (see Section 2.4). This section argues that these predictions are borne out in three well-known case studies. Sections 3.1 and 3.2 focus on two implicational generalizations involving gender syncretism in pronouns. Section 3.3 focuses on the typology of aspect syncretism.

Most probabilities used in the paper are estimated using corpus frequencies. Spoken language is usually considered as more relevant than written language for typological purposes, on the assumption that typology is shaped by communicative interactions between speakers (e.g. Croft 1990, Jäger 2007). For this reason, corpora of spoken speech or written corpora that are closest to speech (e.g. corpora of dialogs, corpora of subtitles) will be given preference. In two cases though (gender in duals and gender in first/second person pronouns), no corpora were available to estimate the relevant probabilities. In these cases, a specific probability value was chosen for concreteness, either by assuming random sampling (see Section 3.1 on duals) or by assuming a very low probability of error for the identification of gender in first and second person pronouns (see Section 3.2).

3.1. Gender syncretism based on number

3.1.1. Greenberg's Universal 45

Greenberg's Universal 45 is probably one of the most famous implicational generalizations involving syncretism. It states that if there are any gender distinctions in the plural of the pronoun, there are some gender distinctions in the singular also (Greenberg 1963: 60). As stated, this generalization is supposed to apply to all pronouns. However, it is typically mentioned in the typological literature about third person pronouns. Also, the quantitative data in Siewierska

		Gender in the plural	
		yes	no
Gender in the singular	yes	42	61
	no	1	256

Table 1: Greenberg's Universal 45: gender distinctions in third person pronouns (Siewierska 2013)

(2013) directly support a narrower version of Universal 45 applying to third person pronouns only: her data do not specify how the expression of gender differs in singular vs. plural first and second persons cross-linguistically.

Siewierska's (2013) survey is summarized in Table 1 as a contingency table classifying languages according to whether they mark gender in singular and plural third person pronouns. These data support Greenberg's Universal 45, in a statistical sense: languages with gender distinctions in plural pronouns only are very few (there is only one such language in her survey) compared to the three other types of languages.[4] The question of how to deal with exceptions in the present model, which only derives absolute implicational generalizations, is left aside for now (see Section 5.2).

Examples of third person pronoun paradigms with different types of gender syncretism in the singular and in the plural are shown in Table 2, going from Spanish (with a gender distinction in both numbers) in Table 2a to Turkish (without any gender distinction in any number) in Table 2d. Syncretic forms are enclosed in boxes, following the convention in Baerman et al. (2005).

In languages that maintain a gender distinction in the plural (e.g. Spanish in Table 2a), one of the two forms is typically used to refer to mixed groups. In Spanish, the masculine is used to refer to both male-only groups and mixed groups. In Buin, the feminine form is used to refer to both female-only groups

[4] The survey contains 378 languages but Table 1 only contains 360 languages. Among these 378 languages, there are 18 languages for which the author does not specify how gender distinctions are distributed in singular vs. plural pronouns. As a consequence, these 18 languages were not included here.

	SG	PL
MASC	él	ellos
FEM	ella	ellas

(a) Spanish

	SG	PL
MASC	han	de
FEM	hun	de

(b) Danish (Plank & Schellinger 1997: 54)

	SG	PL
MASC	ənta	əntaniḍ
FEM	ənta	əntanətiḍ

(c) Tahhagart Tuareg(Plank & Schellinger 1997: 68)

	SG	PL
MASC	o	onlar
FEM	o	onlar

(d) Turkish (Göksel & Kerslake 2005: 230)

Table 2: Examples of third person pronoun paradigms with various degrees of gender syncretism.

and mixed groups (Laycock 2003: xv). But there are also some languages that use a special form to refer to mixed groups (e.g. Vanimo; Plank & Schellinger 1997: 76). This variability in the way mixed groups are treated morphologically will not be addressed further in this paper.[5]

Among the two patterns of contextual syncretism represented by Danish (Table 2b) and Tahaggart Tuareg (Table 2c), the former one is the most common across languages. Indeed, Tahaggart Tuareg (Table 2c) has gender distinctions in the plural only and is therefore an exception to Universal 45. Plank & Schellinger (1997: 62-65) list a handful of additional exceptions which were not part of Siewierska's sample.

3.1.2. Deriving Greenberg's Universal 45

The theory laid out in Section 2 predicts that plural pronouns should favor gender syncretism as compared to singular pronouns if the probability of misidentifying gender is smaller in plurals than in singulars in case of gender syncretism, namely if the following holds (where GENDER_1 and GENDER_2 refer to the two values of the relevant binary gender feature, e.g. MASC and FEM):

[5] Languages like Vanimo that show a tripartite gender system in the plural (masculine, feminine, mixed) are also beyond the scope of the model discussed in Section 2. This model focuses on binary features.

$$\min\{P(\text{GENDER}_1, \text{PL}), P(\text{GENDER}_2, \text{PL})\} < \min\{P(\text{GENDER}_1, \text{SG}), P(\text{GENDER}_2, \text{SG})\} \quad (9)$$

This equation is likely to generally hold and this for two reasons. First, in nominal categories, singular is more frequent than plural cross-linguistically (Greenberg 1966: 32; Croft 1990: 157), that is $P(\text{PL}) < P(\text{SG})$. Second, the probability distribution of gender conditioned on number should be more skewed in the plural than in the singular, that is $\min\{P(\text{GENDER}_1|\text{PL}), P(\text{GENDER}_2|\text{PL})\} < \min\{P(\text{GENDER}_1|\text{SG}), P(\text{GENDER}_2|\text{SG})\}$. The most common gender distinction used across languages is based on sex (male/female; Corbett 2013). In the singular, sex-based gender partitions the set of individuals in roughly two equal groups: assuming random sampling, the probability that any random individual belongs to one of the two groups is roughly equal to 0.5. Beyond sex-based gender, equal likelihood of feature values seems to be a common property of two-gender systems (whether gender is entirely semantic or partly semantic and partly lexical), as noted by Polinsky & van Everbroeck (2003: 359): 'many two-gender systems have a roughly equal balance in both type and token frequency across the two genders.' However this property does not extend to plurals. Indeed, a binary gender predicate that partitions the set of individuals in two equal-sized groups does not partition the corresponding set of groups of individuals into two equal-sized compartments. For instance, sex-based gender partitions a set of individuals including two men and two women in two equal-sized groups but does not partition the corresponding set of groups of individuals in two equal-sized compartments: among the 11 groups that can be formed from this set of four individuals, 10 are in the extension of MASC (assuming a language like Spanish where MASC is the default gender value, covering both only-male groups and mixed groups) and only one in the extension of FEM (i.e. the group that contains the two women).[6]

[6] For languages that treat FEM as the default (e.g. Buin), there would be 10 groups in the extension of FEM and only one in the extension of MASC.

	Freq 1	Freq 2		Freq 1	Freq 2
SG	3417	0.82	FEM (*ella*)	1806	0.53
			MASC (*él*)	1611	0.47
PL	716	0.18	FEM (*ellas*)	105	0.14
			MASC (*ellos*)	671	0.86

Table 3: Frequency of number and frequency of gender conditioned on number in Spanish subject tonic pronouns in SUBTLEX-ESP (corpus size: 41 million words): frequency per million of words (Freq 1) and relative frequencies (Freq 2).

This very general reasoning is also supported by corpus data. Data from Cuetos et al's (2011) corpus of Spanish subtitles (Table 3) are used to illustrate this point. In this corpus, the cumulated frequency of singular pronouns *ella* 'she' and *él* 'he' is larger than the cumulated frequency of the corresponding plural pronouns (*ellas* 'they'-FEM and *ellos* 'they'-MASC), in accordance with the hypothesis that singular pronouns are more frequent than plural pronouns. Moreover, the frequency distribution of gender values is more skewed in the plural than in the singular. Indeed, in the plural, the masculine form is much more frequent than the feminine form (arguably because masculine plural forms refer to both mixed and male-only groups in Spanish). In the singular, the two feature values are almost equally likely, in accordance with Polinsky & van Everbroek's observation on two-gender systems. As a consequence, the prediction in Equation 9 is supported: the probability of gender misidentification is smaller in the plural ($P(\text{PL}) \times \min\{P(\text{FEM}|\text{PL}), P(\text{MASC}|\text{PL})\} = P(\text{PL}) \times P(\text{FEM}|\text{PL}) = 0.18 \times 0.14 \approx 0.03$) than in the singular ($P(\text{SG}) \times \min\{P(\text{FEM}|\text{SG}), P(\text{MASC}|\text{SG})\} = P(\text{SG}) \times P(\text{MASC}|\text{SG}) = 0.82 \times 0.47 \approx 0.39$) and, as a consequence, gender syncretism is correctly predicted to arise preferentially in the plural.

3.1.3. The case of duals

So far, only singular and plural numbers have been considered. What happens in languages that have more than one non-singular number category, e.g. languages with plurals and duals? To the author's knowledge, there is no comprehensive typological survey on gender syncretism across different non-singular numbers.

	SG	PL	DU
FEM	iɲaɪ̀ɲɔ	iɲaɪ̀ɲuaʊ̀ɪ	iaʊ̀ɪɲuaʊ̀ɪ
MASC	imɯe	imakɯ	iaʊ̀ɪmaɪ̀aʊ̀ɪ

(a) Murui Huitoto (Wise 2004: 322)

	SG	PL	DU
FEM	hiya	hunna	humā
MASC	huwa	hum	humā

(b) Classical Arabic third person pronoun (Greenberg 1966: 35)

	SG	PL	DU
FEM	fo	fova	fol
MASC	fona	fova	fonala

(c) Lavukaleve proximal demonstrative (Terrill 2003: 159)

Table 4: Examples of pronoun paradigms with dual number.

Some languages maintain gender distinctions across all three numbers, as illustrated by Murui Huitoto in Table 4a. Furthermore, the two possible patterns of contextual syncretism involving duals and plurals are attested (Tables 4b and 4c). But further typological surveys are needed to establish whether one of these two patterns is more frequent across languages.

Although the typology of gender syncretism in non-singular numbers does not seem to have been investigated in detail, it is still possible to consider what predictions the current model derives. The model predicts that duals should favor gender syncretism as compared to plurals if the following holds:

$$\min\{P(\textsc{fem}, \textsc{du}), P(\textsc{masc}, \textsc{du})\} < \min\{P(\textsc{fem}, \textsc{pl}), P(\textsc{masc}, \textsc{pl})\} \tag{10}$$

Greenberg (1966: 32) notes that the dual is about five times less likely than the plural in a Sanskrit corpus (see Table 5), suggesting that $P(\textsc{du}) < P(\textsc{pl})$. However, the distribution of gender should theoretically be less skewed towards one of the two gender values in the dual than in the plural (i.e. $\min\{P(\textsc{fem}|\textsc{du}), P(\textsc{masc}|\textsc{du})\} > \min\{P(\textsc{fem}|\textsc{pl}), P(\textsc{masc}|\textsc{pl})\}$). Indeed, assuming simple random sampling, a group

Number	Frequency
SG	0.70
PL	0.25
DU	0.05

Table 5: Frequency of singular, plural, and dual nouns in Sanskrit (Greenberg 1966: 32; sample size: 93,277 nouns)

containing two individuals should be more likely to be only-female (or only-male) than a group containing more individuals. For concreteness, if female individuals are as likely to form referents for duals than male individuals, the probability of feminine duals should be equal to $P(\text{FEM}|\text{DU}) = 0.5 \times 0.5 = 0.25$ and the probability of masculine duals to $P(\text{MASC}|\text{DU}) = 1 - 0.25 = 0.75$,[7] and therefore $\min\{P(\text{FEM}|\text{DU}), P(\text{MASC}|\text{DU})\} = 0.25$. This number is larger than what we would expect for plurals referring to groups with size larger than two (under the same sampling assumptions). Also, it is larger than the smallest of the two conditional probabilities of gender in the plural observed in the Spanish corpus ($\min\{P(\text{FEM}|\text{PL}), P(\text{MASC}|\text{PL})\} = 0.14$; see Table 3). But the probability of duals is probably low enough to compensate for this asymmetry in conditional probabilities (see Table 5). Indeed, in case of gender syncretism, the probability of error is still predicted to be larger in the plural ($P(\text{FEM}|\text{PL}) \times P(\text{PL}) = 0.14 \times 0.25 \approx 0.04$) than in the dual ($P(\text{FEM}|\text{DU}) \times P(\text{DU}) = 0.25 \times 0.05 \approx 0.01$). Under the assumptions described above, the theory therefore predicts that gender distinctions in dual pronouns should asymmetrically entail gender distinctions in plural pronouns. Lavukaleve (in Table 4) should then be analyzed as an exception. Note that this prediction is compatible with the classic markedness-based approach according to which duals are more marked and hence less likely to feature gender distinctions than plurals (see Section 5.1 for further discussion).

[7] MASC is here assumed to be the default morph that can refer to both male-only groups and mixed groups in the plural (as in Spanish).

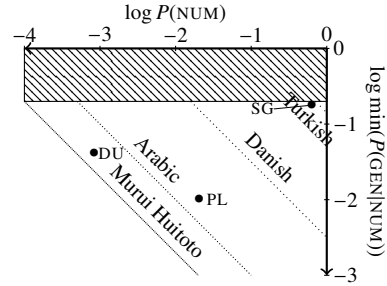


Figure 3: Model predictions for gender syncretism based on number in pronouns.

3.1.4. Summary

The predictions of the theory are plotted in Figure 3. The three numbers (singular, plural, dual) are plotted according to their estimated log probabilities (x-axis; based on the frequencies in Table 5) and the smallest of the two conditional log probabilities of gender (masculine vs. feminine) in the corresponding number (y-axis; based on the frequencies in Table 3 for singulars and plurals and assuming random sampling for duals). The dotted lines correspond to different trade-offs between the size and ambiguity costs: the more densely dotted the line is, the larger the weight on the ambiguity cost is with respect to the weight on the size cost and the richer the morphological paradigm is.

The theory predicts that, as the weight on the ambiguity cost decreases with respect to the weight on the size cost, the masculine/feminine gender distinction should be lost first in the dual (Murui Huitoto vs. Arabic), then in the plural (Arabic vs. Danish) and finally in the singular (Danish vs. Turkish). The prediction concerning singulars vs. plurals is supported by typological data. The prediction concerning plurals vs. duals would need to be further tested, as current typological surveys do not clearly establish which of the two patterns of contextual syncretism (syncretism in plurals only or syncretism in duals only) is more frequent. The few languages that do not conform to Greenberg's Universal 45 (e.g. Tahhagart Tuareg in Table 2c) are not derived by the current model, due to the fact that this model is deterministic (see Section 2). The problem of exceptions will be further discussed in Section 5.2.

		Gender in first/second person	
		yes	no
Gender in third person	yes	18	104
	no	2	254

Table 6: Greenberg's Universal 44' (Siewierska 2013)

3.2. Gender syncretism based on person

3.2.1. Greenberg's Universal 44

Another of Greenberg's implicational generalizations refers to gender syncretism in pronouns. Universal 44 states that if a language has gender distinctions in the first person, it always has gender distinctions in the second or third person or in both (Greenberg 1963). Based on her survey of 378 languages, Siewierska (2013) provides evidence for a slightly different version of this generalization that distinguishes non-third vs. third person pronouns: if a language has gender distinctions in non-third persons (i.e. first or second person), it has gender distinctions in the third person (modulo some exceptions). The data supporting this generalization is shown in Table 6: languages with gender distinctions only in non-third person pronouns are very few (there are only two such languages in her survey) compared to the three other types of languages. Because this generalization differs from Greenberg's original formulation, it is referred to as Universal 44' here.

In her survey, Siewierska (2013) does not distinguish between singular and plural non-third person pronouns. However, she suggests that all 20 languages with gender distinctions in first or second person pronouns have them in the singular but not necessarily in the plural: 'the gender distinctions in question may involve just the singular, or any combination of both singular and non-singular.' Therefore, the data in Table 6 hold for singular first and second person pronouns but not necessarily for plural first and second person pronouns. As a consequence, the remainder of this paper will focus on singular pronouns.

Table 7 provides examples of pronoun paradigms with different types of

	1SG	2SG	3SG
MASC	wn	mən	kər
FEM	ñən	yn	yn

(a) Ngala (Laycock 1965: 133)

	1SG	2SG	3SG
MASC	nī	kai	shī
FEM	nī	kē	ita

(b) Hausa (Newman 2000: 477)

	1SG	2SG	3SG
MASC	kuè	ɔ	imɯe
FEM	kuè	ɔ	iɲaìɲɔ

(c) Murui Huitoto (Wise 2004: 322)

	1SG	2SG	3SG
MASC	ben	sen	o
FEM	ben	sen	o

(d) Turkish (Göksel & Kerslake 2005: 230)

Table 7: Examples of singular pronoun paradigms with different types of gender syncretism based on person.

gender syncretism based on person, going from Ngala (with a gender distinction across all three persons) in Table 7a to Turkish (without any gender distinction) in Table 7d. Among the exceptions to Universal 44', Siewierska (2013) cites Iraqw (not shown here).

3.2.2. Deriving Universal 44'

The theory laid out in Section 2 predicts that singular first/second person pronouns should favor gender syncretism as compared to singular third person pronouns if the probability of misidentifying gender is smaller in non-third person pronouns than in third person pronouns in case of gender syncretism. This prediction is very likely to be correct: because first and second person pronouns refer to discourse participants, it should be easier for a listener to guess the gender of the referent denoted by a given token of first or second person pronoun than the gender of the referent denoted by a given token of third person pronoun (Heath 1975: 96; Corbett 1991: 321; Plank & Schellinger 1997: 65). The reason is that information about the gender of discourse participants is provided by the extra-linguistic context, through visual and auditory cues. Similar cues are not guaranteed to be systematically present to identify the referent's gender in case of third person pronouns. Therefore, for a given pronoun token, the probability distribution of

	Freq 1	Freq 2
1SG	4514	0.43
2SG	2552	0.24
3SG	3417	0.33

Table 8: Frequency of person features in Spanish based on SUBTLEX-ESP (Cuetos et al. 2011; corpus size: 41 million words): frequency per million of words (Freq 1) and relative frequencies (Freq 2).

gender should be much more skewed towards one of the two values in case this pronoun is first or second person than in case it is third person.

Assume the conditional probability of misidentifying the gender value of gender-syncretic pronouns is vanishingly low for non-third person pronouns but close to one half for third person pronouns (as found in Spanish, cf. Table 3). To obtain the overall probability of error for each person, this conditional probability must be multiplied by the probability of each person feature. The frequencies of Spanish pronouns in Table 8 may be used as estimates for the corresponding probabilities. In the absence of strong frequency asymmetries between the three persons, gender syncretism should be mainly driven by the conditional probability of misidentifying gender in the three persons. This probability is expected to be much lower in non-third person pronouns than in third person pronouns, as discussed above. Accordingly, gender syncretism in third person pronouns is predicted to entail gender syncretism in non-third person pronouns. Note that this prediction is made despite the fact that the first person pronoun has the highest frequency in the Spanish corpus. This prediction would not follow in a model that only considers the probability of the context: indeed, first person pronouns are more frequent than second and third person pronouns (at least in the Spanish corpus considered here) and therefore would be incorrectly predicted to favor gender distinctions cross-linguistically (see Section 5 for further discussion).

The predictions of the theory are plotted in Figure 4. For concreteness, a very low probability of 0.01 was hypothesized for incorrect gender identification in non-third person pronouns. The other probabilities were estimated using the

frequencies of Spanish pronouns in Tables 3 and 8. The theory correctly captures Universal 44': as the weight on the ambiguity cost decreases with respect to the weight on the size cost, the masculine/feminine gender distinction disappears first in non-third person pronouns (Ngala vs. Murui Huitoto) and then in third person pronouns (Murui Huitoto vs. Turkish). Because the second person pronoun is less frequent than the first person pronoun (see Table 8) and because of the simplifying assumption that the probability of gender errors is identical in first and second person pronouns in the absence of gender distinctions, the theory currently predicts that a gender distinction in the first person should asymmetrically entail a gender distinction in the second person. This is problematic to derive a language like Hausa, which has gender distinctions in second but not in first person pronouns (Table 7c). However, it is possible that the probability of incorrect gender identification is smaller for first person pronouns, on the reasonable assumption that there are more reliable cues to identify the speaker's than the addressee's gender (e.g. the speaker's voice). In that case, the predictions could be different, with gender distinctions in first person pronouns asymmetrically implying gender distinctions in second person pronouns (as per Greenberg's original Universal 44). Unfortunately, Siewierska (2013) does not specify how many among the 20 languages with gender distinctions in first or second persons in her sample have distinctions in both first and second person pronouns, only in first person pronouns, or only in second person pronouns. The typology of gender marking in first and second person pronouns should be investigated in more details. However, this uncertainty does not affect the main result established in this section: under the reasonable assumption that gender is much easier to identify for discourse participants than for referents that are not present in the discourse context, the theory derives the basic asymmetry between third and non-third person pronouns with respect to gender syncretism.

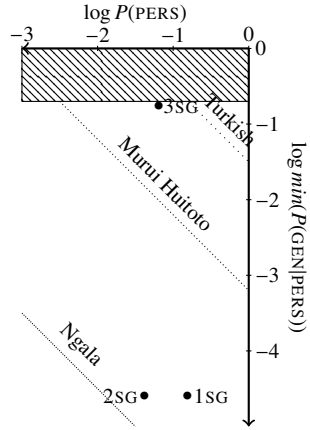


Figure 4: Model predictions for gender syncretism based on person in pronouns.

3.3. Aspect syncretism based on tense

3.3.1. Implicational generalizations

Aspect conveys two kinds of information about a situation: how it is categorized ontologically, as an event or a state, and how it is viewed, as perfective or imperfective (Comrie 1976, Smith 1997). Smith (1997) refers to the former distinction as situation type and to the latter one as viewpoint aspect. This section focuses on viewpoint aspect since this aspectual category shows interesting interactions with tense. With the imperfective aspect (IPFV), an event or state is presented as ongoing or habitual, as illustrated in (2a) with the French imperfective aspect (the imperfect). The ongoing-event reading will be referred to as the progressive reading. With the perfective aspect (PFV), the event or state is presented as complete, as illustrated in (2b) with the French perfective aspect (the simple past).[8]

(2) Imperfective and perfective aspects in French

[8] The perfective aspect is conveyed by the simple past (*je promis*) in the formal register and by the *passé composé* (*j'ai promis*) in conversational French. In both formal and conversational registers, the distinction between imperfective and perfective aspects is expressed morphologically in the past.

(a) À huit heures, je promettais de venir.
at eight hours I promise.IPFV.PAST of come
'At 8 o'clock, I was promising/used to promise to come.'

(b) À huit heures, je promis de venir.
at eight hours I promise.PFV.PAST of come
'At 8 o'clock, I promised to come.'

The imperfective/perfective distinction is widespread across languages: in a sample of 222 languages, Dahl & Velupillai (2013b) found that 101 languagse distinguish the two aspects morphologically. However, the presence of this distinction also depends on the tense of the sentence (Dahl 1985: 81-83; Bybee & Dahl 1989: 83; Malchukov 2009). Tense is a linguistic category that relates the time of a situation to a reference time, usually the utterance time (Comrie 1985). Three basic tenses are usually distinguished: present, past, and future. In its most basic use, the present tense locates a situation as simultaneous with the utterance. In the past tense, the situation is located as anterior to the utterance. In the future tense, it is located as posterior to the utterance.

The interaction between viewpoint aspect and tense can be illustrated with French. In the past, the imperfective form is only compatible with progressive and habitual readings, as shown in (2a). In the future, it is also compatible with the perfective reading (Smith 1997: 78). Sentence (3) can mean that a complete event of Jean singing will happen after Marie enters the office (perfective interpretation). Progressive and habitual readings are also available: sentence (3) can also mean that the event of singing will be ongoing when Marie enters (progressive interpretation) or that Jean will sing or be singing whenever Marie enters the office (habitual interpretation).

(3) Jean chantera quand Marie entrera dans le bureau.
Jean sing.IPFV.FUT when Marie enter.IPFV.FUT in the office
'Jean will sing/be singing when/whenever Marie enters the office.'

Since there is no formal counterpart of the past perfective (or simple past) in the future in French and since imperfective forms are compatible with the three aspectual readings in this tense, French presents a case of aspect syncretism based on tense: aspect is syncretic in the future but not in the past.

	PAST	FUT	PRES
IPFV	obljubljal sem	obljubljal bom	obljubljam
PFV	obljubil sem	obljubil bom	obljubim

(a) Slovenian ('I promise')

	PAST	FUT	PRES
IPFV	iposkhómun	tha ipóskhome	ipóskhome
PFV	iposkhéthika	tha iposkhethó	ipóskhome

(b) Modern Greek ('I promise')

	PAST	FUT	PRES
IPFV	promettais	promettrai	promets
PFV	promis	promettrai	promets

(c) French ('I promise')

	PAST	FUT	PRES
IPFV	versprach	werde versprechen	verspreche
PFV	versprach	werde versprechen	verspreche

(d) German ('I promise')

Table 9: Attested patterns of aspect syncretism across tenses.

Interestingly, the range of attested syncretic patterns is more restricted than what would be expected if the expression of aspect was independent of tense. Among the eight possible patterns of aspect syncretism in the three tenses, only four seem to be attested (Malchukov 2009): languages with a perfective/imperfective distinction in all tenses (e.g. Slovenian; Močnik 2008: 5), languages with a perfective/imperfective distinction in past and future tenses only (e.g. Modern Greek; Holton et al. 1997: 223-228), languages with a perfective/imperfective distinction in the past only (e.g. French; Martinet 1968: 18; see also Bybee & Dahl 1989: 43 for other examples), and languages without perfective/imperfective distinction in any tense (e.g. German; Reyle et al. 2007). The four types of languages are illustrated in Table 9.

In case of aspect syncretism, the imperfective form is generally used as a default: for instance, in Greek and French, the only form available in the present

tense formally corresponds to the imperfective form in the past tense (see Iatridou 2000; Reyle et al. 2007 among others for the interpretation of imperfective as a default).

This typology can be described with two implicational generalizations. The presence of a perfective morph in the present asymmetrically entails the presence of a perfective morph in the future and in the past. The presence of a perfective morph in the future asymmetrically entails the presence of a perfective morph in the past.

3.3.2. Motivating the morphological analysis

In this paper, the lack of perfective morph in the present tense (e.g. in Greek, French, and German in Table 9) is analyzed as a case of morphological syncretism: both perfective and imperfective readings are available in the present tense but the morphology fails to mark this distinction overtly. This analysis is found in Bertinetto & Bianchi (2003: 588), according to whom 'the present [is] an aspectually ambiguous tense' in languages with only imperfective morphology in the present tense.

However, there is an alternative analysis according to which the absence of perfective morph in the present tense in these languages is due to a semantic incompatibility between present and perfective (Giorgi & Pianesi 1997: 160-162). This semantic incompatibility is sometimes referred to as the 'present perfective paradox' in the literature (Malchukov 2009, de Wit 2017). According to this analysis, languages fail to have a perfective morph in the present tense because perfective readings are unavailable in semantically present sentences. This section motivates the morphological analysis assumed in the present paper against the semantic analysis.

First, the semantic analysis cannot easily account for languages where the combination of present temporal reference and semantically perfective aspect does occur and is expressed morphologically through the combination of present morphology and perfective morphology. This possibility is found in a number of

Slavic languages, including Polish, Czech, and Slovenian among others (Dickey 2000, Žagar 2011). For instance, in Slovenian, morphologically present perfective sentences can entail that a complete event in the denotation of the verbal phrase happens exactly at the moment of utterance (Žagar 2011; Močnik 2008: chapter 3). This is the case in performative utterances with perfective and present morphology such as *priznam* 'I admit', *prisežem* 'I swear' and *obljubim* 'I promise'. Second, the semantic analysis cannot easily account for the existence of present perfective readings for morphologically present imperfective sentences in languages without perfective morph in the present (e.g. Greek, French, and German in Table 9). These readings are available in performative utterances (e.g. *I promise to come*) and in so-called reportive contexts (e.g. *Mary wins the race*, as uttered by a sportscaster as Mary crosses the finish line), as acknowledged by many authors (e.g. Dowty 1979: 167, 189-190; Parsons 1990: 30). For instance, in Russian, morphologically imperfective performative sentences do occur, as shown in (4), and their most salient aspectual interpretation is perfective: the speaker is typically understood as having promised to come after uttering (4). This entailment would not go through if, among perfective and progressive interpretations, only the progressive interpretation were available: sentences with achievement predicates like *promise* and progressive morphology (e.g. I am promising to come) do not entail completion of the event (see Dowty 1979: 133).

(4) Ja obescaju prijti zavtra.
I promise.IPFV.PRES come tomorrow
'I promise to come tomorrow.' (Smith 1997: 251)

Finally, based on the semantics of present and perfective, there is no clear reason for why present perfective sentences should never be true, rather than true under very restricted conditions. For instance, Parsons (1990: 31) shows how his semantic system for tense and aspect predicts that a present perfective sentence should be true only if the event described by the sentence coincides temporally with the utterance of this sentence (see also Bary 2012: 41-42). Outside of performative and reportive contexts, this condition is hard to satisfy: as Parsons

puts it,

> The speaker must usually be observing the scene in order to be sure of getting the time right, and the sentence cannot be used in anticipation of the culmination or recapitulation - it must be used exactly once and exactly at the right time. Such uses are rare.

All these problems can be avoided if the lack of perfective morph in the present tense is analyzed as morphological syncretism. Under this analysis, one expects languages to be able to have the imperfective-perfective distinction expressed morphologically in the present tense: hence, the fact that there are languages such as Slovenian is not problematic. In languages without perfective morph in the present, morphologically present imperfective sentences are expected to be aspectually ambiguous: hence, the fact that present tense sentences in Russian are compatible with the event being completed at the moment of utterance is expected. Finally, the fact that semantically perfective present sentences are rare follows from basic pragmatic principles, as proposed by Parsons (1990).

3.3.3. Deriving the implicational generalizations

The theory laid out in Section 2 predicts that present tense should favor aspect syncretism as compared to past and future tenses and that future tense should favor aspect syncretism as compared to past tense if the probability of misidentifying aspect is smaller in the present than in the future and in the future than in the past, that is if the following inequalities hold:

$$\min\{P(\textsc{ipfv}, \textsc{pres}), P(\textsc{pfv}, \textsc{pres})\} < \min\{P(\textsc{ipfv}, \textsc{fut}), P(\textsc{ipfv}, \textsc{fut})\} \quad (11)$$

$$\min\{P(\textsc{ipfv}, \textsc{fut}), P(\textsc{ipfv}, \textsc{fut})\} < \min\{P(\textsc{ipfv}, \textsc{past}), P(\textsc{pfv}, \textsc{past})\} \quad (12)$$

As seen above, the truth conditions of a present perfective sentence are arguably hard to be met. The event in the denotation of the verbal phrase has to coincide with the utterance time for a perfective present sentence to be true (Parsons 1990, Bary 2012). This happens only with verbs that denote very short events whose run time can match the length of the utterance (e.g. in performative

and reportive contexts). In future and past tenses, there is no such constraint: the reference time picked up by the tense morpheme can be large enough to include the run time of virtually any complete event. As a result, the probability of aspect conditioned on tense should be highly skewed towards the imperfective in the present tense and more balanced between perfective and imperfective in the past and in the future. Hence, Equation 11 is likely to hold. In addition, sentences are much more likely to be about past events than about future events, as shown by corpus data (see Josselson 1953 on Russian, Szagun 1978 on English and German). Therefore, Equation 12 is likely to hold too.

Some of these hypotheses were tested quantitatively, using a set of 1,000 sentences randomly sampled from a corpus of French subtitles (New & Spinelli 2013). Each sentence was annotated with its temporal interpretation (past, present, future) as judged by a French native speaker (the author).[9] Sentences with past and present interpretations were further annotated with their aspectual interpretation (perfective, imperfective). This was not done for future sentences because it was harder to determine aspect in this case, due to the absence of morphological aspect marking in French and to the absence of contextual information in New & Spinelli's corpus. These clues were also absent in temporally present sentences but aspect was easier to determine there due to the aspectual restrictions discussed above. For past sentences, sentences in the *imparfait* were annotated as imperfective and sentences in the *passé composé* or simple past as perfective. The results are presented in Table 10.

As a consequence of the difficulty with annotating aspectual information for future sentences, only the hypothesis concerning the asymmetry between past and present (i.e. present tense favors aspect syncretism as compared to past tense cross-linguistically) was tested (Equation 13). It should be noted that this asymmetry is the most robustly documented in the typological literature. Hence,

[9] Formally present sentences with a future interpretation (e.g. *on se revoit demain* 'we'll meet again tomorrow') were annotated as future. In a few cases, it was not possible to determine whether the interpretation was present or future. These sentences were not included in the corpus.

	Count	Frequency		Count	Frequency
PAST	233	0.23	IPFV	69	0.30
			PFV	164	0.70
PRES	646	0.65	IPFV	634	0.98
			PFV	12	0.02
FUT	121	0.12			

Table 10: Count and frequency of temporal and aspectual interpretations in a corpus of French subtitles (New & Spinelli 2013; sample size=1,000 sentences).

it is the most important one to account for.

$$\min\{P(\text{IPFV}, \text{PRES}), P(\text{IPFV}, \text{PRES})\} < \min\{P(\text{IPFV}, \text{PAST}), P(\text{PFV}, \text{PAST})\} \quad (13)$$

In case of aspect syncretism, the probability of aspect misidentification is predicted to be smaller in the present ($P(\text{PRES}) \times P(\text{PFV}|\text{PRES}) = 0.65 \times 0.02 \approx 0.01$) than in the past ($P(\text{PAST}) \times P(\text{IPFV}|\text{PAST}) = 0.23 \times 0.30 \approx 0.07$). As a consequence, present tense is correctly predicted to favor aspect syncretism as compared to past tense. Note that the model derives this prediction despite the fact that present tense is more likely than past tense in the corpus. This prediction would not be derived in a model that only considers the probability of the context (see Section 5 for further discussion). As expected, future interpretations are less frequent than past and present interpretations. However, the absence of aspectual information on future sentences does not make it possible to further test the model's predictions.

The predictions of the theory regarding past and present are plotted in Figure 5. The two contexts are plotted according to their estimated log probabilities (x-axis) and the smallest of the two conditional log probabilities of aspect conditioned on the corresponding tense (y-axis). The theory correctly captures the relevant implicational generalization: as the weight on the ambiguity cost decreases with respect to the weight on the size cost, the perfective-imperfective distinction is lost first in the present (Slovenian vs. Greek) and then in the past (French vs. German).

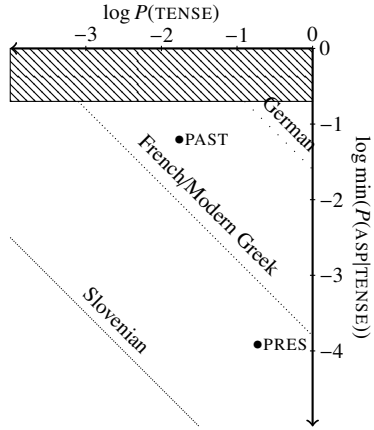


Figure 5: Model predictions for aspect syncretism based on tense.

4. MODEL EXTENSION

The model developed in Section 2 assumes that the computation of the ambiguity cost is independent from context to context. As a consequence, this model is suited to derive patterns where the choice of expressing a grammatical distinction morphologically in a given context does not depend on whether this distinction is expressed morphologically in another context. However, there are cases where the independence assumption clearly does not hold. Case syncretism constitutes a good example. In the Latin sentence (5), the syncretism of nominative and accusative cases for the neuter noun *uerbum* 'word' is harmless: it is possible to deterministically infer that the case is accusative because (i) the masculine *homo* 'man' is unambiguously nominative (the accusative is *hominem*) and (ii) a transitive verb only allows for a single nominative argument.[10] However, if masculine nouns were also syncretic for case in (5), syncretism in neuters would be potentially harmful: in the absence of other clues (e.g. word order, if word order is free, or world knowledge, if the verb is compatible with both neuter and non-neuter subjects and objects), there would be no way to decide deterministically

[10] In this case, it is also possible to disambiguate using world knowledge: a word cannot be the agent of a saying event. However, this type of world knowledge may not always be available.

which argument is the subject and which argument is the object.

(5) Homo uerbum dixit.
man.NOM word.ACC say.PAST.PFV.3SG
'The man said a word.'

Section 4.1 shows how this type of syncretism may be treated in the present model and section 4.2 shows how the typology of case syncretism based on animacy can be derived assuming this extended version of the model. This study on case syncretism adds to a large body of evidence suggesting that case marking is shaped by considerations of communicative efficiency (Comrie 1981, Fedzechkina et al. 2012).

4.1. Extending the model

In a situation where the specific value of a binary grammatical feature T (taking values t_1 and t_2) in a word w_1 deterministically conditions the specific value of this feature in another word w_2 in the same sentence, the ambiguity cost of not expressing T morphologically in contexts c_1 and/or c_2 must be calculated across the two words w_1 and w_2. There are four paradigms to consider: (a) T is syncretic in neither context c_1 and c_2 , (b) T is syncretic only in c_1, (c) T is syncretic only in c_2, and (d) T is syncretic in both contexts.

Table 11 shows the ambiguity and size costs of each paradigm (a)-(d) (before multiplying by language-specific weights). For (a), the ambiguity cost is equal to zero because T is always conveyed unambiguously. The size cost of (a) is equal to two because there is a morphological distinction between t_1 and t_2 in the two contexts. For (b), the ambiguity cost is equal to the probability of misidentifying the values of T for w_1 and w_2 when the context is c_1 in both words.[11] This probability is equal to the probability that the context is c_1 in both words, divided by two (assuming that the two words are as likely to have the feature value t_1

[11] When the context is c_2 in at least one of the two words, the value of T can be inferred deterministically for both words (assuming that the two words must have distinct values for T, as for nominative and accusative in transitive sentences).

	Size cost	Ambiguity cost
(a) No syncretism	2	0
(b) Syncretism in c_1	1	$\frac{P(<w_1,c_1>,<w_2,c_1>)}{2}$
(c) Syncretism in c_2	1	$\frac{P(<w_1,c_2>,<w_2,c_2>)}{2}$
(d) Syncretism in c_1 and c_2	0	$\sum_{i=1}^{2}\sum_{j=1}^{2}\frac{P(<w_1,c_i>,<w_2,c_j>)}{2}$

Table 11: Size cost and ambiguity cost for the four types of paradigms.

as the feature value t_2). For (c), the ambiguity cost is equal to the probability of misidentifying the values of T for w_1 and w_2 when the context is c_2 in both words. This probability is equal to the probability that the context is c_2 in both words, divided by two. For both (b) and (c), the size cost is equal to one because there is a morphological distinction only in one context. For (d), T is always conveyed ambiguously: the ambiguity cost is equal to the sum of the probabilities of misidentification across all contexts.

Paradigms (b) and (c) tie on the size cost. The preference for one or the other paradigm will therefore be entirely determined by their relative ambiguity costs. Assuming that the probabilities for the combinations of grammatical contexts across the two words w_1 and w_2 are the same across languages, only one of (b) or (c) will ever be able to win over the other (whichever one happens to have the smaller ambiguity cost). Which one has the smaller ambiguity cost ultimately depends on the relationship between $P(< w_1, c_1 >, < w_2, c_1 >)$ and $P(< w_1, c_2 >, < w_2, c_2 >)$: for instance, if $P(< w_1, c_1 >, < w_2, c_1 >)$ is smaller than $P(< w_1, c_2 >, < w_2, c_2 >)$, then (b) wins over (c), due to its smaller ambiguity cost. For concreteness assume that, among (b) and (c), (b) has the smaller ambiguity cost. The present approach then predicts that only (a), (b) and (c) should be attested across languages. In other words, it predicts that syncretism of T in c_2 should entail syncretism of T in c_1. As a consequence, the model also predicts implicational generalizations for this more complex case of morphological syncretism. More specifically, it predicts that syncretism of T should happen preferentially in the grammatical context that is least likely to cooccur with both words w_1 and w_2.

	ANIMATE	INANIMATE
NOM	kukkalu (‘dogs’)	iḷḷu (‘houses’)
ACC	kukkalani/nu	iḷḷu

Table 12: Telugu (Krishnamurti & Wynn 1985: 88-89)

4.2. Case syncretism based on animacy

4.2.1. Implicational generalizations

Case refers to the morphological marking of syntactic roles in a verb phrase. There are two broad systems that are found cross-linguistically: nominative-accusative languages and ergative-absolutive languages. In nominative-accusative languages, subjects of transitive and intransitive verbs are marked with nominative case and objects of transitive verbs are marked with accusative case. In ergative-absolutive languages, subjects of transitive verbs are marked with ergative case whereas subjects of intransitive verbs and objects of transitive verbs are marked with absolutive case (Comrie 1981).

The morphological marking of syntactic roles depends on animacy cross-linguistically: it has been observed that higher animacy nominals are more likely to have distinct nominative and accusative forms than lower animacy nominals (Silverstein 1976; Baerman et al. 2005: 40-49; Baerman & Corbett 2013). For instance, in Telugu (Table 12), nominals referring to animates have distinct nominative and accusative forms but nominals referring to inanimates do not (Baerman et al. 2005: 47). The absence (or rarity) of languages with a nominative-accusative distinction only in inanimates points to the following implicational generalization: if a language distinguishes nominative and accusative cases morphologically in inanimates, then it also distinguishes them in animates.

Although lower animacy nominals are more likely to have distinct ergative and absolutive forms in inanimates than in animates (Silverstein 1976, Baerman & Corbett 2013), Baerman et al. (2005: 47) note that there is no language in their sample with an ergative-absolutive distinction in inanimates and no case distinction at all in animates: languages with distinct ergative and absolutive forms

for inanimates either also have this distinction in animates (and then are fully ergative languages) or have distinct nominative and accusative forms in animates (and then are languages with split-ergativity).

However, the implicational generalization as stated above (i.e. distinct nominative and accusative forms in inanimates entails distinct nominative and accusative forms in animates) does not exclude this type of languages (i.e. a language with an ergative-absolutive distinction in inanimates and no case distinction at all in animates). These languages can also be excluded if the implicational generalization is restated specifically in terms of the distinction between subject and object of transitive verbs: if a language distinguishes subject and object of transitive verbs morphologically in inanimates, it distinguishes them in animates. This implicational generalization is satisfied by (i) languages which maintain the two relevant syntactic roles morphologically distinct across animates and inanimates (i.e. fully ergative-absolutive languages, fully nominative-accusative languages, and languages with split ergativity based on animacy), (ii) languages with a nominative-accusative system only for animates and no case distinction for inanimates and (iii) languages which never distinguish subjects and objects of transitive verbs morphologically.

The restatement of the implicational generalization in terms of the distinction between subject and object of transitive verbs is in line with Comrie's (1981) analysis of the typology of case marking systems. As seen above, there are two broad systems of case marking: nominative-accusative and ergative-absolutive systems. In both systems, subjects and objects of transitive verbs are morphologically distinct. The absence of a third system where subjects and objects of transitive verbs are conveyed by the same form whereas subjects of intransitive verbs are conveyed by a distinct form can be understood functionally (Comrie 1981: 120):

> The discriminatory function of case marking will show itself most clearly in the transitive construction, where there is a need to distinguish between [the subject] and [the object], rather than in the intransitive construction, where [the subject] alone occurs.

	MASC	FEM	NEUT
NOM	hi	jü	hat
ACC	ham	har	ham

(a) North Frisian third person pronouns (Mooring dialect in the Bkingharde; Walker 1990: 17)

	MASC	FEM	NEUT
NOM	ille	illa	illud
ACC	illum	illam	illud

(b) Latin demonstrative pronouns (Allen et al. 1903)

	MASC	FEM	NEUT
NOM	wirico	wiricam	je
ACC	wirico	wiricam	je

(c) Wari' emphatic pronouns (Everett & Kern 1997: 295)

Table 13: Attested patterns of case syncretism across neuters and non-neuters.

In some languages, case syncretism is not based on literal animacy but on the neuter/non-neuter distinction, which broadly correlates with animacy. In these languages, neuter nominals typically only refer to inanimate entities whereas non-neuter nominals typically refer to animates, but can also refer to inanimates (Baerman et al. 2005: 47). In these languages, the presence of a morphological distinction between subject and object of transitive verbs in neuters asymmetrically entails the presence of this distinction in non-neuters (i.e. feminine and masculine nominals): there are languages with a nominative-accusative distinction across neuter and non-neuter genders (Table 13a), languages with case distinctions only in non-neuters (Table 13b), languages with no case distinction in any gender (Table 13c), but seemingly no language with a nominative-accusative distinction only in neuters.

4.2.2. Deriving the implicational generalizations

According to the analysis in Section 4.1, inanimates should favor case syncretism if the subject and the object of a transitive verb are less likely to be both inanimates than both animates. This hypothesis is supported by evidence from a corpus of spoken Swedish (Table 14). This corpus is a subset of the corpus 'Samtal i Göteborg' (Conversations in Göteborg): it contains about 60,000 words and was

Subject	Object	Count	Frequency
animate	animate	300	0.10
animate	inanimate	2648	0.84
inanimate	animate	17	0.01
inanimate	inanimate	186	0.06

Table 14: Count and frequency of grammatical function in transitive sentences conditioned on animacy in Swedish (Jäger 2007: 80).

hand-annotated for syntactic roles and animacy by Dahl (2000). In this corpus, inanimates are less likely to co-occur as subject and object in a transitive sentence than animates are ($P(< subject, inanimate >, < object, inanimate >) = 0.06$ vs. $P(< subject, animate >, < object, animate >) = 0.10$). The conclusion drawn from the Swedish data is corroborated by data from a small corpus of spoken Sacapultec (Mayan), where no transitive sentence has both an inanimate subject and an inanimate object (Du Bois 1987: 841).

Similarly, neuters should favor case syncretism as compared to non-neuters if the subject and the object of a transitive verb are less likely to be both neuter than both non-neuter. This hypothesis is likely to be correct. As noted by Baerman et al. (2005: 47), the set of neuter nouns is typically a subset of the set of inanimate nouns and the set of non-neuter nouns a superset of the set of animate nouns. As a consequence, (i) the probability of a neuter subject co-occurring with a neuter object should be smaller than the probability of an inanimate subject co-occurring with an inanimate object and (ii) the probability of a non-neuter subject co-occurring with a non-neuter object should be larger than the probability of an animate subject co-occurring with an animate object. Based on the Swedish data and by transitivity, a neuter subject should therefore be less likely to co-occur with a neuter object than a non-neuter subject with a non-neuter object.

5. Discussion

The preceding sections showed how the theory of syncretism described in Sections 2 and 4.1 can account for the contexts favoring syncretism in four

detailed case studies. Section 5.1 discusses three alternative accounts which do not rely directly on synchronic principles of communicative efficiency. Section 5.2 briefly discusses the problem of exceptions to implicational generalizations and how to generate them in the present framework.

5.1. Alternative theories

5.1.1. Structural approach

In some accounts, syncretism arises preferentially in contexts that are structurally complex. The motivation is a limit on complexity: adding morphological distinctions adds complexity to the grammar. Assuming that there is a language-specific threshold on grammatical complexity, if a language allows an amount *x* of complexity, then it will also allow an amount of complexity smaller than *x*. This general approach is referred to as the 'structural approach' in this paper because typological generalizations about the syntax-morphology interface are derived only from structural properties of the syntax and morphology (without reference to the way they are put to use).

This approach is represented by Harley & Ritter (2002). These authors propose a feature geometry for pronoun structures that is able to capture the universals relative to gender syncretism in pronouns. In their geometry, duals are structurally more complex than plurals and singulars (Harley & Ritter 2002: 492) and plurals are more complex than singulars (Harley & Ritter 2002: 514-515). Combined with the assumption that syncretism in contexts that are structurally less complex entails syncretism in contexts that are more complex, this system predicts that gender syncretism should happen preferentially in duals, then in plurals, and finally in singulars. Also, in their system, non-third person pronouns are structurally more complex than third person pronouns (Harley & Ritter 2002: 488) and this explains why they favor gender syncretism.

How does this theory extend to the other case studies treated in this paper? The authors do not include case features in their feature geometry for pronouns (see Harley & Ritter 2002: 507 for a justification). However, this could a priori be

done. For their analysis to account for the typology of case syncretism based on animacy, inanimates would need to be structurally more complex than animates. For the typology of case syncretism based on the neuter/non-neuter distinction, neuters would need to be structurally more complex than non-neuters. However, these predictions are not directly compatible with the feature geometry assumed by Harley & Ritter (2002: 486): in their geometry, the node Inanimate/Neuter is less marked (i.e. less complex) than the node Animate and therefore inanimates and neuters should allow for richer case distinctions than animates and non-neuters, contrary to fact.

It is unclear how to extend the structural approach to the typology of aspect syncretism. There are several reasons to think that present tense should be the simplest tense structurally and yet it favors aspect syncretism. Present tense is more basic than past and future tenses in the sense that these tenses are defined with respect to the present tense semantically. One way out would be to assume that present and perfective are semantically incompatible and therefore the absence of perfective morph in the present in languages like French or Greek is not due to syncretism. However this approach suffers from a number of problems discussed in Section 3.3. Furthermore, under the structural approach, future should probably be the tense favoring aspect syncretism cross-linguistically, as it is semantically more complex than the other tenses (e.g. Klecha 2014 on the hypothesis that future is a modal operator) and often morphologically more complex: inflectional futures often derive from periphrastic constructions (Dahl & Velupillai 2013a).

5.1.2. Markedness approach

In the markedness approach, morphological distinctions are richer in contexts that are more frequent. This approach is represented by Greenberg (1966), who established markedness scales based on the frequency of grammatical features and used them to explain the typology of morphological syncretism (see Croft 1990: 92-93 for a list of markedness scales extracted from Greenberg's work).

This approach can be illustrated with the typology of gender syncretism based on number (Universal 45). The higher frequency of singulars over plurals provides evidence for a scale where singular is less marked than plural (singular < plural). This scale explains why gender syncretism is favored in plurals (Universal 45).

In both the markedness-based approach and the present approach, the probability of the grammatical context plays a role in deriving typological tendencies in patterns of syncretism. However, in the present approach, the conditional probability of the target grammatical feature given that context also matters. This probability was crucial to derive the typology of gender syncretism based on person (Section 3.2) and the typology of aspect syncretism based on tense (Section 3.3). It is unclear how these patterns can be derived without appealing to the concept of conditional probability: first person pronouns and present tense are highly frequent feature values and therefore, under the markedness approach, should not particularly favor gender syncretism and aspect syncretism, respectively.

5.1.3. Diachronic approach

The present approach assumes that there is an explicit synchronic principle that favors more informative paradigms in speakers' productions (see Fedzechkina et al. 2012 among others for experimental evidence). However, there are alternative approaches where functionalist principles shaping morphological patterns do not play out at the synchronic level but only at a diachronic level (e.g. Croft 1990: chapter 9).

Haspelmath (2006: 48) sketches a diachronic scenario to explain implicational generalizations in syncretic patterns without requiring a synchronic principle favoring small and unambiguous paradigms. In this scenario, syncretism arises as the result of imperfect learning: some morphological combinations are less frequent in the learner's input and therefore harder to remember and, for this reason, get lost over time.

This diachronic scenario explains why aspect syncretism is favored in the

present tense: the combination of perfective aspect and present tense is infrequent (for the pragmatic reasons exposed in Section 3.3) and therefore learners should tend to assume that only the imperfective morph is available in this tense. It also explains why plural pronouns favor gender syncretism. Plural pronouns are less frequent than singular pronouns and therefore gender distinctions get lost in the plural first. For instance, a system with full gender distinctions across numbers (e.g. Murui Huitoto; see Table 4a) should tend to be reinterpreted by learners as a system without gender distinctions in the plural (e.g. Danish; see Table 2b). Pidgin Hausa provides a concrete example of loss of gender distinctions in pronoun paradigms: Standard Hausa has masculine and feminine pronouns *yā* 'he' and *tā* 'she' but Pidgin Hausa generalized the masculine form *yā* to both genders (Heine & Reh 1984: 42), arguably because this form was more frequent. Finally, this diachronic scenario can also account for the preference to neutralize case distinctions in inanimates: as shown in Table 14, inanimate subjects of transitive sentences are less frequent ($P(\text{INANIM}, \text{SUBJ}) = 0.06 + 0.01 = 0.07$) than animate objects of transitive sentences ($P(\text{ANIM}, \text{OBJ}) = 0.10 + 0.01 = 0.11$) and therefore the morphs expressing the combination inanimates + subjects should be harder to memorize by learners.

However, the diachronic approach sketched by Haspelmath fails on the typology of gender syncretism based on person. First person pronouns should be frequent in the learner's input. Assuming that a first person pronoun is roughly as likely to be feminine as masculine in the learner's input, a learner should receive a large amount of evidence for positing gender distinctions in first person pronouns. Why then should gender distinctions be underrepresented in this case?

Also, the analysis based on imperfect learning is only suited to describe cases where a richer morphological system gets poorer, but not cases where a poorer morphological system gets richer. However, such cases are attested. For instance, Ancient Greek had a perfective/imperfective morphological distinction in the past only (Bary 2009). This distinction was later extended to the future tense: Modern Greek distinguishes perfective and imperfective aspects in both past and future

tenses (see Table 9b). Analyses have been proposed to account for the creation of new morphological distinctions without resorting to synchronic principles of communicative efficiency, though. One analysis relies on grammaticalization. Grammaticalization describes a process by which a former lexical element is integrated into a paradigm (Croft 1990: 234). Combinations of lexical elements that are more frequent are expected to be grammaticalized earlier than combinations that are less frequent (Bybee 2006: 719-721). However, like the analysis based on imperfect learning, this analysis fails to account for the typology of gender syncretism based on person. Singular non-third pronouns should be highly frequent under both genders (masculine and feminine) and therefore it is unclear why gender distinctions should not grammaticalize there at least as fast as in third person pronouns.

5.1.4. Summary

To summarize, among the four approaches discussed in this paper, only the approach assuming communicative biases in language users (learners or speakers) makes the correct predictions for all case studies considered here. The structural approach does not generalize well beyond the case studies it was designed to account for. The markedness-based and diachronic approaches make predictions that are closer to the approach based on communicative efficiency but they fail on some case studies, in particular on the typology of gender based on person in pronouns. This failure is crucially due to the fact that these theories give no role to a feature's recoverability in context.

5.2. Exceptions

The implicational generalizations discussed in this paper and whose validity has been tested through large typological surveys (i.e. Greenberg's Universals 45 and 44; see Sections 3.1 and 3.2) admit exceptions. To the author's knowledge, there are no exceptions reported for the other implicational generalizations (i.e. aspect syncretism and case syncretism; see Sections 3.3 and 4.2) but this could

be because they have not been tested as extensively. Also, as shown by Piantadosi & Gibson (2014), although it is in principle possible to find absolute linguistic universals using cross-linguistic surveys and typological statistics, the number of independent languages necessary to do so is generally unachievable. Therefore, even if there were no exception to an implicational generalization in a typological survey, this most likely would not guarantee that this generalization is an absolute universal.

However, as outlined in Section 2, the present approach only derives absolute, exceptionless implicational universals. How to deal with exceptions then? Exceptions are problematic not only for the approach proposed here but also for the other approaches discussed in Section 5.1. Fortunately, there are probabilistic implementations of the kind of weighted grammars used in this paper that predict non-null probabilities for candidates that have zero probabilities in a deterministic framework (e.g. stochastic Harmonic Grammar). In these probabilistic approaches, the algorithm evaluating linguistic forms is no longer conceived as specifying a single winner but a probability distribution over candidate forms. Patterns that violate implicational generalizations can be generated in the probabilistic approach but they remain less likely than patterns that do not violate them (Magri 2018).

6. Conclusion

This paper has shown that, if syncretism is treated as motivated by constraints requiring morphological paradigms to be both small and unambiguous, a number of implicational generalizations fall out. In particular, the account successfully predicts the contexts favoring syncretism in four case studies whereas alternative analyses only account for some of the patterns. The crucial advantage of the current approach is that it takes into account the contextual recoverability of the feature targeted by syncretism.

The approach of morphological syncretism proposed in this paper is conceptually close to the phonetically-based approach of contextual neutralization in

phonology (e.g. Flemming 2002). In Flemming's model, contextual neutralization of phonemes arises as the result of the interaction of three functionally motivated constraints: a constraint favoring large phoneme inventories, a constraint favoring the minimal expenditure of effort in producing phonemes, and a constraint favoring perceptually distinct contrasts. Contextual neutralization arises when the additional effort necessary to maintain a large number of phonemes in a given context is not compensated by a sufficient gain in perceptual distinctiveness. This approach to phonological neutralization is also able to account for phonological implicational universals: for instance, the fact that the presence of contrasts involving consonant place of articulation in word-final positions asymmetrically entails the presence of these contrasts pre-vocalically follows from the hypothesis that place contrasts are universally more distinct pre-vocalically than word-finally (Jun 2004). If speakers are willing to put a lot of effort in order to maintain sufficiently distinct contrasts word-finally, then they should be willing to maintain these contrasts pre-vocalically since it should require less effort to do so in this context.

The present analysis is conceptually similar because it also involves a trade-off between a constraint favoring large inventories (cf. the ambiguity cost) and a constraint favoring small inventories (cf. the size cost). Probabilities of grammatical features play the same role in the analysis of syncretism as perceptual distinctiveness in the analysis of contextual phonological neutralization: they provide a measure of the listener's ability to recover underlying distinctions (grammatical features in the morphological case vs. phonemes in the phonological case) from the linguistic signal. The analysis of syncretism presented in this paper together with the phonetic analysis of phonological neutralization provide a rationale for the long observed parallel between morphological syncretism and contextual phonological neutralization: syncretism and phonological neutralization are similar in their typological manifestations because both arise as a result of language being a way of transmitting information under cognitive (perceptual, articulatory, processing, etc.) constraints.

REFERENCES

Allen, Joseph Henry, J. B. Greenough, Goerge Lyman Kittredge, Albert Andrew Howard & Benjamin L. D'Ooge. 1903. *Allen and Greenough's new Latin grammar for schools and colleges, founded on comparative grammar*. Boston, MA/London: Ginn & Company.

Baayen, R Harald. 2007. Storage and computation in the mental lexicon. In G. Jarema & G. Libben (eds.), *The mental lexicon: Core perspectives*, 81–104. Amsterdam: Elsevier. doi: 10.1163/9780080548692_006.

Baerman, Matthew, Dunstan Brown & Greville G. Corbett. 2005. *The syntax-morphology interface: a study of syncretism*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511486234.

Baerman, Matthew & Greville G. Corbett. 2013. Case syncretism. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*, Leipzig: Max Planck Institute for Evolutionary Anthropology. https://wals.info/chapter/28.

Bary, Corien. 2009. *Aspect in Ancient Greek: A semantic analysis of the aorist and imperfective*. Nijmegen: Radboud University dissertation.

Bary, Corien. 2012. The Ancient Greek tragic aorist revisited. *Glotta* 88. 31–53. doi:10.13109/glot.2012.88.1-4.31.

Bertinetto, Pier Marco & Valentina Bianchi. 2003. Tense, aspect, and syntax. *Linguistics* 41. 565–606. doi:10.1515/ling.2003.019.

Brysbaert, Marc, Paweł Mandera & Emmanuel Keuleers. 2018. The word frequency effect in word processing: An updated review. *Current Directions in Psychological Science* 27(1). 45–50. doi: 10.1177/0963721417727521.

Bybee, Joan L. 2006. From usage to grammar: the mind's response to repetitions. *Language* 82. 711–733. doi:10.1353/lan.2006.0186.

Bybee, Joan L. & Östen Dahl. 1989. The creation of tense and aspect systems in the languages of the world. *Studies in Language* 13. 51–103. doi:10.1075/sl.13.1.03byb.

Comrie, Bernard. 1976. *Aspect*. Cambridge: Cambridge University Press.

Comrie, Bernard. 1981. *Language universals and linguistic typology: Syntax and morphology*. Chicago, IL: University of Chicago Press.

Comrie, Bernard. 1985. *Tense*. Cambridge: Cambridge University Press.

Corbett, Greville G. 1991. *Gender*. Cambridge: Cambridge University Press.

Corbett, Greville G. 2013. Sex-based and non-sex-based gender systems. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*, Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info/chapter/31.

Croft, William. 1990. *Typology and universals*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511840579.

Cuetos, Fernando, Maria Glez-Nosti, Analía Barbón & Marc Brysbaert. 2011. SUBTLEX-ESP: Spanish word frequencies based on film subtitles. *Psicológica* 32. 133–143.

Dahl, Östen. 1985. *Tense and aspect systems*. Oxford: Blackwell.

Dahl, Östen. 2000. Egophoricity in discourse and syntax. *Functions of language* 7. 37–77. doi: 10.1075/fol.7.1.03dah.

Dahl, Östen & Viveka Velupillai. 2013a. The future tense. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*, Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info/chapter/67.

Dahl, Östen & Viveka Velupillai. 2013b. Perfective/imperfective aspect. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*, Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info/chapter/65.

Dickey, Stepen M. 2000. *Parameters of Slavic aspect*. Stanford, CA: CSLI Publications.

Dowty, David R. 1979. *Word meaning and Montague grammar*. Dordrecht/Boston/London: D. Reidel Publishing Company.

Du Bois, John W. 1987. The discourse basis of ergativity. *Language* 63. 805–855.

Everett, Daniel L. & Barbara Kern. 1997. *Wari: the Pacaas Novos language of Western Brazil* Descriptive Grammar Series. London/New York, NY: Routledge.

Fedzechkina, Maryia, Florian T. Jaeger & Elissa L Newport. 2012. Language learners restructure their input to facilitate efficient communication. *Proceedings of the National Academy of Sciences* 109(44). 17897–17902. doi:10.1073/pnas.1215776109.

Flemming, Edward. 2002. *Auditory representations in phonology*. New York: Routledge. doi:10.4324/9781315054803.

Gibson, Edward, Richard Futrell, Steven T. Piandadosi, Isabelle Dautriche, Kyle Mahowald, Leon Bergen & Roger Levy. 2019. How efficiency shapes human language. *Trends in cognitive sciences* 23. 389–407. doi:10.1016/j.tics.2019.02.003.

Giorgi, Alessandra & Fabio Pianesi. 1997. *Tense and aspect: from semantics to morphosyntax*. New York, NY/Oxford: Oxford University Press. doi:10.2307/417406.

Göksel, Asli & Celia Kerslake. 2005. *Turkish: a comprehensive grammar*. London/New York, NY: Routledge. doi:10.4324/9780203340769.

Greenberg, Joseph H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In Joseph H. Greenberg (ed.), *Universals of Human Language*, 73–113. Cambridge, Mass: MIT Press.

Greenberg, Joseph H. 1966. *Language universals, with special reference to feature hierarchies*. The Hague: Mouton.

Harley, Heidi & Elizabeth Ritter. 2002. Person and number in pronouns: a feature-geometric analysis. *Language* 78. 482–526. doi:10.1353/lan.2002.0158.

Haspelmath, Martin. 2006. Against markedness (and what to replace it with). *Journal of Linguistics* 42. 25–70. doi:10.1017/S0022226705003683.

Haspelmath, Martin. 2021. Explaining grammatical coding asymmetries: Form-frequency correspondences and predictability. *Journal of Linguistics* 129. doi:10.1017/S0022226720000535.

Haspelmath, Martin & Andrea Sims. 2010. *Understanding morphology*. London: Routledge. doi:10.4324/9780203776506.

Hawkins, John A. 2004. *Efficiency and complexity in grammars*. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199252695.001.0001.

Heath, Jeffrey. 1975. Some functional relationships in grammar. *Language* 51. 89–104. doi:10.2307/413151.

Heine, Bernd & Mechthild Reh. 1984. *Grammaticalization and reanalysis in African languages*. Hamburg: Buske Helmet Verlag Gmbh.

Holton, David, Peter Mackridge & Irini Philippaki-Warburton. 1997. *Greek: A comprehensive grammar of the modern language*. London: Routledge. doi:10.4324/9780203802380.

Iatridou, Sabine. 2000. The grammatical ingredients of counterfactuality. *Linguistic Inquiry* 31. 231–270. doi:10.1162/002438900554352.

Jaeger, Florian T. & Harry Tily. 2011. On language 'utility': Processing complexity and communicative efficiency. *Wiley Interdisciplinary Reviews: Cognitive Science* 2. 323–335. doi:10.1002/wcs.126.

Jäger, Gerhard. 2007. Evolutionary game theory and typology: A case study. *Language* 83. 74–109. doi:10.1353/lan.2007.0020.

Josselson, Harry H. 1953. *The Russian word count*. Detroit, MI: Wayne University Press.

Jun, Jongho. 2004. Place assimilation. In Bruce Hayes, Robert Kirchner & Donca Steriade (eds.), *Phonetically based phonology*, 58–86. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511486401.003.

Kenstowicz, Michael. 2005. Paradigmatic uniformity and contrast. In Laura Downing, T. A. Hall & Renate Raffelsieffen (eds.), *Paradigms in phonological theory*, 145–169. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199267712.003.0007.

Klecha, Peter. 2014. Diagnosing modality in predictive expressions. *Journal of Semantics* 31. 443–455. doi:10.1093/jos/fft011.

Krishnamurti, Bhadriraju & John Peter Lucius Wynn. 1985. *A grammar of Modern Telugu*. Delhi: Oxford University Press.

Kurumada, Chigusa & Florian T. Jaeger. 2005. Communicative efficiency in language production: Optional case-marking in Japanese. *Journal of Memory and Language* 83. 152–178. doi:10.1016/j.jml.2015.03.003.

Laycock, Donald C. 1965. *The Ndu language family (Sepik District, New Guinea)*. Canberra: Australian National University.

Laycock, Donald C. 2003. *A dictionary of Buin, a language of Bougainville*. Canberra: Pacific Linguistics. doi:10.15144/PL-537.

MacKay, David J. C. 2003. *Information theory, inference, and learning algorithms*. Cambridge: Cambridge University Press.

Magri, Giorgio. 2018. Implicational universals in stochastic constraint-based phonology. In *Proceedings of the 2018 conference on empirical methods in natural language processing*, 3265–3274. Brussels: Association for Computational Linguistics. doi:10.18653/v1/D18-1364.

Malchukov, Andrej. 2009. Incompatible categories: resolving the 'present perfective paradox'. In Lotte Hogeweg, Helen de Hoop & Andrej Malchukov (eds.), *Cross-linguistic semantics of tense, aspect and modality*, 13–33. New York, NY: John Benjamins Publishing Company. doi:10.1075/la.148.02mal.

Martinet, André. 1962. *A functional view of language*. Oxford: Clarendon Press.

Martinet, André. 1968. Neutralisation et syncrétisme. *La Linguistique* 4. 1–20.

Močnik, Maša. 2008. *Slovenian perfective and imperfective explicit performative utterances*. University of Amsterdam MA thesis.

New, Boris & Elsa Spinelli. 2013. Diphones-fr: A French database of diphone positional frequency. *Behavior research methods* 45. 758–764. doi:10.3758/s13428-012-0285-y.

Newman, Paul. 2000. *The Hausa language: An encyclopedic reference grammar*. New Haven, CT: Yale University Press. doi:10.2307/3087661.

Parsons, Terence. 1990. *Events in the semantics of English: A study in subatomic semantics*. Cambridge, MA: MIT Press.

Piantadosi, Steven T. & Edward Gibson. 2014. Quantitative standards for absolute linguistic universals. *Cognitive Science* 38. 736–756. doi:10.1111/cogs.12088.

Piantadosi, Steven T., Harry Tily & Edward Gibson. 2012. The communicative function of ambiguity in language. *Cognition* 122. 280–291. doi:10.1016/j.cognition.2011.10.004.

Plank, Frans & Wolfgang Schellinger. 1997. The uneven distribution of genders over numbers: Greenberg nos. 37 and 45. *Linguistic Typology* 1. 53–101.

Polinsky, Marie & Evra van Everbroeck. 2003. Development of gender classifications: modeling the historical change from Latin to French. *Language* 79. 356–390. doi:10.1353/lan.2003.0131.

Reyle, Uwe, Antje Rossdeutscher & Hans Kamp. 2007. Ups and downs in the theory of temporal reference. *Linguistics and Philosophy* 30. 565–635. doi:10.1007/s10988-008-9027-0.

Siewierska, Anna. 2013. Gender distinctions in independent personal pronouns. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*, Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info/chapter/44.

Silverstein, Michael. 1976. Hierarchy of features and ergativity. In Richard M. W. Dixon (ed.), *Grammatical categories in australian languages*, 112–171. Canberra: Australian Institute of Aboriginal Studies. doi:10.1515/9783110871661-008.

Smith, Carlota S. 1997. *The parameter of aspect, 2nd edition*. Dordrecht: Kluwer.

Smolensky, Paul & Géraldine Legendre. 2006. *The harmonic mind: From neural computation to optimality-theoretic grammar*. Cambridge, MA: MIT Press.

Szagun, Gisela. 1978. On the frequency of use of tenses in English and German children's spontaneous speech. *Child Development* 49. 898–901. doi:10.2307/1128267.

Terrill, Angela. 2003. *A grammar of Lavukaleve*. Berlin: Mouton de Gruyter. doi:10.1515/9783110923964.

Žagar, Igor Ž. 2011. Performativity as tense and aspect. *International Review of Pragmatics* 3. 168–193. doi:10.1163/187731011X597505.

Walker, A. G. H. 1990. Frisian. In Charles V. J. Russ (ed.), *The dialects of Modern German: A linguistic survey*, 1–30. London: Routledge.

Wise, Mary Ruth. 2004. Small language families and isolates in Peru. In Richard M. W. Dixon & Alexandra Aikhenvald (eds.), *The Amazonian languages*, 307–340. Cambridge: Cambridge University Press.

de Wit, Adtrid. 2017. *The present perfective paradox across languages*. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780198759539.001.0001.

Author's address: *Université de Lausanne, CH-1015 Lausanne, Switzerland*
benjamin.storme@unil.ch
ORCID: https://orcid.org/0000-0002-2560-2265