Natalia Levshina*

Communicative efficiency and differential case marking: a reverse-engineering approach

https://doi.org/10.1515/lingvan-2019-0087
Received December 10, 2019; accepted January 7, 2021

Abstract: The use of differential case marking of A and P has been explained in terms of efficiency (economy) and markedness. The present study tests predictions based on these accounts, using conditional probabilities of a particular feature given the syntactic role (cue availability), and conditional probabilities of a particular syntactic role given the feature in question (cue reliability). Cue availability serves as a measure of markedness, whereas cue reliability is central for the efficiency account. Similar to reverse engineering, we determine which of the probabilistic measures could have been responsible for the recurrent cross-linguistic patterns described in the literature. The probabilities are estimated from spontaneous informal dialogues in English and Russian (Indo-European), Lao (Tai-Kadai), N||ng (Tuu) and Ruuli (Bantu). The analyses, which involve a series of mixed-effects Poisson models, clearly demonstrate that cue reliability matches the observed cross-linguistic patterns better than cue availability. Thus, the results support the efficiency account of differential marking.

Keywords: conditional probability; differential argument marking; efficiency; spoken corpora

1 Theoretical background

1.1 Cross-linguistic evidence of differential case marking

Differential case marking (DCM) has received a lot of attention in the literature (e.g., Aissen 2003; Bossong 1985; Comrie 1978; de Hoop and de Swart 2008; Dixon 1979; Silverstein 1976; see also a recent overview in Witzlack-Makarevich and Seržant 2018). The term describes a situation where the same argument has different formal coding depending on its semantic, pragmatic or other properties. This paper focuses on differential marking of A and P arguments of transitive verbs, where A stands for the semantically more agent-like argument and P for the more patient-like one (cf. Dowty 1991), as in the sentence *John (A) broke the window (P)*. One of the languages with differential A marking (DAM) is Qiang, a Sino-Tibetan language. In Qiang, animate A's are formally unmarked, as in (1a), whereas inanimate A's are marked, as in (1b):

(1) Qiang (Sino-Tibetan, LaPolla and Huang 2003: 79–80)
   a. Animate A: unmarked
      the: *qa* *dʑete.*
      3SG 1SG hit
      'He is hitting me.'
   b. Inanimate A: marked
      moʁu-wu *qa* *da-tuə-ʐ.*
      wind-AGT 1SG DIR-fall.over-CAUS
      'The wind knocked me over.'

***Corresponding author: Natalia Levshina**, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands,
E-mail: natalevs@gmail.com

Differential marking of P is known as differential object marking or DOM (see, e.g., Bossong 1985). A language with DOM is Spanish, where animate objects tend to be formally marked with the preposition *a*, as in (2a), while inanimate objects are unmarked, as in (2b). Definiteness, semantics of individual verbs, and some other variables also play a role (see von Heusinger and Kaiser 2007):

(2) Spanish (García García 2018: 211)

a. Inanimate P: unmarked

Pepe	*ve*	*la*	*película*.
Pepe	see.3SG	DEF	film

'Pepe sees the film.'

b. Animate P: marked

Pepe	*ve*	*a*	*la*	*actriz*.
Pepe	see.3SG	OBJ	DEF	actress

'Pepe sees the actress.'

It has often been claimed that the presence or absence of overt markers is constrained by diverse referential effects, which are often represented as scales (e.g. Croft 2003: 130–132; Haspelmath 2021):

(3) a. Person: 1 and 2 > 3
b. Nominality: pronoun > noun
c. Animacy: human (> animal) > inanimate
d. Definiteness/specificity: definite > specific > non-specific
e. Givenness: discourse-given > discourse-new
f. Focus: background (topic) > focus

The scales in (3) should be interpreted as follows. If a language has a coding split in A, more prominent A arguments (i.e. the ones on the left) are usually unmarked, while less prominent ones (i.e. the ones on the right) are marked. For P arguments, the reverse holds. The examples (1) and (2) provided above illustrate this mirror-image pattern: A is marked when it is inanimate, as in Qiang (1), while P is marked when it is animate, as in Spanish (2).

At the same time, empirical data have demonstrated that this elegant account requires several qualifications. First of all, it has been shown that the relationship between A and P marking is not perfectly symmetric (cf. Fauconnier 2011). In particular, differential P marking is more frequent cross-linguistically than differential A marking (e.g. de Hoop and de Swart 2008; Malchukov 2008). The latter is typically observed in ergative-absolutive languages, which are less numerous than nomimative-accusative languages. Moreover, different scales are relevant for different roles. Commonly, animacy and definiteness are regarded as the most relevant features in differential P marking (Malchukov 2008; Sinnemäki 2014). These splits, however, are rarely reported for differential A marking (cf. Comrie 1981:123; de Hoop and Malchukov 2008).

Second, quantitative analyses indicate that the strength of evidence for the hypothesized effects of individual scales vary greatly depending on the geographic area (Bickel et al. 2015). The scale effects are therefore not universal in the sense 'highly probable to be observed in any part of the world'. Rather, they are of the type, 'if there is a split in feature F, then the arguments with this feature are more(less) likely to be marked than the arguments without this feature' (cf. Levshina 2018; Schmidtke-Bode and Levshina 2018).

The previous cross-linguistic research suggests that the universals of the latter type are observed for the following features. As for the A role, we observe splits in grammars involving **nominality** (pronouns are unmarked, while nouns are marked) and **person** (1st and 2nd person subjects are unmarked, while 3rd person subjects are marked). Although categorical **animacy** and **definiteness** splits are untypical for A, one can find, however, some effects of animacy and definiteness in optional marking. For example, inanimate and indefinite A's in colloquial Korean cannot drop their case marking, while animate and definite A's can (Lee 2008).

As far as P is concerned, there are reliable splits in **animacy** and **definiteness** (animate and/or definite/specific objects are marked, inanimate and/or indefinite/non-specific objects are unmarked), as well as in **nominality** (nouns are unmarked, while pronouns are marked). As for **person**, P has many violations of the

person scale (especially in non-singular forms), where 1st and 2nd person objects are unmarked, and 3rd person objects are marked (Schmidtke-Bode and Levshina 2018). Because of that this scale will not be important in the subsequent discussion.

1.2 Explanations of differential case marking

DCM can be interpreted as a result of conventionalization of language users' efficient communicative behaviour. According to Aissen (2003: 438): "the overt marking of a typical object facilitates comprehension where it is most needed, but not elsewhere" (see also de Swart 2005). This behaviour is efficient because it allows the speaker to minimize articulatory costs without jeopardizing the comprehension of the message. This kind of recipient design manifests itself in all kinds of linguistic structures – from the fact that more frequent grammatical categories are usually expressed by shorter or zero markers (Croft 2003: Ch. 4; Greenberg 1966) to the use of zero, pronominal or full lexical referential expressions depending on the accessibility of the referent (Ariel 1990). Other examples are the use of reduced phonological forms in more predictable contexts (e.g. Bell et al. 2009) and the omission of complementizers after verbs that are usually followed by complement clauses (Jaeger 2010). See also Jaeger and Buz (2017) and Gibson et al. (2019), as well as the Introduction to this special issue.

The diachrony of DCM can involve morphological reduction or enhancement (i.e. addition of phonological material). According to de Hoop and de Swart (2008), differential subject marking occurs when the ergative marker is dropped due to economy reasons. At the same time, there are languages in which the case marker has been added. For example, Seržant (2019) shows that the object marker in Old Russian appeared exclusively in those contexts where the old morphological distinctions were lost due to phonological processes, and there was functional pressure for the emergence of DOM for the purposes of disambiguation (i.e. only on animate nouns and some pronouns).

It has also been argued that DOM can be explained by markedness theory (Greenberg 1966; Jakobson 1971 [1932]). The morphology of DOM is overwhelmingly privative: it is zero marking against some audible expression. The markedness of expression is iconically related to the markedness of content, such that semantically marked members of grammatical categories are marked formally, while semantically unmarked ones are also formally unmarked (Aissen 2003). The same holds for subjects.

There also exist other explanations. A popular idea is related to the indexing function of case, according to which only individuated and affected objects receive case (e.g., Malchukov 2008; Siewierska and Bakker 2008: 292). They are more salient in discourse and therefore make 'better' P's. In order to test this hypothesis, we need independent evidence that the degree of individuation, affectedness, salience, etc., influences the probability of marking. It seems that such evidence can only be provided experimentally, which is why this hypothesis is beyond the scope of the present corpus-based study.

Yet another popular hypothesis says that DOM emerges from topicalization of objects (e.g., Iemmolo 2010). Ongoing experimental work based on artificial language learning by Tal et al. (2020) does not find, however, a direct causal link between object givenness and case marking, although the relationship may be indirect: object marking is more frequent in OSV sentences, which are produced more frequently when the object is topicalized. These explanations need further psycholinguistic research.

The present study compares two explanatory factors, which can be operationalized with the help of corpus data: efficiency (economy) and markedness. The next section explains how the relevance of these two factors for DCM can be tested with the help of different types of corpus measures.

2 Different types of conditional probabilities and theoretical predictions

There are two types of probabilities which are relevant for the goals of our study and will be inferred from corpora. One of them represents *cue availability*, a concept from the Competition Model of learning

(MacWhinney et al. 1984). It is the probability that a certain cue is available when a certain grammatical or semantic function is expressed. For example, the English pronoun *you* is more likely to occur as the subject of a sentence than the archaic *thou*. Using the notation from probability theory, this statement will look as follows:

(6) Pr(*you*|Subject) > Pr(*thou*|Subject)

This means that the conditional probability of *you* given the function of a subject is higher than the conditional probability of *thou*. In other words, *you* is more available than *thou*. A similar measure is category validity, or representativeness, in psychology of categorization (see Murphy and Ross 2005 for an overview).

The other type of probabilities is called *cue reliability*, which is also sometimes called cue validity (Rosch and Mervis 1975).[1] Cue reliability is high when the cue is never misleading or ambiguous. It can be expressed as the conditional probability of a grammatical function given the cue. For example, the probability that the pronoun *she* is a subject, and not an object, is higher than the probability that the pronoun *it* is a subject because it can also be used as an object:

(7) Pr(Subject|*she*) > Pr(Subject|*it*)

Therefore, *she* is a more reliable cue with regard to grammatical subject than *it*.

Cues can also be referential properties, such as animate, nominal, definite, etc. In the present study, cue availability is the conditional probability of a referential feature given a syntactic role (A or P), or Pr(Feature|Role), and cue reliability is the conditional probability of a syntactic role given a referential feature, or Pr(Role|Feature).

How do these measures relate to the theoretical explanations of DCM? Let us begin with markedness. There can be marked and unmarked A's, and marked and unmarked P's. Semantically marked categories in morphosyntax have marked forms, while semantically unmarked have unmarked forms. Singular nouns are unmarked (in both senses), while plural ones are marked; positive forms of adjectives are unmarked, while comparative forms are marked; present tense forms are unmarked, while future tense forms are marked, and so on (Greenberg 1966).

Unfortunately, markedness as an explanatory concept is very complex and involves multiple factors, such as frequency of use and semantics (Bybee 2010: 136; Haspelmath 2006). Speaking about markedness as a causal factor on its own is therefore not very informative. We follow here Greenberg's (1966) and Haspelmath's (2006) interpretation of markedness in terms of frequency.[2] According to Greenberg, unmarked members should also be the ones that are more frequent, while marked ones should be the ones that are less frequent.

From all this follows that A's with high Pr(Feature|A) and P's with high Pr(Feature|P) should be considered unmarked. For example, if we take 100 objects and 80 of them are inanimate, then the inanimate objects will be considered unmarked, representing 80/100 = 0.8, or 80% of all objects, whereas animate objects will be marked, representing 20%.[3] These measures reflect cue availability.

As for the efficiency hypothesis, it is based on the assumption that the speaker minimizes the effort, at the same time making sure that the listener interprets correctly who did what to whom. From the perspective of the listener, when hearing an NP, one should decide on its role in the sentence. When the referent has features that suggest a bias towards A or P, it helps the listener to decide. Thus, the semantic, pragmatic and formal properties serve as different cues for predicting the role. If the cues are reliable and strong, then additional marking is redundant. If they are weak or lead to wrong predictions, then the speaker provides additional cues – in

1 In the Competition Model, 'cue validity' is the joint product of cue availability and cue reliability.

2 We cannot exclude that some markedness phenomena are due to semantic markedness, e.g. inanimate objects may be inherently better patients than animate ones. In order to test that, we need experimental evidence, which would test these effects while controlling for frequency asymmetries.

3 Aissen (2003) also speaks about markedness reversal in the sense what is marked for objects is unmarked for subjects, and vice versa. This hypothesis is not pursued in the present study, but see Levshina (2018), who shows that predictions based on markedness reversal are not very well supported by corpus data.

particular, case marking. Therefore, the relevant probabilities here are Pr(Role|Feature), or cue reliability. Consider an illustration. If we take 100 inanimate referents, and 90 of them occur as objects, these 90 objects account for 90/100 = 0.9, or 90% of all inanimate referents. In that case, inanimacy will have high cue reliability with regard to P.

Using the typological claims about DCM summarized at the end of Section 1, we can now formulate the predictions displayed in Table 1. For example, if DOM is explained by iconicity of markedness, we can expect P's to be more frequently indefinite, because indefinite P's are less frequently marked formally than definite P's across the languages. To put it mathematically, we should find that Pr(Indefinite|P) is higher than Pr(Definite|P). These are cue availability measures. In contrast, if DOM is better explained by efficient communicative behaviour, we can expect indefinite referents to be more frequently P's than A's, because indefinite P's are less frequently marked, which suggests that their role should be easier to identify. In other words, we should expect Pr(P|Indefinite) to be higher than Pr(A|Indefinite). Thus, we perform a kind of reverse engineering, testing which of the corpus-based conditional probabilities and theoretical accounts associated with them might be responsible for the cross-linguistic generalizations in DCM.

Note that the predictions related to markedness reflect the asymmetries between A and P marking, whereas the predictions related to efficiency do not because cue reliability implies complementary relations between the proportions of A and P with a particular feature. The efficiency account thus over-generates predictions with regard to the DCM patterns observed in languages of the world. It is less dangerous than incorrect predictions. Moreover, we will see that markedness over-generates typological predictions, as well.

Some remarks are due about the corpus-based operationalization of the theoretical claims. Instead of the nominal versus pronominal distinction discussed in the literature, I will rely on the **lexical** versus **non-lexical** contrast, where 'non-lexical' includes pronominal and implicit arguments. The reason is that the pro-drop rates vary considerably across languages of the world. In languages like Lao, subject and object pro-drop is very frequent, whereas it is restricted in languages like English. This distinction has been fruitfully applied in previous research on preferred argument structure (Du Bois 1987; Du Bois et al. 2003). In addition, I have tested the nominal versus pronominal distinction, and the results lead to the same conclusions as the ones presented below. The reader can find these additional analyses in Appendix 3.

Another adjustment concerns the distinction between given and new referents, which is closely related to definiteness. It was added because definiteness as identifiability of the referent for both the speaker and the

Table 1: Theoretical predictions. A star (*) marks a tentative prediction.

Theoretical explanation	Relevant probability	Predictions based on cross-linguistic generalizations
Markedness	Probability of Feature given Role (cue availability)	– A's are usually non-lexical. – A's are usually 1st or 2nd person. – *A's are usually animate. – *A's are usually definite and given. – P's are usually lexical. – P's are usually inanimate. – P's are usually indefinite and new.
Efficiency	Probability of Role given Feature (cue reliability)	– Non-lexical arguments are usually A's. – Lexical arguments are usually P's. – 1st and 2nd person referents are usually A's. – 3rd person referents are usually P's. – Animates are usually A's. – Inanimates are usually P's. – Definite and given referents are usually A's. – Indefinite and new referents are usually P's.

hearer was often difficult to code. The expectations for givenness are the same as for definiteness. We will see that the results for givenness and definiteness are very similar.

Earlier studies used corpora of diverse languages for identification of referential properties of arguments, e.g. research on preferred argument structure in discourse (Du Bois 1987; Du Bois et al. 2003; Haig and Schnell 2016; see also Dahl 2000). Jäger (2007) provided frequencies from a corpus of spoken Swedish in order to explain differential case marking using a game-theoretic approach. However, this is the first time when the different types of probabilities are systematically compared in several typologically diverse languages with regard to the main features that play a role in marking splits.

3 Data from spoken corpora

Informal conversational interaction represents the primary mode of communication in which most linguistic innovations emerge and spread, although there is evidence that some innovations can also emerge in written communication (e.g., Biber and Gray 2011). This is why I use informal spoken dialogical data to obtain the conditional probabilities of features and roles. Of course, even this type of data may not represent all diverse daily experience of language users, but we can regard it as a reasonable approximation. Another crucial assumption is that the frequency distributions of the referential properties of A and P remain sufficiently stable. Therefore, the data from contemporary corpora of diverse languages can be extrapolated to previously existing language varieties in which differential case marking emerged. This assumption is common in functional typology, where frequency distributions derived from several corpora of several present-day languages are used for the purpose of explaining universal grammatical patterns (e.g. Du Bois 1987; Haig and Schnell 2016; Haspelmath and Karjus 2017). If several typologically diverse languages from different areas of the world – with or without DCM – converge, this can be regarded as a justification for this extrapolation.

The data for this case study represent manual annotations of transitive clauses of informal spoken conversations in five languages: English and Russian (Indo-European), Lao (Tai-Kadai), N||ng (Tuu) and Ruuli (Bantu). The details are provided in Appendix 1. Only the clauses with two-argument active, non-reflexive and non-reciprocal predicates were included. The datasets are available in Supplementary Materials. The annotation for English, Lao and Russian was performed by the author (Levshina 2020). The annotation for N||ng and Ruuli was performed by Alena Witzlack-Makarevich (the database is available upon request).

The arguments A and P were coded for diverse grammatical, semantic and pragmatic variables. The subset of variables and their values (i.e. referential features) used in the present study is as follows:

- Lexicality: lexical (common and proper nouns, adjectival or other nominalizations) or non-lexical (diverse pronouns and implicit arguments);
- Person: 1st, 2nd or 3rd;
- Semantic class: animate (human, animal, kinship term, organization) or inanimate (physical object, abstract entity, event);
- Definiteness (identifiability): definite, indefinite (specific or non-specific);
- Givenness (discourse accessibility): given (mentioned previously or inferable from context) or new.

More details about the variables and the coding procedure are provided in Appendix 1.

4 Quantitative analyses

4.1 Testing markedness: cue availability

This section discusses the measures of cue availability, which are used to test the markedness account of DCM. These are conditional probabilities of a feature given a role, or Pr(Feature|Role). It is computed as the frequency of all instances of a specific role with a specific feature divided by the total frequency of that role in a corpus:

(8) Cue availability $Pr(Feature|Role) = \frac{F(Role, Feature)}{F(Role)}$

The observations with missing values with regard to a particular feature were not included in *F(Role)*.

Figure 1 displays the probabilities of the features given the role A. The proportions related to complementary features (e.g. Non-lexical vs. Lexical, Animate vs. Inanimate) are a mirror image of each other. They are represented here for the sake of comparability with the plots that follow in the next section. The plot shows that A's are nearly exclusively non-lexical, animate, definite and discourse-given. As far as the person is concerned, the results are not conclusive. According to the scales, we would expect the languages to have predominantly 1st and 2nd person A's. This is what we find in Russian, English and N||ng, but Ruuli exhibits nearly equal probabilities of 1st/2nd person and 3rd person A's, while Lao even has more 3rd person A's than 1st/2nd person ones. This may have to do with the fact that the largest text in the Lao corpus can be characterized as gossip about other people. Overall, it seems that this type of probabilities must be highly sensitive to the text type and topic. When people tell stories or gossip, it is natural to use more 3rd person subjects than 1st or 2nd person subjects. Similar topical preferences seem to be responsible for the nearly equal proportions of the 1st or 2nd and 3rd person in Ruuli.

These results are corroborated by mixed-effects Poisson regression models. The dataset and R code are provided in the online supplementary materials. The procedure and results are discussed in Appendix 2. The models support our previous conclusions based on the visual inspection of the aggregate proportions. A's tend to be non-lexical, animate, definite and given, rather than lexical, inanimate, indefinite and new. These tendencies are statistically significant ($p < 0.05$). The results thus support the expectations based on the markedness account. As for the person, A indeed has preference for the 1st and 2nd person referents, but this preference is very weak and not significant at the conventional level of 0.05 ($p = 0.059$). Therefore, the predictions based on the markedness hypothesis are not supported.

Let us now turn to the most typical features of P with high Pr(Feature|P). Figure 2 shows the probabilities of the features given the role P. With the exception of Ruuli, most P's are non-lexical, which contradicts our expectations based on the markedness account. As for the person, the languages are unanimous: the 1st and 2nd person P's are very unlikely, while the 3rd person P's are extremely likely. P also exhibits a bias towards inanimacy in all five corpora, which agrees with the predictions. However, counter to the markedness hypothesis, it is more frequently definite and given than indefinite and new. These descriptive observations are supported by mixed-effects Poisson models with the observed frequencies of P's as the response and the contrasting features as the predictor (see Appendix 2). All these effects are statistically significant.

To summarize, we conclude that the corpus data provide little evidence that the cue availability measures reflecting the markedness hypothesis can be responsible for the cross-linguistic scale effects. Most

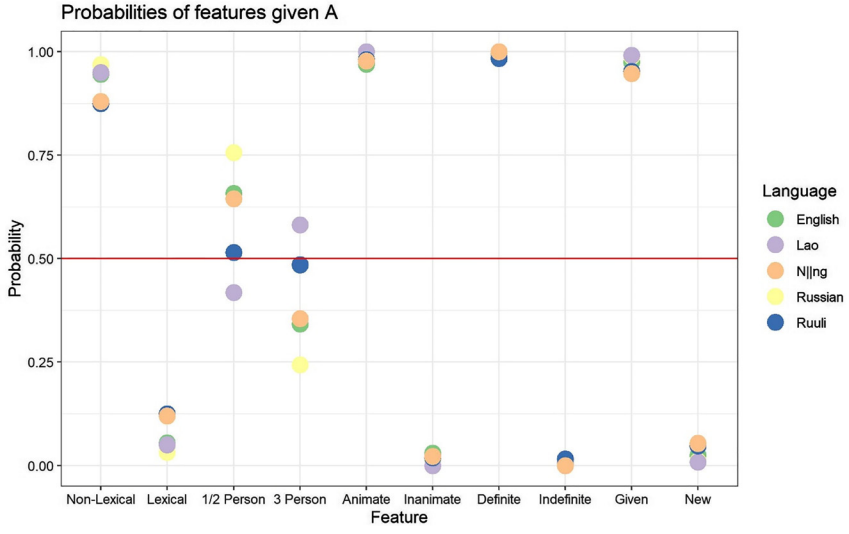

Figure 1: Cue availability: Probabilities of the features given A, or Pr (Feature|A), in five spoken corpora.

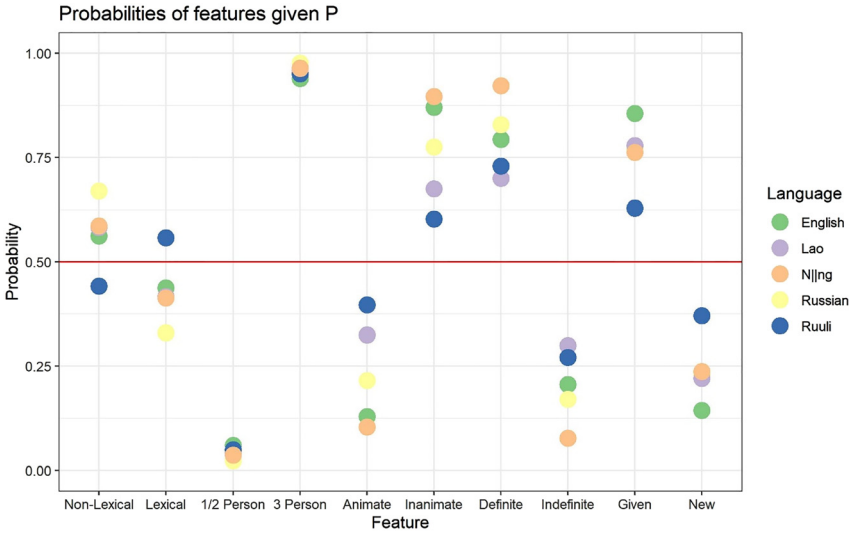

Figure 2: Cue availability: Probabilities of the features given P, or Pr (Feature|P), in five spoken corpora.

importantly, non-lexical, definite and discourse-given P's, which are usually formally marked in DOM, are in fact typical in the sense that they are significantly more frequent than lexical, indefinite and discourse-new P's. We also do not find support for the prediction that A's are usually 1st and 2nd person referents. The markedness account also somewhat over-generates predictions for the typological distribution of the splits. In particular, the preference of P for the 3rd person is very clear; person does not play an important role, however, in DOM across different languages.

4.2 Testing efficiency: cue reliability

This section discusses cue reliability, which represents the conditional probability of a particular role given a feature, or Pr(Role|Feature). It is computed as the frequency of all arguments in a specific argument role (A or P) with a specific feature (e.g. animate or indefinite) divided by the total frequency of the feature:

(9) Cue reliability $Pr\,(\,Role|Feature) = \frac{F\,(Role, Feature)}{F\,(Feature)}$

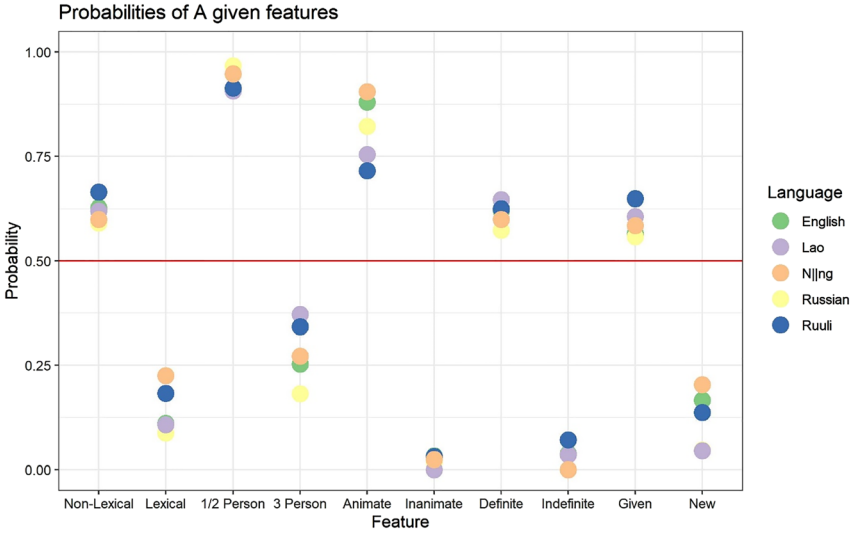

Figure 3: Cue reliability: Probabilities of the role A given features, or Pr (A|Feature), in five spoken corpora.

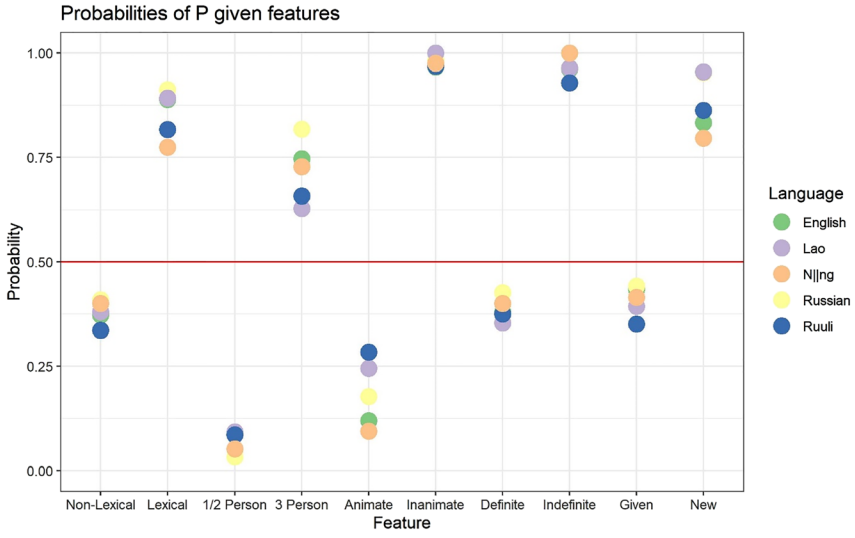

Figure 4: Cue reliability: Probabilities of the role P given features, or Pr(P|Feature), in five spoken corpora.

Figure 3 displays the probabilities of A given the ten features examined in this study. If a referent is non-lexical, first or second person, animate, definite and given, it is more likely to appear as A than P, as one can judge from the probabilities greater than 0.5 (or 50%). Their counterparts (nominal, third person, inanimate, indefinite and new) are all below 0.5. This means that arguments with these features are more likely to be P than A. Figure 4 displays the probabilities of P given the features, which is a mirror image of Figure 3. It is shown for the sake of completeness.

These observations are corroborated by mixed-effects models with individual languages and texts as random effects. All preferences are highly significant with $p < 0.0001$. See Appendix 2 for more details. Therefore, we can say that the cue reliability probabilities reflect the scales in (3), according to which more prominent arguments are more likely to be A, and less prominent ones are more likely to appear as P. Since we do not find any violations of the predictions in Table 1, we can conclude that cue reliability fits the cross-linguistic data.

5 Conclusions and discussion

This study investigated which of two accounts – iconicity of markedness or communicative efficiency – better explains the cross-linguistic effects of referential scales in differential case marking. We compared two types of corpus-based probabilities: conditional probabilities of a feature given a role, or Pr(Feature|Role), labelled as cue availability, and conditional probabilities of a role given a feature, or Pr(Role|Feature), also called cue reliability. Following the logic of Greenberg's (1966) approach to frequency as the central factor in grammatical markedness (see also Haspelmath 2006), cue availability is associated with the markedness account of differential case marking. Cue reliability is relevant for communicative efficiency, where the speaker needs to provide the listener with sufficient cues to understand who did what to whom.

The quantitative analyses suggest that cue reliability (efficiency) outperforms cue availability (markedness) in predicting the patterns observed cross-linguistically in differential case marking of A and P. In particular, languages with differential P marking conditioned on definiteness tend to mark definite objects formally, although definite (and given) P's are more frequent than indefinite (and new) objects in the spoken data and are therefore more typical ('unmarked') in this sense (cf. Jäger 2007). Moreover, the P's in our data are more frequently lexical than non-lexical, contrary to the predictions. We also do not find reliable evidence that A's are more frequently 1st and 2nd person than 3rd person referents, although this is what we could expect if the markedness explanation is correct. The predictions and results are presented in Table 2.

Table 2: Theoretical predictions and results. *If the efficiency account is correct, we can predict that other marking splits can also represents tentative predictions.

Theoretical explanation	Relevant probability	Predictions based on cross-linguistic generalizations	Results
Markedness (following Greenberg and Haspelmath)	Probability of Feature given Role (cue availability)	– A's are usually non-lexical.	Yes
		– A's are usually 1st or 2nd person.	No
		– *A's are usually animate.	Yes
		– *A's are usually definite and given.	Yes
		– P's are usually lexical.	No
		– P's are usually inanimate.	Yes
		– P's are usually indefinite and new.	No
Efficiency	Probability of Role given Feature (cue reliability)	– Non-lexical arguments are usually A's.	Yes
		– Lexical arguments are usually P's.	Yes
		– 1st and 2nd person referents are usually A's.	Yes
		– 3rd person referents are usually P's.	Yes
		– Animates are usually A's.	Yes
		– Inanimates are usually P's.	Yes
		– Definite and given referents are usually A's.	Yes
		– Indefinite and new referents are usually P's.	Yes

In contrast, the cue reliability probabilities, associated with efficiency, closely correspond to the scale effects found in typological data. For instance, indefinite and new referents are more frequently P's than A's, and 1st and 2nd person referents are more frequently A's than P's. Cue reliability is important for efficiency because it shows how much the listener can rely on particular semantic, pragmatic and formal features as cues for a particular syntactic role. If these cues are reliable and lead to correct expectations, the speaker can omit the case marker; if they lead to wrong expectations (e.g. A instead of P), the listener may need an extra cue, such as the case marker. Over time, this efficient behaviour may become conventionalized and grammaticalized.

If the efficiency account is correct, we can predict that other marking splits can also be explained by Pr(Function|Feature), where Features can be of any kind: semantic, pragmatic, lexical, formal, etc., and Function represents any kind of lexical meaning, syntactic role or grammatical category. The same predictions would hold for variable or optional marking. Low Pr(Function|Feature) will predict additional, longer or more consistent marking, while we can expect zero, shorter, or less consistent marking for high Pr(Function|Feature).

Predictability based on previous linguistic experience, which is represented by these probabilistic measures, is not the only type of predictability that can influence language production and comprehension. Another important type is the accessibility of information from immediate context, or common ground (Lestrade and de Hoop 2016). The role of different types of contextual knowledge and other cues, such as word order and verb agreement (Siewierska and Bakker 2008), also requires further investigation.

Sources of corpus data:

Du Bois, John W., Wallace L. Chafe, Charles Meyer, Sandra A. Thompson, Robert Englebretson and Nii Martey. 2000–2005. Santa Barbara Corpus of Spoken American English, Parts 1–4. Philadelphia: Linguistic Data Consortium.

Enfield, N.J. 2007. *A grammar of Lao*. Berlin: Mouton de Gruyter.

Güldemann, Tom, Martina Ernszt, Sven Siegmund and Alena Witzlack-Makarevich. 2007–2014. *A text documentation of Nluu*. Electronic corpus, available at http://elar.soas.ac.uk/deposit/0089.

Witzlack-Makarevich, Alena, Saudah Namyalo, Amos Atuhairwe, Anatole Kiriggwajjo and Zarina. 2017–2019. A corpus of spoken Ruuli. Electronic corpus (available upon request).

Zemskaja, Elena L., and L. A. Kapanadze (eds.). 1978. *Russkaja Razgovornaja Reč. Teksty [Russian colloquial speech. Texts]*. Moscow: Nauka.

Acknowledgements: I am sincerely grateful to the anonymous reviewers for their insightful comments, and to Susanne Flach, who was the responsible Area Editor for this manuscript. Special thanks go to Alena Witzlack-Makarevich for her help with data annotation, and for generously sharing the data related to N||ng and Ruuli. I also gratefully acknowledge funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement n° 670985) and from the Dutch Research Foundation NWO (Gravitation grant Language in Interaction, grant number 024.001.006). All usual disclaimers apply.

Appendices

Appendix 1 Information about the corpora and annotation

Table A1: Corpus data used in the case study.

Language	Corpus	Subset	Number of transitive clauses
American English	Santa Barbara Corpus of Spoken American English (Du Bois et al. 2000–2005)	Samples from files 001, 002, 003, 004, 005, 006, 007 and 011	201
Lao (Tam-Kadai; Laos)	Texts from Enfield (2007)	Texts A, B, C, D, F	120
N||ng (Tuu; South Africa)	Corpus by Güldemann et al. (2012)	Five conversations	225
Russian (Indo-European)	Collection of texts by Zemskaja & Kapanadze (1978)	"A day in a family", "Music and Theatre", "A conversation in the hotel", "At the dinner"	221
Ruuli (Bantu; Uganda)	Corpus by Witzlack-Makarevich et al. (2017–2019)	Five conversations	208

1) Text fragment selection

If the text is large, we select randomly a chunk of text from the beginning, middle or end of a dialogue. The size is to be decided upon individually, depending on the number of observations that we want to extract.

2) Conditions for coding a transitive predicate

We select only transitive predicates with either explicit or implicit A and P arguments. Imperatives and impersonal constructions are allowed. Ditransitives were excluded. Repetitive structures (e.g. *Take it, take it*) are counted only once.

In addition:

- The predicate is active, e.g. *Mary built a house*, and not *The house was built by Mary*.
- The predicate is not a reflexive or reciprocal verb, as in *He washes himself* or *They hugged each other*.
- The meaning of the predicate, A and P should be compositional, e.g. *Mary builds a house*, but not *The house caught fire*.
- The predicate is either simple or complex with the same subject, e.g. *Mary builds a house* or *Mary wants to build a house*, but not *Mary wants John to build a house*.
- The predicate is not a part of a discourse marker, e.g. ***You know**, it looks interesting*.

In Lao, which has a high pro-drop rate, there were many cases with implicit A and P. In that case, we relied on the detailed information in the translation (Enfield 2007), where implicit arguments were provided in parentheses, e.g. "(So the new boss) fired (Taa, you're) saying?" (p. 531).

3) Identification of A and P

- A and P can be explicit or implicit (e.g. A in imperative sentences or in impersonal constructions). If an argument is implicit, it is considered non-lexical. We coded the other features, to the extent this was possible (i.e. they are recoverable from the structure and the context). E.g. for *Do it!* A is coded as second person, animate, definite, given.
- For modal and phasal verbs, we only code the main verb with its P and the A of the matrix verb (e.g. *I can see you*, where *see* is the main verb, *I* and *you* are A and P, respectively). For other complement taking predicates, such as *I see that you are right*, A of the matrix clause (*I*) is a regular A, and P is coded as a clause. If the argument is a clause, it is non-lexical, 3rd person and abstract. Givenness and definiteness were coded as 'NA'.
- Only one pair of core arguments is allowed for one predicate. In the case of homogeneous parts (e.g. Mary and John), only the properties of the first NP are coded.

4) Coding of A and P:

- Lexical versus non-lexical:
  - Lexical (nouns, nominalized adjectives and other non-grammatical words);
  - Non-lexical (pronouns, implicit participants, complement clauses).

Three examples of modifier classifiers used without nouns were detected in the Lao texts. They were coded as non-lexical, as well.

- Grammatical Person:
  - 1st,
  - 2nd,
  - 3rd (including complement clauses as arguments),
  - undefined (difficult to say),
  - not applicable (impersonal constructions).

The coding of implicit arguments is based on context. Animacy:

- animate (human, animal, anthropomorphic, kinship terms, organization, generic animate),
- inanimate (physical object, abstract entity, event, generic inanimate, complement clauses),
- undefined (difficult to say).

Metonymy and metaphor were coded depending on their target semantic domain, e.g. in *I play Schubert*, *Schubert* is abstract. The coding of implicit arguments is based on context.

- Definiteness, or identifiability:
  - definite (identifiable by the speaker and the hearer, unique in its kind, anaphoric or uniquely identifiable from context),
  - indefinite: specific (identifiable only by the speaker, can be replaced by "a certain") or non-specific (not identifiable),
  - generic (e.g. *Doctors treat patients*),
  - undefined (difficult to say),
  - not applicable (impersonal uses, clauses, interrogative, negative and relative pronouns).

Implicit arguments are usually definite (e.g. pro-drop). Givenness, or referent accessibility:

- given (introduced in the previous context or inferable from the context),
- new,
- undefined (difficult to say),
- not applicable (impersonal uses, clauses, interrogative, negative and relative pronouns).

Implicit arguments are usually given.

Undefined and 'other' values, impersonal uses, clauses, interrogative, negative and relative pronouns were disregarded in the analyses, as well as generic uses (definiteness) were treated as missing values and left out of the calculations.

One should make a separate note on the inclusion of sentences with finite and non-finite complements in the role of P, as in the following sentences:

(A1) a. *I was imagining he had broke an arm or something*. (SBC002)
b. *Would you like to string the beans?* (SBC003)

Analyses presented below are based on the data including such cases. Complements were considered to be abstract, 3rd person, and non-lexical; definiteness and givenness were not coded (missing values). In order to make sure that this decision does not distort the results, a subset of contexts without clausal arguments was tested separately. The results, which differ very little from the results presented below, can be found in Appendix 4.

The annotation was performed with the help of the RShiny software for annotation in https://github.com/levshina/CARAT, version 2. Note that the original coding scheme was more detailed. The values were conflated as described above.

Analysis of interrater agreement has been performed on a fragment of the English and Russian data. Several rounds and discussions were necessary to establish the annotation guidelines that produced robust interrater agreement (see above). Because of extreme skewedness of some variables, only the proportion of overlapping coding decisions was computed. The proportions of identical choices made by the annotators were from 84 to 100%.

Appendix 2 Mixed-effects models

1. Probability of referential features given A (cue availability)

The response variable was the number (frequency) of A's with contrasting features (i.e., non-lexical vs. lexical, animate vs. inanimate, and so on). Each model tested which of the features in a contrast, e.g. animate or inanimate, was more frequent in all A's in the data. Since the data are hierarchical and are collected from five languages and diverse texts, we also included random intercepts for each of the languages and for every individual text within a language.

A series of likelihood ratio tests was performed in order to see whether it made sense to include random slopes or not. This turned out to be useful in the models of part of speech and person. As for the other features, the random-slope models had boundary singular fit due to data sparseness. This was also the reason why we did not test the random slopes of the individual texts within the languages. The details are provided in Table A2, which lists the beta coefficients and the corresponding *p*-values. The coefficients are all positive, which means that the features in bold, which correspond to high prominence, have higher chances of occurrence than their counterparts. Therefore, A's tend to be non-lexical, 1st and 2nd person, animate, definite and given, as expected on the basis of the scales in (3). However, for the person this tendency is not statistically significant, as we could already suspect from the descriptive statistics above.

Table A2: Mixed-effects Poisson regression models: A role.

Referential Features	Type of random effects	Beta coefficient for value in bold	*p*-Value of the beta coefficient
Non-Lexical –Lexical	Random intercepts and slopes	2.6	<0.0001
1st&2nd person – 3rd person	Random intercepts and slopes	0.42	0.059
Animate – Inanimate	Random intercepts only	3.97	<0.0001
Definite – Indefinite	Random intercepts only	5.12	<0.0001
Given – New	Random intercepts only	3.38	<0.0001

2. Probability of referential features given P (cue availability)

The response variable was the number (frequency) of P's. As in the models described above, random intercepts for the languages and individual texts were included. The random slopes turned out to improve the models significantly, according to the likelihood ratio tests, with the exception of the person model. The results are displayed in Table A3. The coefficients show if the features in bold, which correspond to high prominence, are more frequent than their counterparts. According to the scales in (3), we would expect the coefficients to be negative because P's are considered to be associated with low-prominence referents. This is the case only for 1st and 2nd person and animate arguments, which are untypical of P's. However, as the descriptive analyses above suggest, non-lexical (although the effect size is very small), definite and given P's are more frequent than indefinite and new ones, contrary to the expectations.

Table A3: Mixed-effects Poisson regression models: P role.

Referential Features	Type of random effects	Beta coefficient for value in bold	*p*-Value of the beta coefficient
Non-Lexical – Lexical	Random intercepts and random slopes	0.28	0.044
1st&2nd person – 3rd person	Random intercepts only	−3.14	<0.0001
Animate – Inanimate	Random intercepts and random slopes	−1.29	<0.0001
Definite – Indefinite	Random intercepts and random slopes	1.43	<0.0001
Given – New	Random intercepts and random slopes	1.19	<0.0001

3. Probability of A or P given referential features (cue reliability)

The response variable was the frequencies of A's and P's. The main results are displayed in Table A4. Random slopes turned out to be important only in the models related to 1st and 2nd person, 3rd person, and animate participants. A positive coefficient indicates that this feature increases the frequency of P, while a negative coefficient means that this feature increases the frequency of A. In all cases, the observed asymmetries, which were described above, remain significant. This means that the languages are unanimous with regard to the probabilities of particular features to be either A or P.

Table A4: Mixed-effects Poisson models predicting the frequencies of A versus P for specific features.

Referential Feature	Type of random effects	Beta coefficient (effect in favour of P and against A)	*p*-Value of the beta coefficient
Non-lexical	Random intercepts only	−0.48	<0.0001
Lexical	Random intercepts only	1.70	<0.0001
1st&2nd person	Random intercepts only	−2.66	<0.0001
3rd person	Random intercepts and slopes	0.94	<0.0001
Animate	Random intercepts and slopes	−1.56	<0.0001
Inanimate	Random intercepts only	3.73	<0.0001
Definite	Random intercepts only	−0.44	<0.0001
Indefinite	Random intercepts only	3.26	<0.0001
Given	Random intercepts only	−0.37	<0.0001
New	Random intercepts only	1.89	<0.0001

Appendix 3 Plots with the pronominal versus nominal distinction, instead of the lexical versus non-lexical one

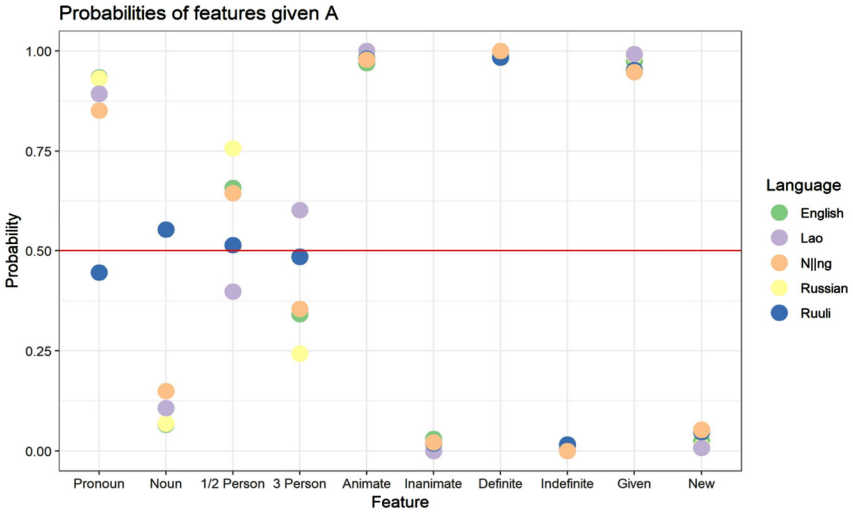

Figure A1: Probabilities of the features given role A, or Pr(Feature|A), in five spoken corpora.

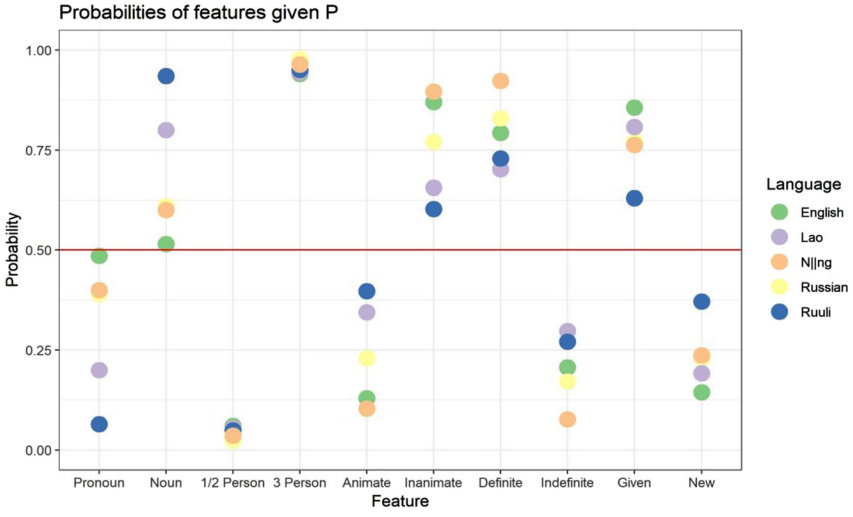

Figure A2: Probabilities of the features given role P, or Pr(Feature|A), in five spoken corpora.

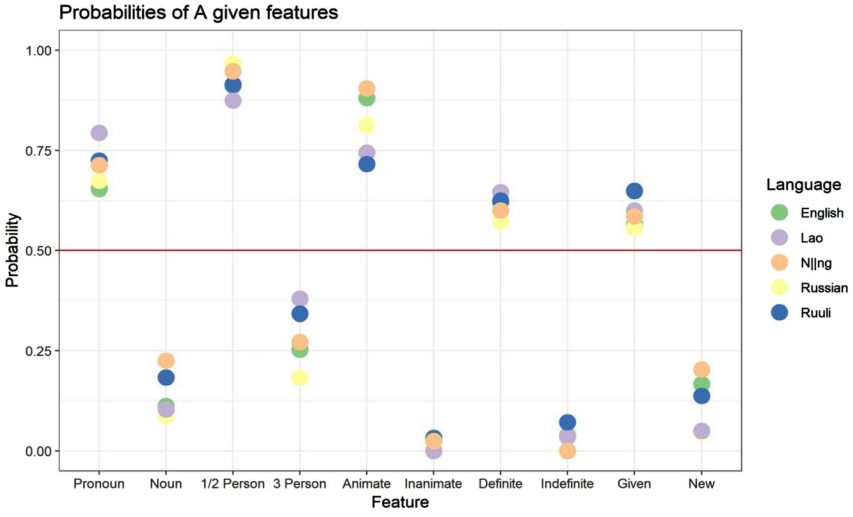

Figure A3: Probabilities of the role A given features, or Pr(A|Feature), in five spoken corpora.

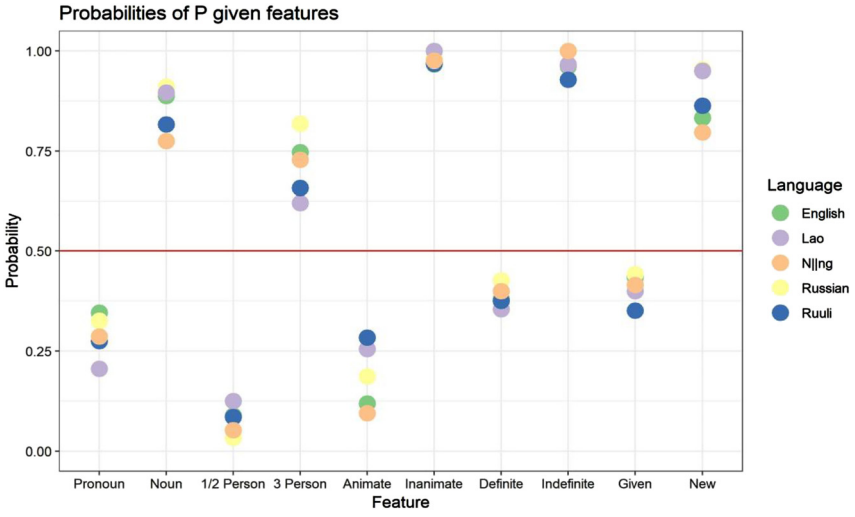

Figure A4: Probabilities of the role P given features, or Pr(P|Feature), in five spoken corpora.

Appendix 4 Corpus-based probabilities without P's expressed by complement clauses

All clauses with finite and non-finite complements were excluded from the analyses.

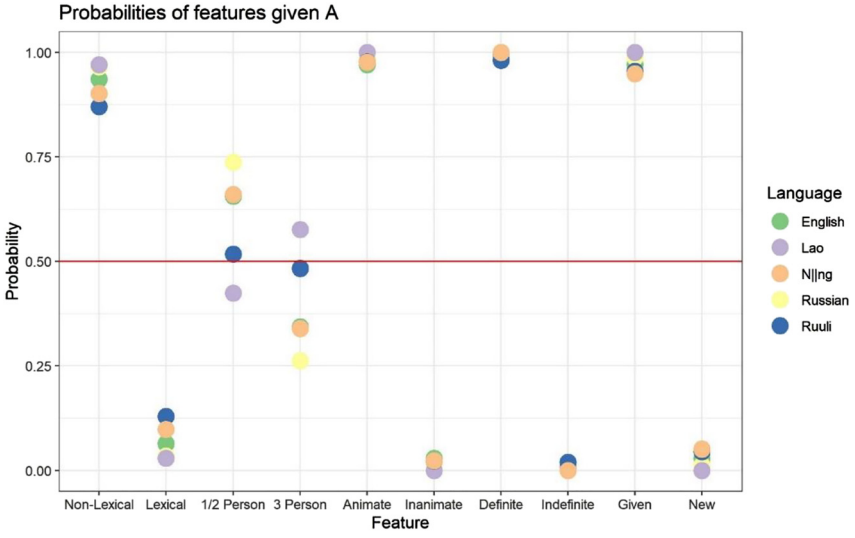

Figure A5: Probabilities of the features given role A, or Pr(Feature|A), in five spoken corpora.

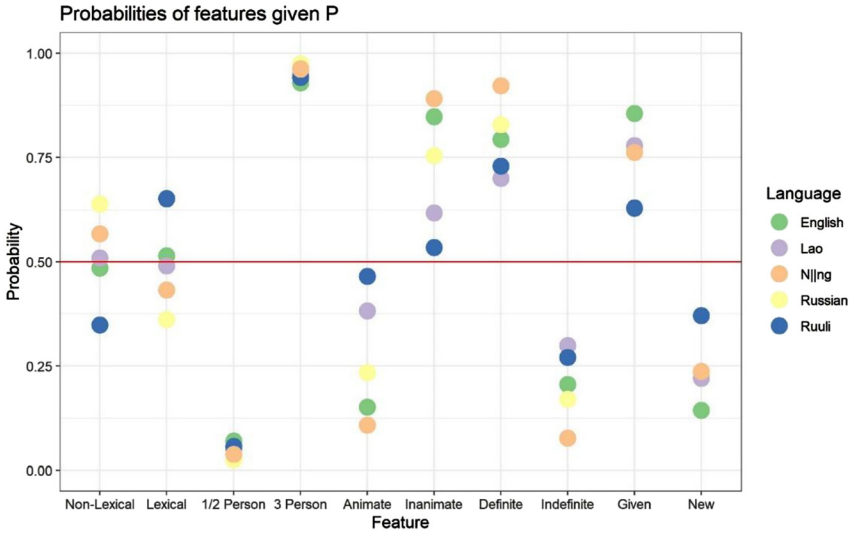

Figure A6: Probabilities of the features given role P, or Pr(Feature|P), in five spoken corpora.

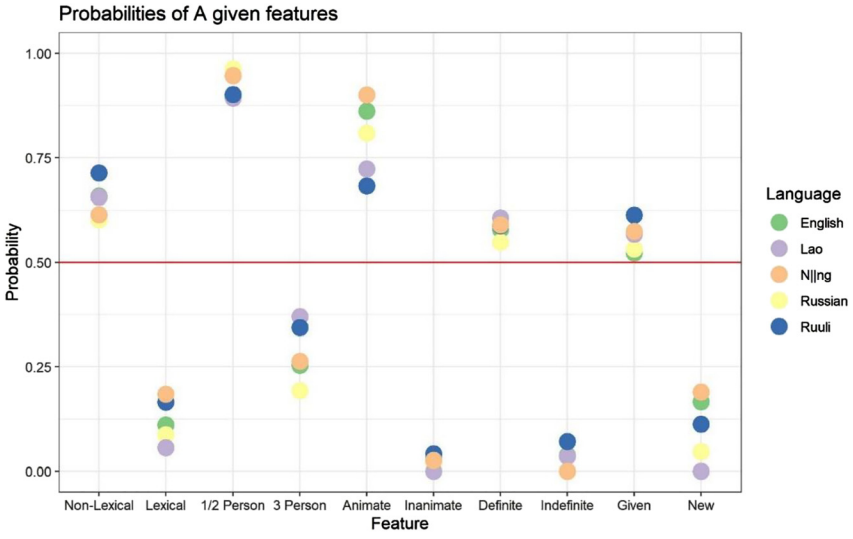

Figure A7: Probabilities of the role A given features, or Pr(A|Feature), in five spoken corpora.

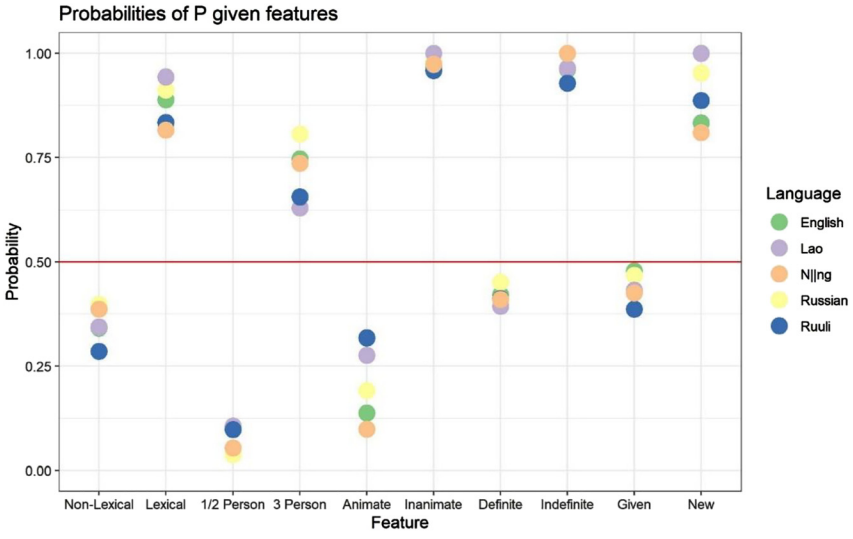

Figure A8: Probabilities of the role P given features, or Pr(P|Feature), in five spoken corpora.

References

Aissen, Judith. 2003. Differential object marking: Iconicity vs. economy. *Natural Language and Linguistic Theory* 21. 435–483.

Ariel, Mira. 1990. *Accessing noun-phrase antecedents*. London: Routledge.

Bell, Alan, Jason Brenier, Michelle Gregory, Cynthia Girand & Dan Jurafsky. 2009. Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language* 60(1). 92-111.

Biber, Douglas & Bethany Gray. 2011. Grammar emerging in the noun phrase: The influence of written language use. *English Language and Linguistics* 15. 223–250.

Bickel, Balthasar, Alena Witzlack-Makarevich & Taras Zakharko. 2015. Typological evidence against universal effects of referential scales on case alignment. In Ina Bornkessel-Schlesewsky, Andrej L. Malchukov & Marc Richards (eds.), *Scales and hierarchies: A cross-disciplinary perspective*, 7–43. Berlin & Boston: De Gruyter Mouton.

Bossong, Georg. 1985. *Empirische Universalienforschung: Differentielle Objektmarkierung in den neuiranischen Sprachen*. Tübingen: Narr.

Bybee Joan. 2010. Markedness: iconicity, economy and frequency. In Jae Jong Song (ed.), *Handbook of linguistic typology*, 131–147. Oxford: Oxford University Press.

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

Du Bois, John W., Lorraine E. Kumpf & William J. Ashby (eds.). 2003. *Preferred argument structure: Grammar as architecture for function*. (Studies in discourse and grammar 14.) Amsterdam: John Benjamins.

Comrie, Bernard. 1978. Ergativity. In W. P. Lehmann (ed.), *Syntactic typology. Studies in the phenomenology of language*, 329–394. Austin: The University of Texas Press.

Comrie, Bernard. 1981. *Language universals and linguistic typology*. Oxford: Blackwell.

Croft, William. 2003. *Typology and universals*. 2nd edn. Cambridge: Cambridge University Press.

Dahl, Östen. 2000. Egophoricity in discourse and syntax. *Functions of Language* 7(1). 37–77.

Dixon, R. M. W. 1979. Ergativity. *Language* 55. 59–138.

Dowty, David. 1991. Thematic proto-roles and argument selection. *Language* 67(3). 547–619.

Fauconnier, Stefanie. 2011. Differential agent marking and animacy. *Lingua* 121(3). 533–547.

García García, Marco. 2018. Nominal and verbal parameters in the diachrony of differential object marking in Spanish. In Ilja Seržant & Alena Witzlack-Makarevich (eds.), *Diachrony of differential argument marking*, 209–242. Berlin: Language Science Press.

Gibson, Edward, Richard Futrell, Steven T. Piantadosi, Isabelle Dautriche, Kyle Mahowald, Leon Bergen & Roger Levy. 2019. How efficiency shapes human language. *Trends in Cognitive Sciences* 23(5). 389–407.

Greenberg, Joseph. 1966. *Language universals, with special reference to feature hierarchies*. (Janua Linguarum, Series Minor, 59.) The Hague: Mouton.

Haig, Geoffrey & Stefan Schnell. 2016. The discourse basis of ergativity revisited. *Language* 92(3). 591–618.

Haspelmath, Martin. 2006. Against markedness (and what to replace it with). *Journal of Linguistics* 42(1). 25–70.

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

Haspelmath, Martin & Andres Karjus. 2017. Explaining asymmetries in number marking: Singulatives, pluratives and usage frequency. *Linguistics* 55(6). 1213–1235.

de Hoop, Helen & Andrej L. Malchukov. 2008. Case-marking strategies. *Linguistic Inquiry* 39(4). 565–587.

de Hoop, Helen & Peter de Swart (eds.). 2008. *Differential subject marking*. Dordrecht: Springer.

von Heusinger, Klaus & Georg A. Kaiser. 2007. Differential object marking and the lexical semantics of verbs in Spanish. In Georg A. Kaiser & Manuel Leonetti (eds.), *Proceedings of the Workshop "Definiteness, Specificity and Animacy in Ibero-Romance Languages", Konstanz*, 83–109. Konstanz: Universität Konstanz, Fachbereich Sprachwissenschaft.

Iemmolo, Giorgio. 2010. Topicality and differential object marking: evidence from Romance and beyond. *Studies in Language* 34(2). 239–272.

Jaeger, T. Florian. 2010. Redundancy and reduction: speakers manage syntactic information density. *Cognitive Psychology* 61 (1). 23–62.

Jaeger, T. Florian & Esteban Buz. 2017. Signal reduction and linguistic encoding. In Eva M. Fernández & Helen Smith Cairns (eds.), *Handbook of psycholinguistics*, 38–81. Wiley-Blackwell.

Jäger, Gerhard. 2007. Evolutionary game theory and typology. A case study. *Language* 83(1). 74-109.

Jakobson, Roman. 1971 [1932]. Zur Structur des russischen Verbums. In: Roman Jakobson. *Selected Writings. Vol. II. Word and Language*, 3–15. Berlin: De Gruyter Mouton.

LaPolla, Randy J. & Chenglong Huang. 2003. *A grammar of Qiang with annotated texts and glossary*. Berlin: De Gruyter Mouton.

Lee, Hanjung. 2008. Quantitative variation in Korean case ellipsis: implications for case theory. In Helen de Hoop & Peter de Swart (eds.), *Differential subject marking*, 41–61. Dordrecht: Springer.

Lestrade, Sander & Helen de Hoop. 2016. On case and tense: The role of grounding in differential subject marking. *The Linguistic Review* 33(3). 397–410.

Levshina, Natalia. 2018. *Towards a theory of communicative efficiency in human languages.* Leipzig: University of Leipzig habilitation thesis.
Levshina, Natalia. 2020. Database of annotated core arguments: English, Lao and Russian (Version 1.0) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.4065523.
MacWhinney, Brian, Elizabeth Bates & Reinhold Kliegl. 1984. Cue validity and sentence interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behavior* 23(2). 127–150.
Malchukov, Andrej L. 2008. Animacy and asymmetries in differential case marking. *Lingua* 118(2). 203–221.
Murphy, Grerogy L. & Brian H. Ross. 2005. Two faces of typicality in category-based induction. *Cognition* 95. 175–200.
Rosch, Eleanor & Carolyn B. Mervis. 1975. Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology* 7. 573–605.
Schmidtke-Bode, Karsten & Natalia Levshina. 2018. Reassessing scale effects on differential case marking: Methodological, conceptual and theoretical issues in the quest for a universal. In Ilja A. Seržant & Alena Witzlack-Makarevich (eds.), *Diachronic typology of differential argument marking*, 509–537. Berlin: Language Science Press.
Seržant, Ilja. 2019. Weak universal forces: The discriminatory function of case in differential object marking systems. In Karsten Schmidtke-Bode, Natalia Levshina, Susanne M. Michaelis & Ilja Seržant (eds.), *Explanation in typology: Diachronic sources, functional motivations and the nature of the evidence*, 149–178. Berlin: Language Science Press.
Siewierska, Anna & Dik Bakker. 2008. Case and alternative strategies. In Andrej Malchukov & Andrew Spencer (eds.), *Handbook of case*, 290–303. Oxford: Oxford University Press.
Silverstein, Michael. 1976. Hierarchy of features and ergativity. In R. M. W. Dixon (ed.), *Grammatical categories in Australian languages*, 112–171. Canberra: Australian Institute for Aboriginal Studies, Canberra.
Sinnemäki, Kaius. 2014. A typological perspective on differential object marking. *Linguistics* 52(2). 281–313.
de Swart, Peter. 2005. Noun phrase resolution: the correlation between case and ambiguity. In Mengistu Amberber & Helen de Hoop (eds.), *Competition and variation in natural languages: The case for case*, 205–222. Oxford: Elsevier.
Tal, Shira, Kenny Smith, Jennifer Culbertson, Eitan Grossman & Inbal Arnon. 2020. The impact of information structure on the emergence of differential object marking: an experimental study. PsyArXiv. https://psyarxiv.com/759gm.
Witzlack-Makarevich, Alena & Ilja Seržant. 2018. Differential argument marking: Patterns of variation. In Ilja Seržant & Alena Witzlack-Makarevich (eds.), *Diachrony of differential argument marking*, 1–40. Berlin: Language Science Press.