Analogy in inflection and morphosyntax

András Bárány, andras.barany@uni-bielefeld.de, Bielefeld University
Jutta-Maria Hartmann, jutta-maria.hartmann@uni-bielefeld.de, Bielefeld University

We discuss analogical extension in complex syntactic structures involving inflection and other morphosyntactic properties in Hungarian and German.

Long-distance object agreement in Hungarian In Hungarian, (mostly) definite direct objects trigger object agreement (Bartos 1999, É. Kiss 2002, Bárány 2017). In addition, Hungarian shows **long-distance agreement (LDA)**, as in (1a), where the transitive matrix verb agrees with the infinitive's object, which is not selected by the matrix verb. Were the infinitive's object an indefinite object such as *egy verset* 'a poem', the matrix verb would not show object agreement.

(1) Anna meg-próbált-a [$_{INF}$ megtanul-ni a vers-et].
Anna try-PST-3SG.SBJ>3.OBJ learn-INF the poem-ACC
'Anna tried to learn the poem by heart.' (Kenesei et al. 1998: 33)

It has long been known that certain **intransitive** verbs (lacking acc objects) also participate in long-distance object agreement (cf. É. Kiss 1987, Kálmán C. et al. 1989, den Dikken 1999, Szécsényi & Szécsényi 2020). While this often treated as exceptional, Bárány (2020) argues that some speakers **analogically extend** and to some degree regularise LDA to **intransitive matrix verbs**, as in (2).

(2) %Igyeksz-itek [$_{INF}$ ez-t a rémálm-ot elfelejte-ni].
strive-2PL.SBJ>3.OBJ this-ACC the nightmare-ACC forget-INF
'You strive to forget this nightmare.' (Bárány 2020: 55)

This analogical process creates paradigmatically novel forms, namely intransitive verbs with transitive agreement inflection. However, the proposed analogy crucially relies on syntactic structure, namely the configuration shown in (3). For most speakers, only transitive finite verbs can (and must) show LDA, although the superordinate verb does not select the embedded object. Extending this pattern to intransitive verbs thus involves reference to the whole structure in (3).

(3) Schematic structure of LDA
[… finite verb+AGR … [INF infinitive OBJ-ACC]]

German long-distance dependencies A second empirical domain concerns **long-distance dependencies (LDDs)** in German. Based on the observation that some (e.g. Southern) German speakers allow certain long-distance wh- and topicalisation-dependencies (as in (4)), we hypothesise that these speakers rate other LDDs, in particular weak islands, as more acceptable than other speakers who generally do not accept them. Higher acceptability of dependencies across islands could be due to surface or structural similarity to LDDs which are not island violations, such as (4).

(4) %[FILLER *Welchen Priester*] *glaubt er, dass der Bischof* **GAP**
which.ACC priest believe.3SG he that the bishop
ermahnt hat?
reprimand.PTCP AUX.3SG
'Which priest does he think that the bishop reprimanded?'

We tested this hypothesis by collecting acceptability judgements of LDDs that are superficially similar but structurally different (psych verbs) as well as structures that are both superficially and structurally similar (factive islands) to LDDs such as (5). Results from this

pilot study suggest that, indeed, speakers who accept (4) rate other LDDs (weak islands) as more acceptable (a strong effect, Cohen's d = 1.08, corresponding to an average difference of 1.5 points on a 7 point Likert scale; see Figure 1) than speakers who do not accept (4). The more permissive speakers rate LDDs with weak islands as worse than (4) but as more acceptable than less permissive speakers rate either type of structure.

Analogy in morphosyntax Both of these phenomena raise the possibility that complex syntactic structures serve as the basis for analogical extension. Moreover, this extension may not just involve superficial similarities. Hungarian LDA with (in)transitive superordinate verbs appears in virtually all possible linear orders, suggesting that **structurally**, not just superficially similar structures are involved in licensing the acceptability of LDA. The acceptability of these novel structures is, on the one hand, supported by analogical, similar structures, but limited by other factors such as the oddity of intransitive verb forms showing object agreement, as in (2), and the well-known difficulties of structures involving islands. We suspect that these opposing pressures limit the productivity and spread of the phenomena in question.

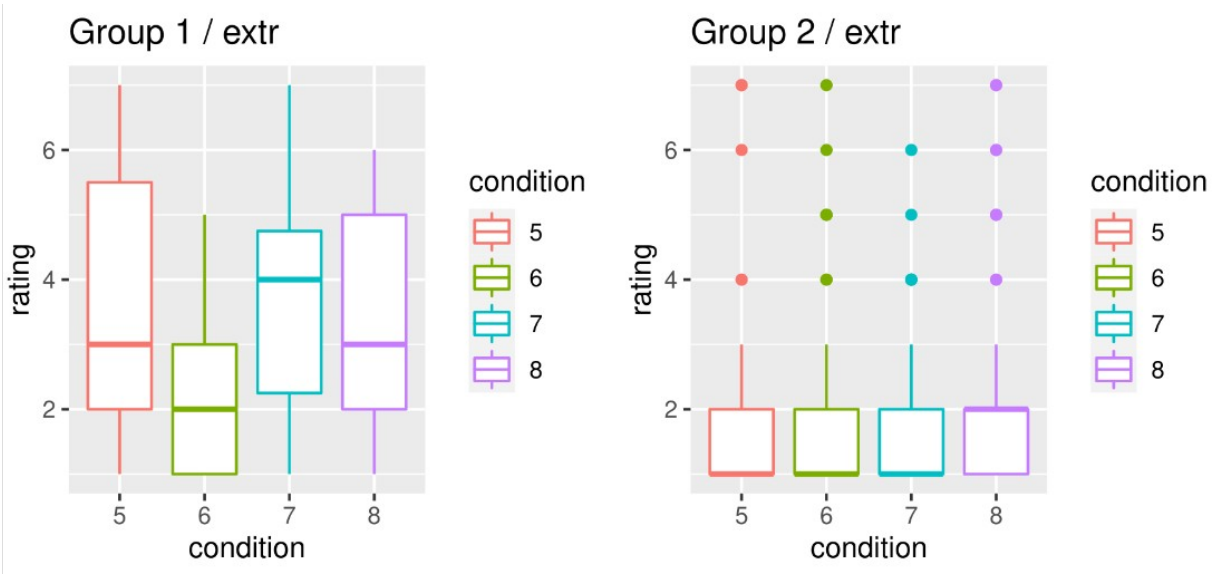


Figure 1: Acceptability ratings (1–7, y-axis) by permissive speakers (Group 1, left) vs. less permissive speakers (Group 2, right) on experimental conditions involving extraction (5–8, x-axis).

Bibliography

Bárány, András. 2017. *Person, case, and agreement*. Oxford: Oxford University Press.

Bárány, András. 2020. Hosszú távú egyeztetés intranzitív igékkel a magyarban [Long-distance agreement with intransitive verbs in Hungarian]. *Általános Nyelvészeti Tanulmányok* 32. 55–68.

Bartos, Huba. 1999. *Morfoszintaxis és interpretáció: A magyar inflexiós jelenségek szintaktikai háttere* [Morphosyntax and interpretation: the syntactic background of Hungarian inflectional phenomena]. Budapest: ELTE dissertation.

den Dikken, Marcel. 1999. On the structural representation of possession and agreement. In *Crossing boundaries*. István Kenesei (ed.). Amsterdam: John Benjamins. 137–178. https://doi.org/10.1075/cilt.182.10dik.

É. Kiss, Katalin. 1987. *Configurationality in Hungarian*. Dordrecht: Reidel.

É. Kiss, Katalin. 2002. *The syntax of Hungarian*. Cambridge: Cambridge University Press.

Kálmán C., György et al. 1989. A magyar segédigék rendszere [The system of Hungarian auxiliary verbs]. *Általános Nyelvészeti Tanulmányok* 17. 49–103.

Kenesei, István et al. 1998. *Hungarian*. London: Routledge.

Szécsényi, Krisztina & Tibor Szécsényi. 2020. Object agreement and locality in Hungarian: Infinitival complement clauses, second person objects and accusative adjuncts. *In Approaches to Hungarian: Volume 16: Papers from the 2017 Budapest conference*. Irene Vogel & Veronika Hegedűs (eds.). Amsterdam: John Benjamins. 166–186. https://doi.org/10.1075/atoh.16.08sze.

The role of analogous morphological representations of previously acquired lexical entries in incidental acquisition in L1and L2

Denisa Bordag, Leipzig University, Germany, denisav@uni-leipzig.de
Andreas Opitz, Leipzig University, Germany, andreas.opitz@uni-leipzig.de

In a psycholinguistic study, we addressed the question whether previously acquired linguistic knowledge concerning existing morphologically structured lexical entries is used for the creation of new, complex lexical entries. We asked whether new mental representations are idiosyncratic in that they contain only grammatical information derived from the linguistic context in which the new word appeared, or whether their establishment is guided also by principles of analogy to previously established lexical entries. Moreover, we asked whether analogy-driven formation of lexical units is moderated by a speaker's particular language proficiency or whether it is relatively independent from it by comparing native (L1) and (advanced but) non-native (L2) speakers.

Empirically, we focussed on a domain that allows the study of complex lexical entries while excluding confounding factors such as form differences. German inflectional/morphological families exhibit several syncretic forms, for example: infinitives (e.g., *SPIELEN* 'to play'), inflected forms for 2nd person plural (*wir SPIELEN* 'we play'), and conversion forms (*das SPIELEN* 'the playing') - all share a form-identical affix (*-en*) that has different grammatical functions. In our study we focussed on the acquisition of conversion which is a borderline phenomenon sometimes considered as a special instance of inflection (cf. e.g., Haspelmath, 1996, who argues for an extended class of inflections also including word-class changing processes if they are regular, general, and productive, but nonetheless transpositional). Recent research (e.g., Bordag & Opitz, 2021) indicates that, psycholinguistically, conversion nouns are nested as word-category-specific subentries under a more basic (probably underspecified) lexical entry that comprises also subentries for verbal representations (e.g., inflected forms). Having these properties, conversion was a suitable phenomenon to explore the role of analogy in establishing new, complex lexical entries. Since all German verbs can be converted into nouns, does a conversion noun need to be encountered in the input so that its representation can be established, or can its establishment be triggered by an encounter with a verb form alone based on analogy with earlier established equivalent lexical entries?

Method: We conducted four self-paced reading experiments: two with native (L1) speakers of German (N=72 in experiment L1a, N=70 in experiment L1b) and two with advanced learners of German (L2) with Czech as their native language (N=72 in experiment L2a, N=68 in experiment L2b). In each experiment, participants read 24 short texts containing a novel word (pseudoword). The pseudowords were presented *either in inflected present tense forms* (experiments L1a & L2a), *or as nominalised forms* (i.e., conversion nouns; in experiments L1b & L2b). After each text, participants read several sentences in a self-paced reading (SPR) manner. One of the sentences always contained the pseudoword in one of four different functions yielding the four conditions of the experiment (see *Examples*). Reading times were measured. The rational was that reading times reflect to which degree participants were able to access the representation of the pseudoword in a given function. We hypothesised that: (i) if participants established an idiosyncratic initial lexical representation only of the form/function immediately present in the input (preceding texts), then only this form should benefit most in terms of reading times; (ii) if participants established more complex initial lexical entries based on analogy, then all other related forms (within the same structured lexical entry) should profit from the repetition in the SPR-sentence to the same degree.

Results (see *Figure*) for **native speakers** indicate involvement of general knowledge about lexical structures and constructing of new entries based on analogy to existing entries. Reading times were equally fast for inflected verbs, conversion nouns, and infinitives, regardless of the grammatical or function in which the new word (inflected verb vs. conversion

noun) was presented in the introductory texts. For **non-native participants**, these results were only partially replicated. While they also showed faster reading times for all forms if the novel word was introduced as an inflected verb in the text (experiment L2a), the reduction/facilitation in reading times was less pronounced than for native participants. Moreover, if the novel word was introduced as a conversion form (experiment L2b), reading times were only shorter for the conversion form itself in the SPR sentences – not for all related (verbal) forms, indicating that L2 participants failed to make use of analogically structured existing entries to the same extend as L1 participants.

Summary: Our findings indicate that although both groups of speakers employ knowledge about existing lexical representations when establishing new, analogically structured lexical entries, native speakers outperform non-native learners. Their newly established representations have more precise structure and are better differentiated from related representations with respect to their grammatical information (cf. fuzzy lexical representation hypothesis, Cook & Gor, 2015;.Gor et al., 2021).

Examples of self-paced reading sentences:

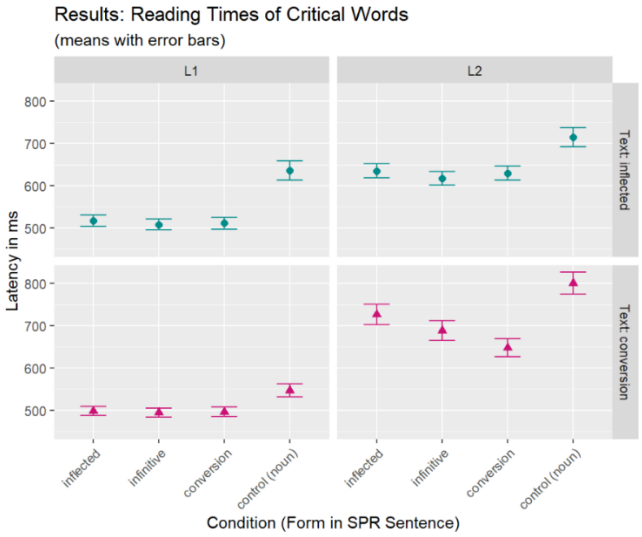
1. Self-Paced Reading of Sentences

a) **infinitive:** VIELE LEUTE WOLLEN NUR BRÖSSEN, ANSTATT […]
"A lot of people just want to [gawk] instead of doing something themselves."

b) **inflected verb:** VIELE LEUTE KOMMEN NUR UND BRÖSSEN, ANSTATT […]
"A lot of people just come and [gawk] instead of doing something themselves."

c) **conversion noun:** VIELE LEUTE KOMMEN NUR FÜR DAS BRÖSSEN, […]
"A lot of people come just for the [gawking] instead of doing anything themselves."

d) **countable noun** (= control condition):
DIE LEUTE KOMMEN NUR FÜR DIE VIELEN BRÖSSEN, ANSTATT […]
"A lot of people come just for all the / for the many XXX instead of […]



Reference

Cook, S. V. & Gor, K. (2015). Lexical access in L2: Representational deficit or processing constraint? *The Mental Lexicon*, 10, 247-270.

Haspelmath, M. (1996). Word-class-changing inflection and morphological theory. In G. Booij & J. van Marle (Eds.), *Yearbook of Morphology 1995* (pp. 43-66). Dordrecht, Netherlands: Springer.

Bordag, D. & Opitz, A. (2021). Representation and processing of overtly identical complex forms in L1 and L2: The case of conversion and its 'siblings'. *Studies in Second Language Acquisition*, 2021:S0272263121000206.

Gor, K., Cook, S., Bordag, D., Chrabaszcz, A. & Opitz, A. (2021). Fuzzy lexical representations in adult second language speakers. *Frontiers in Psychology*, 12:732030.

Quantifying Factors Shaping Analogical Restructuring of the Maltese Nominal System

Sara Court
court.22@osu.edu
The Ohio State University

Micha Elsner
elsner.14@osu.edu
The Ohio State University

Andrea D. Sims
sims.120@osu.edu
The Ohio State University

This study investigates the relative influence of three factors – words' phonological form, distributional semantics, and etymological origin – on the organization of plural inflection classes in Maltese. Using a combination of computational modeling and information-theoretic methods, we quantify the extent to which the form, meaning, and origin of a Maltese singular noun may predict its plural. Over the course of centuries, the Semitic language of Maltese has been shaped by a large influx of Sicilian, Italian, and English words, leading to a restructuring of the Maltese inflectional system. The extension of Semitic morphology to non-Semitic lexical items (and vice versa) is likely an analogical process, but one which is hard to observe over time. However, information-theoretic analyses of the synchronic state of the morphological system provide quantitative evidence about factors that exert pressure on the lexicon. We conclude that the language's original inflectional classes have been redrawn partly on the basis of word form, but that this restructuring is partial rather than complete.

Estimates of the lexical composition of Maltese suggest as much as 70.86 percent of modern Maltese vocabulary derives from non-Semitic (largely Romance) origin (Brincat, 2017; see also Comrie & Spagnol, 2016; Lucas & Čéplö, 2020). According to Nieder et al. (2021), Maltese nouns may take one or more of at least 23 different plural inflectional patterns, including 12 different concatenative plural suffixes of Arabic, Romance, and English origins and 11 non-concatenative ('broken plural') templates deriving from the Arabic substrate (for a discussion see Spagnol, 2011). We run a series of experiments using a Long Short-Term Memory (LSTM) classifier adapted from Williams et al. (2020) to quantify the extent to which a word's form, meaning, and origin may predict its plural inflection. Phonological form is encoded as a unified word-level representation using character embeddings to capture recurrent correspondences across singular-plural pairs in the dataset, while lexical semantics are represented using FastText word vectors (Bojanowski et al., 2016), and etymology is encoded as a binary input: Semitic or non-Semitic. We then quantify the relative contributions of these predictors in terms of the amount of Mutual Information (MI) they share with plural inflection class, using cross-entropy estimates obtained from the model at test time. MI is an information-theoretic measure that encodes the extent to which one or more random variables, e.g. potential word forms, meanings, and/or origins, informs accurate predictions of another, e.g. plural inflection, in terms of a difference in unconditional and conditional entropies over their distributions (see Figure 1, below).

Our results suggest both phonological form and etymological origin inform predictions of a noun's plural inflection type (concatenative or non-concatenative) in Maltese. As seen in Table 1 below, a word's etymology is highly predictable given its phonological form, however both form and origin contribute distinct information shaping the models' predictions of plural type. While phonological similarity is most predictive of plural type, a distinct contribution of etymology suggests a segment of the lexicon has retained its original morphology over time despite form-based analogical pressures. Ongoing work using word vectors will further consider the relative contribution of lexical semantics, while additional models will investigate form, meaning, and origin as predictors of fine-grained plural classes for more detailed insight into the analogical relationships shaping contact-induced morphological integration and change.

Tables and Figures

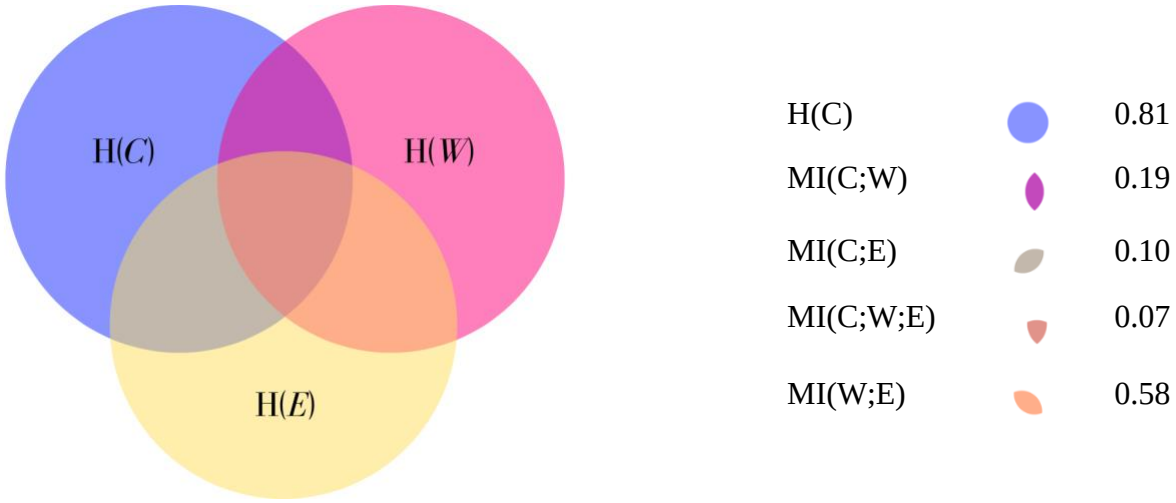
Table 1

Mutual information (in bits) shared by word form, origin, and plural type.

Model Input	Model Output	MI
Word form	Plural type	0.19
Etymology	Plural type	0.10
Word form + Etymology	Plural type	0.07
Word form	Etymology	0.58

Figure 1

Mutual information shared by word form (W), etymology (E), and plural type (C) schematized as the overlap in information conveyed by each system.



$$MI(C;W) = H(C) - H(C \mid W)$$
$$MI(C;W;E) = MI(C;W) - MI(C;W \mid E)$$

References

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2016). Enriching Word Vectors with Subword Information. *ArXiv Preprint ArXiv:1607.04606*.

Brincat, J. M. (2017). Maltese: blending Semitic, Romance and Germanic lexemes. *Lexicographica*, *33*(1), 207–224. https://doi.org/https://doi.org/10.1515/lexi-2017-0011

Comrie, B., & Spagnol, M. (2016). Maltese loanword typology. In G. Puech & B. Saade (Eds.), *Shifts and Patterns in Maltese*. De Gruyter Mouton. https://doi.org/10.1515/9783110496376-012

Lucas, C., & Čéplö, S. (2020). Maltese. In C. Lucas & S. Manfredi (Eds.), *Arabic and contact-induced change* (pp. 265–302). Language Science Press.

Nieder, J., Tomaschek, F., Cohrs, E., & Van De Vijver, R. (2021). Modeling Maltese Noun Plural Classes without Morphemes. *Language, Cognition and Neuroscience*, 1–22.

Spagnol, M. (2011). *A Tale of Two Morphologies Verb structure and argument alternations in Maltese (Doctoral dissertation)*. http://nbn-resolving.de/urn:nbn:de:bsz:352-163722

Williams, A., Pimentel, T., McCarthy, A. D., Blix, H., Chodroff, E., & Cotterell, R. (2020). *Predicting Declension Class from Form and Meaning*. http://arxiv.org/abs/2005.00626

The role of analogy in a paradigm gap in Turkish

Muhammed İleri
Boğaziçi University
muhammed.ileri@boun.edu.tr

Introduction. This study focuses on a gap in the agreement paradigm of -AsI desideratives in Turkish. Morpho-syntactically, -AsI desideratives is like other nominalized clauses. For instance, both -mA clauses and -AsI desideratives have as predicate a verbal stem followed by a nominalizer and an agreement marker reflecting the person/number features of the genitive subject (1).

(1)

a. [(Ben -im) şimdi kahve yap -ası -m] var.
[(I -GEN) now coffee make -DESID -1SG.POS] exist
'I want to make coffee now.' ***DESIDERATIVE CLAUSE***

b. [(Ben -im) şimdi kahve yap -ma -m] lazım.
[(I -GEN) now coffee make -NOM -1SG.POS] necessary
'I need to make coffee now.' ***-MA NOMINALIZED CLAUSE***

-AsI desideratives and regular nominalized clauses (e.g. -mA clauses) share the same possessive agreement paradigm except the 3SG cell (2). While the forms in other cells of the desiderative paradigm include a separate agreement marker following the desiderative marker -AsI, there is no separate marker for agreement in the 3SG cell. To exemplify with the verb *yap* 'to make', it does not have the expected regular form **yapasısı* but the irregular form *yapası* (2b).

(2)

a.

-mA paradigm of *yap* 'to make'	**SG**	**PL**
1	yap*ma***m**	yap*ma***mız**
2	yap*ma***n**	yap*ma***nız**
3	yap*ma***sı**	yap*ma***ları**

b.

-AsI paradigm of *yap* 'to make'	**SG**	**PL**
1	yap*ası***m**	yap*ası***mız**
2	yap*ası***n**	yap*ası***nız**
3	yap*ası*	*yap*ası***ları**/*yapa**ları**

For many speakers, there is no acceptable form for 3PL desideratives although other desideratives are acceptable (2b). Since third person plural suffix is grammatically required in Turkish in pro-drop environments (3) and since there is no grammatical form to fulfill this need, there is a gap in the 3PL cell of the desiderative paradigm. This gap is observed in all desiderative verbs, contrary to the common pattern of lexical restriction reported in the paradigm gap literature (Sims, 2015).

(3)

a. *[(Onlar -ın) şimdi kahve yap **-a** ***/-ası*** -ları] var.
[(They -GEN) now coffee make **-DESID** / ***-DESID*** -3PL.POS] exist
Intended: 'They want to make coffee now.'

b. [(Onlar -ın) şimdi kahve yap -ma -ları] lazım.
[(They -GEN) now coffee make -NOM -3PL.POS] necessary
'They need to make coffee now.'

Method. 12 frequent and 12 infrequent verbs are tested in an acceptability judgment experiment (*N=183*) with a full factorial design: Person (3PL/Other) X Frequency (Freq/Infreq). An ordinal Bayesian model fit to Likert-scale (1-7) responses shows a reliably negative effect of 3PL marker on acceptability ratings of desideratives (*M = -0.91, 95%-CI = [-1.21, -0.59]*). There is some evidence for an interaction between frequency and person-number agreement (M = −0.37, 95%-CI = [−0.77, 0.02]) such that 3PL desideratives with frequent verbal stems reduce acceptability

more than those with infrequent stems, contrary to some findings in the paradigm gap literature (Albright, 2003; Sims, 2006). Further experiments are needed to test the robustness of this effect.

Analysis. I argue that the gap in 3PL desideratives results from an unsolvable Paradigm Cell Filling Problem (Ackerman et al. 2009, p.55). Corpus data (Fig.1) shows that a speaker is likely to have heard many 1SG and 3SG desiderative forms as opposed to a few 1PL, 2SG, and 2PL forms and no 3PL forms prior to their very first production of the desiderative. Then, the problem is to fill the cells of the desiderative paradigm with the appropriate novel forms. The speaker can proceed to make analogy between desideratives and other nominalized forms that are similar, such as -mA clauses. Given that the frequency/availability of a form is a crucial factor in analogical base selection (Bybee 1985) and that around 95% of desideratives occur in 1SG or 3SG forms with no big frequency difference between them (Fig.1), the base can be either 1SG or 3SG desiderative form. However, since the unique 1PL, 2SG, and 2PL forms in the paradigm that are accepted by all Turkish speakers can only be derived by 1SG as the base, I argue that the base for first and second person forms is the 1SG desiderative. Nevertheless, there is not one but two attested forms for 3PL desideratives. Some speakers prefer one form and others prefer the other even though most speakers find even their preferred form unacceptable. I argue that for producing 3PL desideratives, if the speaker chooses 1SG as the base, the predicted form would be *yapasıları* (5a). On the other hand, if 3SG is chosen as the analogical base, the predicted form would be *yapaları* (5b).

(4)

a. 1PL → yapmam : yapmamız :: yapasım : **yapasımız**
b. 2SG → yapmam : yapman :: yapasım : **yapasın**
c. 2PL → yapmam : yapmanız :: yapasım : **yapasınız**

(5)

a. 3PL → yapmam : yapmaları :: yapasım : **yapasıları**
b. 3PL → yapması : yapmaları :: yapası : **yapaları**

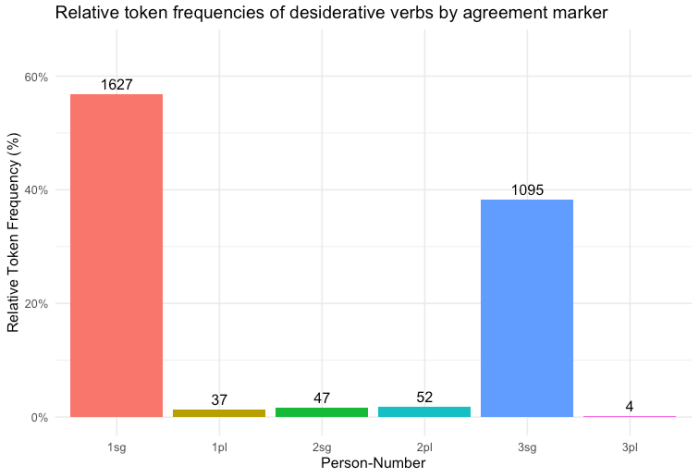


Figure 1: The distribution of token frequencies of desiderative forms by agreement marker based on TS Corpus v2 from https://tscorpus.com/ (Sezer, 2017).

Conclusion. There is no conclusive evidence in favor of the regular 1SG or the irregular 3SG being a better predictor for 3PL desideratives. I argue that this causes uncertainty in the base selection, leading to two competing forms for 3PL desideratives. As a result, uncertainty about the grammatical form of 3PL desideratives prompts the speakers to avoid using the form by means of periphrastic strategies (overt pronoun use, paraphrasing etc.), which eventually leads to the gap.

References. [1] Ackerman, F., Blevins, J. P., and Malouf, R. (2009). Parts and wholes: Implicative patterns in inflectional paradigms. [2] Albright, Adam. (2003). A quantitative study of Spanish paradigm gaps. [3] Bybee, Joan L. (1985). Morphology: A Study of the Relation between Meaning and Form. [4] Sezer, T. (2017). TS Corpus Project: An online Turkish Dictionary and TS DIY Corpus. [5] Sims, Andrea D. (2006). Minding the gaps: Inflectional defectiveness in paradigmatic morphology. [6] Sims, Andrea D. (2015). Inflectional Defectiveness. Cambridge University Press.

Solving the PCFP: results from LSTM neural network models

Tamar Johnson[1], Micha Elsner[2] and Kenny Smith[3]

University of Amsterdam[1], Ohio State University[2], University of Edinburgh[3]

Language users must constantly produce inflected forms for which they have variable degrees of familiarity. Their ability to predict appropriate forms determines how they produce previously unseen inflected forms (the Paradigm Cell Filling Problem or PCFP, Ackerman et al. [2009]), but prediction, along with memory, is also implicated in their ability to recall previously observed forms [Baayen et al., 1997]. Both tasks can be solved by analogy with other known inflected forms [e.g. Ackerman et al., 2009, Blevins, 2006]. Ackerman and Malouf [2013] suggest that the averaged conditional entropy of the inflectional paradigm (also called i-complexity) reflects the difficulty of solving the PCFP, since it captures the implicative structure of the paradigm; in paradigms with low i-complexity forms are predictive of one another.

Previous work in these two settings finds conflicting evidence for the predictive power of i-complexity. In an artificial language learning experiment, Seyfarth et al. [2014] show that it predicts the ability of human learners to generalize the paradigm to unknown forms. Johnson et al. [2021] test how quickly human learners and LSTM neural network models learn encountered inflected forms. I-complexity has only a slight effect on learning in LSTMs; in human learners there is no evidence for an effect. Moreover, they show that in both learners, another property of the paradigm, its average cell entropy (ACE) [Ackerman and Malouf, 2013], better reflects learners' difficulty in learning the trained word forms. ACE captures the distribution of exponents over inflection classes and crucially, does not capture the inter-predictability of forms in the paradigm.

While the results of Johnson et al. cast some doubt on the predictive power of i-complexity, they have some limitations. They used a small vocabulary (9 words) to ensure that their human subjects could learn the language, but language learners tend to generalize and abstract more as vocabularies grow [e.g. Schuler et al., 2021]. Moreover, their subjects learned the inflectional sys- tem and the specific lexical items simultaneously, while for language users rather than learners, the PCFP describes generalization of a known system to new words [Erdmann et al., 2020]. We extend their computational simulations by pre-training the inflectional system and by using much larger vocabularies, but our findings remain comparable, showing that i-complexity remains a secondary factor in modeling learning.

In Experiment 1, we replicate Johnson et al. [2021] with a 1000-form vocabulary. We train LSTM neural networks to predict the correct suffix for nouns in artificial languages which differ either in their averaged conditional entropy (i-complexity) or ACE (Table 1). Enlarging the vocabulary does not alter the result: networks trained on languages with low ACE learn faster than networks trained on other languages, including the low i-complexity languages (Figure 1a). In Experiment 2, we train the model in a two-stage process which better simulates the PCFP: in phase 1, the network learns the inflection system from a sample of 50 items, and in phase 2, the network learns 100 previously unseen items. While phase 1 shows a slight effect of i-complexity (as before), this result is further *reduced* in phase 2; only ACE seems to matter (Figure 1b).

These results suggest that, at least in simulation, the distribution of exponents over noun classes (ACE) plays a greater role in learners' word form productions than i-complexity, both for initial learning of the system and for generalization. Understanding how analogy and implicative structure contribute to learning may require a more nuanced understanding of

predictability within paradigm structure which looks beyond affix-to-affix probabilities [Wilmoth and Mansfield, 2021].

	Singular	Dual	Plural
noun class 1	-op	-um	-ib
noun class 2	-at	-oc	-el
noun class 3	-op	-um	-ek

(a) System A - low-i/low-ACE

	Singular	Dual	Plural
noun class 1	-op	-um	-ib
noun class 2	-at	-um	-el
noun class 3	-op	-oc	-ek

(b) System B - high-i/low-ACE

	Singular	Dual	Plural
noun class 1	-op	-um	-op
noun class 2	-at	-ib	-el
noun class 3	-op	-oc	-ek

(c) System C - low-i/high-ACE$_1$

	Singular	Dual	Plural
noun class 1	-op	-um	-el
noun class 2	-at	-ib	-op
noun class 3	-op	-oc	-ek

(d) System D - low-i/high-ACE$_2$

Table 1: Example inflection systems for each language type as tested in Johnson et al. [2021]

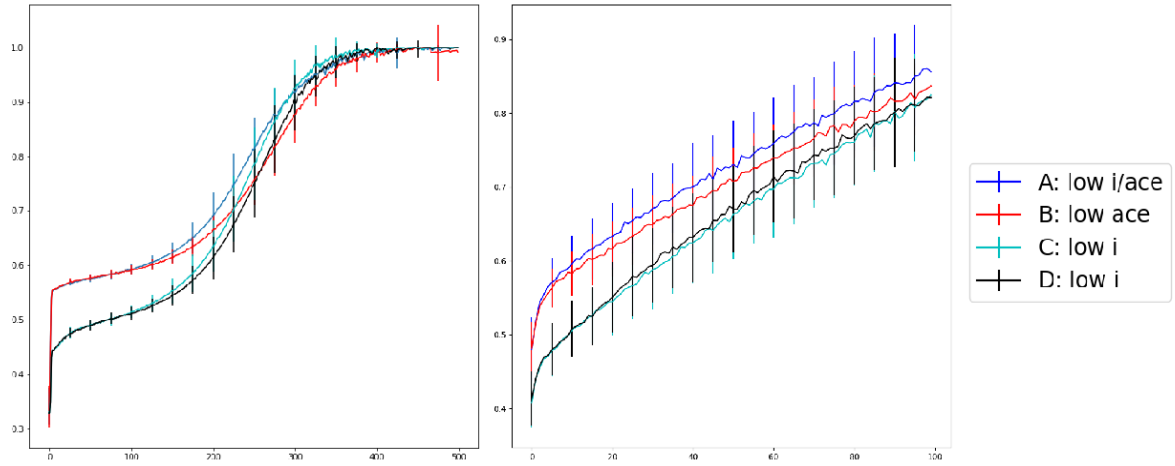


Figure 1: Acc. by epoch (avg. 50 runs, 25 unit LSTM): (a) Exp. 1: learning 1k words. Low ACE systems (A and B) learn faster early, though C shows a secondary effect of i-complexity. (b) Exp. 2: Generalization to 100 new words, after learning 50 words. Systems with low ACE (A and B) generalize much faster.

Farrell Ackerman, James P. Blevins, and Robert Malouf. Parts and wholes: Patterns of relatedness in complex mor- phological systems and why they matter. In James P. Blevins and Juliette Blevins, editors, *Analogy in grammar: Form and acquisition*, pages 54–82. Oxford University Press, Oxford, 2009.

R Harald Baayen, Ton Dijkstra, and Robert Schreuder. Singulars and plurals in Dutch: Evidence for a parallel dual- route model. *Journal of Memory and Language*, 37(1):94–117, 1997.

James P. Blevins. Word-based morphology. *Journal of Linguistics*, 42(3):531–573, 2006.

Farrell Ackerman and Robert Malouf. Morphological organization: The low conditional entropy conjecture. *Language*, 89(3):429–464, 2013.

Scott Seyfarth, Farrell Ackerman, and Robert Malouf. Implicative organization facilitates morphological learning. In *Annual meeting of the Berkeley Linguistics Society*, volume 40, pages 480–494, 2014.

Tamar Johnson, Kexin Gao, Kenny Smith, Hugh Rabagliati, and Jennifer Culbertson. Predictive structure or paradigm size? investigating the effects of i-complexity and e-complexity on the learnability of morphological systems. *Journal of Language Modelling*, 9(1), 2021.

Kathryn Schuler, Charles Yang, and Elissa Newport. Testing the tolerance principle: Children form productive rules when it is more computationally efficient. 2021.

Alexander Erdmann, Micha Elsner, Shijie Wu, Ryan Cotterell, and Nizar Habash. The paradigm discovery problem. In *Proceedings of 58th ACL*. Association for Computational Linguistics, 2020.

Sasha Wilmoth and John Mansfield. Inflectional predictability and prosodic morphology in Pitjantjatjara and Yankun- ytjatjara. *Morphology*, pages 1–27, 2021.

Recovering from accidental syncretisms with word2vec analogies

Gilles Boyé, Université Bordeaux-Montaigne/CLLE (CNRS UMR 5263) gilles.boye@u-bordeaux-montaigne.fr

Camille Lavigne, Université Bordeaux Montaigne camille.lavigne@etu.u-bordeaux-montaigne.fr

Gauvain Schalchli, Université Bordeaux-Montaigne/CLLE (CNRS UMR 5263) gauvain.schalchli@u-bordeaux-montaigne.fr

Systematic syncretism has received a lot of attention since Aronoff (1994) and the introduction of morphomes. Morphologists have questioned its stability (Maiden, 1992), its orientation (Baerman *et al.*, 2005), its structure (Bonami and Boyé, 2003). With the emergence of the paradigm cell filling problem, PCFP, in (Ackerman *et al.*, 2009) as formulated below in (1), systematic syncretism and the morphomes associated with it have become part a central part of the discussion.

(1) What licenses reliable inferences about the inflected (and derived) surface forms of a lexical item?

With a morphomic paradigm (Boyé and Schalchli, 2019), systematic syncretisms are captured by equating contrasting forms and their morphosyntactic distribution. In the absence of accidental syncretism, a form definesamorphomiccell. Forexample, inEnglish, the8differentformsof BE definethecellsofthemorphomic paradigm for all verbs. In this perspective, cells are not defined by a single set of morphosyntactic properties being realized as a form but rather by a form associated with the contexts in which it appears, a distribution.

For example, in an ideal system, the distribution of the forms of BE would provide a set of key distributions with which to classify the distribution of the forms of any other verb. But when accidental syncretism occurs, this straightforward association between a verb form and one key distribution is obscured. Rather than corresponding to a key, the syncretic form distribution results from the union of two or more key distributions. In this paper, we propose to look at what we call the paradigm cell association problem (PCAP).

(2) Assuming that speakers have identified a morphomic paradigm shape based on contrasting forms, how can they place forms in the morphomic paradigm?

We propose they can use the key forms they have already placed and compare their distribution with the ones under scrutiny using all the combinations possible to identify accidental syncretism.

To test out this idea on a small scale, we look at syncretic forms of English preterite and past participle in a sub-corpus of 6.5 million words in the BNC. We extracted triplets constituted of the base form, the preterite and the past participle with at least 50 occurrences each (around 200 triplets). In this sample, 10 verbs had different forms for their preterite and their past participle, giving us 20 discriminating forms (10 preterites, 10 past participles), while the other 190 had syncretic forms for those cells.

To capture the distribution of the forms, we used a word2vec model that gave us 20 reference vectors for the discriminating forms and 190 ambiguous vectors to be classified. We added the 10 compound vectors resulting from the sum of the preterite and the past participle of the discriminating verbs to the model (e.g. *ate*+*eaten*). For every syncretic verb, we calculated which of these 30 reference vectors is the most similar and associated the form with the cell/cells of the reference vector. As shown in Table (1), 162 of our 190 syncretic forms were associated with a compound vector (c_1+c_2). For example, *ate*+*eaten* was the closest reference vector to *found*. We controlled the classification by using a second word2vec model with tagged forms and calculated the most similar vectors to the classified verb preterite vector v_1 summed

with the compound difference: $v_1+c_2-c_1$. In the case of *found*, for example, the composition found$_1$+eaten-ate had the expected found$_2$ as the 3rd closest neighbour. Figure 1 presents the precision of our prediction depending on the size of neighbourhood considered to get the right form.

The classification presented here allows to calculate separate vectors for syncretic forms in the original word2vec model. If the form vector, v, is indeed a compound of v_1 and v_2, we know that $v_1+v_2=v$ and if the classification is right, we also know that $v_2-v_1 \simeq c_2-c_1$. So $v_1 \simeq (v+c_1-c_2)/2$, and $v_2 \simeq (v+c_2-c_1)/2$. This allows to recover from a simple accidental syncretism between two cells. The same method could be applied to larger combinations involving more accidentally syncretic cells.

distribution	syncretic forms	neighbouring rank	number
ate+eaten	found, used, provided, included, produced, bou...	51, 242, 682, 841, 287, 8, 156, 79, 98, 248, 2...	26
blew	opened	38	1
blew+blown	shot	2	1
did	wanted, let, refused	10, 68, 90	3
did+done	needed, helped, tried, meant, understood, paid...	60, 108, 20, 6, 65, 235, 53, 45, 191, 7, 282, ...	14
drew+drawn	made, brought, set, held, obtained, sought, in...	32, 1, 11, 66, 222, 229, 607, 104, 4, 88, 406,...	21
drove	started, stopped, stayed, returned, watched, hit	42, 15, 55, 35, 11, 7	6
drove+driven	moved, followed, offered, called, built, creat...	7, 16, 436, 24, 30, 529, 99, 616, 23, 378, 10,...	21
forgot	had, said, felt, talked	18, 43, 11, 2	4
forgot+forgotten	thought, told, liked, believed, remembered, he...	6, 25, 12, 20, 7, 0, 330, 56, 116, 271, 19, 31...	40
grew+grown	developed, remained, increased, tended, result...	353, 64, 775, 124, 130, 36	6
sank+sunk	formed, reached, settled, dropped, fought, dis...	176, 18, 22, 1, 75, 8, 6	7
spoke	seemed, lived, appeared, met, played, joined, ...	14, 36, 14, 20, 155, 47, 23, 218, 223, 55, 106	11
spoke+spoken	asked, read, described, shared, referred, cont...	48, 11, 731, 260, 145, 568, 267, 12, 23, 405, ...	14
threw	looked, stood	16, 14	2
threw+thrown	kept, put, carried, turned, caught, placed, la...	3, 1, 19, 1, 8, 18, 4, 20, 1, 0, 3, 6	12

Table 1: Syncretic forms classification

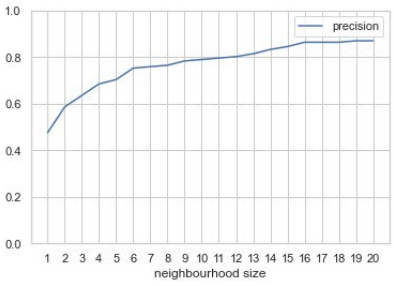


Figure 1: Precision of the prediction of form$_2$ according to the size of the neighbourhood

References

Ackerman, F., Blevins, J. P., and Malouf, R. (2009). Parts and wholes: Implicative patterns in inflectional paradigms. In J. P. Blevins and J. Blevins, editors, *Analogy in grammar: Form and acquisition*, pages 54–82. Oxford Scholarship Online.

Aronoff, M. (1994). *Morphology by itself*. M.I.T. Press, Cambridge.

Baerman, M., Brown, D., and Corbett, G. G. (2005). *The Syntax-Morphology Interface: A Study of Syncretism*. Cambridge University Press.

Bonami, O. and Boyé, G. (2003). Supplétion et classes flexionnelles dans la conjugaison du français. *Langages*, **152**, 102–126.

Boyé, G. and Schalchli, G. (2019). Realistic data and paradigms: the paradigm cell finding problem. *Morphology*, **29**(2), 199–248.

Maiden, M. (1992). Irregularity as a derterminant of morphological change. *Journal of Linguistics*, **28**(2), 285–312.

Analogy across classes and languages: Regularization in Frisian 2SG past forms

Anne Merkuur, Fryske Akademy, amerkuur@fryske-akademy.nl

This paper presents and evaluates recent developments in the inflection of 2SG past forms of Frisian verbs. With the aid of an analysis along the lines of Distributed Morphology (Halle & Marantz, 1993), we specifically test whether we can explain the developments in Frisian verbal inflection using a model of categorical productivity: the Tolerance Principle (Yang, 2016).

Frisian has two classes of regular verbs (class I with an infinitive ending in *-ə*, class II ending in *-jə*), both with their own inflectional paradigms. The past tense of class I verbs is similar to Dutch verbs, and exhibits a dental suffix *-te/-de* in the simple past. Class II verbs on the other hand show a different pattern with no dental suffix in the simple past:

(1) Paradigms of class I and II

	Person/Nr	**Class I**	**Class II**
PRS	1SG	bak ø	wurk j e
	2SG	bak st	wurk e st
	3SG	bak t	wurk e t
	PL	bak e	wurk j e
PST	1SG	bak te ø	wurk e ø
	2SG	bak te st	wurk e st
	3SG	bak te ø	wurk e ø
	PL	bak te n	wurk e n

A dialect survey in which inflections of class II verbs were elicited (n=271) (Merkuur, 2021) reveals an ongoing change: younger participants tend to inflect one form of class II verbs more and more with the dental suffix *-te/-de*, i.e. as if they were class I verbs: the second person singular (2SG hereafter) in the past tense. *Do wurkest* thus becomes *do wurk̲test*:

(2) Development of 2SG past of regular class II verbs

	norm	**alternative**
wurkje	*do wurk-e-st*	*do wurk-te-st*
work.INF	you work-PSTclassII-2SG	you work-PSTclassI-2SG

A similar and earlier development has been observed in Hoekstra (2013) on the past tense of 2SG forms of strong and irregular verbs:

(3) Development of 2SG past of strong and irregular verbs

	norm	**alternative**
sjen	*do seach-st*	*do seach-de-st*
to see.INF	you see.pst-2SG	you see-PSTclassI-2SG

First of all, I argue that this is a similar developmen: the element *-te/-de* is the same past tense suffix on all verbs, and not a past tense suffix on the regular verbs and an empty element on

the strong and irregular verbs, as proposed by Hoekstra (2013).

Another striking similarity between the class II verbs and some of these strong/irregular verbs is that without the dental past tense suffix, the 2SG past tense form is homophonous with the 2SGsg present tense forms, and that this homophony is dissolved by adding the dental past tense suffix *-te/-de* to the past tense forms. As such, from a functional perspective, it seems clear why both these changes occurred: attaching the dental suffix dissolves this ambiguity between the past and present forms. However, such an explanation cannot completely account for the changing forms, as it does not solve the actuation problem (Weinreich, Labov & Herzog, 1968) of these particular changes: What initiated them? Why did they change at a certain point in time, whereas apparently, the homophony was not a problem/sufficient trigger in earlier stages? And, why didn't all classes change at the same time? Furthermore, it does not explain why the homophony is dissolved with the aid of the dental suffix?

The analysis put forward in this paper shows that, indeed, due to both the homophony between the past and present tense as well as the availability of the two past tense suffixes of class I (*-e)* and class II (*-te/-de),* Frisian offers evidence for more than one grammatical analysis for the 2SG past forms. Moreover, the analysis shows that Frisian is inconclusive about which rule to apply to the 2SG past forms, because there is no productive and non-homophonous rule available. Usually, such an indecisive situation is destined to result in the forms falling back to the elsewhere rule, or by lack of an elsewhere rule, to result in a lexical gap (Yang, 2016). Clearly, in the case of Frisians 2SG past forms, the forms do not fall back on their elsewhere rule (since they change), but there is no lexical gap either. Following Hulk & Müller (2000) I argue that such a deadlock situation enables cross-linguistic influence to occur: Because Frisian is inconclusive about which rule to apply, and Dutch only exhibits the past tense suffix *-te*/*-de,* Dutch reinforces the use of the *-te*/*-de* suffix. Such an explanation linked to Dutch also accounts for why the changes regarding the homophonous second person singulars actuated in roughly the last 100 years. In this period, Frisian lost some of its relative isolation (Klinkenberg, 2017) and the bilingual relationship with Dutch intensified.

Halle, M. & Marantz, A. (1993). Distributed morphology and the pieces of inflection. In *The view from building 20: Essays in linguistics in honor of Sylvain Bromberger*. (pp. 111-76). Cambridge, MA: MIT Press.

Hoekstra, J. (2013). Do seachdest der idd aardig wasted út ja. De 2de persoan iental doetiid fan'e sterke tiidwurden yn it Frysk fan'e jongerein. *Us Wurk, 62*(1-2), 1-20.

Hulk, A., & Müller, N. (2000). Bilingual first language acquisition at the interface between syntax and pragmatics. *Bilingualism: language and cognition*, *3*(3), 227-244.

Klinkenberg, E. (2017). Rapportaazje: Taalsosjologyske Survey:'Taal yn Fryslân, de folgjende generaasje'. Leeuwarden: Fryske Akademy.

Merkuur, A.A. (2021). Changes in Modern Frisian verbal inflection. Amsterdam: LOT publications.

Weinreich, U., Labov, W., & Herzog, M. (1968). *Empirical foundations for a theory of language change* (Vol. 58). Austin: University of Texas Press.

Yang, C. (2016). *The price of linguistic productivity: How children learn to break the rules of language*. Cambridge MA: MIT press.

Analogical rule extension underlying novel requisitive imperatives in web Japanese

Akiko Nagano and Masaharu Shimada
University of Shizuoka and University of Tsukuba
nagano.9@u-shizuoka-ken.ac.jp; shimada.masaharu.fu@u.tsukuba.ac.jp

In terms of illocutionary force, imperative clauses may convey order/command/instruction or request/entreaty, depending on the speaker beneficiality. In English, *please* functions to mark the latter force; it is added to a plain imperative when the addressee's following the order is expected to result in something favorable or advantageous on the part of the speaker. In Japanese, plain imperatives minimally differ from requisitive imperatives in the latter's composite morphology [verb + *kureru* 'give']. This dimension intersects with politeness morphology, which in the case of imperative clause is sensitive to the honorification toward the (non-overt) subject. Thus, (1a, b) illustrate plain imperative with and without subject honorification, while (2a, b) illustrate requisitive imperative with and without subject honorification. The lexical verb inflects for these categories in (1), but *kureru* inflects in (2).

(1) "Open the window." Imperative Force: order
a. *mado-o akero.* [non-honorific]
window-ACC open.IMPERATIVE.
b. *mado-o akenasai.* [honorific]
window-ACC open.HONORIFIC.IMPERATIVE.

(2) "Please open the window." Imperative Force: request
a. *(dooka) mado-o akete-kure.* [non-honorific]
(please) window-ACC open-give.imperative
b. *(dooka) mado-o akete-kudasai.* [honorific]
(please) window-ACC open-give.HONORIFIC.IMPERATIVE

The force difference is confirmed by the distribution of *dooka* 'please,' which is naturally added to the clause-initial positions of (2), but not to (1) (cf. Hasegawa 2011: 94).

Naya (2017) observes that since around 2009, new requisitive imperatives like the following have begun to be found in casual writings on internet community sites, blogs, or twitters:

(3) *(dooka) mado-o akero-kudasai.* [request/honorific]
(please) window-ACC open.IMPERATIVE-give.HONORIFIC.IMPERATIVE

No previous study discusses how the verb form in (3) has emerged, but we propose that it is caused by analogy between two pertinent inflectional classes: Table 1 and Table 2. Table 1 accommodates the four verb forms in (1, 2). Table 2 introduces imperative forms which we have not discussed; they are based on the *o*-prefixed adverbial form of the base verb. This form, which we call honorific stem, is employed not only in imperative but also other clausal types; also, it is not restricted to subject honorification (see Table 3). The shortest form (*mado-o*) *o-ake* "Open the window!" in Table 2 is more polite than the shortest form *akero* in Table 1, but less polite than *o-ake-nasai* in Table 2.

What is crucial for the emergence of (3) is different derivation of *-kudasai* form in the two classes. In Table 1, *v-kudasai* and *v-kure* are related and derived from v-KURERU, whereas in Table 2 the honorific requisitive is paradigmatically related to the plain imperatives. Table 4 schematizes the different formal dependency between the pertinent cells A, C, and D. Given these inflectional classes, it is possible to regard the new form in (3) as derived by extending the established mapping rule between A and D (blue-colored) to the non-honorific stem. By this rule extension, *-kudasai* is attached to the form in Cell A (*akero*), yielding *akero-kudasai.*

Our analysis predicts the possibility of the other analogical extension, i.e., applying the C-D mapping rule (red-colored) to the honorific stem. Indeed, we find such requisitive imperatives as illustrated by the bold face in Table 5. They are found on the same type of media as (3), i.e., internet community sites, blogs, or twitters. Comparing Table 2 and Table 5, we observe that the analogical formation fills in the originally defective cell in the inflectional class using the honorific stem.

In sum, analogy in inflection is synchronically observable in current web Japanese.

Table 1. Imperative forms based on the non-honorific stem

	−Subject honorification	+Subject honorification
−Speaker beneficial	*akero*	*akenasai*
+Speaker beneficial	*akete-kure*	*akete-kudasai*

Table 2. Imperative forms based on the *o*-prefixed honorific stem

	−Subject honorification	+Subject honorification
−Speaker beneficial	*o-ake*	*o-ake-nasai*
+Speaker beneficial	※	*o-ake-kudasai*

※defective because there is no verb like *O-AKE-KURERU (vs. AKETE-KURERU)

Table 3. The use of the honorific stem in declarative clauses

	Examples of declarative clauses
Subject honorification	*sensee-ga mado-o* ***o-ake****-ni naru* teacher-nom window-acc hon-open-dat become "The teacher opens the window(s)."
Object honorification	*Taro-ga kyooinshitsu-no mado-o* ***o-ake*** *suru* Taro-nom teacher's room-gen window-acc hon-open do "Taro opens the window(s) of the teacher's room."

Table 4. Cell D related to cell C in Table 1, while Cell D related to A in Table 2

	−Subject honorification	+Subject honorification
−Speaker beneficial	A	B
+Speaker beneficial	C	D

Table 5. "(Please) forgive (something/someone)" based on honorific stem *o-yurushi*

	−Subject honorification	+Subject honorification
−Speaker beneficial	*o-yurushi*	*o-yurushi-nasai*
+Speaker beneficial	***o-yurushi-kure***	*o-yurushi-kudasai*

References:

Hasegawa, Nobuko. 2011. Toogo koozoo to hatsuwa no chikara (syntactic structure and utterance force). In: M. Takeuchi and H. Sato (eds.) *Hatsuwa to bun no modaritei* (Utterance and Sentential Modality), 89-114. Hituzi Shobo: Tokyo.

Naya, Ryohei. 2017. "Shirokudasai" no goyooron (The pragmatics of *shirokudasai*). Paper read at the 18th annual meeting of the Society of Japanese Grammar, University of Tsukuba.

A typological view of analogy in morphology: some issues and possible solutions

Matías Guzmán Naranjo, Universität Tübingen, mguzmann89@gmail.com

Introduction Several proposals exist for representing and automatically inferring proportional analogies. However, these proposals have not been designed with the wide spectrum of inflectional typology in mind, and have mostly focused on a very limited selection of morphophonological phenomena. As a consequence, most formalisms for proportional analogies cannot capture a very large number of inflectional phenomena found in the worlds languages. At the same time, several of these phenomena which cannot be captured by proportional analogies are abstractions made by linguists, and analogical models of inflection can model the inflectional system in question if we give up on the idea of capturing all possible abstractions and adopt a more brute-force approach. In this talk I will argue that (1) we need a more powerful formalism for analogy, and (2) that not all abstractions are needed.

Background Multiple proposals for representing and implementing proportional analogies exist, and computational implementations have been around for over two decades (Federici, Pirrelli, and Yvon, 1995). The traditional method (I will call the X-notation) for representing proportional analogies is based on named variables and phonemes: Xo⇋Xas which captures an alternation like (1sg.pres.ind) *canto-cantas* (2sg.pres.ind, 'sing'). This approach relies on something akin to string unification (Calder, 1989), though it is rarely spelled out. Beniamine (2017) improves on this approach by using contextual information to represent the analogies, and to constraint the positions of unification. For example, for the French alternation (1.sg.prs) *amɛn-amøne* (2.pl.prs), instead of XɛY⇋XøYe, one can write _ɛ_ ⇋_ø_e / am_n_. Variations on these systems have been applied to, and argued for, from the perspective of multiple languages (James P Blevins, 2016, 2007), but mostly from an informal perspective. It is unclear to what degree these symbolic models of proportional analogy can account for different inflectional patterns in the worlds languages.

Challenges The issue with the formalisms mentioned above is that they cannot properly capture situations in which a pattern affects parts of segments, suprasegmental phenomena, or situations in which segments change relative positions. The first two are mostly due to a string-like representation of words, and could arguably be overcome by enriching the systems with feature structures. The later is more serious and cannot be solved without further technical changes to the way in which we represent proportional analogies. The following is a small selection of phenomena which cannot be captured properly with either the X-notation (an in some cases neither with Beniamine's system). This is not an exhaustive list of but it should illustrate some of the problems.

• Hungarian: shortening/deletion (Paschen, 2018)
– (sg.nom) mada:r-madarak (pl.nom) ('birds')
– (sg.nom) bokor-bokrok (pl.nom) ('bushes')
– (sg.nom) terem-termek (pl.nom) ('halls')

• Russian: metathesis (personal knowledge)
– (dat.sg) avós'ke - avós'ek (gen.pl)

• Nahuatl: reduplication in singular-plural (Karttunen, 1992)
– (sg) chichitōn - chichitotōn (pl) ('a small dog')
– (sg) pilpīl - pīpilpipīl (pl) ('a child or youth')

• Gã downstep in verbal inflection (Paster, 2003)
– (2.sg.pres) e-hulú - é-!húlú (2.sg.subj) (‘to jump’)
• Free-positions (Chintang) (Bickel et al., 2007)
– a-ma-im-yokt-e (2.neg.sleep.neg.pst) (‘you didn’t sleep’)
– ma-a-im-yokt-e (neg.2.sleep.neg.pst) (‘you didn’t sleep’ identical as above)

Other types of abstractions are also not captured, even though a system could be built to ignore them. Some examples are phenomena like vowel or consonant harmony which cannot be easily expressed using simple X-notation style formalisms, but which can be modelled as different inflection classes. Similarly, analyses which decompose an inflectional systems into multiple independent dimensions can be modelled directly, but the independent dimensions must be collapsed into one.

A new formalism While solving the limitations on suprasegmental patterns is relatively straightforward using feature structures, addressing some of the segmental patterns is much more difficult. For reasons of space I cannot discuss all the reasons why the examples above represent problems for formalisms of proportional analogies, but I will focus on the main challenge: having more than one free variable. This is a problem for proportional analogies because the string unification produces multiple possible solution. To give a trivial example we can simply look at infixes. An alternation like *pancas-pancos* could be captured with XaY⇋XoY. However, this alternation does not produce a unique solution, because the <a> in the pattern can match either of the two <a>s in pancas. More challenging even are cases of metathesis, which would require three free variables, e.g. for *catar-catra* XYZ⇋XZY.
A formalism which solves most of these issues needs to impose restrictions on the matching potential of the free variables. Additionally, we limit the umber of variables which can match arbitrarily many segments to one on each side of the proportion. We express variables as <NAME, n-matches> with * meaning arbitrarily many, and numbers meaning a fixed number of segments. We can then represent a proportion for the infix example above as follows: [<X,*> a <Y,1> ⇋ <X,*> o <Y,1>]. Metathesis can also be easily captured because [<X,*><Y,1><Z,1> ⇋ <X,*><Z,1><Y,1>]. While this is a relatively simple system, it can capture most phenomena found in inflectional morphology. I will present a computational implementation of this system, and show that it works according to our expectations on most inflectional systems. I will argue that the limitation on having only one variable be able to match an arbitrary number of segments seems to be a robust cross-linguistic generalization and that there are no inflectional patterns which require patterns with the form: [<X,*> a <Y,*> ⇋ <X,*> o <Y,*>].
While this approach solves most of the phenomena mentioned above, others remain challenging like free marker position like in Chintang or postion-dependent morphosyntax like in Swahili. I will discuss what these cases mean for the idea of having a purely analogical approach to inflectional morphology.

Concluding remarks In order to be able to evaluate analogical models of inflection we need to test them on as many different inflectional systems as possible. The proposed system solves some of the problems mentioned in this abstract (though not all of them), and its computational implementation can efficiently infer and parse these patterns. However, to pursue a purely analogical model of inflection we might have to give up on capturing some traditional abstractions found in morpheme-based models.