Analogical extensions of suppletive stems in the verb 'go' in Gallo-Romance

Xavier Bach

Christ Church College, University of Oxford

The verb 'go' is notoriously suppletive in the Romance languages (Aski 1995; Adams 2007:792-820; Maiden 2018). Here I used dialectal data taken from the *Atlas linguistique de la France* (Gilliéron & Edmont 1902-1910) to map types of suppletive analogical reformations in this verb in Gallo-Romance varieties. Atlas data is rather detailed as to spatial variation, but contains limitations as to what forms of the verb are available: it is only possible to reconstruct a very partial paradigm. Mostly, there is at least one form for every morphomic space where suppletion obtains: are available the infinitive, the present indicative forms of the singular plus 2PL, 2SG and 3SG future, 1SG present subjunctive, 2PL imperative, and 3PL compound past with both the form of the auxiliary and the past participle. The phenomena that are revealed need to be confronted to other sources to ascertain their full paradigmatic extent. But a number of phenomena are clearly visible, such as the extension of N-pattern roots into the 1PL in Limousin because of systematic syncretism, as well as in Gascony because of paradigm levelling, or the existence of Francoprovençal future forms in *adre*. I will here focus mostly on two phenomena: the suppletive collision of two roots into one suppletive form in the future in Franche-Comté (viré < v- + ir-) and Languedoc (anirei < an- + ir-), and the phenomenon of overlapping suppletion in the compound tenses (where the past participle of *go* in Oïl varieties is taken from *be*, or even *have*).

Aski, Janice. 1995. Verbal suppletion: an analysis of Italian, French, and Spanish *to go*. *Linguistics* 33: 403-432.

Adams, J. N. 2013. *Social variation and the Latin language*. Cambridge: Cambridge University Press.

Gilliéron, Jules & Edmond Edmont. 1902-1910. *Atlas linguistique de la France*. Online digital version: Cartodialect 5.4.6. http://lig-tdcge.imag.fr/cartodialect5/#/

Maiden, Martin. 2018. *The Romance verb. Morphomic structure and diachrony.* Oxford: Oxford University Press.

Analogical evidence for morphological paradigms as complex implicational networks

In Gévaudan varieties of Occitan (Gallo-Romance), exceptionless syncretism between preterite and imperfect subjunctive forms arises in the first and second person plural (e.g. *faguessiam* 'do.PRET/IPF.SBJV.1PL', *faguessiatz* 'do.PRET/IPF.SBJV.2PL'). As the attested forms are identical with historically expected imperfect subjunctive forms, the syncretism pattern superficially appears to be a compact and localised case of analogy in which the imperfect subjunctive directly influences or replaces the preterite. Yet reconstructing the historical emergence of the pattern reveals instead that it is crucially dependent on multiple and diverse implicational relationships of form, inferred and productively exploited by speakers: in particular, inherited identity between preterite and imperfect subjunctive stems; and identity between imperfect indicative forms of *èstre* 'be' and preterite or imperfect subjunctive desinences. The Gévaudan case offers empirical support for a view of inflectional analogy as informed by intricate paradigmatic and implicational structure, of a type which can only be adequately captured within 'abstractive', word-based theories of inflection.

LOUISE ESHER
Centre national de la recherche scientifique
Langage, langues et cultures d'Afrique (Llacan, UMR 8135)

Analogical classification and the paradigm cell filling problem

Matías Guzmán Naranjo - Université de Paris

1 Introduction

Recent work on paradigm organization has focused on the question of how speakers can deduce the complete paradigm of a lexeme given that they only encounter a limited number of inflected forms of that lexeme (Ackerman et al., 2009). This is also known as the paradigm cell filling problem (PCFP). These studies have proposed entropy as a measurement of paradigm complexity/predictability. The basic idea is that one can calculate the conditional entropy between two cells of a paradigm, which measures the predictability between cells. Bonami & Beniamine (2016) have even expanded this approach to work for multiple cells.

Entropy-based approaches have a some limitations, however. First, entropy is not a normalized metric, which makes it unreliable for comparing different systems/languages. Second, many studies have convincingly shown that the inflection class of a lexeme is predictable from its phonology and semantics (Matthews, 2005), which is information that entropy cannot easily take into account (Although Bonami & Beniamine (2016) offer a possible alternative).

Using the Russian nominal inflection system as an example, I will argue that analogical classification (i.e. class assignment on the basis of similarity) offers a convincing solution to the PCFP, and that accuracy metrics are a better measurement of predictability/paradigm complexity than entropy.

2 Materials and methodology

From the Grammatical Dictionary of Russian by Zaliznyak (1977), I extracted all nouns (43412) with their complete paradigm (including the prepositional case). I then converted the extracted forms to a phonological transcription using epitran (Mortensen et al., 2018). This phonological transcription is not perfect but it is a reasonable approximation of the Russian system. For the present study I did not consider stress but this feature could be easily included into the models.

Many accounts of Russian nominal inflection have been proposed in the literature, each suggesting a different analysis of the inflection classes found in the Russian system. To sidestep these discussions, I extracted the inflection class of each noun automatically with a surface-based method described in Guzmán Naranjo and Becker (forth.) and Guzmán Naranjo (forth.).

To include semantic information I used the pre-trained Word2Vec semantic vectors provided by Kutuzov & Kuzmenko (2017) using word2vec. To match a lexeme to a semantic vector I used the NOM.SG cell. From the dataset I only kept those nouns for which there was a semantic vector.

In order balance the dataset I only considered I excluded inflection classes with fewer than 20 nouns. This yielded a total of 35329 nouns across 108 inflection classes.

On the resulting dataset I trained multiple analogical models using a multilayer perceptron following Guzmán Naranjo (2019). For every cell in the paradigm I trained models predicting that cell from: (i) one other cell, (ii) two other cells, and (iii) one cell and stem information.[1]

3 (Selected) results

Table 2 shows the results for the analogical models considering one additional cell, phonology and semantics: `cell-1-marker` ∼ `cell-2-marker + final_segment_1 + final_segment_2 + …+ X1 + …+ X300`, where `final_segment_n` are the final segments of the stem, and `Xn` are the individual dimensions of the semantic vectors. The mean accuracy for the other models (not shown here) were as shown in Table 1.

Seen in absolute terms, we can say that just knowing one form of a Russian noun (including the stem) is enough to give almost perfect predictive accuracy for 7 cells (INS.PL, INS.SG, NOM.PL,

[1] The stem information consisted of the last four segments of the stem plus the semantic information in the semantic vectors.

model	mean accuracy
`cell-1-marker` ∼ `cell-2-marker`	0.81
`cell-1-marker` ∼ `cell-2-marker + phonology`	0.93
`cell-1-marker` ∼ `cell-2-marker + semantics`	0.8
`cell-1-marker` ∼ `cell-2-marker + phonology + semantics`[2]	0.96

Table 1: Mean accuracy for each model

	Predictor												
Predicted	NOM.SG	GEN.SG	DAT.SG	ACC.SG	INS.SG	PRE.SG	NOM.PL	GEN.PL	DAT.PL	ACC.PL	INS.PL	PRE.PL	mean
NOM.SG	1	0.99	0.99	1	0.99	0.9	0.91	0.98	0.89	0.93	0.89	0.89	0.95
GEN.SG	0.98	1	0.99	0.98	0.99	0.9	0.9	0.97	0.88	0.92	0.88	0.88	0.94
DAT.SG	1	1	1	0.99	1	0.91	0.91	0.98	0.89	0.93	0.89	0.89	0.95
ACC.SG	0.99	0.99	0.99	1	0.94	0.83	0.84	0.93	0.82	0.93	0.82	0.82	0.91
INS.SG	0.99	0.99	0.99	0.99	1	0.9	0.9	0.98	0.89	0.92	0.85	0.88	0.94
PRE.SG	0.99	1	1	1	1	1	0.99	0.99	0.98	0.99	0.98	0.98	0.99
NOM.PL	0.99	0.99	0.99	0.99	0.98	0.98	1	0.99	0.98	1	0.99	0.98	0.99
GEN.PL	0.99	0.99	0.99	0.99	0.99	0.91	0.91	1	0.9	0.94	0.9	0.88	0.95
DAT.PL	1	1	1	1	1	1	1	1	1	1	1	1	1.00
ACC.PL	0.93	0.92	0.92	0.97	0.92	0.91	0.92	0.92	0.9	1	0.9	0.9	0.93
INS.PL	1	1	1	1	1	1	1	1	1	1	1	1	1.00
PRE.PL	1	1	1	1	1	1	1	1	1	1	1	1	1.00
mean	0.99	0.99	0.99	0.99	0.98	0.94	0.94	0.98	0.93	0.96	0.92	0.92	0.96

Table 2: Cell predictability including phonology and semantics.

NOM.SG, PRE.PL, PRE.SG and DAT.PL), it gives a reasonable certainty for three cells (DAT.SG, GEN.PL and GEN.SG) and it gives a some certainty for the remaining two (ACC.PL and ACC.SG). It is an interesting result that these two final cells, ACC.SG and ACC.PL, are the hardest to predict from the other cells and at the same time ACC.SG is the best predictor of other cells in average.

These results show that analogical classification is a solid alternative for solving the PCFP. Especially interesting is that both semantics and phonology play a role in in the analogies. Finally, this method allows us to calculate accuracy metrics, which are normalized and allow for easy model comparison.

References

Ackerman, Farrell, James P. Blevins & Robert Malouf. 2009. Parts and wholes: Implicative patterns in inflectional paradigms. In James P. Blevins & Juliette Blevins (eds.), *Analogy in grammar: Form and acquisition*, 54–82. Oxford, New York: Oxford University Press.

Bonami, Olivier & Sacha Beniamine. 2016. Joint predictiveness in inflectional paradigms. *Word Structure* 9(2). 156–182.

Guzmán Naranjo, Matías. 2019. *Analogical classification in formal grammar* Empirically Oriented Theoretical Morphology and Syntax. Language Science Press.

Kutuzov, Andrey & Elizaveta Kuzmenko. 2017. WebVectors: A Toolkit for Building Web Interfaces for Vector Semantic Models. In Dmitry I. Ignatov & et. al. (eds.), *Analysis of Images, Social Networks and Texts: 5th International Conference, AIST 2016, Yekaterinburg, Russia, April 7-9, 2016, Revised Selected Papers*, 155–161. Cham: Springer International Publishing.

Matthews, Clive A. 2005. French Gender Attribution on the Basis of Similarity: A Comparison Between AM and Connectionist Models. *Journal of Quantitative Linguistics* 12. 262–296.

Mortensen, David R, Siddharth Dalmia & Patrick Littell. 2018. Epitran: Precision G2P for many languages. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC-2018)*, .

Zaliznyak, Andrey Anatolyevich. 1977. *Grammatical dictionary of the Russian language.* Moscow: Russkij Jazyk.

The Romance S-morphome and the interaction of token frequency and predictive structure in the paradigm

The stem alternation patterns of Romance verbal inflection have been analyzed extensively in recent years (see Maiden 2018 for a compendium). The stem alternation patterns N, L, and P(YTA) have been identified as the most widespread and influential ones in the family. This does not mean, however, that no other patterns exist, or that more minor patterns cannot also provide invaluable insights into the workings of analogy in inflectional paradigms. This is the case of the pattern I refer to as the 'S-morphome' (from Short-stem morphome). Stem shortenings occur in various high frequency verbs and paradigm cells across Romance. Notice, for example, how some forms (red) of the reflexes of *sapiō* 'know', and *dīcō* 'say' below lack the etymologically expected stem coda (i.e. show C(V) rather than CVC stems). Although occasional mentions to these exist (Malkiel 1977, Maiden 2004:237, Mariño Paz 2019, Dubert García 2021), these stem alternations have never been subject to a systematic pan-Romance investigation.

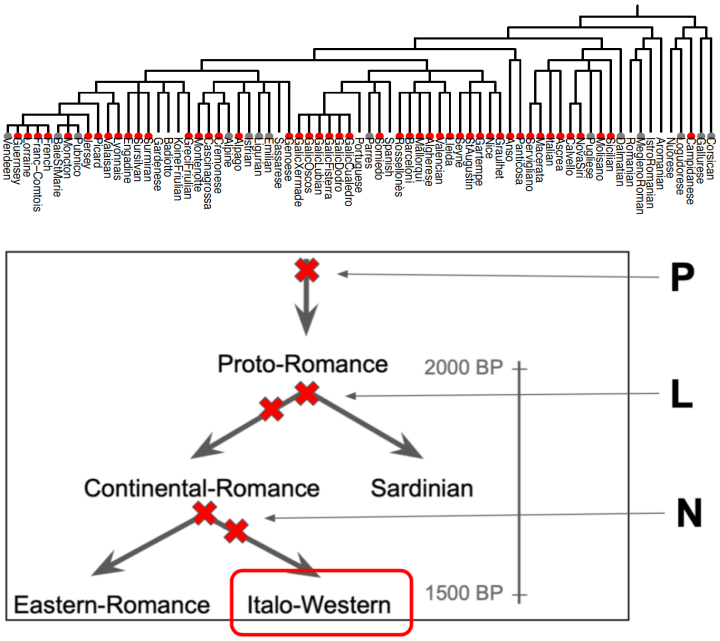
	Reflexes of Latin *sapiō* 'know'			Reflexes of Latin *dīcō* 'say'		
	Friulian	Molisano	Seyne	Fisterra	Gartempe	Ascrea
1SG.PRS.IND	saj	ˈsattʃə	ˈsabu	ˈdixo	dizə	ˈiço
2SG.PRS.IND	saːs	sje	ˈsabes	dis	dize	ˈiçi
3SG.PRS.IND	sa	sa	ˈsabe	di	di	ˈiçe
1PL.PRS.IND	saˈvin	sɛˈpemə	saˈben	diˈsimos	dizã	iˈçemo
2PL.PRS.IND	saˈveːs	sɛˈpetə	saˈbe	diˈsiðes	dize	iˈçete
3PL.PRS.IND	san	ˈsannə	ˈsabun	diŋ	dizã	ˈiçu

Relying on the data in the Oxford Database of Romance Verb Morphology (Maiden et al. 2010), I explore the lexical and paradigmatic range of these short stems. An initial qualitative inspection reveals that short stems are most common in the reflexes of Latin *faciō, dīcō, habeō, sapiō*, *possum,* and *volō*, all extremely high frequency lexemes (among the 20 most frequent ones in Latin according to Delatte et al. 1981).

For these lexemes, in addition, and for every single cell in their paradigm in every Romance variety in the database (73), I registered whether the form contained a long or a short stem (or whether the form/lexeme was undocumented or nonexistent). Short stems were found in 2773 forms in total (18.67%), while long stems appeared in 12082 forms (81.33%). Regarding their concrete paradigmatic distribution, short stems were found to be most frequent (representing usually >50% of the attested cases) in the cells 3SG.PRS.IND, 2SG.PRS.IND, 3PL.PRS.IND, 2SG.IMP, INF, and 1SG.PRS.IND. These forms are also extremely high in frequency within the Romance verbal paradigm (top 10 in the Latin corpus according to the frequency dictionary of Delatte et al. 1981).

The lexical and paradigmatic scope of the S-morphome, thus, seems to be clearly specialized into high frequency contexts. This, together with the well-known (e.g. Zipf 1935, Bybee 2006, Gahl et al. 2012) relations between length of expression and frequency of use, makes it clear that token frequency must be an important part of the story of the S-morphome. It cannot be, however, the whole story. The exact lexical, paradigmatic, and phylogenetic scope of the S-morphome arises from the interaction of token frequency with paradigmatic predictability and cross-lexeme analogies. With respect to the latter, patterns of crosslinguistic colexification (between do and say, see e.g. CLICS, Rzimski &Tresoldi et al. 2019), attested diachronic semantic shifts, and morphological changes (e.g. contaminations) unrelated to the S-morphome suggest that *faciō* and *dīcō*, *habeō* and *sapiō*, and *possum* and *volō* constitute tightly-integrated morphological-cum-semantic gangs that very probably drive stem allomorphy. With respect to the paradigmatic distribution of the S-morphome, its core domain are those cells that partake in the N-morphome domain but not in the L-morphome domain (i.e. N-L, see Herce 2019).

	PRS.IND	PRS.SBJV
1SG	cueg-o	cueg-a
2SG	cuez-es	cueg-as
3SG	cuez-e	cueg-a
1PL	coz-emos	cog-amos
2PL	coz-edes	cog-ades
3PL	cuez-en	cueg-an
INF	coz-er	
GER	coz-iendo	
2SG.IMP	cuez-e	



Consider the illustrative partial paradigm of Old Spanish cozer 'cook'. It is a lexeme that partakes in both the N (ue/o) and the L (g/z) stem alternations, which produces 4 different stems in the paradigm. The paradigmatic subdomain in red is the one where short stems have been found to prevail. 'Stem spaces' (Pirrelli & Battista 2000) like this constitute domains of perfect stem-allomorphic interpredictability that language users must perceive and actively use to solve the Paradigm Cell-Filling Problem (Ackerman et al. 2009). This explains that, when innovations arise (e.g. a short stem in an isolated cell [e.g. 2SG.IMP *fac* > fa]), alternations tend to be "pushed to the borders" between these domains.

The distribution of the S-morphome within Romance provides additional evidence that this story must be on the right track. Short stems have been found to be most prevalent in Italo-Western Romance (see above) and much less so in Insular or Balkan Romance. The latter constitute the earliest branches to split from the rest of the family. Because of this, they took part only partially in the sound changes leading to L and N alternations. This provides the final clue regarding the 'when', 'where' and 'why' of the S-morphome.

References

Ackerman, Farrell, James P. Blevins & Robert Malouf. 2009. Parts and wholes: implica- tive patterns in inflectional paradigms. In James P. Blevins & Juliette Blevins (eds.), *Analogy in grammar*, 54-82. Oxford: Oxford University Press.

Bybee, Joan. 2006. *Frequency of use and the organization of language*. Oxford: Oxford University Press.

Delatte, Louis, Étienne Evrard, Suzanne Govaerts & Joseph Denooz. 1981. *Dictionnaire fréquentiel et index inverse de la langue latine*. L.A.S.L.A, Liege.

Dubert García, Francisco. 2021. Las formas reducidas de «facĕre» en el noroeste de la península ibérica. Convergencia y divergencia geolingüísticas." *Studia linguistica romanica* 6: 163-191.

Gahl, Susanne, Yao Yao, and Keith Johnson. 2012. Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of memory and language* 66, 4: 789-806.

Herce, Borja. 2019. Morphome interactions. *Morphology* 29, 1: 109-132.

Maiden, Martin. 2004. When lexemes become allomorphs: on the genesis of suppletion. Folia Linguistica 38: 227-256.

Maiden, Martin. 2018. The Romance verb: Morphomic structure and diachrony. Oxford: Oxford University Press.

Maiden, Martin, John Charles Smith, Silvio Cruschina, Marc-Olivier Hinzelin & Maria Goldbach. 2010. Oxford online database of Romance verb morphology. Available at: http://romverbmorph.clp.ox.ac.uk/.

Malkiel, Yakov. 1977. Editorial Post-Script: Old Spanish far, fer, fazer. *Romance Philology* 31, no. 2: 257-262.

Mariño Paz, Ramón 2019. O cambio *dizes* > *dis*, *diz* > *di*, *dizen* > *din* nas fontes do galego medio. Xosé A. Fernández Salgado, Aquilino S. Alonso Núñez (eds.). *Non haberá illa, pro hai o nome. Homenaxe a Antón Palacio*. Vigo: Universidade de Vigo, 97-113. https://secretaria.u.vigo.gal/uv/web/publicaciones/public/show/342.

Pirrelli, Vito & Marco Battista. 2000. The paradigmatic dimension of stem allomorphy in Italian verb inflection. *Italian Journal of Linguistics* 12, 2: 307-380.

Rzymski, Christoph and Tresoldi, Tiago et al. 2019. The Database of Cross-Linguistic Colexifications, reproducible analysis of cross- linguistic polysemies.

Zipf, George Kingsley. 1935. The Psycho-Biology of Language: An Introduction to Dynamic Philology. Boston, MA: Houghton Mifflin Company.

CONTAMINATION IS NOT THE MECHANISM UNDERLYING MORPHOLOGICAL ANALOGY
EVIDENCE FROM DIALECTS OF INDO-EUROPEAN

Eugen Hill
(Universität zu Köln)

Two different mechanisms of "analogical change" are securely established for languages of different genetic and areal affiliations. The first mechanism, which is traditionally called "proportional analogy", consists in replacing of an inherited word-form as a whole by a new word-form, created on the model of pre-existing paradigmatic relations in similar inflectional patterns (cf. most recently Hill 2007, 2020, Garrett 2008, Fertig 2013, 2015). The second mechanism works by modification of inherited words which makes them more similar to other words existing in the language (cf. Campbell 2004: 111–113). So, in (1a) given below, the inherited 2SG.IMPV MHG *bis* 'be!' has been replaced with MoG *sei*, created on the model pattern inf. *werden*, *bleiben* etc. ~ 2SG.IMPV *werde*, *bleibe* etc. The new 2SG.IMPV *sei* clearly replaced the inherited *bis* and cannot be considered its mere modification. By contrast, in (1b) Elean Greek *optṓ* 'eight' cannot be explained as created entirely anew. For the lack of a model, this *optṓ* must be the successor of the inherited *oktṓ*, not replaced but merely modified in a way making it more similar to inherited *heptá* 'seven'.

(1a)	Middle High German			Modern German	
	INF	2SG.IMPV		INF	2SG.IMPV
	sīn	*bis*	>	*sein*	***sei***
(1b)	Homeric Greek			Elean dialect	
	'seven'	'eight'		'seven'	'eight'
	heptá	*oktṓ*	>	*heptá*	***op**tṓ*

Both mechanisms – the "proportional analogy" in (1a) and the "contamination" given in (1b) – can be responsible for change in inflection. In the case of "proportional analogy" this is obvious. As for "contamination", cf. the well-known change in inflectional properties of Latin *senātus* given in (2).

(2)		pre-Cicero Latin		Cicero
	GEN.SG	*senāt**ūs** populī=que Romānī*	>	*senāt**ī** populī=que Romānī*

It has been suggested (Fertig 2016) that "contamination" is a major factor in inflectional change, potentially responsible for virtually all innovations in which modification of inherited words and/or word-forms cannot be excluded. The proposed talk is aimed at identifying ways to test this assumption against the evidence of actually attested changes in inflection.

It is well known that most secure cases of "contamination" are found in words and word-forms which are located near each other in frequently repeated pieces of speech, for instance in numerals immediately preceding and/or following each other in counting (such as in 1b), in legal formulas (such as in 2) etc. (cf. Campbell 2004: 113). It follows that evidence for "contamination" as a mechanism underlying morphological analogy should be expected, first of all, in those cases of inflectional change where the words in question frequently occur near each other in the utterance, i.e. in verbs in a serial verb construction or in recurrent constituents of complex noun phrases.

The proposed talk will investigate the latter situation in languages belonging to one and the same Indo-European language family. As shown by the Indo-European languages with most ancient attestation, in the common ancestor of the whole family deictic pronouns followed an inflectional pattern shared by no other class of nominals. This situation is basically preserved in the ancient Indo-Iranian languages, the attestation of which begins as early as in the late 2nd millennium BCE, cf. (3a). In the other Indo-European languages, many adjectives and nouns gradually adopted case markers which originally belonged to deictic pronouns, cf. (3b) and (3c).

(3a)	Vedic Sanskrit	(late 13th c. BCE)	
	té	*sánās*	*áśvās*
	this.NOM.PL.M	old.NOM.PL.M	horse.NOM.PL
(3b)	Gothic	(late 4th c. CE)	
	þai	*mikil**ai***	*bagmōs*
	this.NOM.PL.M	big.NOM.PL.M	tree.NOM.PL
(3c)	Old Russian	(early 11th c. CE)	
	ti	*dŭlz**i***	*pĭrst**i***
	this.NOM.PL.M	long.NOM.PL.M	finger.NOM.PL

The investigation reveals that in all relevant languages the analogical transition from the inherited pattern given in (3a) to the innovative patterns given in (3b) and (3c) can hardly be attributed to "contamination" but rather progressed by means of "proportional analogy". This follows from two facts:

- in all Indo-European languages, only those adjectives and nouns are capable of secondarily adopting the pronominal case endings which already share with the deictic pronouns some endings in other cases (such as accusative singular and plural, etc.),
- in all relevant Indo-European languages, the development does not follow the inherited default linear order of NP constituents (pronoun – noun – adjective) but rather begins with the adjectives which share with deictic pronouns a more substantial part of the inflectional paradigm than the nouns.

It follows that – despite the fact that the relevant words must have been frequently located near each other in numerous utterances produced by speakers – "contamination" played no role in the analogical change of inflectional patterns described in (3). Accordingly, "contamination" cannot be considered a major factor of inflectional change in those cases where replacing of inherited words and/or word-forms cannot be excluded. Rather, analogical change in inflection seems to proceed by means of extending pre-existing encodings patterns, i.e. by means of replacing whole word-forms, also in those cases where, on the surface, modification seems a reasonable alternative.

The second conclusion which can be drawn from the investigation regards the *modus operandi* of the "contamination" itself which turns out to be sensitive to the syntactic context in which the relevant words are embedded. "Contamination" is securely established for syntactically coordinated elements such as numerals in counting (in 1b) or nouns sharing syntactic function in the same clause (in 2). By contrast, "contamination" seems to avoid elements of subordinated constructions, such as the constituents of hierarchically structured NPs given in (3). This difference between "proportional analogy" on the one hand and "contamination" on the other might be instrumental in further delimiting the operational domains of both processes.

References

Campbell, Lyle. 2004. *Historical Linguistics. An introduction*. Second edition. Cambridge (Massachusetts): MIT Press.

Fertig, David. 2013. *Analogy and Morphological Change*. Edinburgh: Edinburgh University Press.

Fertig, David. 2015. Analogy and morphophonological change. In: Honeybone, Patrick & Joseph Salmons (eds.). *The Oxford Handbook of Historical Phonology*, 205–218. Oxford: Oxford University Press.

Fertig, David. 2016. Mechanisms of paradigm leveling and the role of universal preferences in morphological change. *Diachronica* 33.4 423–460.

Garrett, Andrew. 2008. Paradigmatic uniformity and markedness. Historical convergence and universal grammar. In: Good, Jeff (ed.). *Linguistic Universals and Language Change*, 125–143. Oxford: Oxford University Press.

Hill, Eugen. 2007. Proportionale Analogie, paradigmatischer Ausgleich und Formerweiterung: ein Beitrag zur Typologie des morphologischen Wandels. *Diachronica* 24.1 81–118.

Hill, Eugen. 2020. Morphological analogy: Modification or whole word replacement? *Language* 96.1 e38–e58.

Judith Meinschaefer
judith.meinschaefer@fu-berlin.de

Analogical patterns in Romance stress assignment

In many Romance languages, stress assignment to inflected verb forms cannot be straightforwardly accounted for by rules or constraints that govern stress assignment to words of other categories. Rather, verbal stress seems to be subject to different principles, and may even appear altogether irregular (Harris 1989). At the same time, it has often been observed that the distribution of allomorphs in Romance verbal paradigms, such as the PYTA roots (Maiden 2017), or the so-called N-pattern (Maiden 2021), correlate with the locus of stress (Burzio & DiFabio 1994). When considering inflectional paradigms of verb lexemes, stress assignment and paradigmatic structure seem to be closely related, at least in Romance. Yet, this may be a historical accident, without any implications for the cognitive representation or processing of the language. The present study addresses the question whether stress and paradigm structure interact also in storage and processing of verb forms. In an artificial language learning experiment, we examine whether the presence of patterns in the distribution of stress in the paradigm facilitates (L2-) learning of an artificial miniature language, as compared to verb forms in which stress is randomly distributed over the forms of the paradigm. We study the influence of the Romance N-pattern and the U-pattern on learning stress distributions, in speakers of Spanish, a language to which these patterns are relevant, as compared to speakers of German, in which neither the N-pattern nor the U-pattern is relevant.

References

Burzio, Luigi & Elvira DiFabio. 1994. Accentual Stability. In Michael L. Mazzola (ed.), *Issues and Theory in Romance Linguistics*, 19–34. Washington, D.C.: Georgetown University Press.
Harris, James W. 1989. How Different is Verb Stress in Spanish? *Probus* 1(3). 241–258.
Maiden, Martin. 2018. New thoughts on an old puzzle. The Italian alternation type *dissi, dicesti; feci, facesti*. *Revue Romane*. 53(2). 217–260.
Maiden, Martin. 2021. The Morphome. *Annual Review of Linguistics* 7. 89–108.

Syntactically relevant morphology: The Coastal Marind fourth gender

Bruno Olsson and Matthew Carroll
Australian National University

A common assumption in most approaches to morphosyntax is that purely morphological facts, such as inflectional class membership or patterns of syncretism, do not have effects in the syntax (Zwicky 1987, 1992). Corbett (2016: 77–84) reviews some tantalising, but ultimately inconclusive, examples of purely morphological patterns seemingly having an impact on syntax. In this paper, we argue that a more convincing counterexample to morphology-free syntax is provided by syncretism in the exponence of gender agreement in Coastal Marind, a Papuan language of the Anim family.

Coastal Marind has a system of four genders (labelled with the Roman numerals I–IV; Olsson 2017). Nouns are assigned to a gender on the basis of animacy and sex: male humans are in Gender I, female humans and all animals in Gender II; inanimates are assigned to Genders III and IV. Assignment to the two inanimate genders is largely arbitrary: most inanimate nouns belong to Gender III (70%), and the residue is assigned to Gender IV (30%). The relevant morphological pattern is found in targets that agree according to gender. Table 1 illustrates exponence of gender agreement on an adjective. The important observations are (i) that only the animate genders (I and II) trigger a distinction between singular and plural, and (ii), that the agreement form of the plural of the animate genders is syncretic with Gender IV (cf. the shaded cells). The syncretic form here is *akik*; other adjectives differ in the details of exponence, but the pattern of syncretism remains constant across lexemes. Outside the nominal domain, the same pattern is found in the indexing of patient-like participants on the verb stem, as illustrated in Table 2. Exponents of patient indexing (which are unreated to those found in the nominal domain) occur either as prefixes, infixes or suffixes depending on the inflectional subclass of the verb, but the conflation of Gender IV and the 3rd person plural is consistent in a system otherwise rife with lexical exceptions. This systematic syncretism is **morphomic** (Aronoff 1994), most clearly because there are no semantic features grouping the categories (Gender IV and animate plural) that map onto the shared set of exponents. We discuss, and reject, descriptions in term of semantic commonalities, which would have deprived the pattern of its morphomic status, as morphomes per definition have no semantic basis.

We then demonstrate some unexpected repercussions of the morphomic pattern in two areas of Coastal Marind grammar: **verb suppletion** according to participant number and the availability of **applicative constructions**. Several verbs employ unrelated stems depending on whether the absolutive participant is singular or plural, as in the pair *man* 'one come' vs. *nayam* 'many come'. We show that the morphomic syncretism reoccurs with such verbs, and we systematically find the suppletive plural stems triggered by absolutive arguments belonging to Gender IV, as in example (1). This alternation is remarkable as the semantic concept of participant number is a well-known trigger of suppletion cross-linguistically, but here the verb suppletion turns out to obey a completely morphomic pattern found elsewhere in the grammar. We then turn to two applicative constructions used to add a comitative participant to a motion verb. One construction is used to add an animate object, while the other adds an inanimate object. We show that in addition to involving unrelated morphology, these constructions differ structurally and semantically and must be treated as separate syntactic configurations (rather than different variants of the same construction). Surprisingly, we find that the construction associated with inanimates is unavilable to nouns in Gender IV, and that despite their inanimate status, these nouns select for the construction associated with an added animate object—again, as predicted by the morphome grouping Gender IV nouns with (plural) animates. The data is remarkable, and we argue that it constitutes a serious challenge to the claim that syntax is morphology-free.

Tables and examples

	‘light (not heavy)’	
	SG	PL
I	*akek*	*akik*
II	*akuk*	
III	*akak*	
IV	*akik*	

Table 1: Exponents of gender agreement

		SG	PL
I/II	1	*hwaga⟨n⟩ib*	
	2	*hwaga⟨y⟩ib*	*hwaga⟨h⟩ib*
	3	*hwag⟨ø⟩ib*	
III		*hwag⟨ø⟩ib*	
IV		*hwaga⟨h⟩ib*	

Table 2: Patient indexing in *hwagib* ‘put away’

(1) a. *nggawil-yahun* *ø-a-* *man*
motorcycle(III) NEUT-3.A- come
‘A motorcycle came.’

b. *lahwalah-yahun* *ø-a-* *nayam*
airplane(IV) NEUT-3.A- many.come
‘An airplane came.’

References

Aronoff, Mark. 1994. *Morphology by Itself*. Linguistic Inquiry Monographs 22. Cambridge, Massachusetts: MIT Press.

Corbett, Greville G. 2016. “Morphomic splits”. In: *The Morphome Debate*. Ed. by Ana R. Luís and Ricardo Bermúdez-Otero. Oxford: Oxford University Press, pp. 64–88.

Olsson, Bruno. 2017. “The Coastal Marind language”. PhD thesis. Singapore: Nanyang Technological University.

Zwicky, Arnold M. 1987. “Slashes in the passive”. In: *Linguistics* 25.4, pp. 639–669.

— . 1992. “Some Choices in the Theory of Morphology”. In: *Formal Grammar: Theory and Implementation*. Ed. by Robert Levine. Vancouver Studies in Cognitive Science 2. New York: Oxford University Press, pp. 327–371.

The distribution of suppletive stems over inflectional paradigms, at the various stages of the life-cycle of suppletion

Frans Plank
Somerville College, Oxford

My aim here is to examine possible generalisations about the paradigmatic distribution of suppletive stems. To this end, it is advisable to distinguish important stages in the life-cycle of suppletion: (i) the birth of suppletively inflecting lexemes; (ii) their coming of age; (iii) their demise, usually through morphological regularisation, but also lexeme loss.

As to (i), suppletion can come about (most commonly) through phonological differentiation or lexeme combination – and this is a distinction that matters insofar as lexeme-combination origins guarantee that suppletive stems are distributed along morphological lines, whereas phonological differentions don't need to respect paradigmatic structures. Further, to the extent that suppletions owe their origins to defective paradigms (hence Gabelentz's term "Defectivwesen", as opposed to Osthoff's "Ergänzungswesen"), the question is the incidence of paradigmatic gaps and their semantic, phonological, or random nature.

As to (ii), once lexemes have become suppletive, the subsequent paradigmatic distributions of the stems involved can be observed sometimes to be diachronically stable and sometimes rather unstable, constantly undergoing re-distributions. One might think that re-distributions are likelier if the original distributions didn't make much morphological sense – but this expectation isn't always met, and it is not obvious (to me) why there should be such differences in stability.

As to (iii), when morphological irregularities such as suppletion are levelled, one would assume that such developments are guided by morphological principles (including markedness, paradigm structure conditions and such).

For present purposes I hope to be able to draw together work of mine on such matters, partly published and partly in progress (or also long stationary).

References (somewhat egocentric: apologies)

Plank, Frans. 1981. *Morphologische (Ir-)Regularitäten: Aspekte der Wortstruktur-theorie.* Tübingen: Narr.
Plank, Frans (ed.). 1991. *Paradigms: The economy of inflection*. Berlin: Mouton de Gruyter.
Plank, Frans. 1991. Rasmus Rask's dilemma. In: *Paradigms*, 161-196.
Plank, Frans. 1991. Of abundance and scantiness in inflection: A typological prelude. In: *Paradigms*, 1-39.
Plank, Frans. 1994. Homonymy vs. suppletion: A riddle (and how it happens to be solved in ...). In: *Agreement Gender Number Genitive &* (*EUROTYP Working Papers* VII/23), 81-86.

Plank, Frans. 2015. The temporal nature of universals: Is the paradigmatic distribution of suppletion reined in by timeless laws or laws of change? Various presentations, slides at https://ling.sprachwiss.uni-konstanz.de/pages/home/plank/fp_presentations.html

Plank, Frans. 2016. Vom Suppletiv(un)wesen, in Beziehung zu Paradigmenstruktur und in besonderer Rücksicht der historischen Natur beschränkter Möglichkeiten. In Andreas Bittner & Klaus-Michael Köpcke (eds.), *Regularität und Irregularität in Phonologie und Morphologie: Diachron, kontrastiv, typologisch*, 1-28. Berlin: Walter de Gruyter.

Plank, Frans. 2016. Review of Jonathan David Bobaljik, *Universals in comparative morphology* (Cambridge MA: MIT Press, 2012). *Linguistic Typology* 20. 687-713.

Plank, Frans. 2016. The Indo-European demonstrative pronoun **to-/*so-* 'that': A unique suppletion pattern explained, with far-reaching implications. Various presentations, slides at https://ling.sprachwiss.uni-konstanz.de/pages/home/plank/fp_presentations.html

Plank, Frans. 2016. Re-distributions of suppletive stems over inflectional paradigms. Various presentations, draft at https://ling.sprachwiss.uni-konstanz.de/pages/home/plank/fp_unpublished.html

Plank, Frans & Nigel Vincent. 2019. Suppletion: Questions for history and theory. *Transactions of the Philological Society* 117 (3). 319-337. (Issue on suppletion, Ed. Vincent & Plank.)

Evolving morphomes

Erich Round, Louise Esher & Sacha Beniamine

One of the most striking facts of inflectional morphology is the ubiquity of organisational principles that are morphomic, *sensu* Aronoff (1992). In inflectional class systems, the lexicon is arbitrarily partitioned. Within the partitions (classes), lexemes pattern similarly in terms of the content found in their paradigm cells. A second kind of morphomic organisation is what Round (2015) terms 'metamophomes', of the kind familiar from Romance (Maiden 2018), where it is the paradigms themselves that are arbitrarily partitioned. Within the partitions, sets of cells contain common forms such as a stem allomorph, although, while these partitions *per se* recur throughout the lexicon, the specific pieces of form that fill the partitions may be distinct for each lexeme. Here we ask, can such morphomic structures arise emergently, through simple, evolutionary, analogical steps? We present three new findings.

We begin with the analogical model of Ackerman & Malouf (2015). Here, the simple evolutionary steps revolve around predicting the content of one paradigm cell, c_1, from the implicative relationships that hold between the forms found in c_1 and those found in some other cell, c_2, throughout the lexicon. Though the model can give rise to fleeting lexical class structure, that structure is lost rapidly as the model tends towards cell-wise uniformity. As Ackerman & Malouf emphasise, the key macro-level dynamic here is evolution towards low paradigm entropy, a property which holds distinct advantages for language learners and users. We show that when the predictive relationships governing analogy are extended beyond just c_2, to all cells other than c_1 — arguably more linguistically plausible — the evolution to low paradigm entropy is more rapid, though multiple inflectional classes still struggle to survive.

We move next to a model by Esher (2015) where analogy hinges not on forms per se, but on patterns of (non)identity among cells within paradigms. The model was conceived to test whether metamorphomes would emerge. The results are doubly surprising. Firstly, metamorphomes do not emerge, rather the system moves rapidly towards lexeme-wise uniformity. Secondly, the entropy trajectory of the inflectional system begins not by falling but by rising. However, it crests soon thereafter and subsequently falls to a very low level more rapidly than in Ackerman & Malouf's model, raising the theoretically interesting observation that if the ultimate goal is low entropy, a somewhat circuitous path can be more effective than a direct one.

Finally, we shift away from asking whether metamorphomes can *emerge* from randomness via analogy, to examining instead their propensity to survive and persist, when they are present in the initial state — corresponding, for example, to the situation just after a sound change has created a widespread, common partitioning within lexemes' paradigms. Here, the model based on Ackerman & Malouf (2015) causes the initial morphomic patterns to disintegrate, whereas the model based on Esher (2015) indeed preserves them better. Our conclusion is that the modelling of morphome dynamics in terms of simple evolutionary, analogical steps is highly promising. Research has barely begun, yet intriguing, early results are now beckoning for more to be done.

Beyond proportional analogy:
the case of Slavic *vĭsĭ* 'all' and *sicĭ* 'this, such'

Florian Wandl (University of Zurich)

Due to regular fronting of back vowels after palatal consonants, Slavic nominal and pronominal declensions are divided into two types (cf. Vaillant 1958, Lunt 2001). Nouns and pronouns with stem-final palatals triggering fronting inflect according to the so-called soft type while others belong to the hard type. An exception to this distribution is, however, constituted by the two pronouns represented by Old Church Slavonic *vĭsĭ* 'all' and *sicĭ* 'this, such'. Despite having a soft stem-final consonant resulting from velar palatalization, they show hard-type endings in the instr. sg. m./n. and the gen., dat., instr., loc. pl. of all three genders. It is generally assumed that these endings cannot have intruded into the paradigm secondarily because of the lack of a clear proportional model (cf. Meillet 1900, Vaillant 1953, Vermeer 2008). The purpose of the present paper is to show that this claim is unjustified. The curious paradigms of *vĭsĭ* and *sicĭ* can be explained by considering morphological changes other than proportional analogy.

In my talk, I will present two mechanisms by which the unexpected endings can be explained: (1) product-oriented innovation and (2) syntagmatic assimilation (Wandl 2020). Product-oriented innovation is a concept proposed by Bybee & Slobin (1982, cf. also Albright & Hayes 2003) to account for morphological changes that are not based on relations between input and output forms, but on output forms solely. The output form is created to resemble a specific schema constituted by phonological and semantic factors. In the case of the Slavic pronominal endings under scrutiny the creation of forms resembling other pronouns displaying a stem final consonant resulting from velar palatalization in the corresponding case forms can be identified as the driving factor behind the innovation. Syntagmatic assimilation consists in rapprochements of often co-occurring forms (Andersen 1980, Fertig 2013). As is shown by Andersen (1980), such innovations can result in the introduction of new paradigmatic forms. Since the pronoun *vĭsĭ* regularly occurs together with the hard-type pronoun *tŭ* 'this' to form a syntagma with the meaning 'all these', I propose that it was here or in similar constructions where the pronoun could have adopted hard stem endings. The curious fact that the innovation did not spread beyond the instr. sg. m./n. and the gen., dat., instr., loc. pl. can be explained due to phonotactic constraints.

Selected Bibliography

Albright, A. & Hayes, B. (2003): Rules vs. analogy in English past tenses: a computational/experimental study, in: *Cognition* 90, 119–161.
Andersen, H. (1980): Morphological change: towards a typology, in: *Historical Morphology*, ed. by J. Fisiak, 1–50. The Hague.
Bybee, J. L. & Slobin, D. I. (1982): Rules and schemas in the development and use of the English past tense, in: *Language* 58/2, 265–289.
Fertig, D. (2013): Analogy and morphological change. Edinburgh.
Lunt, H. G. (2001): Old Church Slavonic grammar. Seventh revised edition. Berlin – New York.
Meillet, A. (1900): Notes sur quelque faits de morphologie, in: *Memoires de la Société de linguistique de Paris* 11, 6–21.
Vaillant, A. (1958): Grammaire compareé des langues slaves. Tome II, Morphologie, Première partie, Flexion nominal. Deuxième partie, Flexion pronominal, Lyon.
Vermeer, W. (2008): Pedersen's chronology, in: *Studies in Slavic and General Linguistics* 34, 503–571.
Wandl, F. (2020): On the relative chronology of the II regressive and the progressive palatalization of Slavic, in: *Russian Linguistics* 44, 79–108.