Paradigmatic complexity metrics as signals of phylogenetic relatedness:
a proof of concept in Romance and Pamean diachrony

Borja Herce & Balthasar Bickel
University of Zurich

Abstract

Morphological complexity metrics like entropy, and notions like the Paradigm Cell-Filling Problem, have recently (re)gained popularity for the synchronic analysis of inflectional systems. The potential of these quantitative approaches for diachronic research, however, remains largely untapped. This paper constitutes a first exploration of whether and if so, how these methods can be used profitably in this domain. We first use Romance to establish the diagnostic value of complexity metrics for phylogenetic relatedness under a best-case scenario, with rich historical knowledge, and then apply them to Pamean to show that the same metrics help diagnose the phylogenetic relatedness of tenses and inflection classes even when, as is the case in this family, most of the morphological material has been replaced or altered beyond recognition. Results suggest that complexity metrics can successfully diagnose phylogenetic relatedness over extended periods of time and fruitfully complement traditional qualitative approaches.

Keywords: paradigm structure, complexity metrics, diachrony, relatedness, cognacy

1 Introduction

In the comparative method, the relatedness of languages is determined through lexical elements that have the same or similar meaning and show sound correspondences (e.g. English *to, two, ten, tongue, tooth, toe* vs German *zu, zwei, zehn, Zunge, Zahn, Zehe*). If there are enough of these to demonstrate regularity in the correspondences (regular meaning shifts and sound laws), this strongly suggests that the languages descend from a common ancestor. This inference rests on three widely accepted assumptions. First, the pairing of a lexical meaning with a form is mostly arbitrary, so much so that similarities are usually better attributed to common descent than to convergent evolution. Second, the state space, i.e. the total range of possible phonological forms for a given concept, is so large that similarities are not expected to recur by chance. Third, regular sound correspondences are unlikely to emerge from borrowing, which in fact disrupts them (e.g. English *pound* gives away its borrowed status because it disrupts the *p:f* correspondence with, say, Latin [cf. *pondō*]).

Lexical roots, however, are just one of many aspects that make up a language. Another aspect of language for which the same three assumptions are potentially valid is the structure of morphological paradigms (i.e. semantically related word forms like Spanish *amo amas ama* etc.), specifically the amount of complexity, (ir)regularity, and idiosyncrasy in the relations between the cells of a paradigm, e.g. the extent to which plural forms are predictable from singulars, or which affixes syncretize which categories. The choices that languages make in these regards are also largely arbitrary, thus defining a space of possible patterns that is large enough that chance recurrence is unlikely. Paradigms, and inflectional

morphology, are also commonly believed to resist borrowing (Matras 2015) and to change slower than other aspects of grammar (Greenhill et al. 2010).

While all of these assumptions have limits, as they do in the lexicon (e.g. Blasi et al. 2016 and Erben Johansson et al. 2020 for arbitrariness, and Tadmor 2009 for borrowing), paradigmatic structures often seem to be probative of the phylogenetic relatedness between languages. Meillet (1925, 1958) was among the first to notice this and to suggest that unusual paradigmatic structures like the Indo-European Ablaut are, like lexical items, highly arbitrary and idiosyncratic traits that reflect common inheritance when they are shared across multiple languages. Other cases where this has been noted are for example. Macro-Gunwinyguan TAM conjugations (Alpher et al. 2003) or Romance morphemes (Maiden 2018). Nichols (2014) uses derivational paradigms to uncover ancient phylogenetic relatedness. The accumulation of many specific aspects of paradigms across languages seems extremely unlikely by chance alone (Nichols 1996), making it more plausible to occur in separate languages as a result of common descent than independent development. In other words, focusing on paradigms, closely-related inflectional systems are expected to be more similar in their paradigmatic structures than far-related or unrelated systems.

While plausible, the diagnostic value of paradigms has not been empirically tested. This is largely due to the fact that most aspects of paradigmatic structure and morphological predictability have only recently become reliably quantifiable in a uniform way across languages. At the core of this are various information-theoretic and set-theoretic metrics of (ir)regularity and inter-predictiveness of the cells in a paradigm, such as conditional entropy (Ackerman & Malouf 2013) and principal parts (Stump & Finkel 2013).

In their presentation of a tool to calculate paradigmatic complexity metrics (the Principal Parts Analyzer [PPA]), Stump & Finkel (2015) analyzed the nominal declensions of Latin and Sanskrit. They noted a large degree of similarity of the two systems across multiple metrics, which becomes easy to gauge (see Figure 1) if we compare them against the variability they observed in the same metrics cross-linguistically in their previous 2013 monograph.
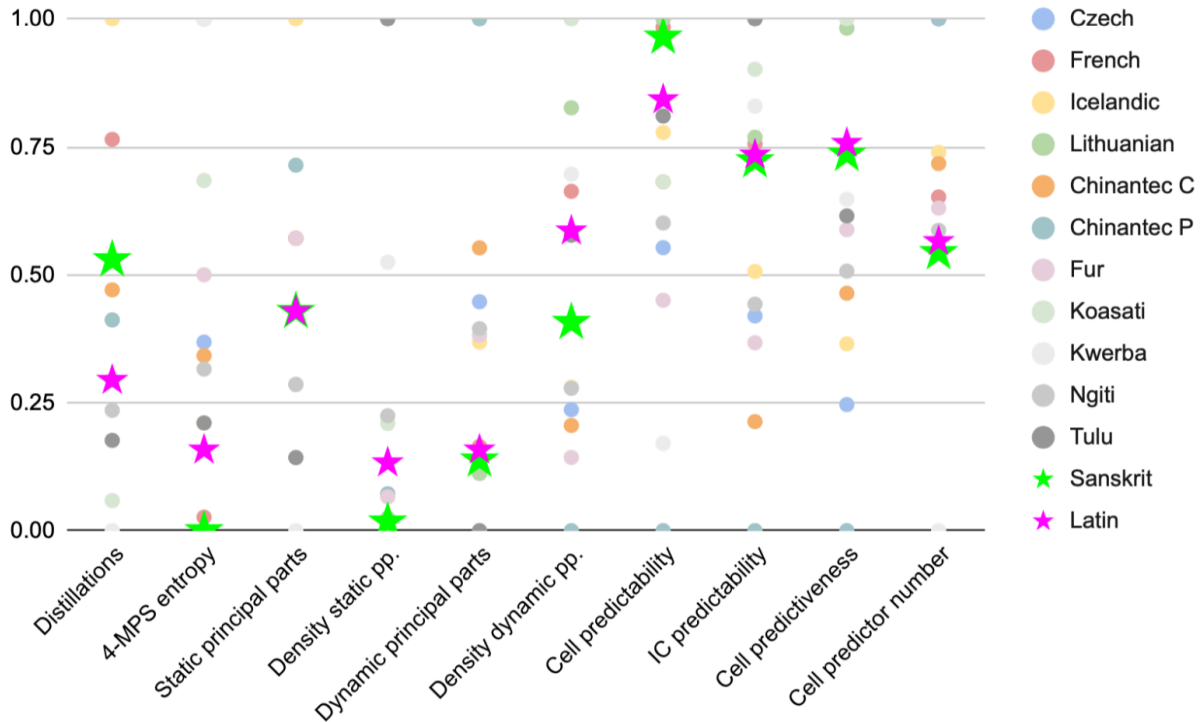
Figure 1: Quantitative profile of Latin (large green star) and Sanskrit (large fuchsia star) nominal declension in the sample of inflection class systems from Stump & Finkel 2013.[1]

Figure 1 shows that despite some variability between metrics, Latin and Sanskrit inflection class systems are always close or (nearly) identical in the values they take. Although in their 2015 paper Stump and Finkel did not pursue this idea, it is plausible that the similarity of the Sanskrit and Latin declensions might be due to their homology. We know that Latin and Sanskrit descended from a common ancestor, Proto-Indo-European, that must have been spoken a few millennia before these languages. As a result, the quantitatively very similar profile of the two systems may simply follow from their more-or-less faithful continuation of the characteristics of the ancestral Proto-Indo-European declension.

In this paper we assess whether the morphological complexity metrics of paradigms can be successfully harnessed for diachronic research. Our focus is on probing phylogenetic relatedness and not on reconstructing phylogenetic trees, with finely resolved topologies and time estimates. While an extension to tree modelling may be possible, we find it critical to first establish the extent to which complexity metrics can in fact recover whether inflectional systems are phylogenetically related at all.

The structure of this paper is as follows: In Section 2 we introduce the quantitative metrics from Stump & Finkel (2013) that we will work with in the rest of this paper. In Section 3 we assess the extent to which the metrics recover phylogenetic relatedness where this is known. More specifically, we quantify the paradigm structures of 100 verb cognates in 7

---

[1] The numerical range of the metrics has been transformed to have the minimum attested value as 0 and the maximum attested one as 1. Thus, for example, the number of distillations varied (in Stump & Finkel's 2013 original 12-language sample) between 4 and 21. By subtracting the minimum from the original values and dividing the result by the new maximum (17), we transformed all values into a comparable 0-1 scale (e.g. Latin's 9 distillations are transformed into a value 0.294: (9-4)/17 = 0.294)

different Romance languages. The idea is that a language family with a comparatively well-understood history can provide a benchmark to assess the stability and/or historical informativity of different aspects of (the predictive relations within) an inflection class (IC) system. Our results suggest that indeed these metrics tend to be inheritable enough that homologous inflectional systems, homologous tenses, and homologous classes, tend to be significantly more similar than unrelated ones. Section 4, in turn, explores whether these metrics might inform diachronic research in un(der)studied language families as well, for example by providing a signal that historical linguists can "pick up" even after (almost) all cognate morphology has disappeared or changed beyond recognition. The verbal inflectional paradigms of two Oto-Pamean languages (Otomanguean, Mexico) are explored with a mixed quantitative and qualitative approach that appears to successfully identify shared ancestry between tenses and classes in these languages. Section 5 summarizes the paper and its results, and proposes avenues for future further research.

2 Paradigm complexity metrics

The last 10-15 years have witnessed a "turn" (Blevins 2013) in morphological research by which quantitative approaches to inflectional systems have quickly increased in popularity. How language users of Estonian, Russian, Latin, and other highly-inflecting languages predict and produce all forms in the paradigm, often on the basis of an incomplete input, has long been a question of interest. This has been so not only in the Word and Paradigm tradition (e.g. Matthews 1965), but also in "traditional" pedagogical grammars, which have been capitalizing for centuries on the fact that a small subset of word forms (so-called 'principal parts' like Latin *amo, amare, amavi, amatum*, see Finkel & Stump 2009) usually suffice to predict the complete paradigm. The Paradigm Cell Filling Problem (Ackerman et al. 2009), hence, was and continues to be one of the most pressing issues in morphological research. Information Theory (with its core notion of entropy) has provided the theoretical background for a lot of empirical research in recent years (e.g. Milin et al. 2009, Ackerman & Malouf 2013, Bonami & Beniamine 2016, Sims & Parker 2016, Cotterell et al. 2019).

Inflectional systems are not reducible to any one metric, as they are complex, multifaceted entities. Different metrics capture different aspects of an inflectional system. Some of the most widespread and most straightforward metrics involve simply counting. Simple count metrics have been used for a long time in typological research: phonological inventories differ in their number of phonemic distinctions; inflectional paradigms differ in the number of features, values, and cells involved; inflection class systems differ in the number of classes and their number of members, etc. Different values can usually be ranked according to their relative degree of 'complexity'. Thus, for example, a paradigm with 8 cases and 3 numbers for 24 total different cells is considered more complex, *ceteris paribus*, than one with a simple SG/PL morphological distinction and thus just two cells. Similarly, at the whole-system level, having 38 inflection classes should be considered more complex than having only three. These metrics correspond to what Ackerman & Malouf (2013) call "enumerative complexity".

Information- and set-theory provide metrics with higher resolution (Sagot 2013). The structuredness of an inflectional system, for example, can be approximated by its minimal

description length (i.e., what is known as Kolmogorov Complexity). From another perspective, one can also probe the orderliness of an inflectional system by assessing the uncertainties involved in predicting unobserved datapoints (e.g. in a new lexeme). Shannon's (1948) entropy is the core notion for this. These metrics capture what Ackerman & Malouf (2013) call "integrative complexity", in contrast to enumerative complexity.

Couched in this newly-emerging field of research, Stump and Finkel also introduced, in their 2013 monograph 'Morphological Typology: from Word to Paradigm', a series of metrics that capture different aspects of an inflectional class system's complexity (i.e. of how easy/difficult it is to infer some forms in the paradigm on the basis of other forms). These included, among others, metrics like static and dynamic principal parts, predictability, predictiveness, etc.[2] A web interface (the Principal Parts Analyzer [PPA], freely-accessible at https://www.cs.uky.edu/~raphael/linguistics/analyze.html) was also made available where these metrics can be calculated online. Much as their simpler 'counting' predecessors, different values in each of these metrics can also be ranked for their relative degree of complexity. This is something that will not constitute the focus of the present paper, however, which will be more concerned with assessing the relative degree of (dis)similarity of the different possible values of a given metric. In what follows, we briefly introduce the metrics explored in Stump & Finkel (2013), before their diachronic potential is explored in Romance in Section 3, and in Pamean in Section 4.

2.1 Distillations

A distillation is a set of paradigm cells whose morphology is isomorphic, in the sense that the corresponding word forms are either identical, or perfectly predictable from each other. Here and with the rest of the metrics, a suitable subset of the Romance inflectional data from Section 4 can serve to illustrate what it is exactly that the different metrics capture.

	INF	SG.IMP	1SG.PRS	2SG.PRS	3SG.PRS
Poner 'put'	er	Ø	go	es	e
Correr 'run'	er	e	o	es	e
Caer 'fall'	er	e	jgo	es	e
Amar 'love'	ar	a	o	as	a
Sudar 'sweat'	ar	a	o	as	a
Fumar 'smoke'	ar	a	o	as	a

Table 1: A subset of the Spanish conjugation (suffixes only)

---

[2] It is beyond the scope of this paper to motivate the Set-Theoretic or Information Theoretic bases and formulae that underlie these different notions, which we take from Stump & Finkel (2013) "as is", without addressing possible improvements in this paper.

If we had an inflectional class system like the one in Table 1, its 5 cells would be reduced to 3 distillations. The forms INF, 2SG.PRS and 3SG.PRS would be part of the same distillation because there is a perfect bidirectional predictability relation between the three cells (2SG.PRS and INF in this case just adding /s/ and /r/ respectively to the 3SG.PRS form). Inflection class systems, thus, differ in their number of distillations, with the total paradigm size (i.e. number of cells) constituting a logical upper limit. The rest of the metrics that will be described in this section will build on distillations (i.e. {INF, 2SG.PRS, 3SG.PRS},[3] SG.IMP, 1SG.PRS in Table 1) rather than on individual cells.

2.2 n-MPS entropy

Shannon Entropy is an information-theoretic metric that captures uncertainty. The uncertainty involved in two equally probable values/outcomes (e.g. the flip of a coin, or a choice between a 3SG.PRS -e or -a in Table 1) is 1 bit. Where there is more uncertainty (e.g. choosing between -Ø, -e, or -a for the SG.IMP), the entropy value becomes larger (in this case 1.46), and in the absence of uncertainty (e.g. in the absence of allomorphy), entropy equals 0.

n-MPS[4] entropy is the average conditional entropy of a cell, given every set of paradigm cells (excluding the predicted cell) with up to n members. Thus, the 4-MPS entropy in Stump & Finkel (2013) gives the average entropy of a cell, given all different 0, 1, 2, 3, and 4-cell sets of other possible cells. The larger the number of cells that are used for inference, the lower the remaining entropy/uncertainty. In our illustrative system in Table 1, a 1-MPS entropy, for example, gives a value of 0.8 and a 2-MPS entropy a value of 0.64.[5]

2.3 Static vs dynamic principal parts

A set of principal parts is any set of cells in a lexeme's paradigm from whose realizations one can reliably deduce the forms of every remaining cell in the paradigm. As Stump & Finkel (2015:106) explain, however, there are several different conditions one may choose to adhere to in relation to this, most notably: does it have to be the same set of cells that act as principal parts across all lexemes? When the answer to this question is 'yes', then we are speaking of the traditional (or 'static') principal parts.[6] When the answer to this question is

---

[3] In the rest of this section, this distillation will be simply referred to as INF for the sake of brevity. The isomorphic cells 2SG.PRS and 3SG.PRS (columns 4 and 5) can be safely ignored for the purposes of the rest of the metrics that will be presented in this section.

[4] MPS (Morphosyntactic Property Specification) is another term for what is usually referred to as a (content) paradigm cell.

[5] Stump & Finkel (2013, 2015) and the Principal Parts Analyzer report these and other entropy values multiplied by 100 (i.e. as 80 and 64 in this case).

[6] In Latin conjugation, for example, first person singular present active indicative, the present active infinitive, the first person singular perfect active indicative, and the supine (e.g. *fero ferre tuli latum* for the verb 'carry') were traditionally provided for all verbs.

'no', we are free to pick and choose the most informative cell(s) of every lexeme and we are then speaking of 'dynamic' principal parts.

Applied to our illustrative sub-system in Table 1, two static principal parts (e.g. SG.IMP and 1SG.PRS) are minimally needed to predict the whole paradigm. Any cell by itself (e.g. a SG.IMP in -e, or a 1SG.PRS in -o) would fail to reliably predict the right form (e.g. -o vs -jgo, -e vs -a) in other parts of the paradigm. In the case of 'dynamic' principal parts, however, a single cell suffices to identify the inflection class and full paradigm of most verbs in Table 1. A SG.IMP in -Ø, for example, or a 1SG.PRS in -jgo unmistakably predicts the remaining forms. Only the inflection class represented by *correr* requires two dynamic principal parts. In the subsystem of Table 1, thus, the average number of dynamic principal parts needed is 1.25 ([1+1+1+2]/4).

2.4 Density (of static and dynamic principal parts)

A fact about principal parts that we did not address in Section 3.3 is that there need not be only one optimal (i.e. minimal) set of principal parts. When two static principal parts were needed in Table 1, for example, we mentioned the cells SG.IMP and 1SG.PRS. However, the alternative two-cell set of INF+1SG.PRS is an equally effective choice of cells to predict the complete paradigm in our toy example. In Table 1 there are three possible two-cell sets, two of which suffice to predict the complete paradigm. This is the density of static principal parts: the proportion of sets of cells of the same size that serve as static principal parts, in our case 2/3 = 0.666, i.e. 66%.

The same reasoning applies to dynamic principal parts. As noted earlier, *correr* requires two dynamic principal parts, and the other lexemes just one. However, which cells exactly these are was not mentioned and, the same as with static principal parts, more than one option is often available: for example, in *correr*, both INF+1SG.PRS, and SG.IMP+ 1SG.PRS (i.e. again two out of three logically possible 2-cell sets) can predict the whole paradigm. The same happens with *poner*, where SG.IMP, and 1SG.PRS both suffice to predict the whole paradigm (2/3). In the inflection class of *amar*, both INF and SG.IMP can serve as dynamic principal parts (2/3). In *caer*, by contrast, 1SG.PRS is the only possible dynamic principal part (1/3). The density of dynamic principal parts, thus, is the ratio of actual to possible dynamic principal parts, here (2+2+2+1)/(3+3+3+3) = 0.583, i.e. 58.3%.

2.5 Cell predictability

The predictability of a cell in a paradigm is the proportion of subsets of other cells in the paradigm that reliably predict the cell at stake. Focusing again on our toy example in Table 1, the INF form of *correr* can be predicted by its SG.IMP *-e*, but not by the 1SG.PRS form *-o*, nor by an empty set (note that in a cell that lacked allomorphy, [or in an inflectional system with no inflection classes] even an empty set could predict the form of a cell). The INF form of *amar* in Table 1 can also be predicted from a SG.IMP in *-a* but not from a 1SG.PRS in *-o* or from an empty set. The infinitive forms of *poner* and *caer*, by contrast, are predictable from both their SG.IMP and their 1SG.PRS forms (but still not from an empty set of cells). As

a result, the average predictability of the infinitive cell in Table 1 is ([2/3]+[2/3]+[1/3]+[1/3])/4 = 0.5. Due to their higher overall allomorphy, the other two cells of Table 1 turn out to be a bit more difficult to predict (both with 0.25 average cell predictability. Averaging across cells ([0.5+0.25+0.25]/3) one can reach a paradigm-wide metric of predictability, in our case 0.333.

Of course, in our small toy-example of Table 1, it does not make sense to consider, for a cell's predictability, sets of predictor cells larger than 1,[7] but in the quantitative exploration of larger inflectional systems one can choose the maximum size of the sets one will consider for the calculation of a cell's (or a system's overall) predictability. However, because, as Stump & Finkel (2015) mention, very large sets of cells will almost always be able to predict a given form, it is useful to look at smaller sets only. Stump & Finkel (2013, 2015) chose to look only at sets of maximally 4 cells. This (arbitrary) limit will also be adopted here for comparability in Sections 4 and 5.

2.6 Inflection class predictability

A similar *modus operandi* as the one described for cell predictability applies to inflection class predictability, which refers to the proportion of sets of cells (of a given size) that reliably predict the inflection class the forms belong to. Focusing, as before, on one-cell sets in our toy example, the inflection class of *correr* is not unmistakably identified by either an INF in -*er*, a SG.IMP in -*e*, or a 1SG.PRS in -*o* (0/3). In *poner*, by contrast, a SG.IMP in -*Ø* and a 1SG.PRS in -*go* (but not an INF in -*er*) identify the inflection class they belong to (2/3). The class of *amar* is similar in that 2 out of its 3 distillations (INF -*ar*, and SG.IMP -*a*) can serve for inflection class identification. In *caer*, this is found for one (1SG.PRS -*jgo*) of its forms. The average inflection class predictability is thus [(0/3)+(2/3)+(2/3)+(1/3)]/4 = 0.417.

In the same way as for cell predictability in Section 3.5, one can freely choose the maximum size of the sets of cells one would like to use for prediction. As before, our toy example is too small for larger sets to make sense, but in larger inflectional systems as in Sections 4 and 5, it will be useful to set a larger limit for these sets. As before, we will follow, for comparability, Stump & Finkel's (2013, 2015) choice of 4 as the limit.

2.7 Cell predictiveness

Cell predictiveness is the reverse of cell predictability, i.e. the proportion of other cells in the paradigm whose form is reliably predicted from the form of a cell. In our illustrative subparadigm from Table 1, for example, the IMP.SG predicts all (2/2) of the other distillations in the paradigm in the inflection classes of *poner* and *amar*, but only half (1/2) of the cells of *correr* and *caer* (-*e* can predict INF -*er*, but cannot distinguish between -*o* or -*jgo* as the 1SG.PRS form). The predictiveness of the IMP.SG cell is thus ([2/2]+[2/2]+[1/2]+[1/2])/4 =

---

[7] There are only 2 other distillations in the subparadigm of Table 1 besides the infinitive one, and this extra set (SG.IMP+1SG.PRS) always turns out to be a perfect predictor of the infinitive. If we added this 2-cell set, then, this would simply increase the overall predictability: ([3/4]+[3/4]+[2/4]+[2/4])/4 = 0.625

0.75. Performing the same calculation for our subparadigms' other two distillations would give us a predictiveness of 0.25 for the INF and 0.5 for the 1SG.PRS. The average of all three distillations (i.e. the average cell predictiveness for the whole system) is thus (0.25+0.5+0.75)/3 = 0.5.

2.8 Cell predictor number

Cell predictor number is the average lowest number of dynamic principal parts needed to determine the realization of a cell in the paradigm, averaged across all cells. In our toy example from Table 1, there is always just one other cell one could possibly need to determine the realization of a cell (i.e. cell-predictor number = 1). This number would change (<1) if, for example, there was a cell with no allomorphy (i.e. if 1SG.PRS was always -o, then this could be predicted on the evidence of zero dynamic principal parts). The cell predictor number would increase (>1) if there was extensive Blur (Carstairs 1994) between inflection classes that prevented any cell in the paradigm of an inflection class from being predicted on the basis of just another cell.[8]

2.9 Other metrics

The above list of inflection class system complexity metrics does not exhaust all of the potentially interesting ones, and metrics could be improved and expanded. Indeed, various others have already been proposed in the literature. Stump & Finkel's (2013) tool can calculate various other ones too, most notably adaptive (rather than static or dynamic) principal parts (relying on an alternative way to choose principal parts; see Finkel & Stump 2007), and average entropies (conditional on another cell or unconditioned, weighted by token frequency of the inflection classes or unweighted). Simpler count-type metrics like the number of different inflection classes, number of distinct exponents (see Bickel & Nichols 2007), or number of different signatures (i.e. patterns of syncretism, two in our illustrative Table 1) continue to be relevant and will also be provided. There is little reason to believe *a priori* that any one of the notions discussed throughout this section will be more important than the others, or in what concerns this paper, more resilient and diachronically informative. While morphological change has been explored for over a century, this literature (e.g. Kuryłowicz 1945, Mańczak 1957) does not easily translate into quantifiable predictions nor do they map straightforwardly to the metrics we explore. This is an empirical matter that will be addressed by looking at the Romance conjugation system in different languages in the following Section 4.

3 The quantitative profile of Romance conjugations

This section constitutes the first of the two main pillars of this paper and are intended as a first step to empirically assess the diachronic resilience of different aspects of inflection class

---

[8] Imagine, for example, a fictional Spanish verb with the forms *-ar -Ø -jgo -as -a* in Table 1. No cell in that verb would be predictable from just one other form.

systems. To do this, we focus on a language family which has i) a rich enough system of inflection classes continued independently in different languages, ii) rich comparative synchronic data from several well-described languages, and iii) a well-understood diachronic development, to be able to establish links between the quantitative findings and the history and finer-grained philological details of different languages. The verbal inflectional system of Romance constitutes one of the best ones[9] according to these criteria and has been therefore selected as the "fruit fly" for this study.

We created a database with the full paradigms (in phonological form) of 100 cognate verbs across 7 different Romance languages.[10] These are, from west to east, Portuguese, Spanish, Catalan, French, Romansh (Surselvan), Italian, and Romanian. Depending on data availability and their lexical conservativeness, different languages ended up with a different number of analyzed cognates (always above 50). Cognate sets also vary in their number of reflexes (i.e. derivatives in modern Romance), with the majority being attested, however, in 4 or more languages. Latin etyma were chosen according to data availability (i.e. those with more reflexes in the Romance Verbal Inflectional Dataset 2.0 of Beniamine et al. 2020). The data in this database were supplemented with the reflexes of the most frequent Latin verbs in LatInfLexi 1.1 (Pellegrini & Passarotti 2018) with the exclusion of the verb 'be', until reaching the pre-established goal of 100 cognate sets.

---

[9] Other possibilities come to mind that could be similarly suited, for example the nominal case-number inflectional system of (Balto-)Slavic languages, which could provide an ideal opportunity to replicate the findings of the present paper's Section 3 in a very different type of inflectional paradigm.

[10] A full list of the analyzed cognate sets in the different languages is provided in the appendix. The Romance Verbal Inflection Dataset 2.0 (Beniamine et al. 2020) was used as the backbone. The data in this database were supplemented with the following other sources: Flexique for French (Bonami et al. 2014, in red in the appendix), LeFFi for Italian (Pellegrini & Cignarella 2020), in green in the appendix), VeLeRo for Romanian (Herce & Pricop 2024, in blue in the appendix), Unimorph 2.0 (Wiktionary) for Portuguese (Kirov et al. 2018, in orange in the appendix), La Flexió Verbal en els Dialectes Catalans for Catalan (Perea & Ueda 2010, in pink in the appendix), and with the first author's native-speaker knowledge for Spanish (in yellow in the appendix).

Forms in all sources were checked for mistakes and transcription inconsistencies. Among the latter, for example, was the transcription of the imperfect ending -*ea* of Romanian after a palatal. In this context, this form was transcribed inconsistently in the Romance Verbal Inflection Dataset 2.0 (e.g. 1SG.IPF.cook *coceam* was transcribed as /koˈʧam/ while 1SG.IPF.make *făceam* was transcribed as /fəˈʧe̯am/. Similarly, 1SG.IPF.burn *încingeam* was transcribed as /intʃinˈʤe̯am/, while 1SG.IPF.scatter *spărgeam* was transcribed as /spərˈʤam/. These inconsistencies would distort the metrics explored here by spuriously multiplying the number of inflection classes. They were therefore manually corrected.

Transcription choice differences were also ironed out when integrating data from different sources. The Romance Verbal Inflection Dataset 2.0, for example, reports a more conservative pronunciation of French â (/ɑ/) than the Flexique of Bonami, which transcribes these as /a/. A form like *chantâmes* was thus transcribed as /ʃɑ̃tɑm/ in one source but as /ʃɑ̃tam/ by the other, thus leading again to the multiplication of forms if this had not been amended.

The forms were also inspected manually for isolated errors which were corrected (e.g. French 3PL.PRET.sew *cousirent* was transcribed as /kusiʁ/ in The Romance Verbal Inflection Dataset 2.0 rather than correct /kuziʁ/, and Spanish GER.die was coded as /moˈrjendo/, rather than correct /muˈrjendo/).
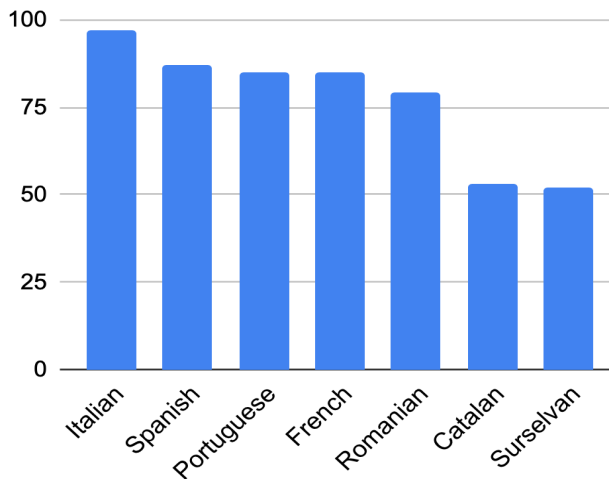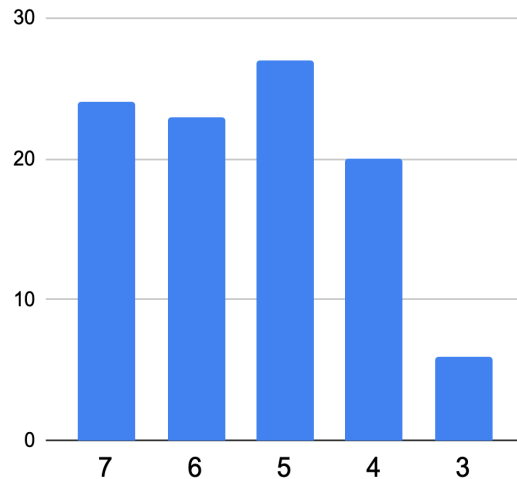
Figure 2: Analyzed cognates per language

Figure 3: Number of reflexes per item

To further ensure that we are comparing like with like, only homologous tenses and forms were taken into account.[11] These included the reflexes (whatever their contemporary semantic value) of the following Latin forms and tenses: infinitive, gerund, imperative (2SG and 2PL), present indicative, present subjunctive, and imperfect indicative. Former *perfectum* tenses were not surveyed because they have been often lost in individual languages, and because they were/are frequently built on a stem different from the *infectum* forms, which would raise problems in the identification of the lexical vs inflectional segmental material. The same reasoning applies to the past participle, a highly irregular form in most contemporary Romance languages.

Before the quantitative profile of the verbal conjugations in these languages can be analyzed one needs to take some more executive decisions in relation to the issue of how the inflectional material will be distinguished from the lexical material. This is obviously not trivial, particularly in Romance inflectional morphology, and a vast literature exists on the general challenges of segmentation (Snyder & Barzilay 2008, Spencer 2012, Bank 2017, etc.). Furthermore, abundant examples exist that lexical material can be reanalyzed as inflectional, and *vice versa,* and cases of this are rife in Romance diachrony. Contributing to this literature is not among the goals of this paper. However, and precisely because of the challenging nature of the stem-suffix segmentation in Romance, clearcut criteria need to be adopted to ensure this is performed in the same way across languages, thus ensuring the subsequent comparability of the data.

For the purposes of what follows, the following segmentation criteria were adopted: Segments that appear in the vast majority of the paradigm (defined as being absent from 5 cells at most) were considered part of the stem, and more variable ones part of the inflectional material. Stem vowel alternations (involving similar qualities and often correlated

---

[11] In the course of diachronic evolution, new tenses may appear (e.g. the Romance future and conditionals, grammaticalizing from periphrastic constructions with the verb 'have'), or disappear (e.g. the synthetic passive tenses from Latin). This is likely to skew the inflection-class quantitative profiles of individual languages in unexpected ways, as some (e.g. Portuguese) end up having significantly larger paradigms than others (e.g. Catalan, Surselvan). It should be the aim of a separate piece of research to analyze to what extent this is the case and whether or not comparing whole inflectional systems blindly (i.e. without considering tense or lexeme homology) would change the results reported here.

with stress differences) and "trapped" morphology (i.e. segments that do exhibit variability but are surrounded by invariable segments) were an exception to this, in that they were considered lexical even when they were not present in the required number of cells. Below are some examples that illustrate the results of this "blind" segmentation procedure in particularly challenging cases, the ones where these executive criteria are most necessary.

	INF	2SG.IMP	1SG.PRS.IND	3SG.PRS.IND	1SG.PRS.SBJV
Sp. 'can'	pod-er	pwed-e	pwed-o	pwed-e	pwed-a
It. 'want'	vo-lere	vɔ-ʎʎi	vɔ-ʎʎo	vwɔ-le	vɔ-ʎʎa
Port. 'fit'	tɾɐ-zeɾ	tra-ʃ	tra-gu	tra-ʃ	tra-gɐ
Fr. 'want'	vu-lwaʁ	vœ-j	vø-Ø	vø-Ø	vœ-j
Cat. 'can'	pu-gɛ	pɔ-t	pu-k	pɔ-t	pu-gi
Rom. 'lie'	zə-tʃa	za-tʃ	za-k	za-tʃe	za-k
Sur. 'know'	sa-ver	sa-piəs	sa-i	sa-Ø	sa-pi

Table 2: Some segmentation examples in highly irregular verbs

As Table 2 illustrates, the segmentation instructions above result sometimes (particularly in highly irregular verbs) in seemingly unorthodox choices like It. *vo-lere* or Surselvan *sa-ver*. In general, this particular segmentation procedure is obviously quite generous when conceding inflectional status and, unlike the most canonical analyses of Romance inflection, where a lot of the morphological "action" is assigned to the stem (see e.g. Montermini & Bonami 2013), it has the obvious result of allocating more morphological variation to suffixes. It must be mentioned, in any case, that which exact segmentation procedure is adopted (e.g. one which unlike this one relegates more alternations to the stem) has not been found to change the findings reported in the present paper.[12]

Using these segmentation instructions, we generated full tables of suffixes (or of what Stump & Finkel 2013 call more neutrally "distinguishers"). We then explored the predictability relations between the resulting forms by calculating the metrics presented in Section 2 using the Principal Parts Analyzer web interface provided by Finkel and freely available online at https://www.cs.uky.edu/~raphael/linguistics/analyze.html. The results are found in Table 3.

Language	Distillations	Number of exponents	Inflection classes	Signa-tures	Static P.P.	Density static P.P.	Dynamic P.P.	Density dynamic P.P.	Adaptive P.P.
Romanian	12	148	37	22	6	0.00216	1.42	23.31	3
Italian	12	192	29	8	6	0.00325	1.66	28.2	3

---

[12] The same metrics that will be presented in the remainder of this section were also calculated on the basis of a completely different (much more canonical/conservative) segmentation, and this resulted in no substantial differences in the extent to which Romance conjugations resemble each other. This can be found in the supplementary materials (https://gitlab.uzh.ch/borja.hercecalleja/ic-complexity-metrics-in-diachrony).

Surselvan	11	156	25	10	3	0.00606	1.41	30.28	3
French	12	77	28	12	5	0.00126	1.64	16.91	3
Catalan	12	137	32	14	6	0.00649	1.67	18.63	4
Spanish	11	140	24	4	5	0.00216	1.5	19.62	3
Portuguese	11	190	25	2	5	0.01299	1.21	32.15	3

Table 3: Romance inflection class metrics (most complex in gray, least complex in white)

Language	Cell predictab.	IC predictab.	Cell predictiv.	Cell predictor number	Unconditioned entropy of cell	Conditional entropy of cell	4-MPS entropy
Romanian	0.831	0.518	0.486	1.11	3.39	0.68	19
Italian	0.81	0.531	0.457	1.1	2.94	0.71	21
Surselvan	0.765	0.592	0.365	1.13	2.5	0.77	26
French	0.819	0.392	0.538	0.96	2.53	0.82	22
Catalan	0.731	0.457	0.305	1.1	2.32	0.97	31
Spanish	0.772	0.466	0.455	1.09	2.56	0.61	26
Portuguese	0.867	0.659	0.526	1.04	3.28	0.64	15

Table 3: Romance inflection class metrics (continued)

These data allow us to make within-Romance comparisons as for the relative degree of complexity of the different systems according to different properties. According to the average number of dynamic principal parts required to predict the whole paradigm, for example, Catalan (1.67) counts as the most complex of the analyzed Romance languages, and Portuguese (1.21) as the most simple. Across metrics, Catalan and Romanian, for example, appear to be on the complex side, while Portuguese and Spanish seem to be on the simpler side.

Interesting (and contentious) as this may be, the goal of this section is a different one, namely the assessment of the relative degree to which these metrics are similar enough to identify the relatedness of inflectional systems. Although they do seem similar, the values in Table 3 tell us little without a cross-linguistic contextualization.[13] We cannot say, for example, if the variability (3～6) found in Romance for static principal parts is high or low, since we do not know what the variability is of this trait outside this family. For this reason we develop a

---

[13] Within Romance one could in theory check whether these metrics allow us to recover the phylogenetic topology more specifically (e.g. the early branching of Romanian, or the closer affinity of Spanish and Portuguese). The comparatively low number of metrics analyzed here, alongside the comparatively rake-like diversification of Romance, make it unlikely that such fine-grained details can be recovered in this particular case. Hierarchical clustering on the basis of the metrics in Table 3 shows French as the most divergent Romance language, followed by Catalan. It seems, therefore, that shared retentions might be driving within-Romance similarities more than shared innovations. As noted in the introduction, we see it as a future opportunity to explore the potential of extended morphological data and metrics for estimating tree topologies.

comparative baseline of unrelated inflection class systems from other families and other parts of the world.

As a point of departure for the baseline we take the 11 non-Romance languages analyzed in Stump & Finkel 2013. We supplemented these with 26 others from around the world from other sources: Russian (russ1263), Greek (mode1238), Võro (voro1231), Kadiwéu (kadi1248), Nuer (nuer1246) and Seri (seri1257) from Sims & Parker (2016), Chichimec (chic1272), Malinaltepec Me'phaa (mali1285), Tilapa Otomi (tila1239), and Xochapa Mixtec (xoch1238) from the Otomanguean Inflectional Class Database (Feist & Palancar 2015), Arapesh (arap1279) and Amele (amel1241) from Hein & Muller (2009), Gaadudju (gaga1251) from Harvey (2011), Jawoyn (djau1244) from Merlan (n.d.), Rembarnga (remb1239) from Saulwick (2003), Mawng (maun1240) from Singer (2006), Limilngan (nucl1327) from Harvey (2001), Wandarang (wand1263) from Heath (1980a), Marra (mara1385) from Heath (1981), Jaminjung (jami1236) from Schutze-Berndt (2000), Chiquihuitlán Mazatec (chic1250) from Jamieson (1982), Ritarungo (rita1239) from Heath (1980b), Nangikurrunggurr (nang1259) from Reid (1990), Pite Saami (pite1240) from Wilbur (2014), and Wadjiginy (wadj1254) from Ford (1990). All languages' data and metrics can be found in the supplementary materials. In total 36 inflection class systems from 22 different stocks were analyzed quantitatively, with the same metrics that have been described for Romance, to provide a standard for comparison.
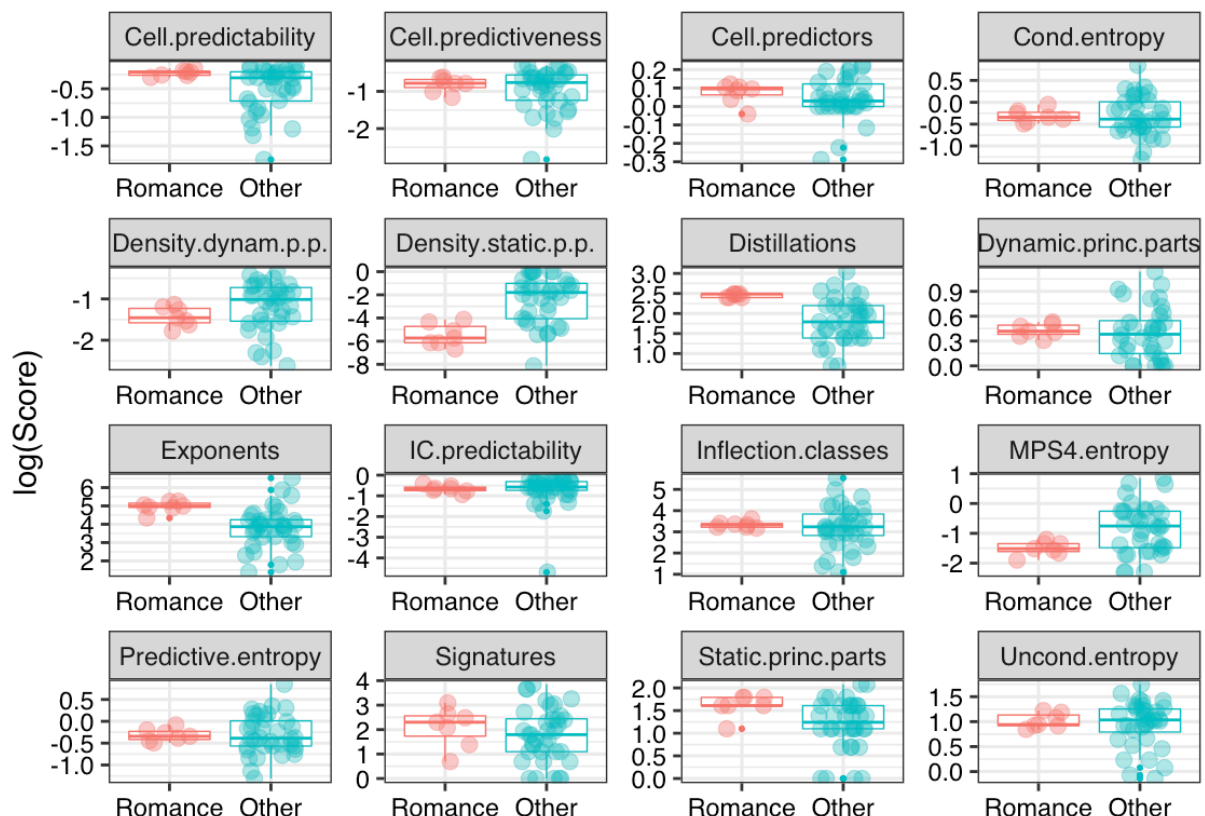

Figure 4: Romance IC's traits against the general cross-linguistic variability

Figure 4 shows that different aspects of an inflection class system appear to be quite variable regarding their observed range of values within Romance, and hence regarding their estimated within-family similarity (which should correlate to diachronic stability).

Number of signatures appears to be the least stable of the surveyed variables, as the 7 Romance languages surveyed already cover most of the observed cross-linguistic range. By contrast, the number of inflection classes, number of distillations, or cell predictability among others appear to be very stable, with the Romance range representing only a small proportion of the cross-linguistically observable one.

Abstracting away from this variability, Romance seems to differ overall significantly from the control sample in showing a lower variance across metrics, occasionally, together with a different mean. Principal component analysis (Figure 5) ratifies these differences at the whole-system level, and shows that Romance inflection class systems are comparatively similar to each other and cover only a small fraction of the cross-linguistically attested design space, which suggests that these metrics might be able to preserve some signal of phylogenetic relatedness. While some unrelated inflectional systems (see Chichimec, Rembarnga and Czech) are similar to the Romance verbal systems across these metrics, the vast majority of unrelated systems are far away from the Romance cluster.
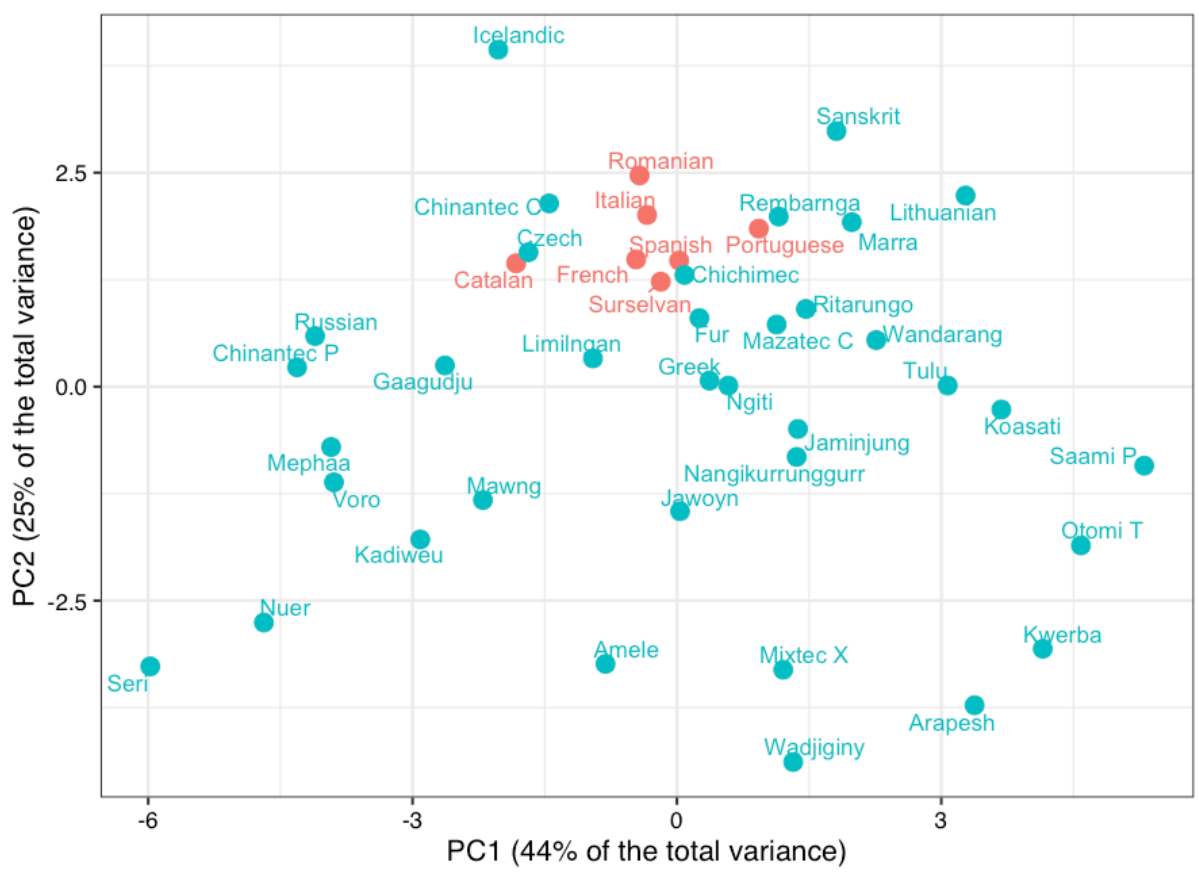
Figure 5: PCA of Romance (red) and cross-linguistic (blue) IC systems' traits

In order to quantify the evidence for the Romance cluster we turn to what is known as "distributional" or "scale/location" models (Bürkner 2018), where differences means and variances can be assessed simultaneously. We fit the models in a Bayesian framework using Stan (Carperter et al. 2015), via the brms (Bürkner 2017) package in R (R Core Team 2020) (see Supporting Information for technical and statistical details). Data were fit with Lognormal, Beta, or negative-binomial (Gamma-Poisson) models depending on the nature of

the variable (Log-normal: positive continuous, e.g. unconditioned and conditional entropy; Beta: proportions, e.g. density of static principal parts, and inflection class predictability; Negative-Binomial: counts, e.g. number of static principal parts and number of inflection classes.). The Romance 7-language sample was compared via repeated sampling (B=10) to same-size samples from the non-Romance group. Results suggest that, as expected, whereas the mean values of Romance were not found to be different from the control sample in any consistent direction (see Figure 6), variances were consistently lower than in the control samples (see Figure 7).
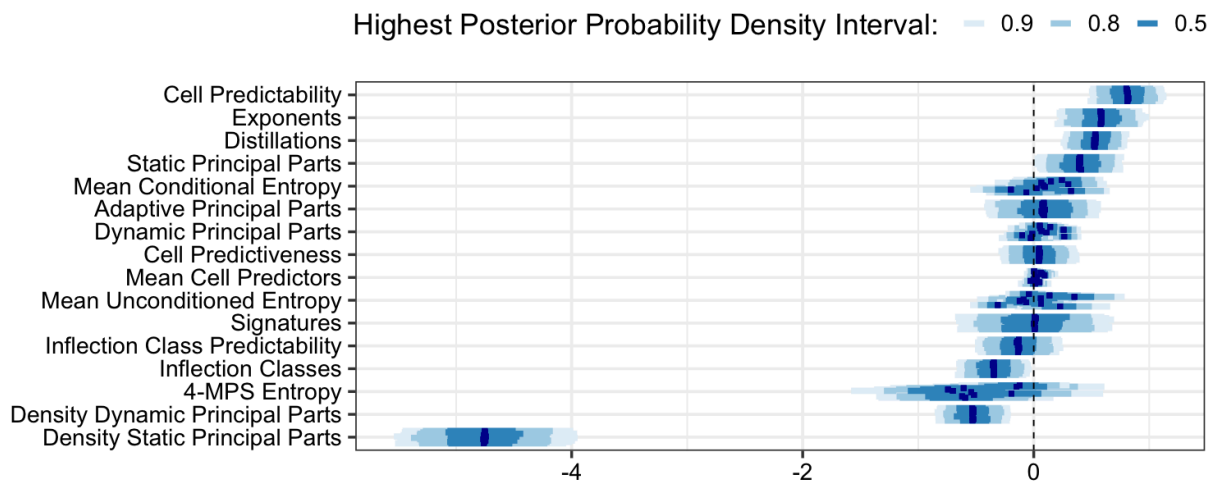
Figure 6: Differences in means between Romance and non-Romance systems (each bar in each metric represents one random sample of non-Romance languages. Dark blue dots represent the median posterior estimate, and shades represent different posterior density intervals, as per the legend. Metrics are ranked by their median posterior estimate.)
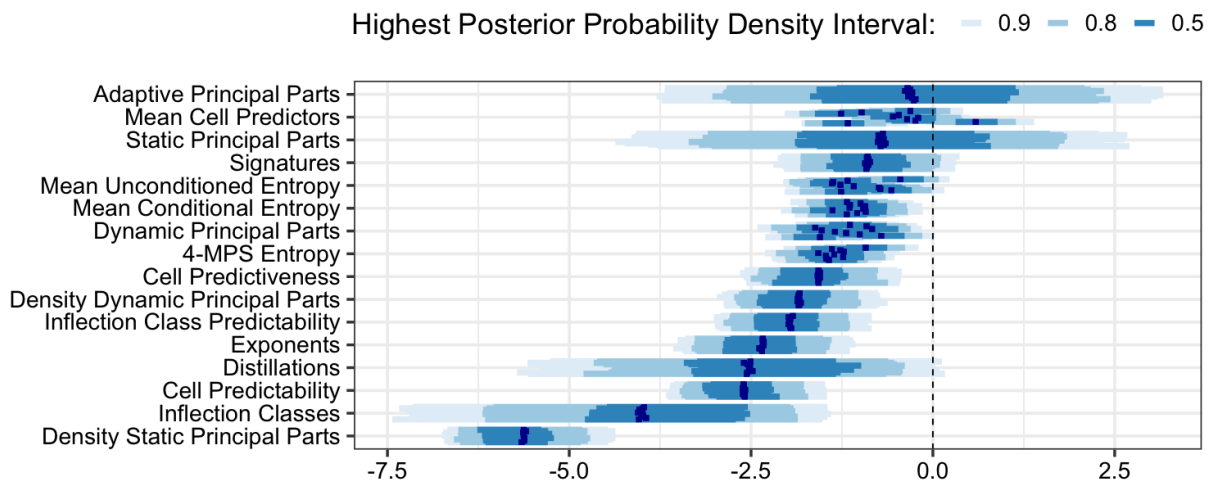
Figure 7 Differences in variances between the Romance and non-Romance systems (same conventions as in Figure 6)

In the case of variances, 95% of the posterior probabilities for most variables lie below 0, suggesting variances are decisively smaller within than outside Romance. At the same time, important differences can be found between metrics, with some (e.g. density of static principal parts, or number of inflection classes) having a decisively smaller variance in Romance compared to others like e.g. mean unconditioned entropy, which are not as

decisively different. Others, like adaptive and static principal parts are not different at all, with even the middle 50% of the posterior including 0. These differences between metrics suggest that they can be ranked with respect to their relative degree of diachronic stability, approximately as in Figure 7. Differences in means (Figure 6) between Romance and the control sample don't reflect the stability of different aspects, and might be expected to be random (inherited in this case from the Latin inflectional system's idiosyncrasies) but could be equally useful to identify related systems.

As an alternative method to evaluate this quantitative Romance signature, we fitted a Gaussian Mixture Model (GMM) to the data using the R package mclust (Scrucca et al. 2016). In datasets built from the aggregation of different subpopulations with different properties, multivariate GMMs try to recover the membership and traits of the original subpopulations that went into a dataset. In a two-component solution (i.e. Romance vs Other), our data is split into 23 vs 20-member language groups, with all Romance languages consistently in the first one. Our data is unusual from a GMM perspective in that one of our groups (namely the 'Other' one) is not a coherent subpopulation like Romance, but rather a control sample made up mostly of unrelated IC systems. As a result, a higher number of components is also an option worth exploring. According to the Bayesian Information Criterion, the optimal number of components for our dataset would be 7. Under this solution, all 7 Romance conjugation systems are once again grouped together in the largest (19 language) cluster, while the 'Other' group is split into various smaller components with 8, 8, 4, 2, 1, and 1 languages respectively, all of them containing non-Romance languages exclusively.

Our results are of course preliminary; in need of replication in a different language family, and subject to important confounds. Part of the reason for the clustering of certain properties in Romance, for instance, may not be diachronic inheritance exclusively. Paradigm size, for example, is one of the variables that can be assumed to have some impact in some of these metrics (although paradigm size is, of course, also likely to be highly inheritable). Larger paradigms demand more regularity or would otherwise be unlearnable in the context of a Zipfian input (Cotterell et al. 2019, cf. Johnson et al. 2020). As a result of our efforts to control for homology, we have also trimmed Romance paradigms into identical size (22-cell) ones. One might therefore like to assess the impact of paradigm size in the variables that are being analyzed here.

In relation to this, we note, impressionistically, that a correlation holds in our sample between paradigm size and some of the metrics for very small paradigms (<7 cells) exclusively. Thus, if we compare Romance conjugations against IC systems of a similar paradigm size only (e.g. those between 12 and 48 cells [N=12]), the obtained results do not differ by much. The variance observed within Romance would be still less than the one observed in unrelated IC systems of a similar size, except for the least stable metrics (static and adaptive principal parts, signatures, and distillations).

Looking at cross-linguistic IC systems of a similar size is one, but not the only option to disentangle the effects of phylogenetic relatedness from those of paradigm size. An alternative is to explore the properties of individual tenses or individual inflection classes within our Romance dataset, rather than the entire inflectional systems. If the quantitative correlates of relatedness can also be observed in individual subsystems, this would also

open the door to probe on the relatedness of these subsystems as well by applying these metrics to languages and families that lack the historical records of Romance (i.e. Latin) and/or a comparable philological tradition.

All Romance tenses, for example, constitute 3x2=6 person-number cell subparadigms. but differ in their relatedness. Thus, the common descent of the Catalan PRS.IND and the Romanian PRS.IND should be reflected in a more similar quantitative profile of these tenses, compared to, say, Catalan PRS.IND and Romanian IPF.IND. Figure 8 shows the metrics for related Romance tenses. Along with the hierarchical clustering analysis of Figure 9, it suggests that phylogenetic relatedness is reflected in the greater similarity of related tenses to each other. Looking at the metrics in Figure 8, one can appreciate that PRS.IND tenses tend to be the most complex across languages (e.g. most distillations, most exponents, most dynamic principal parts) while IPF.IND ones tend to be the least complex, with PRS.SBJV tenses falling usually in-between. In terms of the similarity between the individual tenses in different Romance languages, hierarchical clustering (hclust, ward.D, prenormalized, euclidean) shows again (Figure 9) that homologous tenses tend to be most similar. Observe, for example, that tenses descended from the Latin present indicative are usually grouped with each other (bottom cluster in Figure 9), and so are most present subjunctives and imperfect indicatives. Furthermore, in the case of outliers (e.g. the Romanian PRS.SBJV clustering with PRS.IND tenses rather than with other present subjunctives), plausible historical language-particular reasons can be identified as to why this might be the case.[14]
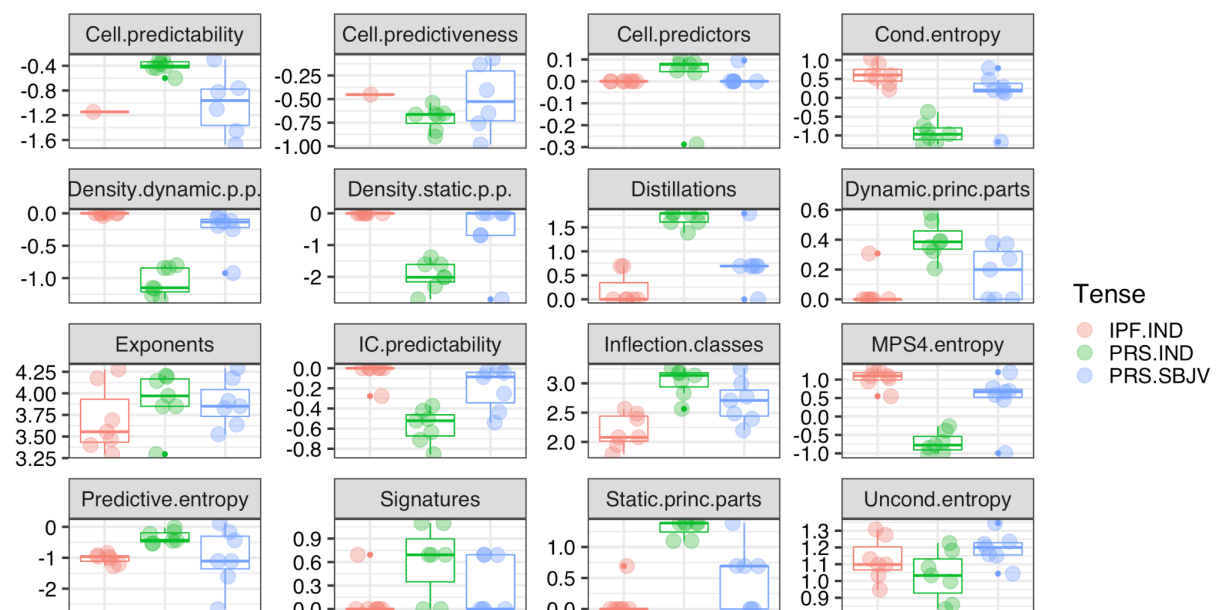
Figure 8: Quantitative complexity profile of Romance tenses in different languages

---

[14] The Romanian present subjunctive forms, for example, only continue Latin present subjunctive forms in the third person. First and second person are syncretic with (and constitute etymological continuants of) the present indicative. Thus, the quantitative similarity of this tense with the present indicatives is, as a matter of fact, signalling (partial) homology.
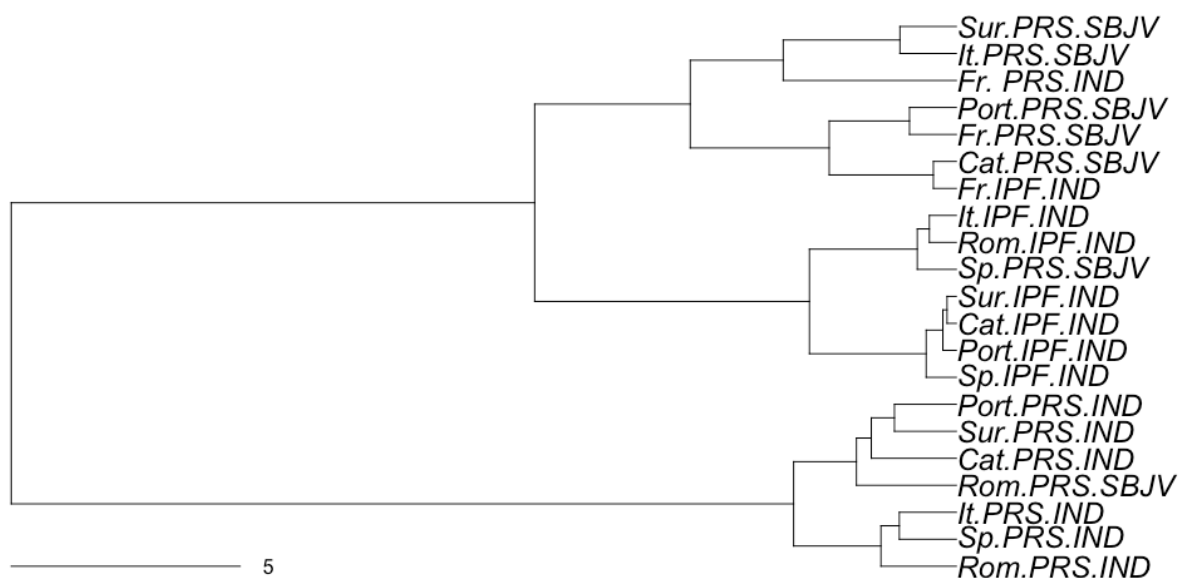
Figure 9: Hierarchical clustering of the tenses in different Romance languages

Alongside tense homology, IC homology is another relation of interest to historical linguists and one worth reflecting on quantitatively. homologous ICs (like the First Conjugation of French [e.g. *crever* 'burst', *semer* 'sow', *tousser* 'cough'] and the First Conjugation of Spanish [e.g. *entrar* 'enter', *lavar* 'wash', *gozar* 'enjoy'], or the Second Conjugation of French [e.g. *finir* 'finish', *salir* 'exit', *jouir* 'enjoy'] and the Third Conjugation of Spanish [e.g. *partir* 'split', *servir* 'serve', *dirigir* 'direct']) should also be expected to have a greater (quantitative) affinity than nonhomologous inflection classes. Even though individual lexemes have sometimes migrated from one IC to another (e.g. *tousser* < Lat. 4th *tussīre*, *gozar* and *jouir* < Lat. 2nd *gaudēre*, *dirigir* < Lat. 3rd *dīrigere*), IC homology relations can still be identified that allow the historical linguist to identify the Spanish and the French First Conjugations as continuants of the Latin First Conjugation, and the Spanish Third Conjugation and the French Second, as continuants of the Latin Fourth Conjugation. In line with the exploration of tense homology (Figures 8-9), and whole-system homology (Figures 4-5), therefore, the interest lies in assessing whether these shared ancestries can be recovered to some extent from the metrics that this paper is exploring.

It is worth mentioning at this point that, although the notions introduced in Section 2 applied to whole IC systems, some of them did so only by averaging the relevant metric across all the different ICs. These aspects/metrics (n-MPS entropy, dynamic principal parts and their density, cell and IC predictability, and predictive entropy, together with number of members) apply to individual inflection classes unproblematically, and their variability can therefore be explored, in the same way as it has been done at the whole-system and at the tense level.
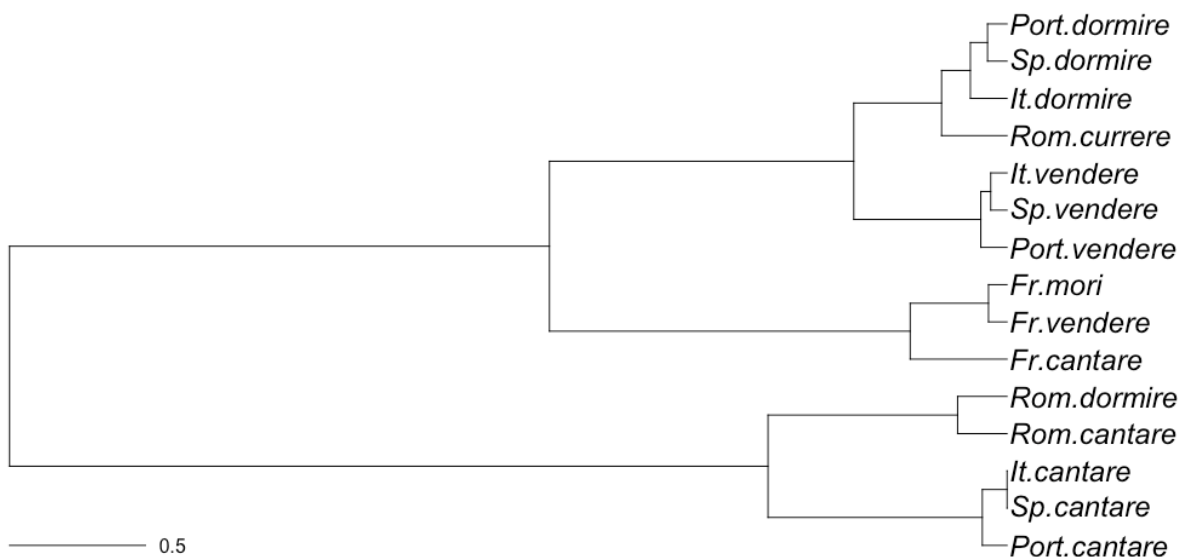
Figure 10: Clustering of the largest ICs[15] across Romance languages

Figure 10 shows the quantitative similarity between the most direct and largest ICs descended from Latin conjugations 1st (represented with reflexes of *cantāre* 'sing'), 3rd (reflexes of *vendere* 'sell' and *currere* 'run'), and 4th (reflexes of *dormīre* 'sleep' and *mori* 'die'). The same as with tense homology in Figure 9, we can see that homologous inflection classes in Romance also tend to preserve similar quantitative profiles. Although some classes are, again, grouped incorrectly (e.g. French conjugations not being grouped with their homologous conjugations in other Romance languages), it remains the rule that homologous ICs tend to be most similar to each other. Most notably, Figure 10 shows that the homologous ICs of Spanish, Portuguese, and Italian, the most morphologically conservative languages in our sample, are always the closest quantitatively.

Taken together, our results suggest that the information- and set-theoretic quantitative fingerprint of inflectional systems, inflection classes, and/or individual tenses, preserves a signal of shared ancestry in Romance. This finding is consistent with the tradition that emphasizes the diagnostic importance of paradigm structure in historical linguistics (Meillet 1958, Nichols 1996). What the quantitative approach contributes is a principled and replicable way of assessing and comparing paradigm structure. As such it bears promise for languages that are less researched and for which no historical data are available. We turn to one such case in the following Section 4.

4 Diagnostic use of IC complexity metrics in Oto-Pamean languages

4.1 Quantitative matching of Pame homologous tenses and ICs

The results in Section 3 suggest that most metrics of paradigmatic predictability and complexity preserve a signal of (sub)system relatedness in Romance. This language family,

---

[15] The metrics were calculated for the three largest ICs in our dataset (see Appendix 1) ignoring the other (more irregular) classes.

however, and our etymologically-matched dataset in particular, are a best-case scenario. We would like to know if these metrics can be leveraged to inform language prehistory in "noisier" more realistic cases too, for example, by providing an additional source of evidence for determining the relatedness of inflection classes, tenses, lexemes, etc. in different languages, and maybe even to uncover phylogenetic relatedness between languages in families which have inflection classes and are deeper and/or more imperfectly understood than Romance.

The Oto-Pamean branch of Otomanguean provides a good testing ground for the feasibility of this approach, as it is characterized by numerous inflection classes, a considerable time depth (much more than Romance), and little extant research on the family's prehistory (see Kaufman & Justeson 2009). Different phylogenetic structures have been proposed for the family (see Figure 11), with the relevant languages here being Chichimec (chic1272) and (Central) Pame (cent2145). While some sources (Campbell 2017) argue that two languages form a clade, some other sources (e.g. Glottolog and Ethnologue) do not find enough evidence for the closer affinity of Chichimec and Pame. This is largely irrelevant to the present paper. In either case, the two languages are unquestionably related but have been diverging much longer than Romance. Estimates vary widely, with the split between Chichimec and Central Pame dated to around 6000BP by Manrique (2000:84-85), 5500 BP by Josserand et al. (1984), and 3600 BP by Kaufmann (2006:819).
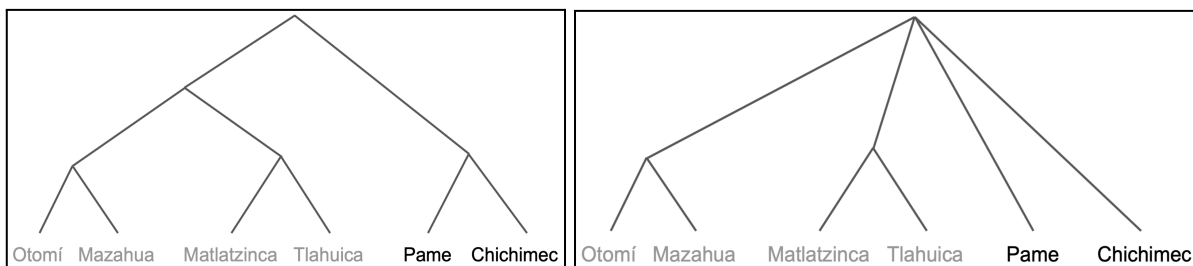
Figure 11: The Oto-Pamean subfamily of Otomanguean (Campbell 2017 vs Glottolog)

More specifically, we will focus here on the languages' verbal inflectional systems as described in Angulo (1933) and Olson (1955) respectively. The overall structure of verbal inflection in the two languages runs uncannily parallel, with stem-initial consonant alternations (particularly in the 3PL), stress-tonal changes, object person-number suffixes following the verb stem, and most importantly for the current purposes, with cumulative prefixes distinguishing person-number (1INC, 1EXCL, 2, 3; SG, DU, PL) and 5-7 TAM values, and showing very robust inflection class distinctions between different verbs.

These prefixes are particularly interesting for the purposes of this paper because they have changed so much (probably due to both analogy, sound changes, and secondary prefixations) that very little transparently cognate morphology has been preserved. In the most comprehensive source available to date on the reconstruction of Proto-Oto-Pamean (Bartholomew 1965), verbal inflection classes were not dealt with because this was considered too daunting a task.[16]

---

[16] Bartholomew (1965:145) explains that while "[t]he reconstruction of these prefixes (...) [would] constitute an interesting study in sorting out the effects of regular sound change, morphophonemic rules affecting the coalescences of compound prefixes, and analogic reshaping[, s]uch a study would be too extensive to incorporate in [her] reconstruction."

C. Pame	Real Perf			Unreal Perf.			Real Progr.			Unreal Progr.			Potential		
obey	no	ni	ndo	nto	mni	mdo	to	to	lo	ndo	ŋgi	nda	lo	ki	la
prepare	ni	ni	ni	mni	mni	mni	ti	ti	li	nta	nta	nda	ta	ta	la
eat	ta	ki	ko	nta	ŋgi	mba	la	ki	Ø	mba	ŋko	mba	Ø	ko	Ø
hate	no	na	na	nto	mna	mna	to	la	wa	ndo	nda	nda	lo	la	la
Chichimec	Present			Anterior Past			Potential			Immed. Past			Future		
see	e	ki	e	tu	ki	u	nu	mi	mu	u	i	zu	ga	ki	ga
bet	tu	ka	u	tu	ka	u	mu	ma	ma	u	e	e	ga	ka	ga
enter	ti	si	i	ta	sa	ta	na	za	na	i	i	i	ta	sa	ta
dream	e	ki	e	ta	ki	u	ma	mi	ma	ka	ki	ku	ta	ki	ga

Table 4: Prefixes in partial paradigms of some Chichimec and Central Pame verbs[17]

Consider the small sample of prefixes from the two languages given in Table 4. It shows the 1SG, 2SG, and 3SG person prefixes for different tenses in verbs from different inflection classes in Central Pame (above) and Chichimec (below). Despite the aforementioned structural parallels, there are remarkably few morphological similarities between the two languages' prefixes that would reveal which Central Pame tense is related to which Chichimec tense, and which Central Pame IC is related to which Chichimec IC. Even when the label given by the authors to a tense is the same (e.g. Potential) it is unclear whether the tenses are cognate, as the forms of the prefixes in the two languages (e.g. *lo ki la, ta ta la* vs *nu mi mu, mu ma ma*) bear hardly any resemblance, and forms can easily change functions in language history.

Due to the shortness of forms (often just CV- or V-) and the limited size of the prefixal lexicon, it would be impossible to distinguish cognate forms from accidental resemblances. Furthermore, due to the large number of possible correspondences, matching the tenses by hand (i.e. evaluating the plausibility of every single combination) would be quite a time-consuming and error-prone task. Only with the 5 tenses in Table 4 (and there are more of them), there would be 120 possible pairings. Still more acute is the problem to identify IC homology, as even more classes can be identified in each language. Because of these difficulties, (Oto-)Pamean can provide an ideal test-case for whether the IC metrics that this paper has dealt with might be harnessed for diachronic research, to guide the linguist, for example, in assessing the merit of potential tense and IC homology relations. A quantitative similarity analysis of Pamean tenses and ICs, parallel to the one of Romance in Section 3, follows in the remainder of this section.

---

[17] The ordering of the tenses and verbs in Table 4 is not meaningful (i.e. is not indicative of possible homology relations). In general, the order and labels of tenses of the original sources (Olson 1955 and Angulo 1933) has been preserved. Three more tenses were provided for Chichimec (called 'Recent', 'Sequential' and 'Negative'), which have not been included in Table 4 for reasons of space and because they appear to have the weakest correspondences to the Central Pame tenses in terms of their labels and forms (and also in terms of their quantitative profiles, as will be shown in Figure 10).
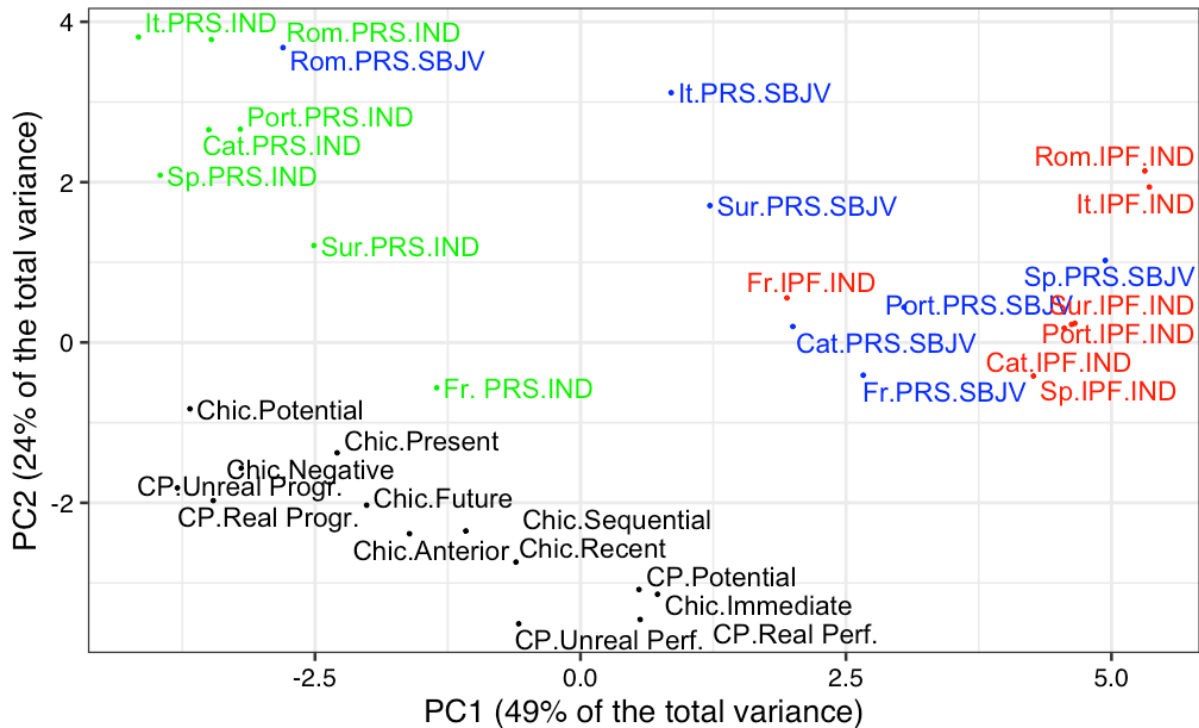
Figure 12: Quantitative complexity affinities of Pamean (vs Romance) tenses

	CP.Real Progr.	CP.Real Perf.	CP.Unreal Perf.	CP.Potential	CP.Unreal Progr.
Chic.Present	**4.862014**	6.3783767	7.3339023	6.5615021	4.9653213
Chic.Anterior	5.7837383	4.8432383	5.8881843	4.9241071	4.897575
Chic.Potential	4.442861	8.1687678	7.7376053	8.2169728	*4.052366*
Chic.Sequential	5.0379267	3.8019968	3.7621928	4.1018044	5.0485345
Chic.Future	5.0616813	5.2064876	6.2225212	*5.206488*	4.3254614
Chic.Immediate	7.6407287	**1.993344**	4.6781422	2.1824971	7.8244046
Chic.Recent	7.6971155	4.0111514	6.6395191	3.7040134	6.9100819
Chic.Negative	6.5586136	7.5136975	7.3796565	7.2461219	5.2475831

Table 5: Euclidean distances, across all variables, between the Chichimec and Central Pame tenses (low values in white, high values in gray)

Figure 12 is a PCA based on the IC complexity metrics of the individual Chichimec and Central Pame (and Romance) tenses as separate self-standing subsystems (see Figures 8 and 9 before). The first thing to note, along with the similarity of homologous Romance tenses already highlighted in Section 3, is that Pamean tenses are quantitatively different from Romance ones, which suggests again that the potential of these metrics to reflect shared ancestry is high.

Focusing on the similarities between Pame tenses exclusively, Table 5 shows the Euclidean distances (based on the metrics in Figure 8 after normalization) between all tenses in the two languages. Across all metrics, the Chichimec Immediate Past tense and the Central

Pame Real Perfective are the absolute closest ones.[18] Quantitatively most similar to the Chichimec Present is the Central Pame tense Real Progressive. Subsequent qualitative exploration (see Section 4.2) has shown that these correspondences (in bold) seem to correspond to *bona fine* homologies. Curiously enough, the absolute farthest tenses in Table 5, are the ones labelled 'Potential' in both languages. A qualitative inspection of the prefixes and contrasts in the two tenses also suggests that these two tenses are, despite their identical label (and hence possibly similar semantics) almost certainly not homologous.

All these "good" pieces of information notwithstanding, it must be mentioned that some tense similarities (such as between Chichimec Recent and Central Pame potential, and Chichimec Sequential and Central Pame Unreal Perfective) do not seem to correspond to homologies under qualitative analysis. After extensive qualitative comparison, tense-homology relations could not be determined with certainty beyond the two aforementioned ones in bold in Table 5. Somewhat more uncertain but plausible homologies are indicated in bold italics, one of which (Chichimec Potential vs Central Pame Unreal Progressive) is also signalled by the lowest Euclidean distance between the tenses.

More (comparative) research would certainly be needed to assess these less certain homologies between the tenses in the two languages. Also more research would be needed into which specific paradigmatic configurations favour or hamper the quantitative detection of homology. Here, it appears that, while different Romance tenses tend to be characterized by quite different predictive complexity profiles (see Figures 8 and 12), Pame tenses are more homogeneous in this regard. As these Figures show, the more frequent Romance tenses like present indicative (also forms like 2SG imperative and participle) are characterized by a high overall degree of complexity (larger number of allomorphic distinctions, high unpredictability,[19] etc.), while the least frequent tenses like the imperfect indicative often lack unpredictable suffixal allomorphy altogether. In Pamean, by contrast, robust IC differences are present in the morphology of all tenses. Furthermore (as the following qualitative Section 4.2 will show), the distinctions and (un)predictability relations in different tenses are isomorphic (see Table 8) to a large extent. This difference means that, as Figure 12 shows, the overall information-theoretic profile of different Pame tenses is more similar than that of different Romance tenses. This reduced variance may be hampering the quantitative detection of tense homology in Pamean to some extent.

Alongside the detection of tense homology, the detection of IC homology is the other remaining relationship that presents a great challenge in Pamean. In the same way as with Romance before, IC complexity metrics can provide valuable information to diagnose IC homology even when most of the transparent cognate morphology has disappeared.

---

[18] Excluded have been, for obvious reasons, all pairs of tenses belonging to the same language, as these could not ever be homologous.

[19] The literature on so-called 'stem spaces' (see e.g. Boyé and Cabredo-Hofherr 2006, Montermini & Bonami 2013) identifies 3-5 different ones in the present indicative but a more reduced number of them (1-2) in the least frequent tenses. This configuration (inherited from Latin to a great extent, see Pellegrini 2020) must be helping to keep the tenses more information-theoretically distinct in Romance.

	A Chich.see	B Chich.loaded	C Chich.ascend	D Chich.last	E Chich.hit3	F Chich.remember
I C.Pame.belittle	**0.6446378**	1.0166573	0.655437	0.9883873	0.7966584	0.9017169
IV C.Pame.obey	0.8042586	1.3599883	0.7228474	**0.2469974**	1.004805	0.353474
VI C.Pame.lose	0.9781648	1.2961693	0.6022016	0.2609839	0.865209	**0.1861345**
II C.Pame.appease	1.1060461	*1.1106062*	1.0848559	0.8460317	1.1861187	0.8573917
C.Pame.prepare	1.1435491	0.7229636	0.849756	1.0133473	0.8244318	0.9106
C.Pame.lose2	1.1208732	0.9920062	1.0219992	0.8718452	1.0801275	0.8388239
III C.Pame.eat	1.9276541	0.7047445	*1.5600909*	1.9916998	1.1185212	1.7963349
V C.Pame.allow	1.1265325	0.9933917	0.4466877	0.7447684	**0.383442**	0.5052044

Table 6: Euclidean distances, across all variables, between the different Chichimec and Central Pame ICs (low values in white, high values in gray)[20]

Table 6 shows the Euclidean distances between the metrics of all Chichimec and Central Pame ICs.[21] Most of the resulting pairings, particularly the closest ones like the Chichimec IC of 'last' and the Central Pame IC of 'obey', that of the Chichimec IC of 'hit3' and of Central Pame 'allow', that of the Chichimec IC of 'see' and of the Central Pame class of 'belittle', and that of Chichimec 'remember' and Central Pame 'lose' (in bold) appear to constitute *bona fine* homology relations, as will be shown in the coming qualitative Section 4.2. Although the homology of the latter two classes could not be confirmed via cognate lexemes, it looks very plausible from the qualitative inspection of the prefixal (and stem) morphology of the classes.

Conversely, and despite these successes of the quantitative approach pursued here, a few homology relations that appear to be relatively certain from qualitative inspection (namely Chichimec IC of 'loaded' = Central Pame IC of 'appease', and Chichimec IC of 'ascend' = Central Pame IC of 'eat') were not flagged by particularly close values of the IC metrics in the two languages. It is nonetheless striking that, class by class (i.e. column by column in Table 6), the lowest Euclidean distance value corresponded to a *bona fide* homology relation in 4 out of 6 cases, which is vastly above the number of matches expected by chance.

Overall, these results are encouraging. In slightly over half of the cases, the complexity metrics succeeded in recovering homologies. This is all the more remarkable as the languages have evolved independently over about four millenia and there is a very large number of logically possible pairings. We next show that metrics outperfom a quantitatively uninformed, traditional "manual" comparison.

4.2 Qualitative matching of Pame homologous tenses and ICs

The quantitative findings and the evaluation of their success in the previous section is obviously in need of an independent source of verification. To assess, as we have done throughout Section 4.1, whether the quantitative tense and IC comparison yielded correct or

---

[20] Only non-singleton ICs (i.e. those with more than a single member) were included in the analysis. There were 8 such classes in Central Pame and 6 in Chichimec, each of them with 4+ members.

[21] As indicated before, for ICs these are a subset from the metrics in Section 2 (namely n-MPS entropy, dynamic principal parts and their density, cell and IC predictability, and predictive entropy, as well as number of members) because not all metrics can be calculated over individual inflection classes.

incorrect pairings, a qualitative comparative approach was pursued separately to identify related tense and ICs between the two languages in a more traditional way.

The large number of lexemes (100+) and paradigm cells (40+), the large number of possible IC and tense pairings (500+), and the seemingly complete absence of transparent prefixal equivalents between the two languages (Table 4) make it impossible to discover shared ancestry by just eye-balling the paradigms in an uninformed fashion. Because of this, the quest to identify homologous tenses and ICs started from the identification of cognate lexemes in our two datasets: Angulo (1933) [and derived sources like Palancar & Avelino 2019], and Olson (1955).

As a word class, verbs in Pame constitute a smaller, somewhat less productive one than in most Indo-European languages. In addition, one can expect more basic and frequent verbs to be more likely to be documented than more infrequent or less basic ones. These factors increase the chances to find a certain number of cognate verbs in the independently-compiled datasets currently available to us for each language. Comparing the ICs of cognate lexemes maximizes our chances to spot IC homology relations. Even if some lexemes had changed their IC, as in Romance, we should still be able to find enough overlap to allow us to recover IC homology relations provided a sufficient number of cognate lexemes are found in the two datasets.

The search of possible cognates proceeded by identifying as such those verbs in our datasets that had very similar or identical meaning and stems[22] in the two languages. It yielded the 16 likely cognates in Table 7, whose respective ICs (reported with letters and Roman numerals in the table) could then be compared with each other.

Chichimec			Central Pame		
IC	Gloss	Stem	IC	Gloss	Stem
A	stand	me	VI	stand	maʔi
A	put2	hu	I	put down	hũcʔ
A	scold	co	I	scold	cu
A	touch	ta	I	touch	taho
A	receive	tec	I	receive	tahič
A	ask	ʔan	I	ask	ʔahodn
A	hear	ʔo	I	hear	ʔuʔ
A	give	ʔe	I	give	ʔiugŋ
A	bury	ʔa	I	bury2	ʔailʔ
B	laugh	ter	I	laugh	tæhædnʔ
B	married	ter	II	get married	tẽhẽʔt
C	fall	co	III	fall	cuʔ
C	die	rų	III	die	ttõ
D	greet	ga	IV	greet	ngãõʔ
D	do	ʔę	IV	do so	ʔẽĭ
E	cry	gwe	V	cry	wai

Table 7: Pame cognates and IC equivalences (Chichimec classes indicated by letters and Central Pame ones by Roman numerals)

---

[22] Stem morphology, unlike prefixal morphology, appears to be comparatively stable in the family.

As the color-coding in Table 7 illustrates, in our set of cognate verbs, IC distinctions in the two languages map almost perfectly to each other, which suggests that, to our advantage, IC membership must be remarkably stable in the family, impressionistically much more so than in Romance.[23] A majority of the cognate lexemes analyzed (9 in each language, largely the same set, in yellow) belong to the same inflection class, which is also the largest one in the two languages by number of members. The two classes (i.e. the IC 'A' of Chichimec [exemplar 'see' in Figure 13] and 'I' of Central Pame [exemplar 'belittle']) can thus be classified as homologous. This is a relation, I remind, which was successfully spotted by the quantitative approach pursued in this paper (Table 6). Also quantitatively preserved was the special affinity of Chichimec class D (exemplar 'last') and Central Pame IC IV (exemplar 'obey'), and that of Chichimec IC E (exemplar 'hit3') and Central Pame V (exemplar 'allow').

Not associated with high quantitative similarity were the homology of Chichimec classes B and Central Pame II, and that of Chichimec class C and Central Pame III. The homology of those ICs that happen not to be represented in our set of cognates in Table 7 (the Chichimec class of 'remember', Central Pame 'lose', Central Pame 'prepare' and Central Pame 'lose2') needs to be assessed in an alternative way. An exploration of their prefix forms and morphological contrasts suggests that Chichimec 'remember' and Central Pame 'lose' are almost certainly homologous classes, something anticipated as well by the classes' quantitative similarity (the highest one across all possible pairs in Table 6).

Identifying tense homology between the two languages is the other main challenge. Comparing the forms that we now know belong to homologous ICs, we should be in a better position to identify qualitatively tense homology as well. Unlike in the quantitative approach leading to Table 5, a qualitative approach can make use of both the morphological prefix similarities between different tenses (or of systematic differences, like in the Comparative Method), as well as of the semantic information provided by the tense labels in the two languages, and their usage descriptions in the extant sources. Despite this, tense homology has turned out to be very difficult to identify in such a way. Only the status of Chichimec Present and Central Pame Real Progressive (more-or-less transparently pointed at by the tense labels) isimmediately apparent from inspection of the cognate lexemes. For the rest of the tenses, the forms from different inflection classes appear to often contradict each other and, given the number of possible pairings involved, it is difficult to find an optimal arrangement qualitatively that would maximize the similarity of the prefixes or the regularity of their correspondences across all ICs. Tentative pairings were proposed initially (Chichimec Potential = Central Pame Unreal Perfective, Chichimec Sequential = Central Pame Potential, Chichimec Future = Central Pame Unreal Progressive, and Chichimec Anterior = Central Pame Real Perfective) which coincide with some of the ones proposed independently by Bartholomew (1965:299) but appeared to be incorrect upon more detailed inspection and comparison with the quantitative similarity results. Further exploration of these (Table 5) reveals that some of those pairings (specifically Chichimec Immediate Past = Central Pame Real Perfective, and Chichimec Future = Central Pame Potential) have considerably more merit than the tense equivalences that had been proposed initially in a quantitatively uninformed way.

---

[23] It would be interesting to explore in future research why this might be so. One reason might be that, whereas inflection class distinctions in Romance are hardly correlated to any extramorphological property, they often correspond to different types of verbs (e.g. transitive vs intransitive) in Pamean.

Table 8 shows the most likely[24] tense and IC homologies between Chichimec (left) and Central Pame (right). The Chichimec Present tense (in shades of yellow, left), for example, is found to descend from the same Proto-tense as the Central Pame Real Progressive (in shades of orange, right). The Chichimec conjugation A, in turn, seems homologous with the Central Pame Conjugation I. While the first two tenses' homologous status is clear in view of their morphological (e.g. ki, ti, to~tu, ko~ku) and structural similarities, the last two tense pairings remain more speculative. That the two clearest tense-to-tense homology relations could be uncovered quantitatively is remarkable and suggests, again, that the quantitative method delivers useful results. Overall, as explained in Tables 5 and 6, 60% of the shared ancestry relations in Table 8 (2 out of 4 tense homologies and 4 out of 6 IC homologies) could be successfully diagnosed through their complexity metrics.

| | | Chichimec Present | | | | | | Central Pame Real Progressive | | | | | | | |
		1SG	1DU	1PL	2SG	3SG	3PL	1SG	1DU	1PL	2SG	3SG	3PL		
A	touch	e	e	e	ki	e	e	la	ta	ta	ki	wa	∅	hear	I
C	die	e	e	u	ki	e	e	la	ta	la	ki	∅	∅	fall	III
D	greet	tu	tu	tu	su	u	u	to	to	to	to	lo	wa	greet	IV
B	married	ti	ti	ti	si	i	i	ti	ti	ti	ti	li	ti	get married	II
F	remember	tu	tu	tu	su	u	e	to	to	to	to	lo	wo	lose	VI
E	hit3	tu	tu	tu	ka	u	u	to	to	to	la	wa	wa	allow	V
		Chichimec Immediate Past						Central Pame Real Perfective							
A	touch	u	u	u	i	zu	zu	no	no	no	ni	ndo	ndo	hear	I
C	die	ka	ki	ki	ki	ku	ku	ta	ti	i	ki	ko	ko	fall	III
D	greet	u	u	u	i	zu	zu	no	no	no	ni	ndo	ndo	greet	IV
B	married	i	i	i	i	i	i	ni	ni	ni	ni	ni	ni	get.married	II
F	remember	u	u	u	i	zu	zu	no	no	no	ni	ndo	ndo	lose	VI
E	hit3	u	u	u	e	e	e	no	no	no	na	na	na	allow	V
		Chichimec Future						Central Pame Potential							
A	touch	ga	ga	ga	ki	ga	ga	la	la	la	ki	la	la	hear	I
C	die	ta	ti	gu	ki	ga	ga	∅	∅	∅	ko	∅	∅	fall	III
D	greet	gu	gu	gu	ki	ga	ga	lo	lo	lo	ki	la	la	greet	IV
B	married	ta	ti	ti	sa	ta	ta	ta	ti	∅	ta	ta	ta	get.married	II
F	remember	gu	gu	gu	ki	ga	ga	lo	lo	lo	ki	la	la	lose	VI
E	hit3	ga	ga	ga	ka	ga	ga	lo	lo	lo	la	la	la	allow	V
		Chichimec Potential						Central Pame Unreal Perfective							
A	touch	nu	nu	nu	mi	mu	mi	nda	nda	nda	ngi	nda	nda	hear	I
C	die	ma	ma	mu	mi	ma	mi	mba	mba	mba	nko	mba	mba	fall	III
D	greet	nu	nu	nu	mi	mu	mi	ndo	ndo	ndo	ngi	nda	nda	greet	IV
B	married	na	ni	ni	za	na	na	nta	nti	nta	nta	nda	nta	get.married	II
F	remember	nu	nu	nu	mi	mu	mi	ndo	ndo	ndo	ngi	nda	nda	lose	VI
E	hit3	mu	mu	mu	ma	ma	ma	ndo	ndo	ndo	nda	nda	nda	allow	V

Table 8: Prefixal forms of Pamean homologous tenses and ICs. Different shades indicate morphological contrasts or syncretisms parallel in the two languages.[25]

---

[24] After the deployment of the quantitative and qualitative approaches together, the homology of other tenses remains speculative at best. Of course the possibility cannot be ruled out that some tenses do not have homologs in the other language, and this might be the case of the ones not found in Table 8.

[25] The comparison of stem alternation patterns in cognate verbs provides additional support for the tense pairings in Table 8. Chichimec present and Central Pame progressive resemble each other in showing no stem-onset alternations (other than glottalization in the 3PL, which appears in all tenses).

After a successful qualitative-cum-quantitative exploration of the homology relations, and even after arranging the tabular presentation of forms into homologous ICs and tenses in the two languages, it is remarkable just how much the forms of the prefixes in Table 8 differ from one language to the other. Crucially, the morphological trait that appears to be most resilient is not the form of prefixes per se, but rather the morphological oppositions between person-number cells, tenses and ICs. The morphological affinity (i.e. shared allomorphs) of Classes A/I and C/III in the present/real progresive tense, for example, has been preserved, as well as the morphological contrast of these two classes with all the others. The allomorphic uniqueness of ClassB/II is also shared by the two languages. Note that IC affinities are different in other tenses. For example, in the Immediate Past/Real Perfective, Class A/I resembles classes D/IV and F/VI instead, while class C/III is allomorphically unique in this tense. These traits seem to have been preserved despite considerable morphological change that has occurred separately in the two languages.

Within a single tense, person-number oppositions and syncretisms tend to be preserved as well. Observe, for example, that in the Immediate Past/Real Perfective, both languages are characterized by complete underspecification of person-number in the IC B/II, by a 1 vs 2/3 partition in IC E/V, and by a 1 vs 2 vs 3 partition in ICs A/I, D/IV, and F/VI. This structural continuity is precisely the reason why quantitative approaches to the predictive structure of paradigms can succeed where the comparative method cannot, and why it holds great potential to contribute to research on diachronic morphological change and proto-language reconstruction. Paradigmatic-structural properties seem to be sometimes more stable than the actual morphological forms that instantiate them (see also Herce 2021). Because of this, the quantitatively-driven detection of shared ancestry like the one in this paper can sometimes outperform qualitative assessments by the human linguist. A combination of both approaches is likely to be, of course, the optimal one.

5 Conclusion

This paper has assessed the value of quantitative paradigmatic complexity metrics for exploring the historical relatedness of different inflectional (sub)systems. After introducing these metrics and notions in Section 2, Section 3 presented an analysis of Romance verbal inflectional paradigms compared to unrelated inflection class systems from around the world. Our results show that, even after two millennia of (largely) independent evolution, Romance conjugations have managed to preserve a level of synchronic variation that, for most metrics, is significantly lower than among phylogenetically unrelated IC systems. Although metrics differ in the extent to which they preserve a signal for shared ancestry, all quantitative aspects of an IC system can in principle be inherited to a larger or smaller extent. Therefore, they could prove very useful in less well understood or more diversified language families, e.g. to inform shared ancestry as we do in this paper, but in the long term

---

Meanwhile, Chichimec immediate past and Central Pame real perfective, and also Chichimec Potential and Central Pame Unreal Perfective, both show stem changes (e.g. -h- > -nh-) in the second person, while Chichimec Future and Central Pame Potential, show the same stem change in both 2 and 3SG. All remaining tenses have one of these last two patterns, so these parallels are not incontrovertible evidence for homology, but do constitute an additional clue supporting the tense equivalences in Table 8.

also to reconstruct phylogenetic trees or proto-paradigms. We showed that not only whole systems, but also subsystems (tenses and ICs) which are homologous tend to show greater quantitative similarity than unrelated subsystems. This fact could be used diagnostically in less well-researched language families.

The second part of this paper (Section 4) constitutes a proof of concept of precisely this possibility. The Pamean languages Chichimec and Central Pame constitute a particularly challenging case by virtue of the morphological complexity of their paradigms (large number of cells, tenses, and ICs) and the diachronic distance and vast morphological differences between the two languages. In Section 4.1 we focused on identifying tense and IC homology relations between the two languages by comparing the similarity between their predictive complexity metrics (see Tables 5 and 6). A subsequent qualitative comparison of the inflectional systems (in Section 4.2) showed that the quantitative metrics often pointed to shared ancestry relations that appear highly plausible under qualitative analysis, sometimes even detecting relatedness that had been missed by previous qualitative work. Although a few relations appear to have been missed and others turned out to be spurious, the overall success rate of the quantitative approach (around 60%) was still hugely above chance levels considering the number of logically possible relations in the system. We propose that paradigm complexity metrics could be useful to help narrow the search space during or prior to more time-consuming qualitative analysis.

Our findings, and the 'diachronization' of quantitative research on paradigm structure more generally, raise many further questions. For example: does the relative stability ranking of the different metrics that we reached for Romance (Figure 7) apply to other families and IC systems? A quantitative exploration of nominal declension in Balto-Slavic languages might be an ideal ground for replicating the present findings. Can information-theoretic metrics contribute to (some) debates on the genetic (un)relatedness of different languages? The languages in Australia's Top End, for example, all feature TAM-based inflectional classes but there is no consensus regarding which of these languages (aka [Macro-]Gunwinyguan) are phylogenetically (un)related. Can we identify general change trends in morphology, or evolutions typical of different sociolinguistic situations? Language contact, and a large numbers of adult L2 learners, for example, have been argued to cause morphological simplification (Kusters 2003, Trudgill 2011). More research is needed, however, to explore whether this should be expected across all aspects of an inflectional system (i.e. in all the metrics analyzed here) or whether it would be limited to only some of them (Widmer et al. 2021). Bantu nominal inflection classes might constitute a good dataset to explore this (Verkerk & di Garbo 2022). Are there correlations between different, logically independent quantitative aspects of IC systems that would point toward cross-linguistic cognitive biases? Can we model and to some extent predict the direction in which a given inflection class system is likely to develop (at the micro- [e.g. Herce 2020] or the macro-levels [e.g. Carstairs-McCarthy 2010]) based on its synchronic properties? These and other questions at the intersection between paradigm structure, evolutionary morphology, and linguistic cognition will be left for future research.

Appendix 1 Explored verbal cognates (infinitives) in Romance languages

Cognate	Portuguese	Spanish	Catalan	French	Surselvan	Italian	Romanian
habere	ɐˈver	aˈber	əˈβɛ	avwaʁ	aˈve	aˈvere	aˈvea
facere	feˈzer	aˈθer	ˈfe	fɛʁ	ˈfar	ˈfare	ˈfatʃe
posse	puˈder	poˈder	puˈɣɛ	puvwaʁ	puˈder	poˈtere	puˈtea
dicere	diˈzer	deˈθir	ˈdi	diʁ	ˈdir	ˈdire	ˈzitʃe
uenire	ˈvir	beˈnir	bəˈni	vəniʁ	veˈɲir	veˈnire	veˈni
sapere	seˈber	saˈber	səpiˈɣe	savwaʁ	saˈver	saˈpere	
videre	ˈver	ˈber	ˈbɛwrə	vwaʁ	ˈvezər	veˈdere	veˈdea
tenere	ˈter	teˈner	təˈni	təniʁ		teˈnere	ˈtsine
uelle			buˈlɛ	vulwaʁ	vuˈler	voˈlere	ˈvrea
stare	iʃˈtar	esˈtar	əsˈta		ˈʃtar	ˈstare	ˈsta
dare	ˈdar	ˈdar			ˈdar	ˈdare	ˈda
bibere	biˈber	be'ber	ˈbɛwrə	bwaʁ	ˈbeibər	ˈbere	ˈbea
cantare	kẽˈtar	kanˈtar	kənˈta	ʃɑ̃te	kanˈta	kanˈtare	kinˈta
mori	muˈʁer	moˈrir	muˈrir	muʁiʁ	muˈri	moˈrire	muˈri
dormire	durˈmir	dorˈmir	durˈmi	dɔrmiʁ	durˈmi	dorˈmire	dorˈmi
debere	diˈver	deˈber	ˈdɛwrə	dəvwaʁ	duˈer	doˈvere	
sentire	sẽˈtir	senˈtir	sənˈti	sɑ̃tiʁ	senˈtir	senˈtire	simˈtsi
coquere	kuˈzer	koˈθer	ˈkɔwrə	kɥiʁ	ˈkuər	ˈkwɔtʃere	ˈkoatʃe
credere	ˈkrer	kreˈer	ˈkrɛwrə	kʁwaʁ	ˈkrer	ˈkredere	ˈkrede
uendere	vẽˈder	benˈder		vɑ̃dʁ	ˈvendər	ˈvendere	ˈvinde
finire			fiˈni	finiʁ	fiˈnir	fiˈnire	fi'ni
currere	kuˈʁer	koˈrer	ˈkorə	kuʁiʁ	ˈkuorər	ˈkorrere	ˈkurdʒe
ualere	veˈler	baˈler	bəˈlɛ	valwaʁ		vaˈlere	
placere				plɛʁ		pjaˈtʃere	pləˈtʃea
scribere	iʃkriˈver	eskriˈbir	əsˈkriwrə	ɛkʁiʁ	ˈʃkrivər	ˈskrivere	ˈskrie
trahere	treˈzer	traˈer	ˈtrɛwrə	tʁɛʁ	ˈtrer	ˈtrarre	ˈtradʒe
cognoscere	kuɲiˈser	konoˈθer	kuˈnɛʃə	kɔnɛtʁ		koˈnoʃʃere	kuˈnoaʃte
legere	ˈler	leˈer	ʎəˈʒi	liʁ		ˈlɛddʒere	
perdere	pirˈder	perˈder	ˈpɛrdrə	pɛʁdʁ	ˈpɛrdər	ˈpɛrdere	ˈpjerde
mittere	miˈter	meˈter		mɛtʁ	ˈmɛtər	ˈmettere	
cadere	keˈir	kaˈer		ʃwaʁ		kaˈdere	kəˈdea
colligere	kuˈʎer	koˈxer	kuˈʎi	kœjiʁ		ˈkɔʎʎere	kuˈledʒe
fugere	fuˈʒir	uˈir	fuˈʒi	fɥiʁ	fuˈʝir	fudˈdʒire	fuˈdʒi
ridere	ˈʁir	reˈir	ˈriwrə	ʁiʁ	ˈrir	ˈridere	ˈride
ponere	ˈpor	poˈner	ˈpondrə	põdʁ		ˈporre	ˈpune
consuere	kuˈzer	koˈser	kuˈzi	kudʁ	ˈkweʒər	kuˈtʃire	ˈkoase
portare	purˈtar	porˈtar		pɔʁte	purˈta	porˈtare	purˈta
uiuere	viˈver	biˈbir	ˈbiwrə	vivʁ		ˈvivere	viˈa
nasci	neʃˈser	naˈθer	ˈneʃə	nɛtʁ		ˈnaʃʃere	ˈnaʃte
audire	owˈvir	oˈir	uˈi	wiʁ		uˈdire	auˈzi
crescere	kriʃˈser	kreˈθer		kʁwatʁ	ˈkreʃər	ˈkreʃʃere	ˈkreʃte
sequi	siˈgir	seˈgir		sɥivʁ		seˈgwire	
tacere				tɛʁ		taˈtʃere	təˈtʃa
capere	keˈber	kaˈber	ˈkaβrə			kaˈpire	inkəˈpea
parere	peˈir			paʁɛtʁ		paˈrere	pəˈrea
prehendere	prẽˈder	prenˈder	ˈprɛndrə	pʁɑ̃dʁ	ˈprendər	ˈprɛndere	ˈprinde
lauare	leˈvar	la'bar		lave		laˈvare	ˈla
leuare	liˈvar	jeˈbar		ləve		leˈvare	luˈa
quaerere	kiˈrer	keˈrer				ˈkjɛdere	ˈtʃere
recipere	ʁisiˈber	reθiˈbir		ʁəsəvwaʁ	rəˈtʃeivər	riˈtʃevere	

Latin	1	2	3	4	5	6	7
conducere	kõduˈzir	konduˈθir		kõdɥiʀ		konˈdurre	konˈdutʃe
plicare	tʃɟˈgar	jeˈgar		plwaje		pjeˈgare	pleˈka
rumpere	ʀõˈper	romˈper		ʀõpʀ	ˈrumpər	ˈrompere	ˈrupe
aperire	eˈbrir	aˈbrir	uˈβri	uvʀiʀ	ˈarvər	aˈprire	
molere	muˈer	moˈler	ˈmɔldrə	mudʀ	ˈmolər		
ferire	fɨˈrir	eˈrir			ˈfiərər	feˈrire	feˈri
intendere	intẽˈder	entenˈder	ənˈtendrə	ãtãdʀ		inˈtɛndere	inˈtinde
uestire	viʃˈtir	besˈtir		vɛtir		veˈstire	
tussire	tuˈsir	toˈser	tuˈsi	tuse		tosˈsire	tuˈʃi
sugere	suˈgar			syse		sukˈkjare	ˈsudʒe
stringere		estreˈɲir		ɛtʀɛ̃dʀ		ˈstrindʒere	ˈstrindʒe
spargere		esparˈθir				ˈspardʒere	ˈspardʒe
rogare	ʀuˈgar	roˈgar				roˈgare	ruˈga
remanere			ruˈmandrə			rimaˈnere	rəˈmine
manducare	mẽˈʒar			mãʒe		manˈdʒare	mənˈka
incendere		enθenˈder	ənˈsendrə			intʃenˈdjare	ˈintʃindʒe
implere	ẽˈʃer		umˈpli	ãpliʀ		emˈpire	ˈumple
iacere	ʒeˈzer	jaˈθer	ˈdʒɛwrə	ʒɛziʀ	ˈʒer	jaˈtʃere	zəˈtʃa
crepare	kɨˈbrar	keˈbrar		kʀɛve		kreˈpare	krəˈpa
comparare	kõpeˈrar	komˈprar		kõpaʀe			kumpəˈra
timere	tɨˈmer	teˈmer	ˈtemə		teˈmer	teˈmere	ˈteme
sudare	suˈar	suˈdar	suˈa	sɥe	suˈa	suˈdare	asuˈda
serrare		seˈrar		sɛʀe	seˈra	serˈrare	
separare	sipeˈrar	sepaˈrar	səpəˈra	sɛvʀe	tsaˈvra	sepaˈrare	sepaˈra
mulgere		meˈθer	muˈɲi		ˈmulʒər	ˈmundʒere	ˈmuldʒe
monstrare	musˈtrar	mosˈtrar		mõtʀe	muˈsa	moˈstrare	musˈtra
laborare	leˈvrar	laˈbrar		labuʀe	luˈvra	lavoˈrare	
laudare	lowˈvar	loˈar		lwe	luˈda	loˈdare	ləuˈda
gaudere	guˈzar	goˈθar		ʒwiʀ	guˈde	goˈdere	
dirigere	diriˈʒir	diriˈxir		diʀiʒe	ˈdɛrʒər	diˈridʒere	ˈdrege
demandare	dɨmẽˈdar			dəmãde	dumanˈda	domanˈdare	
movere	muˈver	moˈber	ˈmɔwrə	muvwaʀ		ˈmwovere	
amare	eˈmar	aˈmar		ɛme		aˈmare	
servire	sɨrˈvir	serˈbir		sɛʀviʀ		serˈvire	serˈvi
intrare	ẽˈtrar	enˈtrar		ãtʀe		enˈtrare	inˈtra
exire	ɐrˈder		əˈʃi			uʃˈʃire	ieˈʃi
claudere					ˈklaudər	ˈkjudere	
ardere	ɐrˈder	arˈder			ˈardər		ˈarde
adducere		aduˈθir				adˈdurre	aˈdutʃe
seminare		semˈbrar		səme		semiˈnare	seməˈna
partire	perˈtir	parˈtir	pərˈti	paʀtiʀ		parˈtire	
salire	seˈir	saˈlir		saliʀ		saˈlire	səˈri
plangere			ˈplaɲə	plɛ̃dʀ		ˈpjandʒere	ˈplindʒe
computare	kõˈtar	konˈtar		kõte		konˈtare	kumpəˈta
mordere	murˈder	morˈder		mɔʀdʀ		ˈmordere	
cooperire	kuˈbrir	kuˈbrir		kuvʀiʀ	kuˈviərər	koˈprire	akopeˈri
bullire		buˈjir	buˈʎi	bujiʀ		bolˈlire	
fumare	fuˈmar	fuˈmar		fyme	fiˈma	fuˈmare	fuˈma
pensare	pẽˈsar	penˈsar		pãse		penˈsare	penˈsa
fundere	fũˈdir	funˈdir		fõdʀ		ˈfondere	

Appendix 2   Languages, affiliations, and sources of cross-linguistic IC systems

Language	Stock	Source	Distill-ations	Expon-ents	Inflection classes	Principal parts
Portuguese	**Indo-European**	**Kirov et al. 2018**	**11**	**190**	**25**	**5**
Spanish	**Indo-European**	**Personal knowledge**	**11**	**140**	**24**	**5**
Catalan	**Indo-European**	**Perea & Ueda 2010**	**12**	**137**	**32**	**6**
French	**Indo-European**	**Bonami et al. 2014**	**12**	**77**	**28**	**5**
Surselvan	**Indo-European**	**Beniamine et al. 2020**	**11**	**156**	**25**	**3**
Italian	**Indo-European**	**Pellegrini & Cignarella 2020**	**12**	**192**	**29**	**6**
Romanian	**Indo-European**	**Herce & Pricop 2024**	**12**	**148**	**37**	**6**
Ngiti	Central Sudanic	Stump & Finkel 2015	8	7	10	3
Tulu	Dravidian	Stump & Finkel 2015	7	157	6	2
Fur	Furan	Stump & Finkel 2013	9	50	19	5
Gaagudju	Gaagudju	Harvey 2011	6	70	70	5
Kadiweu	Guaicuruan	Sims & Parker 2016	5	40	58	4
Jawoyn	Gunwinyguan	Merlan n.d.	4	30	20	4
Rembarnga	Gunwinyguan	Saulwick 2003	9	99	30	6
Russian	Indo-European	Stump & Finkel 2013	9	52	79	6
Greek	Indo-European	Stump & Finkel 2013	6	65	48	3
Czech	Indo-European	Stump & Finkel 2013	13	46	26	5
Icelandic	Indo-European	Stump & Finkel 2013	21	680	146	8
Lithuanian	Indo-European	Stump & Finkel 2013	9	141	18	3
Sanskrit	Indo-European	Stump & Finkel 2013	13	261	38	4
Mawng	Iwaidjan	Singer 2006	5	35	29	5
Kwerba	Kwerba	Stump & Finkel 2015	4	10	4	1
Limilngan	Limilngan-Wulna	Harvey 2001	7	54	42	4
Wandarang	Mangarrayi-Maran	Heath 1980a	7	66	19	2
Marra	Mangarrayi-Maran	Heath 1981	9	110	30	4
Jaminjung	Mirndi	Schutze-Berndt 2000	4	55	26	3
Koasati	Muskogean	Stump & Finkel 2014	5	88	12	2
Nuer	Nilotic	Sims & Parker 2016	6	4	25	5
Chichimec	Otomanguean	Feist & Palancar 2015	15	18	11	3
Mephaa	Otomanguean	Feist & Palancar 2015	6	20	60	5
Otomi T	Otomanguean	Feist & Palancar 2015	5	56	3	1
Mixtec X	Otomanguean	Feist & Palancar 2015	3	6	5	3
Mazatec C	Otomanguean	Jamieson 1982	5	356	105	3
Chinantec C	Otomanguean	Stump & Finkel 2013	12	42	67	5
Chinantec P	Otomanguean	Stump & Finkel 2013	11	12	46	6
Ritarungo	Pama-Nyungan	Heath 1980b	8	67	18	3
Seri	Seri	Sims & Parker 2016	4	40	254	4
Nangikurrunggurr	Southern Daly	Reid 1990	4	42	21	3
Arapesh	Torricelli	Hein & Muller 2009	2	41	26	1
Amele	Trans New Guinea	Hein & Muller 2009	3	27	23	3
Voro	Uralic	Sims & Parker 2016	8	28	24	8
Saami P	Uralic	Wilbur 2014	4	70	8	1
Wadjiginy	Wadjiginy	Ford 1990	2	17	14	2

References

Ackerman, Farrell, James P. Blevins & Robert Malouf. 2009. Parts and wholes: Patterns of relatedness in complex morphological systems and why they matter. In James P. Blevins & Juliette Blevins (Eds.), *Analogy in grammar: Form and acquisition:* 54-82.

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

Alpher, Barry, Nicholas Evans, and Mark Harvey. 2003. Proto Gunwinyguan verb suffixes. In *The Non-Pama-Nyungan languages of northern Australia: Comparative studies of the continents most linguistically complex region*. Pacific Linguistics.

De Angulo, Jaime. 1933. The Chichimeco language (Central Mexico). International Journal of American Linguistics 7, no. 3/4: 152-194.

Bank, Sebastian. 2017. Segmentation: a remark on the Syncretism Principle. *Morphology* 27, 1: 1-20.

Bartholomew, Doris. 1965. The reconstruction of Otopamean (México). PhD dissertation, University of Chicago.

Beniamine, Sacha, Martin Maiden & Erich Round. 2020. Opening the Romance Verbal Inflection Dataset 2.0: a CLDF Lexicon. In *Proceedings of The 12th Language Resources and Evaluation Conference*, pp. 3027-3035.

Blevins, James P. 2013. The information-theoretic turn. *Psihologija* 46, 4: 355-375.

Bonami, Olivier; Gauthier Caron & Clément Plancq. 2014. Construction d'un lexique flexionnel phoné- tisé libre du français. In Franck Neveu, Peter Blu- menthal, Linda Hriba, Annette Gerstenberg, Judith Meinschaefer and Sophie Prévost, editors, *Actes du quatrième congrès mondial de linguistique fran- çaise*: 2583–2596.

Bickel, Balthasar, and Johanna Nichols. 2007. Inflectional morphology. In Matthew S. Dryer and Timothy Shopen (Eds.), *Language typology and syntactic description*: 169-240. Cambridge: Cambridge University Press.

Blasi, Damián E., Søren Wichmann, Harald Hammarström, Peter F. Stadler, and Morten H. Christiansen. 2016. Sound–meaning association biases evidenced across thousands of languages. *Proceedings of the National Academy of Sciences* 113, 39: 10818-10823.

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

Boyé, Gilles, and Patricia Cabredo-Hofherr. 2006. The structure of allomorphy in Spanish verbal inflection. *Cuadernos de Lingüística del Instituto Universitario Ortega y Gasset* 13: 9-24.

Bürkner, Paul-Christian. 2017. brms: An R package for Bayesian multilevel models using Stan. *Journal of statistical software* 80, 1: 1-28.

Bürkner, Paul-Christian. 2018. Advanced Bayesian multilevel modeling with the R package brms. The R Journal 10. 395–411.

Campbell, Eric W. 2017. Otomanguean historical linguistics: Exploring the subgroups. *Language and Linguistics Compass* 11, 7.

Carpenter, Bob; Andrew Gelman; Matthew D. Hoffman; Daniel Lee; Ben Goodrich, Michael Betancourt; Marcus Brubaker; Jiqiang Guo; Peter Li & Allen Riddell. 2017. Stan: A probabilistic programming language. *Journal of statistical software* 76, 1: 1-32.

Carstairs-McCarthy, Andrew. 1994. Inflection classes, gender, and the principle of contrast. *Language* 70: 737-788.

Carstairs-McCarthy, Andrew. 2010. *The evolution of morphology*. Orford: Oxford University Press.

Cotterell, Ryan; Christo Kirov, Mans Hulden, and Jason Eisner. 2019. On the complexity and typology of inflectional morphological systems. *Transactions of the Association for Computational Linguistics*, 7.

Erben Johansson, Niklas, Andrey Anikin, Gerd Carling, and Arthur Holmer. 2020. The typology of sound symbolism: Defining macro-concepts via their semantic and phonetic features. *Linguistic Typology* 24, 2: 253-310.

Feist, Timothy & Enrique L. Palancar. 2015. *Oto-Manguean Inflectional Class Database*. University of Surrey. http://dx.doi.org/10.15126/SMG.28/1

Finkel, Raphael & Gregory Stump. 2007. Principal parts and morphological typology. *Morphology* 17, 1: 39-75.

Finkel, Raphael & Gregory Stump. 2009. What your teacher told you is true: Latin verbs have four principal parts. *Digital Humanities Quarterly* 3, 1.

Ford, Lysbeth Julie. 1990. The phonology and morphology of Bachamal (Wogait).

Greenhill, Simon J., Quentin D. Atkinson, Andrew Meade, and Russell D. Gray. 2010. The shape and tempo of language evolution. *Proceedings of the Royal Society B: Biological Sciences* 277, 1693: 2443-2450.

Harvey, Mark. 2001. *A grammar of Limilngan: A language of the Mary River region, Northern Territory, Australia*. Pacific Linguistics, Research School of Pacific and Asian Studies, the Australian National University.

Harvey, Mark. 2011. *A grammar of Gaagudju*. Vol. 24. De Gruyter.

Heath, Jeffrey. 1980a. *Basic materials in Warndarang: grammar, texts and dictionary*. Dept. of Linguistics, Research School of Pacific Studies, The Australian National University.

Heath, Jeffrey. 1980b. *Basic materials in Ritharngu: Grammar, texts and dictionary*. Dept. of Linguistics, Research School of Pacific Studies, The Australian National University.

Heath, Jeffrey. 1981. *Basic materials in Mara: Grammar, texts, and dictionary*. Dept. of Linguistics, Research School of Pacific Studies, Australian National University.

Hein, Johannes & Gereon Müller. 2009. Quantitative and qualitative aspects of paradigm economy in lesser studied languages. *Morphology of the World's Languages, Universität Leipzig.*

Herce, Borja. 2020. Alignment of forms in Spanish verbal inflection: the gang *poner, tener, venir, salir, valer* as a window into the nature of paradigmatic analogy and predictability. *Morphology* 30, 2: 91-115.

Herce, Borja. 2021. Stem alternations in Kiranti and their implications for the morphology–phonology interface. *Journal of Linguistics* 57, 2: 321-363.

Herce, Borja & Bogdan Pricop. 2024. VeLeRo: An inflected verbal lexicon of Standard Romanian and a quantitative analysis of morphological predictability. Language Resources and Evaluation.

Josserand, J. Kathryn, Marcus Winter & Nicholas A. Hopkins. 1984. *Essays in Otomanguean culture history*. Vanderbilt University.

Jamieson, Carole Ann. 1982. Conflated subsystems marking person and aspect in Chiquihuitlán Mazatec verbs. *International Journal of American Linguistics* 48, 2: 139-167.

Johnson, Tamar, Kexin Gao, Kenny Smith, Hugh Rabagliati, and Jennifer Culbertson. 2020. Predictive structure or paradigm size? Investigating the effects of i-complexity and e-complexity on the learnability of morphological systems.

Kaufman, Terrence. 2006. Oto-Manguean languages. In Keith Brown (Ed.), Encyclopedia of Language & Linguistics, vol. 9, 118-124. Oxford: Elsevier.

Kaufman, Terrence & Justeson, John. 2009. Historical linguistics and pre-columbian Mesoamerica. *Ancient Mesoamerica*. 20 (2): 221–231.

Kirov, Christo, Ryan Cotterell, John Sylak-Glassman, Géraldine Walther, Ekaterina Vylomova, Patrick Xia, Manaal Faruqui et al. 2018. UniMorph 2.0: universal morphology. *arXiv preprint arXiv:1810.11101*.

Kolmogorov, Andrei N. 1965. Three approaches to the quantitative definition of information. *Problems of information transmission* 1, 1: 1-7.

Kuryłowicz, Jerzy. 1945. La nature des procès dits «analogiques». Acta linguistica, 5, 1: 15–37.

Kusters, Wouter. 2003. *Linguistic complexity*. Netherlands Graduate School of Linguistics.
Maiden, Martin. 2018. *The Romance Verb: Morphomic Structure and Diachrony.* Oxford: Oxford University Press.
Manrique Castañeda, Leonardo. 2000. Lingüística histórica, en Manzanilla, Linda y Leonardo López Lujan (coords.), *Historia antigua de México, Volumen I: El México antiguo, sus áreas culturales, los orígenes y le horizonte Preclásico*, México, INAH-UNAM-PORRUA: 53-93.

Markos, Angelos, Alfonso Iodice D'Enza & Michel van de Velden. 2019. Beyond tandem analysis: Joint dimension reduction and clustering in R. *Journal of Statistical Software* 91, 10.

Mańczak, Witold. 1957. Tendances générales des changements analogiques. Lingua, 7: 298–325.

Matthews, Peter H. 1965. The inflectional component of a word-and-paradigm grammar. *Journal of Linguistics* 1, 2: 139-171.

Matras, Yaron. 2015. Why is the borrowing of inflectional morphology dispreferred. In Francesco Gardani, Peter Arkadiev, and Nino Amiridze (Eds.), *Borrowed morphology*: 47-80. De Gruyter.

Meillet, Antoine. 1925. *La méthode comparative en linguistique historique.* Paris: Champion.

Meillet, Antoine. 1958. *Linguistique historique et linguistique générale.* Société Linguistique de Paris, Collection Linguistique, 8. Librairie Honoré Champion, Paris.

Merlan, Francesca MS. n.d. Jawoyn grammar, texts and dictionary. *Manuscript held in Australian Institute of Aboriginal and Torres Strait Islander Studies*.

Milin, Petar, Victor Kuperman, Aleksandar Kostic, and R. Harald Baayen. 2009. Paradigms bit by bit: An information theoretic approach to the processing of paradigmatic structure in inflection and derivation. *Analogy in grammar: Form and acquisition*: 214-252.

Montermini, Fabio & Olivier Bonami. 2013. Stem spaces and predictability in verbal inflection. *Lingue e linguaggio* 12, 2: 171-190.

Nichols, Johanna. 1996. The Comparative Method as heuristic. In Durie, Mark, and Malcolm Ross (Eds.), *The comparative method reviewed: Regularity and irregularity in language change*: 39-71.

Nichols, Johanna. 2014. Derivational paradigms in diachrony and comparison. In Martine Robbeets and Walter Bisang (Eds.) *Paradigm Change in the Transeurasian Languages and Beyond*: 61-88. Amsterdam: John Benjamins.

Olson, Donald. 1955. Mode-aspect-person inflection in Pame. Unpublished manuscript, SIL.

Palancar, Enrique L., and Heriberto Avelino. 2019. Inflectional complexity and verb classes in Chichimec. *Amerindia* 41: 323-360.

Pellegrini, Matteo & Marco Passarotti. 2018. LatInfLexi: an Inflected Lexicon of Latin Verbs. In *CLiC-it*.

Pellegrini, Matteo, and Alessandra Teresa Cignarella. 2020. (Stem and Word) Predictability in Italian verb paradigms: An Entropy-Based Study Exploiting the New Resource LeFFI. In *Proceedings of CLiC-it 2020*: 1-6.

Perea, Maria-Pilar, & Hiroto Ueda. 2010. Applying Quantitative Analysis Techniques to La flexió verbal en els dialectes catalans. *Dialectologia et Geolinguistica* 18, 1: 99-114.

R Core Team. 2020. *R: A language and environment for statistical computing. R Foundation for Statistical Computing*. Retrieved from http://www.R-project.org/

Reid, Nicholas John. 1990. Ngan'gityemerri: A language of the Daly River region." PhD diss., Australian National University.

Sagot, Benoît. 2013. Comparing Complexity metrics. In : Computational approaches to morphological complexity. Paris, France.

Saulwick, Adam. 2003. *Aspects of the verb in Rembarrnga: A polysynthetic language of northern Australia: Grammatical description, texts and dictionary*. PhD Dissertation, University of Melbourne.

Schultze-Berndt, Eva Friederike. 2000. *Simple and complex verbs in Jaminjung: A study of event categorisation in an Australian language*.

Scrucca, Luca; Michael Fop; T. Brendan Murphy and Adrian E. Raftery. 2016. mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. *The R Journal*, 8, 1: 289–317. https://doi.org/10.32614/RJ-2016-021.

Shannon, Claude E. 1948. A mathematical theory of communication. *The Bell system technical journal* 27, 3: 379-423.

Sims, Andrea D. & Jeff Parker. 2016. How inflection class systems work: On the informativity of implicative structure. *Word Structure* 9, 2: 215-239.

Singer, Ruth. 2006. Agreement in Mawng: Productive and lexicalised uses of agreement in an Australian language. PhD dissertation..

Snyder, Benjamin & Regina Barzilay. 2008. Unsupervised multilingual learning for morphological segmentation. In *Proceedings of acl-08: hlt*, pp. 737-745.

Spencer, Andrew. 2012. Identifying stems. *Word Structure* 5, 1: 88-108.

Stump, Gregory & Raphael A. Finkel. 2013. Morphological typology: From word to paradigm. Vol. 138. Cambridge: Cambridge University Press.

Stump, Gregory & Raphael A. Finkel. 2015. The complexity of inflectional systems. *Linguistics Vanguard* 1, 1: 101-117.

Tadmor, Uri. 2009. Loanwords in the world's languages: Findings and results. In Martin Haspelmath & Uri Tadmor (Eds.), *Loanwords in the world's languages: A comparative handbook:* 55-75. De Gruyter.

Trudgill, Peter. 2011. *Sociolinguistic typology: Social determinants of linguistic complexity*. Oxford: Oxford University Press.

Verkerk, Annemarie, and Francesca Di Garbo. 2022. Sociogeographic correlates of typological variation in northwestern Bantu gender systems. *Language Dynamics and Change* 1: 1-69.

Widmer, Manuel, Mathias Jenny, Wolfgang Behr, and Balthasar Bickel. 2021. Morphological structure can escape reduction effects from mass admixture of second language speakers: evidence from Sino-Tibetan. *Studies in Language* 45, 4: 707-752.

Wilbur, Joshua. 2014. *A grammar of Pite Saami*. Berlin: Language Science Press.