Preprint of the paper published in the Studies in Language, 2022 (John Benjamins)
DOI: 10.1075/sl.21002.bon

Derivation predicting inflection
A quantitative study of the relation between derivational history and inflectional behavior in Latin*

Olivier Bonami[1] and Matteo Pellegrini[2]

[1]Laboratoire de Linguistique Formelle – CNRS – Université de Paris
[2]CIRCSE Research Centre – Università Cattolica del Sacro Cuore, Milano

Abstract

In this paper, we investigate the value of derivational information in predicting the inflectional behavior of lexemes. We focus on Latin, for which large-scale data on both inflection and derivation are easily available. We train boosting tree classifiers to predict the inflection class of verbs and nouns with and without different pieces of derivational information. For verbs, we also model inflectional behavior in a word-based fashion, training the same type of classifier to predict wordforms given knowledge of other wordforms of the same lexemes. We find that derivational information is indeed helpful, and document an asymmetry between the beginning and the end of words, in that the final element in a word is highly predictive, while prefixes prove to be uninformative. The results obtained with the word-based methodology also allow for a finer-grained description of the behavior of different pairs of cells.

Keywords: morphology, derivation, inflection, Latin

*The whole paper is the result of the close collaboration of both authors; however, for academic purposes, Olivier Bonami is responsible for §§ 3.1 and 4, Matteo Pellegrini for §§ 1, 2, 3.2, 3.3 and 5.

1 Introduction

The aim of this paper is to investigate the extent to which derivational information helps to predict the inflectional behavior of lexemes. Variation in the way morphosyntactic property sets are realized in the inflected wordforms of different lexemes is usually modelled by resorting to the notion of inflection(al) classes. These can be defined as sets of lexemes that are inflected in the same way, cf. Aronoff (1994: 64): "An *inflectional class* is a set of lexemes whose members each select the same set of inflectional realizations". In a less strict definition, the members of an inflection class are required to have similar, but not necessarily identical inflectional behavior: "lexemes that share similar morphological contrasts", cf. Brown & Hippisley (2012: Chapter 1, fn. 3).[1]

Within the framework of Canonical Typology (Corbett 2005, Brown, Chumakina & Corbett 2013), inflection classes are considered to be fully canonical if they are purely morphological, i.e. if they are not motivated by external factors, be they phonological, syntactic, pragmatic, or semantic, thus satisfying the criteria subsumed under the principle of "independence" in Corbett (2009: 5 ff.). However, as pointed out by Corbett (2009: 8 f.) himself, inflection classes are on their turn a non-canonical phenomenon, constituting a deviation from the canonical situation where the inflectional material is expected to be the same across lexemes (Corbett 2009: 2). Therefore, the canonical ideal is rarely met by actual instances of inflection classes. Indeed, there is growing evidence that inflection classes are usually not completely arbitrary, although they are rarely categorically predictable (cf. Stump 2015: 127).

Among the external factors that have been shown to be at least partly predictive of inflection classes, there is firstly the phonological shape of the stem. Guzmán Naranjo (2019) shows that phonological properties of the stem are good predictors of inflection class in a variety of inflection systems, by training analogical models to predict the inflection class of lexemes from its stem phonology. Using the same methodology, he shows that broad lexical semantics is predictive of Kasem nominal classes, as is gender for Latin and Romanian. Gender has also been proven to be informative on the inflectional behavior of lexemes with different methodologies, namely by Stump & Finkel (2013: Chapter 5) for Sanskrit using Principal Part Analysis, and by Pellegrini (2020: Chapter 5) for Latin with an information-theoretic, entropy-based methodology. Most recently, Guzmán Naranjo (2020) and Guzmán Naranjo & Bonami (2021) both show how phonology, gender and semantic information encoded in distributional vectors combine to predict the inflectional behavior of nouns in Russian and Czech respectively.

Another piece of information that can be useful to predict the inflection class of a lexeme is its derivational history. Previous work clearly documents on the one hand cases where the inflection of a lexeme is determined by the derivational process by which it is formed: for instance, Bonami & Boyé (2006) account for

[1] A comparable distinction is made by Dressler (2002) between "micro-classes" and "macro-classes". See Beniamine, Bonami & Sagot (2017) for a contemporary discussion.

the irregular behavior of French denominal adjectives in *-eux* and *-eur*, with feminine forms in *-euse* and *-rice*, respectively (e.g. RAGEUR 'rageful', F.SG *rageuse*, DIRECTEUR 'directorial', F.SG *directrice*) by having the derivational rule that creates them specifying directly two distinct forms. On the other hand, there are many cases where a lexeme inherits – at least partly – the inflectional behavior of the base from which it derives, as has been shown to happen for a wide range of languages by Stump (2001: 98). Here, it suffices to exemplify this case by means of the English verb UNDERSTAND, with the irregular past (participle) *understood*, exactly like in the base verb STAND, with past (participle) *stood*.

More generally, several aspects of the derivational history of lexemes are potentially informative on their inflection. Our aim in this paper is to provide a quantitative evaluation of the actual predictive value of (at least some of) these different aspects, and the interplay between them. To do so, we follow a methodology similar to the one of Plaster, Polinsky & Harizanov (2013), that use decision trees to identify a small number of semantic and formal factors that are able to account for gender/noun class assignment in Tsez. We collect a large dataset of Latin verbs and nouns with detailed information on their derivational history (Section 2), and we train boosting trees to predict the conjugation class of verbs and the declension class of nouns, using different pieces of derivational information as predictors (Section 3).

In a recent trend in morphological theory, the inflectional behavior of lexemes is analyzed dispensing with the notion of an inflection class system, in a fully word-based, abstractive (Blevins 2016) approach. In this perspective, the task whose complexity needs to be evaluated does not consist in identifying an abstract inflection class from abstract lexeme-level properties, but rather in inferring the actual wordform filling the paradigm cell of a lexeme from the actual wordform filling another cell of that same lexeme, in what Ackerman, Blevins & Malouf (2009) call the "Paradigm Cell Filling Problem" (PCFP). Uncertainty in the PCFP can be modelled by means of the information-theoretic notion of conditional entropy (cf. Ackerman, Blevins & Malouf 2009, Bonami & Boyé 2014, Beniamine 2018), and it has been shown by Pellegrini (2020: Chapter 6) for Latin that taking derivational information into account allows for a reduction of entropy values, i.e. for easier predictions. In Section 4, we recast the problem that we investigate in a word-based and abstractive perspective, training our classifiers to predict wordforms from wordforms, rather than the inflection class of lexemes. Lastly, in Section 5 we draw our conclusions and we suggest directions for future research.

2 Data collection and annotation

This paper investigates the predictive value of derivation on inflection by focusing on Latin, a well-documented language for which both inflectional and derivational information is easily available from different sources. In this section we detail how our datasets were compiled.[2]

In Section 3.2, the task we face is guessing the conjugation class of verbs. To collect the sample that we use, firstly we take all the 3,348 verbal lexemes of LatInfLexi (Pellegrini & Passarotti 2018; see also later in this paper): the size of the dataset appears to be large enough to allow for reliable generalizations. Importantly, the selection of lexemes in LatInfLexi, and consequently in our sample, is based on frequency: the selected verbs are the ones recorded in the *Dictionnaire fréquentiel et index inverse de la langue latine* (Delatte et al. 1981), a frequency lexicon of Classical Latin, *de facto* excluding lexemes with low token frequency.

We then tag these verbs for the conjugation they belong to. We take the relevant information from the database of Lemlat, a large, recently renewed morphological analyzer of Latin (Passarotti et al. 2017). We thus follow Lemlat in relying on the traditional description of Latin verb inflection, that identifies four conjugations distinguished on the basis of the theme vowel displayed in the present infinitive – *ā* in the 1st, *ē* in the 2nd, *e* in the 3rd, *ī* in the 4th – and a fairly large group of heteroclitic lexemes inflected like 3rd conjugation verbs in some cells (e.g. PRS.ACT.INF) and like 4th conjugation verbs in other cells (e.g. PRS.ACT.IND.3PL), that we call the 'mixed conjugation', following Dressler (2002). We also use LatInfLexi to obtain the shape of the stem of each lexeme in phonemic transcription, by taking the phonological transcription of PRS.ACT.INF provided by LatInfLexi and stripping the exponent (usually constituted by theme vowel + *re*). Table 1 summarizes the outcome of this classification, showing the number of verbs in each conjugation in our dataset and providing examples to illustrate the relevant distinctions between classes. One example is given also for the few verbs that fall outside this classification, and are therefore simply tagged as irregular, like SUM 'be'.

[Table 1 about here.]

As for derivational information, we extract it from the *Word Formation Latin* database (WFL; Litta & Passarotti 2020). Using this source, we tag the verbs in our sample for several aspects of their derivational history, as shown in Table 2 below.[3]

[Table 2 about here.]

First of all, we obviously code the derivational affixes – prefixes and suffixes – that appear in each lexeme. We also provide information on the derivational family to which each lexeme belongs, coded by means of its "ancestor" – i.e., the base from which all the members of the family are ultimately derived. For instance, CREO, RECREO and CRESCO are all in the family labelled as the simple verb CREO. In cases of compounding, where there are two ancestors (e.g. the

[2] The datasets are available at `https://osf.io/z4kyp`.

[3] Some modifications have been made to the derivational information provided by WFL, in order to make it more compatible with the purpose of the present research. For a list of these changes, the reader is referred to Pellegrini (2020: 179-180).

adverb BENE 'well' and the verb DICO 'say' for the compound verb BENEDICO), we select the rightmost one.

Indeed, because of the suffixal nature of Latin inflection, one might assume that the right side of wordforms is what matters most in general, and that the inflectional behaviour of lexemes is mostly predictable from the last derivational morph in linear order. From a qualitative inspection of Latin data, this expectation appears to be overwhelmingly confirmed. For instance, a suffixed verb like CRESCO is assigned to the 3rd conjugation, like other lexemes formed by means of the derivational suffix *-sc-*, which is the last derivational morph appearing in the lexeme. Conversely, in RECREO the last element in linear order before inflectional markers is the stem of its base CREO, and indeed the derived lexeme is a 1st conjugation verb like its base, and the same holds for most prefixed verbs in Latin (but see §3.2 below for exceptions). To be able to perform a quantitative investigation of this aspect, we thus code the last non-inflectional morph appearing in each of the lexemes of our dataset, i.e. either the last derivational suffix – if there is any – or the ancestor – for simple and prefixed lexemes, and also for cases of conversion and parasynthesis.

It is important to keep in mind that the linear order of morphs does not always coincide with the order of morphological operations in the derivational history. To clarify this, let us compare two complex verbs that display both a prefix and a suffix, like CONDOLESCO and INARDESCO. While the suffix *-sc-* is the last morph in linear order in both cases, the derivational history of the two lexemes is very different: CONDOLESCO can be analyzed as derived by suffixation of *-sc-* to CONDOLEO 'suffer greatly', rather than as a prefixed verb, since the putative base to which the prefix should be added, *DOLESCO, is not attested. Conversely, INARDESCO is best analyzed as derived by prefixation of *in-* to ARDESCO 'take fire', because *INARDEO is not available as a base for the suffixation of *-sc-*. Since also this structural order might play a role, we add information on the last derivational operation performed to obtain each lexeme in our sample.[4] Verbs formed by means of different compounding processes are all treated as having a generic "compounding" operation as the last one. On the other hand, for cases of conversion and parasynthesis, we make a distinction according to the lexical category of the base on which the operation is performed. Hence, we distinguish between noun-to-verb (e.g. CORONA_{N} 'crown'$\rightarrow$ CORONO_{V}) and adjective-to-verb (e.g. DIGNUS_{A} 'worthy'$\rightarrow$ DIGNO_{V}) conversion, and we distinguish both from cases of verb-to-verb converted frequentatives/intensives selecting the Third Stem (e.g. RAPIO_{V} 'seize', S3 *rapt-* $\rightarrow$ RAPTO_{V}); and we also distinguish between prefixes creating verbs from nouns (e.g. PILUS_{N} 'hair' $\rightarrow$ DEPILO_{V}) vs. adjectives (e.g. PRAVUS_{A} 'crooked' $\rightarrow$ $\text{DEPRAVO}_{\text{V}}$).

Regarding nouns, we cannot take our data from LatInfLexi directly, since only 1,048 nominal lexemes are included there, which proves to be too small

[4]Note that things are not always as straightforward as in the examples provided here, and in some cases both analyses are possible in principle: cf. Budassi & Litta (2017) for a detailed discussion of this issue. In our classification, we simply follow the choice made by WFL in this respect.

a number to allow for solid generalizations. Therefore, we start from scratch to obtain our dataset of nouns. We want the selection of lexemes to be based on frequency, as it is in our sample of verbs. As a source of information on this respect, we use the frequency lists of LatinISE, a large diachronic corpus of Latin (McGillivray & Kilgarriff 2013), rather than the *Dictionnaire fréquentiel et index inverse de la langue latine*, as in LatInfLexi. The information extracted from the former is much more noisy than the one taken from the latter, but it is also readily available in machine-readable format. Furthermore, we reduce the amount of noise since we only keep the lexemes that are also recorded as nouns in the database of Lemlat. We set our threshold of frequency at 4, so as to obtain a sample whose size – 2,982 nouns – is roughly comparable to the one of verbs. We also use the database of Lemlat to extract information on the phonological shape of the stem – by semi-automatically transcribing into IPA the stems recorded in there under the label of "Lexical Segment" (LES) – and on the declension to which each noun belongs. Again, the classification in inflection classes is the traditional one, identifying 5 declension, according to the exponent used for GEN.SG, as shown in Table 3, where we also display the number of lexemes in each class.

[Table 3 about here.]

The derivational information provided for nouns is the same as the one described above for verbs, and also the source is always the WFL database. Additionally, we also code the gender (masculine, feminine or neuter) of each nominal lexeme, as the examples of Table 4 illustrate.

[Table 4 about here.]

Finally, for the word-based predictions of Section 4 we simply use the data of LatInfLexi, that provides the phonological transcription of the wordforms filling all the non-defective and non-periphrastic cells in the paradigms of 3,348 Latin verbs. To shorten the time of the computation, we do not analyse full paradigms. We rather focus on a 15-cell "distillation" (cf. Stump & Finkel 2013) – i.e., a reduced version where cells that are trivially predictable from one another are conflated in a single zone, keeping only one cell for each zone, as in the description of Pellegrini (2021).

3 Predicting inflection classes

3.1 Methodology

Our goal in this section is to assess whether various phonological and morphological properties of lexemes function as partial predictors of the inflection class they belong to. Following previous work such as Plaster, Polinsky & Harizanov (2013), Guzmán Naranjo (2019) and Guzmán Naranjo & Bonami (2021), we propose to use methods from supervised machine learning to this end. A CLASSIFIER is an algorithm that attempts to predict the value of some categorical

DEPENDENT VARIABLE, here the inflection class, on the basis of a number of other explanatory variables, representing here various pieces of phonological, morphological and morphosyntactic information. It does so by capturing statistical associations between the relevant variables in training data, and relying on these associations to make a prediction as to the value of the dependent variable given the values of the explanatory variables for test datapoints that were not part of the training data. The basic hypothesis for this line of research is that the accuracy of a classifier is indicative of the predictive power of the explanatory variables. If adding an explanatory variable to the classifier improves accuracy, this entails that the information encoded in that variable does have some predictive power in determining the value of the dependent variable.

The validity of this hypothesis is predicated on the quality of the classification method. All classifiers presented in this paper rely on gradient boosting (Friedman 2001, Mason et al. 2000) applied to decision trees, a method that is known to capture efficiently complex interactions between variables, is not too sensitive to spurious predictors, and has been successfully applied to prediction of inflectional behavior in Czech by Guzmán Naranjo & Bonami (2021). All models were trained and tested using the Python implementation of gradient boosting in the scikit-learn package (Pedregosa et al. 2011). All results below were obtained by performing 10-fold cross-validation: first, the dataset is partitioned randomly in 10 equally-sized folds; then 10 classifiers are trained successively on training data consisting of 9 of the folds and tested on the remaining, 10th fold. This allows one to get an accuracy result for each datapoint based on a classifier that has not seen this data. Reported accuracies are the proportions of correct predictions across all 10 classifier results.

Note that we choose to use a uniform classification method and hyperparametrization for all datasets.[5] This has the practical advantage of reducing the number of models to be trained, and the conceptual advantage of allowing for meaningful comparisons across predictor sets. However, as each classification method has its unique inductive biases, it is entirely possible that the choice of a different method would lead to different conclusions.[6]

3.2 Predicting the conjugation of verbs

In this section, we discuss the results obtained by training our classifier to predict the conjugation of Latin verbs using the different pieces of derivational information coded in our dataset as predictors. Table 5 summarizes our results.

[5] All models used 500 boosting stages, a learning rate of 0.1, and a maximum tree depth of 10. These hyperparameters were optimized by grid search over a small subset of the 1101 classification problems we report on in this paper.

[6] We also tried using Classification and Regression Trees as well as Random Forests with various parametrizations to the classification problems presented in this section. This led to qualitatively equivalent results, but a counterintuitive situations where spurious predictors commonly led to a degradation of classification accuracy. We decided to report the boosted tree results in the interest of presentational clarity, because spurious predictors tend to have no effect on accuracy rather than a negative effect. We did not try any other classification algorithm on the word-based classification problems we report on in Section 4.

[Table 5 about here.]

Table 5 reports the estimated accuracy of models trained with different (combinations of) predictors and the 95% confidence intervals within which we can reasonably assume the true accuracy to lie. Models are sorted by increasing accuracy. The accuracy of the baseline classifier, that simply assigns everything to the most frequent class, is also given for comparison. For the same reason, the phonological shape of the final part of the stem – using only the last three segments – is also included among our predictors. As we will see, this is also useful to be able to disentangle the purely morphological predictive power of derivational suffixes from information that can be inferred by final phonology alone.

One model can be considered as clearly outperforming another one if their confidence intervals do not overlap. Therefore, to evaluate the impact of different (combinations of) predictors, it is useful to group together models that clearly outperform the same set of other models, as indicated in the last column of Table 5. In Table 5, the different groups of predictors obtained in this way are separated by a dashed line. The best models – the ones in the last group, i.e. (16), (17) and (18) – all rely on some combination of final phonology and family information. Among them, the optimal model can be reasonably considered to be the one with fewer predictors, i.e. `phonology+family`.

Let us now move to a more detailed evaluation of the contribution of different predictors. Prefixes prove to be very poorly informative of the inflectional behavior of lexemes. On their own, they do not even beat the baseline. When combined with other predictors, in the best-case scenario they do not change the situation: for instance, adding prefix information to the model using the classification in families as predictor, we obtain similar results, with overlapping confidence intervals. In other cases, however, using prefixes even appears to confuse the classifier, making accuracy values decrease: for instance, the combination of phonological and prefix information is clearly outperformed by the model using only phonology, and similarly `prefix+suffix` yields lower accuracy values than `suffix` alone. The limited predictiveness of derivational prefixes is not unexpected, given the inflectionally suffixal nature of Latin. Lexemes formed by means of the same prefix can be assigned to any conjugation, as shown by the quantitative data of Table 6 – where only the 10 prefixes with the largest number of derivatives are displayed, while the other prefixes are merged together in the last line. This happens because the inflection class of the base lexeme is almost always preserved in such cases, as the examples in (1) illustrate.

[Table 6 about here.]

(1)
a. $\text{STO}_{\text{V:1}^{\text{st}}}$ ‘stand’ → $\text{EXSTO}_{\text{V:1}^{\text{st}}}$ ‘stand out’
b. $\text{TORQUEO}_{\text{V:2}^{\text{nd}}}$ ‘twist’ → $\text{EXTORQUEO}_{\text{V:2}^{\text{nd}}}$ ‘twist out’
c. $\text{CLUDO}_{\text{V:3}^{\text{rd}}}$ ‘shut’ → $\text{EXCLUDO}_{\text{V:3}^{\text{rd}}}$ ‘shut out’
d. $\text{POLIO}_{\text{V:4}^{\text{th}}}$ ‘smooth’ → $\text{EXPOLIO}_{\text{V:4}^{\text{th}}}$ ‘smooth off’
e. $\text{CAPIO}_{\text{V:mix}}$ ‘take’ → $\text{EXCIPIO}_{\text{V:mix}}$ ‘take out’

For the same reason, the derivational family of a lexeme is expected to be a very good predictor of its inflectional behavior: while knowing the prefix by which the lexemes of (1) are formed does not help to infer their conjugation, knowing the base from which they are ultimately derived allows for a very safe guess. Some verbs that come from the 1st conjugation verb DO 'give', but are nevertheless assigned to the 3rd conjugation are the only exception to this generalization. This is due to an analogical levelling of their paradigm triggered by wordforms whose surface shape happens to be the same as if they were assigned to the 3rd conjugation, because of the outcome of a phonological process of weakening of front vowels in non-initial syllables that was active in Old Latin, yielding e.g. **ob-damus* > *ob-dimus* (like in 3rd conjugation verbs, e.g. *leg-imus*). Indeed, among single derivational predictors,[7] family is the one with the highest accuracy values, and it is present in all optimal combinations of predictors (the ones in the last group of Table 5).

The outermost derivational operation in the lexemes of our dataset is most frequently prefixation, as shown in Table 7. Hence, it does not come as a surprise that the predictor `outermost` displays a behaviour comparable (although not identical) to the one of `prefix`: adding it as a predictor is never useful to infer inflection class, and on its own it is outperformed even by the baseline.

[Table 7 about here.]

Differently than prefixes, derivational suffixes assign inflection class to the lexemes they form, as clearly emerges from the data in Table 8. For instance, all lexemes formed by means of the suffix *-sc-* belong to the 3rd conjugation, regardless of the (nominal, adjectival or verbal) inflection of the base: hence, we have for instance CRESCO$_{V:3^{rd}}$ 'grow' from CREO$_{V:2^{nd}}$ 'create', HISCO$_{V:3^{rd}}$ 'open' from HIO$_{V:4^{th}}$ 'open', FRONDESCO$_{V:3^{rd}}$ 'become leafy' from FRONS$_{N}$ 'leafy branch' and DURESCO$_{V:3^{rd}}$ 'grow hard' from DURUS$_{A}$ 'hard'.

[Table 8 about here.]

Indeed, suffixes prove to be partly informative on the inflectional behaviour of lexemes in our dataset. However, from our results the information provided by suffixes appears to be redundant with information on the phonological shape of the last part of the stem: adding `suffix` to `phonology` does not yield a relevant increase in accuracy, while the converse is true. Furthermore, the combination of predictors `phonology+family` is the optimal predictor (no set of predictors leads to significantly higher accuracy, and it is the simplest of the set of predictors at that level of accuracy), and it clearly outperforms `family+suffix`. This suggests that by using stem phonology it is possible to capture regularities that go beyond the presence of a suffix, while the opposite is false.

It is clear why phonological information is preferable to information on suffixes in this dataset. By far the most prevalent suffixes are *-sc-*, assigning verbs to the 3rd conjugation, and *-it-*, assigning them to the 1st conjugation. As

[7] The only other predictor with comparable accuracy is the phonological shape of the stem.

it happens, verbs suffixed in *-sc-* form the vast majority of verbs with stems ending in /sk/, and even nonsuffixed verbs with stems in /sk/ mostly fall in the 3rd conjugation. Based on phonology alone, the classifier hence assigns all verbs with stems in /sk/ to the 3rd conjugation, leading to perfect accuracy for suffixed verbs. On the other hand, relying on morphological information alone misses the generalization that verbs with stems in /sk/ tend to fall in the 3rd conjugation even when they are unsuffixed. The same holds, *mutatis mutandis*, for stems in /it/ and the 1st conjugation.

A similar behaviour is displayed by the predictor `last`. The value of this category in our dataset corresponds to the value of suffix if there is one, to the value of family otherwise. Hence, it is somewhat surprising that `last` proves to be not as good a predictor as the combination of `family+suffix`. This suggests that information on the derivational family is partly helpful even for suffixed lexemes.

To confirm this, and to have a clearer picture that allows to abstract away from the quantitative relevance of different derivational processes, it is useful to have a look at the situation that arises when taking into account only verbs that display a suffix on the one hand, only verbs that display a prefix on the other hand (cf. Tables A1 and A2 in the Appendix). In the former case, any combination including either `phonology`, `suffix` or `last` is categorically predictive, as is expected since suffixes assign inflection class to the lexemes they form; however, also family proves to have some predictive power, beating at least the baseline: this confirms that family information plays a role even for suffixed verbs, a fact that cannot be inferred from a qualitative analysis of Latin data. In the latter case, the same observations that could be made on the full dataset are still valid: thus, prefixes prove to be a spurious predictor even when considering only verbs that do display a prefix. Importantly, `phonology` is still a relevant predictor for nonsuffixed verbs, even in combination with `family`. This indicates that there are phonological generalizations over stem shapes that are partially predictive of inflectional behavior.

To sum up, for verbs the best derivational predictor of inflection class is the family to which they belong. The optimal model is obtained by adding this factor to the phonological shape of the last part of the stem. Furthermore, and less expectedly, family information appears to be partly relevant even for suffixed verbs. Also suffixes appear to play a role, as does the last derivational morph of lexemes in linear order, but their role is limited and – most importantly – redundant with the information already provided by final stem phonology. Lastly, prefixes and the outermost derivational operation applied to lexemes prove to be not informative at all on the inflectional behavior of the verbs of our sample.

3.3 Predicting the declension of nouns

In this section, we turn to nominal inflection, training our classifiers to predict the declension of Latin nouns in our dataset. Results are summarized in Table 9 below, again displaying the accuracy, 95% confidence intervals and the (combinations of) predictors that are clearly outperformed – with no overlap in confidence intervals – by each model.

[Table 9 about here.]

To the predictors used for verbs, we add gender, that is potentially informative on the inflection of a noun, because of the uneven distribution of nominal lexemes among genders in different declensions, as summarized in Table 10. For instance, assuming to know that a noun is neuter, it is possible to exclude the 1^{st} and 5^{th} declension from the ones to which it can be assigned, since there are no neuter nouns in those inflection classes. Indeed, this factor is confirmed to be a good predictor of inflection class assignment: it beats the baseline on its own, adding it to other predictors always yields a significant increase in accuracy values, and it is always present in the combinations of predictors used by the best models.

[Table 10 about here.]

Our results for nouns are similar to the ones obtained for verbs in showing no predictive power of prefixes: again, the model using only prefixes does not even beat the baseline, and adding `prefix` to a model usually either leaves the situation unchanged or it actually makes accuracy values decrease. Another confirmation concerns the role of suffixes, again mirrored by information on the last derivational morph in linear order: adding `suffix` or `last` to a model improves accuracy. Differently than what happened for verbs, for nouns suffixal information appear to be complementary, and not redundant with final stem phonology, as is shown by the fact that if `suffix` or `last` are added to `phonology`, there is a significant increase in accuracy.

These facts are hardly surprising, since the same observations made above for verbs on the inflection class assigning nature of derivational suffixes on the one hand and on the transparency of derivational prefixes to the inflection class of the base on the other hand also hold for nouns, as is clearly shown in Tables 11 and 12. Nouns that display the same prefix belong to different declensions. Conversely, nouns displaying the same suffix are all in the same declension, the only exception being constituted by diminutives like *-cul-* and *-ll-* that are transparent to the gender of the base, and are accordingly assigned to different declensions – the 1^{st} if the base is feminine, the 2^{nd} if it is masculine or neuter, as the examples in (2) illustrate.

[Table 11 about here.]

[Table 12 about here.]

(2) a. AURIS$_{\text{N:F-3}^{\text{rd}}}$ ‘ear’ → AURICULA$_{\text{N:F-1}^{\text{st}}}$ ‘ear-lap’ (lit. ‘small ear’)
b. FASCIS$_{\text{N:M-3}^{\text{rd}}}$ ‘bundle’ → FASCICULUS$_{\text{N:M-2}^{\text{nd}}}$ ‘small bundle’
c. OS$_{\text{N:N-3}^{\text{rd}}}$ ‘mouth’ → OSCULUM$_{\text{N:N-2}^{\text{nd}}}$ ‘small mouth’

On the other hand, there is a remarkable difference regarding the role of family, that played such an important role for verbs. For nouns, family on its own does not even outperform the baseline, and adding it to other (combinations of) predictors is rarely helpful in a relevant way. In particular, there are models with maximal predictiveness that do not include `family` as a predictor.

Another difference that jumps to the eye with respect to the results obtained for verbal lexemes is the importance of the outermost derivational operation performed on the nominal lexemes of the sample. This was found to be a spurious predictor for verbs, while for nouns it proves to be the best single predictor, and it is present in the optimal combination of predictors – the one with fewer predictors among the ones of the last group of Table 9, namely `phonology+outermost+gender`.

The explanation of these differences can be reasonably found in the different quantitative relevance of the processes that are applied lastly in the derivational history of lexemes in the sample of nouns and verbs. For verbs, the outermost derivational operation is overwhelmingly prefixation, hence the models including outermost as a predictor often use information on prefixes, that have been shown to be poorly informative on the inflectional behavior of lexemes. Conversely, lexemes where the outermost operation is suffixation are the most frequent in the noun sample. In that case, the models including outermost thus exploit information on the suffix displayed by lexemes, a factor that is shown to be predictive of inflection class.

Similarly, the limited role of the derivational family of nouns is likely to be due to two combined factors. First, there are comparatively fewer prefixed nouns, and we saw that family information is most useful for prefixed items. Second, in our dataset, nouns belong to comparatively smaller families than verbs. 37% of nouns are alone in their family. This entails that for 37% of the nouns, family information is absent from the training data, because there is no other lexeme from the same family to learn from. The same holds only for 12% of verbs. In the same vein, the median family size is 2 for nouns and 6 for verbs, entailing that classifiers for verbs tend to have a lot more information on properties of the family to build on.

[Table 13 about here.]

Results obtained on the samples including only prefixed and suffixed lexemes (cf. Tables A3 and A4 in the Appendix) confirm the conclusions drawn on the full dataset, that can be summarized as follows. For nouns, among derivational predictors information on suffixes is the most useful to help predicting inflection class, with the last morph in linear order – most often, a suffix – consequently playing a comparable role. Conversely, prefixes are poorly informative, and so is information on the derivational family of lexemes. The optimal model for nouns include two non-derivational predictors that prove to be strongly predictive of the inflectional behaviour of lexemes, namely their gender and stem phonology, together with the outermost derivational operation, where information on both the suffix of the noun – if there is one – and the derivational family to which it belongs are encapsulated.

4 A word-based alternative: predicting verb forms

4.1 Rationale

In the previous part of this paper, the task we focused on was predicting the inflection class of lexemes, starting from an abstract lexeme identifier, represented by their citation form. While this is the most usual way of investigating the inflectional behaviour of lexemes, it is not without its problems. On the one hand, by framing the issue in this way, in a sense there is an overestimation of the difficulty that is actually faced by speakers in discourse. This is due to the fact that inflection class indexes are meant as summarizing the overall inflectional behaviour of a lexeme: knowing that a lexeme belongs to a given inflection class is tantamount to being able to fill its whole paradigm. However, the task that speakers have to fulfil is simpler than that: it can be conceptualized as consisting in producing the appropriately inflected wordform – as required by the syntactic context – of a lexeme they already know, having encountered it at least once, i.e. having been exposed to at least one of its other forms. On the other hand, there is also a sense in which the difficulty of the task is underestimated by the use of inflection classes, that very often do not capture all relevant aspects of the inflectional behaviour of lexemes – especially when they are defined more laxly as "macro-classes" of lexemes that inflect similarly rather than identically (see the discussion in Section 1), as is usually done in traditional descriptions like the ones we relied on in the previous section. This kind of simplification is particularly relevant in Latin conjugations, that only refer to variation in the selection of endings in imperfective cells, as was summarized above in Table 1. However, if we also look at perfective cells and some nominal forms, like the perfect participle or the supine, the endings are always the same for lexemes of all conjugations, but there is remarkable allomorphy in the formation of the stems on which these wordforms are based – the so called Perfect Stem and Aronoff's (1994) Third Stem, respectively. Table 14 exemplifies some of the attested patterns, among which we find reduplication (e.g. in the Perfect Stem *cucurr-*), vowel lengthening (e.g. in the Perfect Stem *vēn-* vs. Present Stem *ven-*), apophonic alternation (e.g. in the Perfect Stem *cēp-* vs. Present Stem *cap-*), and full suppletion in the case of FERO. All this variation is simply not considered by the traditional classification of conjugations.

[Table 14 about here.]

Of course, these issues potentially have an impact on the results presented in Section 3. Indeed, at least in some cases, by limiting the investigation to the traditional conjugations, some relevant details regarding the role of derivational information are clearly neglected. Consider, for instance, the data of Table 15.

[Table 15 about here.]

Knowing that the verbs of Table 15 all belong to the derivational family of FACIO, they can be assigned to the same conjugation as the base (i.e., the 3rd) with no uncertainty. However, if we focus on the more realistic task of guessing PRS.ACT.INF from PRF.ACT.IND.1SG, things are not so straightforward: while in CALEFACIO the same pattern of stem allomorphy found in the base lexeme is applied (Perfect Stem *calefēc-* vs. Present Stem *calefac-*), in INFICIO there is a different pattern (Perfect Stem *infēc-* vs. Present Stem *infic-*). This is due to the weakening of front vowels in non-initial syllables already mentioned in §3.2, that was active in Old Latin, yielding *in-ficio* < **in-facio*, but not in Classical Latin, leaving more recent formations like *cale-facio* unaffected. In any event, this variation generates uncertainty in the Paradigm Cell Filling Problem: given a PRF.ACT.INF in *-fēcī*, two possible alternatives are available for PRS.ACT.INF, that can end in *-facere* and *-ficere*. This uncertainty cannot be captured by focusing on traditional conjugations only, since all these verbs are assigned to the 3rd conjugation, because the endings of imperfective cells are always the same.

To evaluate the impact of such facts on our results, in this section we recast the problem of the impact of derivational information on inflectional predictions in fully-word based, abstractive terms, following a methodology that we will outline in the rest of this section.

4.2 The information-theoretic approach to the PCFP

In this subsection we review information-theoretic, word-based methods addressing the Paradigm Cell Filling Problem (PCFP). These will then serve as a crucial inspiration to the approach developed in this paper.

Ackerman, Blevins & Malouf (2009) introduced the idea of using conditional entropy to evaluate the difficulty of predicting what happens in an output paradigm cell c' knowing the form occupying paradigm cell c: the shapes of wordforms occupying cells c and c' are classified into types accross the lexicon (typically characterized in terms of affixal exponents), allowing for the definition of corresponding random variables C and C' capturing the probability of an arbitrary lexeme fitting into each of the possible types. Probabilities are then evaluated on the basis of type frequency information computed over a large lexicon,[8] and the conditional entropy $H(C' \mid C)$ computed from these estimates how much uncertainty there is on average when trying to guess the type of form filling cell c' knowing what type of form fills cell c. For instance, imagine that the Latin verbal lexicon consists solely of the 5 familiar verbs relisted in Table 16, and suppose that we are satisfied by indentifying form types as indicated. From this we can estimate the difficulty of predicting the present infinitive from the first-person singular of the indicative as $H(\text{PRS.ACT.INF} \mid \text{PRS.ACT.IND.1SG}) = 0.8$, while in the opposite direction predicting the first-person singular from the infinitive is easier, as $H(\text{PRS.ACT.IND.1SG} \mid \text{PRS.ACT.INF}) = 0.4$. The asymmetry is due to the fact that 4 out of 5 verbs have in the present 1SG endings which

[8]The earliest studies using this methodology (Ackerman, Blevins & Malouf 2009, Ackerman & Malouf 2013) made the simplifying assumption that all inflection classes are equiprobable, leading to a very poor estimation of probabilities of inflectional types. Bonami & Boyé (2014) and Sims (2015) were the first studies to rely on actual type frequency information.

neutralize a binary distinction relevant in the infinitive (/-oː/ maps to either /-aːre/ or /-ere, /-ioː/ maps to either /-iːre/ or /-ere/, while only 2 out of 5 do so in the opposite direction (/-ere/ maps to /-oː/ or /-ioː/; all three other infinitive endings predict perfectly the present 1SG ending).

[Table 16 about here.]

This simple method was criticized by Bonami & Boyé (2014) and Bonami & Beniamine (2016) for its reliance on a pre-existing classification of wordform shapes whose choice is left to the analyst, leading to concerns on the exact nature of what is measured and in some cases the possibility of circularity. A simple illustration of the problem is presented in Table 17, where three extra Latin verbs are added to the picture. It turns out that some first conjugation verbs have a stem ending in /e/ (e.g. COMMEO 'visit') or /i/ (e.g. GLACIO 'freeze'), and one third conjugation verb in our dataset has a stem ending in /i/ (IMMEIO 'make water into'). These verbs present a challenge for the application of Ackerman et al.'s method. Clearly the exponent of first person is just /-o:/, and the preceding vowel belongs to the stem, as it is constant accross the paradigm. If we hence take the appropriate type of present 1SG to be /-o:/ in these three cases, then our evaluation of the difficulty of predicting the infinitive is essentially unchanged, with $H(\textsc{prs.act.inf} \mid \textsc{prs.act.ind.1sg}) \approx 0.86$.[9]

[Table 17 about here.]

But clearly we are missing something here, if what we are interested in is really how speakers can predict one form from another: while knowledge of the rest of the paradigm allows us to infer that /e/ and /i/ are part of their respective stems, a speaker exposed to just the PRS.ACT.IND.1SG form cannot know that: speakers do not hear morph boundaries. Hence, realistically, looking at a word such as *glacio*, the expected behavior of a speaker should be hesitation rather than certainty: this could be a first, third, fourth, or mixed conjugation verb, because there are verbs in all four conjugations whose present 1SG form ends in /-io:/.

Bonami & Boyé (2014) argue that the way out of this conundrum is to not make any unchecked assumption about the classification of forms filling a cell, but infer these in a principled way from an examination of the surface alternations present in the data and directly observable by speakers.[10] Hence a basic building block of a proper approach to the PCFP is an algorithm for classification of pairs of words into alternation types. The PCFP can then be recast as the problem of predicting which alternation type relates the forms filling two

[9]The numbers are not exactly the same as before, because by adding three items to our toy lexicon we modified the type frequency of each inflection class.

[10]See Malouf (2017) for a completely different way of taking into account Bonami and colleague's critique: Malouf addresses the PCFP by showing that a recurrent network can learn to accurately produce the unseen form of a lexeme in cell *c* from just a lexeme identifier, on the basis of training data that contain other forms of that lexeme and the forms of other lexemes in cell *c*.

cells c and c', given what knowledge of the form in c allows one to infer on which alternations may be applicable. As a case in point, Table 18 illustrates how the problem of predicting the present infinitive from the PRS.ACT.IND.1SG is recast.

[Table 18 about here.]

In this simple example the alternations amount to substitutions of endings. Once these alternations have been identified, predictor shapes (here the PRS.ACT.IND.1SG) are classified by identifying which alternations they could potentially lead to given their phonological makeup. The set of such applicable alternations is given in the last column. For instance, the set of applicable alternations for the form /lawdoː/ is $\{a_1, a_3\}$: this form ends in /oː/ but neither in /eoː/, which is required for application of alternation a_2, nor in /ioː/, which is required for the application of a_4 or a_5. By contrast, the set of applicable alternations for the form /lakio/ is $\{a_1, a_3, a_4, a_5\}$, because that form ends both in /oː/ and /ioː/. Hence we capture the fact that two wordforms of lexemes exhibiting the same alternation (and hence belonging to the same inflection class) may have different predictive power: /glakio/ is uninformative as to what could happen in the infinitive in a way that /lawdoː/ is not. If we go on to compute the conditional entropy of the alternation actually relating to forms given the class of the input form, we find the value to be 1.5. This confirms that relying blindly on a pre-existing classification of wordforms into shape types led to underestimate the difficulty of the PCFP in this particular instance.

4.3 The PCFP as a classification problem

Our goal in this section is to build on the information-theoretic literature on the PCFP to further our understanding of the predictive power of derivational information for inflectional behavior. To this end, following a precedent set by Guzmán Naranjo (2020), we recast the PCFP as a classification problem: classifier accuracy takes the place of conditional entropy as our evaluation of predictability. From Section 3 we retain the idea of constructing multiple classifiers using various combinations of morphological predictors as input, evaluate their performance through cross-validation, and use this to assess the predictive power of these various predictors. Although it would have been possible to expand more directly on the information-theoretic approach, by adding derivational predictors to the set of conditioning variables in evaluations of conditional entropy, there are a number of distinct advantages to the classification-based approach. First, a classifier can be directly interpreted as a model of the PCFP as it arises for speakers: just as a speaker, the classifier is trained on partial knowledge of the system, and then applied to unseen data. Second, the information-theoretic approaches outlined above are predicated on unrealistic omniscience of speakers on the relevant probability distributions. While this may be innocuous as long as one is dealing with macroscopic phenomena for which large frequency allows for accurate estimation of probability, such as inflectional macro-classes, it is unsatisfactory when dealing with microscopic phenomena such as derivational families: in the typical situation where a speaker has been exposed to a handful of members of a family, they cannot be expected to have an accurate estimation of the probability of a member of that family exhibiting a particular inflectional behavior.[11] The present approach has a built-in safeguard against this omniscience problem, as the classifier's generalizations can only be as good as can be extracted from the incomplete sample of data seen in training. Third and finally, the classification approach allows us to deploy exactly the same machine learning algorithms used in Section 3, and hence allows for meaningful quantitative comparison of our results.

From the information-theoretic literature on the PCFP reviewed in the previous subsection we retain two main ideas. First, we want to predict word-level properties from word-level properties, rather than lexeme-level properties (inflection class) from lexeme-level properties (stem phonology and morphology); this is the essence of a word-based approach. Second, we model the morphophonological aspects of prediction using the insights of Bonami & Boyé (2014) and Bonami & Beniamine (2016): our word-based classifiers predict the pattern relating two wordforms on the basis of the relevant morphophonological properties of the predictor form. Specifically, we use the algorithm introduced by Beniamine (2018) for automatic classifications of pairs of paradigmatically-related wordforms into alternation types.[12]

Concretely then, for each pair of paradigm cells of interest, we trained a number of gradient boosting classifiers, using the same hyperparameters described in Section 3. The dependent variable is always the type of alternation relating the wordforms filling the two cells. In the simplest, `vanilla` classifiers, there is a single predictor variable, which is the morphophonological class of the predictor wordform, assessed as indicated in Section 4.2 through the set of applicable alternations: the accuracy of that classifier is our estimation of the difficulty of the PCFP taking only morphophonological information into account, and the direct analogue of the information-theoretic assessment of the PCFP in Bonami & Beniamine (2016) or Pellegrini (2020). It can also be compared to the classifiers with only `phonology` as predictor in the setup of Section 3: as for these, only phonological information is used for prediction, although now it is information on the overall shape of a predictor word rather than just the stem. We compare these vanilla classifiers to classifiers using extra dependent variables encoding various types of derivational information. In the interest of time, we focused on those variables that have been shown to have some predictive power

[11] As a reviewer notes, our sampling process is not realistic inasmuch as token frequency is not into account. In this we follow the lead of Bybee (1995), Pierrehumbert (2001) and Albright & Hayes (2003) who all argue that morphological patterns are extended on the basis of type frequency only. See however Boyé & Schalchli (2019) for a general discussion of the proper sampling of morphological data for the assessment of the PCFP.

[12] Although the toy data we use for illustration is simple enough that alternations can always be seen as substitution of suffixes, one of the virtues of Beniamine's algorithm is to capture more complex types of alternations, including nonconcatenative alternations resulting from infixation, root-and-pattern morphology, and suprasegmental exponence. This is crucial to capturing some of the patterns in the Latin data, including complex stem alternations such as those presented in Table 14. In addition, the patterns encode knowledge on the phonotactic contexts within which each alternation is found.

in Section 3: family information, suffix information, a combination of the two, and the last non-inflectional morph in linear order.

We hence have five predictor combinations that we would in principle want to evaluate on each of the $254 \times 253 = 64,262$ pairs of cells in the Latin verbal paradigm. Since this is computationally unrealistic, as was hinted above in Section 2 we follow the common practice of focusing instead on a *distillation* of the paradigm in the sense of Stump & Finkel (2013), a set of cells each of which is representative of a zone of perfect interpredictability. We use the same 15-cell distillation Pellegrini (2021) derived by hunting for null conditional entropy values between pairs of cells.

4.4 Results

In this section we present the results obtained with the word-based methodology outlined above. Figure 1 reports the accuracy values of all classifiers using the 'vanilla' configuration of predictors, that is, with no derivational information. There are 210 such classifiers, as we are trying to predict what happens in each of the 15 cells in the distillation from any of the remaining 14 cells. It is worth noting that we find here for Latin the same situation documented for various other languages using comparable quantitative methods to assess the PCFP: prediction varies from trivial (accuracy of 1) to mildly difficult (minimum accuracy 0.54), with some cells being overall good or bad predictors (compare PRS.ACT.IND.1SG and PRS.ACT.IND.3PL rows), and easy or hard to predict (compare PRS.PTCP and PRF.ACT.INF columns), with no simple correlation between the two properties (e.g. the last two cells, based on the third stem, are strongly interpredictable but badly connected with any other cells). This picture closely corresponds to the entropy-based description of Latin verb inflection sketched by Pellegrini (2021), to which the reader is referred for further details.

[Figure 1 about here.]

This set of 'vanilla' models will serve as the baseline for evaluation of the value of different kinds of derivational information. For a first, coarse-grained evaluation, in Table 19 we report the average accuracy values of all 210 classifiers for each configuration of predictors.

[Table 19 about here.]

These results again suggest that the most useful derivational predictor is the derivational family of lexemes, that alone is able to yield a relevant increase in the accuracy of the vanilla model. Conversely, suffix information does not seem to be very useful: it contributes nearly nothing in average accuracy improvement, when added to other sets of predictors. As expected, the predictiveness of the last morph appears to be exactly the same as the one obtained combining family and suffix information.

Although these results are interesting, averaging over all classifiers of the same type may hide interesting patterns: for instance, it may be the case that identical averages hide asymmetries where one type of classifier performs well in a corner of the paradigm and poorly in another corner, while another type of classifier does the opposite. It is precisely a virtue of a detailed word-based approach such as that developed here that we can make a more detailed assessment and spot such situations.

To that effect, the 6 plots in Figure 2 provide a more detailed comparison of 6 pairs of classifiers. In each of the plots, each point indicates, for some pair of cells in the distillation (A, B), the accuracy of two classifiers predicting B from A using different sets of predictors. The diagonal line materializes the location where all the dots should be if the two classifier types had exactly the same performance. Points materialized by dark circles document cases where the difference between the accuracy of the two classifiers is statistically significant (the 95% confidence intervals around the accuracy value do not overlap), while lighter crosses correspond to cases where it is not.

[Figure 2 about here.]

We comment the plots in turn, starting from the cleared cases. Looking at the bottom right plot, we see that the last morph has exactly the same predictive power as the combination of family and suffix information (bottom right plot): all points are closely clustered on the diagonal, and there is a significant difference for only 2 out of 210 classifiers.[13] This is unsurprising, since the last morph information is almost the same thing as the combination of family and suffix information. However, this result contrasts with what we saw in §3.2, where, using different methodology, we documented a lower accuracy for prediction from phonology and last morph than from phonology, family and suffix. The remaining plots visually seem to fall into two classes: in three of the plots, points are really close to the diagonal, and very few pairs of classifier are significantly different, whereas in the two remaining ones, most points are markedly higher than the diagonal, and differences are almost always significant. This indicates that in the first three cases the classifier types have very similar performance, whereas in the last two the more complex classifier type performs markedly better than the simpler one. This confirms that the derivational family information is overall highly predictive of inflectional behavior, while suffix information is not.

However, it is notable that there are a few combinations of predictor and predicted cells for which suffix information is predictive. Looking at this more closely, it turns out that suffixes are usually helpful when predicting from the PRF.ACT.INF, which is the only cell in the distillation corresponding to the Perfect Stem. The reason for this can be plausibly found in the fact that the most frequent derivational suffix in our dataset, namely the inchoative *-sc-*, only surfaces in the Present Stem: in cells based on the Perfect, verbs displaying that suffix either are defective, or they use the same stem as in the base they derive

[13] That the points are not *exactly* aligned is to be expected, as there are nondeterministic aspects of classifier training; hence even the exact same classifier trained on the exact same data will not give rise to exactly the same results from one run to the next.

from (e.g. ARDESCO 'take fire', with Perfect stem *ars-*, as in the base ARDEO). Therefore, when faced with perfective wordforms of similar verbs, speakers can exploit their knowledge of the fact that they belong to the series of inchoative verbs – information that they can be assumed to be able to infer at least partly on semantic grounds – to guess the presence of the segment *-sc-* in wordforms based on the Present Stem, despite the fact that such segment does not surface in the phonological makeup of perfective wordforms. Hence, in such cases the classification based on suffixes supplies information that is not redundant with phonology, thus explaining the sharp increase in accuracy values.[14]

Now that we have observed exceptions to the nonpredictiveness of suffixes, it seems worthwhile to worry about the opposite situation, namely pairs of cells where the addition of family information does not lead to a significant improvement in accuracy. Examining the plots more closely, however, this turns out to be a purely mechanical effect: the very few nonsignificant points in the relevant two graphs (top right and middle left) also correspond to the most extreme values of accuracy, where the simpler model type has accuracy above 0.98: there is basically no remaining accuracy gain possible starting from such a high value. More generally, this is an extreme case of the visible general tendency that, as the accuracy of the simpler model rises, the difference in accuracy between the two models decreases; again this is a mechanical effect of the fact that the higher the baseline accuracy, the lower the possible gain obtained by adding predictors. Hence, while we do document an overall strong predictive power of family information across the paradigm, we do not document differential effects of that predictive power for all combinations of predictor and predictee cells.

To sum up, results obtained with the word-based methodology outlined in §4.2 essentially confirm the findings of §3.2, based on stems and inflection classes: knowing the derivational family to which verbs belong always helps to predict their inflectional behaviour, while suffix information is mostly redundant with phonology. However, this abstractive procedure also allows for interesting observations on more or less systematic differences in the impact of derivational information across pairs of cells.

5 Conclusions

In this paper, we have offered a quantitative assessment of the relation between derivational information and inflectional behavior in Latin. In the first part of the study, we have done so by resorting to the notion of inflection classes. In many cases, our results confirm on a more principled ground the findings of qualitative observations on Latin data. For instance, derivational prefixes have been proven to be uninformative across the board, both for nouns and for verbs. This is hardly surprising, since lexemes derived by means of the

[14] A similar situation holds for cells based on the Third Stem, where, however, the former option of defectiveness is much more frequent, explaining why the impact of suffixes is less relevant.

same prefix can be assigned to any of the available inflection classes. On the other hand, the derivational suffix displayed by lexemes and the derivational family to which they belong both have a role to play, since the former normally assigns inflection class to suffixed lexemes, while the latter is informative on the behaviour of prefixed lexemes. Suffixation has the lion's share in Latin nominal derivational morphology: therefore, it comes as no surprise that derivational suffixes are the most useful predictor of inflection when focusing on the noun dataset. Conversely, prefixation is the most usual way of deriving new verbs in Latin, which explains why the derivational family is most predictive when focusing on the verb dataset. Furthermore, and less expectedly, our results show that family information is partially predictive even for suffixed verbs, although not as dramatically as suffix identity. This shows that our quantitative approach is also capable of uncovering facts that could not be inferred from a qualitative inspection of Latin inflection alone.

In the second part of this study, we have recast the problem in abstractive terms. This word-based methodology has basically confirmed the findings of the stem-based study. The two approaches, however, offer complementary perspectives on the point at issue. The word-based approach is more realistic in that it models a prediction task similar to the one that speakers actually have to face in discourse, i.e. the so-called Paradigm Cell Filling Problem. Furthermore, it allows for finer-grained distinctions about variation in the predictiveness of derivational information across different (pairs of) cells: as a case in point, we showed how suffixal information is relevant when predicting from the perfective subpart of the paradigm, but not when predicting from other cells. The stem-based approach, however, is still useful in that it provides a more holistic view of structural aspects of the system, by framing the question in more familiar terms, relying on the notions of inflection classes – that encapsulate a lot of information on the inflectional behaviour of lexemes – and stems – that cannot be identified in a principled way in our word-based procedure.

This study focused on Latin data for concreteness. It would of course be highly relevant to replicate it on other languages, and hopefully be able to draw typological lessons on the interaction of inflection and derivation. Unfortunately such replications are quite costly, as they rely on the availability of a large machine-readable lexicon[15] with phonemically transcribed full paradigms and detailed documentation of the derivational history of lexemes. Currently we do not have at our disposal such a resource for any other language. Despite this, we cautiously draw some conclusions for morphological theory and the structure of morphological systems.

Our first conclusion concerns the Paradigm Cell Filling Problems and attempts at evaluating it quantitatively. What this paper shows is that, at least for Latin, knowledge of derivational information dramatically increases the predictability of inflected wordforms; this suggests that predictability measures as reported in the literature overestimating the difficulty of the PCFP in a direc-

[15] How large depends on the intricacies of the morphological systems; but in our experience one should aim for thousands of lexemes.

tion that was not previously observed. Of course, it remains to be investigated to what extent the detailed derivational knowledge encoded in our dataset is available to speakers. First, while in some cases derivational relatedness is obvious, it can be more or less opaque. To take an extreme example from our dataset, while PROHIBEO is in the derivational family of HABEO, speakers are unlikely to be aware of that fact: the meaning of PROHIBEO ('restrain' or 'prevent') is not obviously related to the one of the base ('have') and prefix *pro-* ('in front/favor of') of which it is composed; and the form relationship is also less than fully transparent. Hence our method may lead to overestimate the ability of speakers to draw reliable inferences from derivational information. To remedy for this simplification, a promising possibility is to modulate the predictive power of derivational information on the basis of the transparency of the relationship between words, where transparency could be evaluated using computational methods from distributional semantics and phonological distance. Second, even where derivational relatedness is transparent, it remains to be seen whether speakers do indeed rely on that information in concrete prediction tasks. This calls for experimentation with speakers, obviously on a language other than Latin.

Our second conclusion concerns the relationship between the position of elements in the linear order of morphs and their impact on inflectional predictions. While we have seen that prefixes, that are located on the left side of wordforms, never help to predict inflection, right-side elements – namely, suffix and family information – have been shown to be informative, at least to some extent. Indeed, the last element in linear order proves to be predictive both for verbs – where, however, it is redundant with phonological information – and for nouns, while the structural order of application of different derivational procedures is relevant only for nouns – but in that case the operation applied lastly usually happens to coincide with the last element in linear order, since suffixation is the more frequent strategy for deriving nouns in Latin. This state of affairs is not really surprising: it appears to be the norm also in familiar Western Indo-European languages with (mainly) suffixal inflection. This raises the question of whether this situation is due to some historical accident, or is the outcome a more general tendency. To provide an answer, it would be interesting to extend the investigation to a larger sample, including other languages with suffixal inflection that, however, do not belong to the Indo-European family, but also to languages with prefixal inflection, or displaying templatic, root-and-pattern, morphology; a topic that we have to leave for future research.

Funding

This work was partially supported by a public grant overseen by the French National Research Agency (ANR) as part of the "Investissements d'Avenir" program (reference: ANR-10-LABX-0083).

Acknowledgements

A preliminary version of this work was presented at the 19th International Morphology Meeting in Vienna (2020). In addition to the audience at that event, we thank Sacha Beniamine and Matías Guzmán Naranjo for various forms of advice, and two reviewers for Studies in Language for their insightful remarks that we hope led to an improved paper. All remaining errors are our own.

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. Oxford: Oxford University Press.

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

Albright, Adam C. & Bruce P. Hayes. 2003. Rules vs. analogy in english past tenses: a computational/experimental study. *Cognition* 90. 119–161.

Aronoff, Mark. 1994. *Morphology by itself: Stems and inflectional classes*. Cambridge: MIT press.

Beniamine, Sacha. 2018. *Classifications flexionnelles. Étude quantitative des structures de paradigmes*. Université Sorbonne Paris Cité-Université Paris Diderot dissertation.

Beniamine, Sacha, Olivier Bonami & Benoît Sagot. 2017. Inferring inflection classes with description length. *Journal of Language Modelling* 5(3). 465–525.

Blevins, James P. 2016. *Word and paradigm morphology*. Oxford: Oxford University Press.

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

Bonami, Olivier & Gilles Boyé. 2006. Deriving inflectional irregularity. In Stefan Müller (ed.), *Proceedings of the 13th International Conference on Head-Driven Phrase Structure Grammar, Varna*, 361–380. Stanford, CA: CSLI Publications. `http://cslipublications.stanford.edu/HPSG/2006/bonami-boye.pdf`.

Bonami, Olivier & Gilles Boyé. 2014. De formes en thèmes. In Florence Villoing, Sarah Leroy & Sophie David (eds.), *Foisonnements morphologiques. Études en hommage à Françoise Kerleroux*, 17–45. Paris: Presses Universitaires de Paris-Ouest.

Boyé, Gilles & Gauvain Schalchli. 2019. Realistic data and paradigms: the paradigm cell finding problem. *Morphology* 29. 199–248.

Brown, Dunstan, Marina Chumakina & Greville G. Corbett. 2013. *Canonical morphology and syntax*. Oxford: Oxford University Press.

Brown, Dunstan & Andrew Hippisley. 2012. *Network morphology: A defaults-based theory of word structure*. Cambridge: Cambridge University Press.

Budassi, Marco & Eleonora Litta. 2017. In trouble with the rules. Theoretical issues raised by the insertion of-sc-verbs into Word Formation Latin. In *Proceedings of the Workshop on Resources and Tools for Derivational Morphology (DeriMo)*, 15–26.

Bybee, Joan. 1995. Regular morphology and the lexicon. *Language and Cognitive Processes* 10(5). 425–455. https://doi.org/10.1080/01690969508407111. http://dx.doi.org/10.1080/01690969508407111.

Corbett, Greville G. 2005. The canonical approach in typology. In Zygmunt Frajzyngier, Adam Hodges & David S. Rood (eds.), *Linguistic diversity and language theories*, 25–49. Amsterdam: John Benjamins.

Corbett, Greville G. 2009. Canonical inflectional classes. In Fabio Montermini, Gilles Boyé & Jesse Tseng (eds.), *Selected proceedings of the 6th Décembrettes: Morphology in Bordeaux*, 1–11. Somerville: Cascadilla Proceedings Project.

Delatte, Louis, Étienne Evrard, Suzanne Govaerts & Joseph Denooz. 1981. *Dictionnaire fréquentiel et index inverse de la langue latine*. Liège: LASLA.

Dressler, Wolfgang U. 2002. Latin inflection classes. In A. Machtelt Bolkestein, Caroline H. M. Kroon, Harm Pinkster, H. Wim Remmelink & Rodie Risselada (eds.), *Theory and description in Latin linguistics*, 91–110. Leiden-Boston: Brill.

Friedman, Jerome H. 2001. Greedy function approximation: a gradient boosting machine. *Annals of statistics*. 1189–1232.

Guzmán Naranjo, Matías. 2019. *Analogical classification in formal grammar*. Berlin: Language Science Press.

Guzmán Naranjo, Matías. 2020. Analogy, complexity and predictability in the Russian nominal inflection system. *Morphology* 30(3). 219–262.

Guzmán Naranjo, Matías & Olivier Bonami. 2021. Overabundance and inflectional classification: quantitative evidence from Czech. *Glossa* 6. https://doi.org/https://doi.org/10.5334/gjgl.1626.

Litta, Eleonora & Marco Passarotti. 2020. (When) inflection needs derivation: a word formation lexicon for Latin. In Nigel Holmes, Marijke Ottink, Josine Schrickx & Maria Selig (eds.), *Lemmata Linguistica Latina. Volume 1. Words and Sounds*, 224–239. Berlin: De Gruyter.

Malouf, Robert. 2017. Abstractive morphological learning with a recurrent neural network. *Morphology* 27(4). 431–458.

Mason, Llew, Jonathan Baxter, Peter L Bartlett & Marcus R Frean. 2000. Boosting algorithms as gradient descent. In *Advances in neural information processing systems*, 512–518.

McGillivray, Barbara & Adam Kilgarriff. 2013. Tools for historical corpus research, and a corpus of Latin. *New Methods in Historical Corpus Linguistics* (3). 247–257.

Passarotti, Marco, Marco Budassi, Eleonora Litta & Paolo Ruffolo. 2017. The Lemlat 3.0 package for morphological analysis of Latin. In Gerlof Bouma & Yvonne Adesam (eds.), *Proceedings of the NoDaLiDa 2017 Workshop on Processing Historical Language*, 24–31. Gothenburg: Linköping University Electronic Press.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot & E. Duchesnay. 2011. Scikit-learn: machine learning in Python. *Journal of Machine Learning Research* 12. 2825–2830.

Pellegrini, Matteo. 2020. *Predictability in Latin inflection. An entropy-based appproach*. Università di Bergamo-Università di Pavia dissertation.

Pellegrini, Matteo. 2021. Patterns of interpredictability and principal parts in Latin verb paradigms: an entropy-based approach. *Journal of Latin Linguistics* 19(2). 195–221.

Pellegrini, Matteo & Marco Passarotti. 2018. LatInfLexi: an inflected lexicon of Latin verbs. In *Proceedings of the Fifth Italian Conference on Computational Linguistics (CLiC-it 2018)*.

Pierrehumbert, Janet. 2001. Stochastic phonology. *Glot international* 5(6). 195–207.

Plaster, Keith, Maria Polinsky & Boris Harizanov. 2013. Noun classes grow on trees. In Balthasar Bickel, Lenore A. Grenoble, David A. Peterson & Alan Timberlake (eds.), *Language typology and historical contingency: In honor of Johanna Nichols*, 153–170. Amsterdam: John Benjamins.

Sims, Andrea. 2015. *Inflectional defectiveness*. Cambridge: Cambridge University Press.

Stump, Gregory T. 2001. *Inflectional morphology: A theory of paradigm structure*. Cambridge: Cambridge University Press.

Stump, Gregory T. 2015. Inflection classes. In Matthew Baerman (ed.), *The Oxford handbook of inflection*, 113–139. Oxford: Oxford University Press.

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

Appendix

Predictor	Accuracy	95% CI	Outperformed predictors
(1) prefix	0.295	(0.276, 0.314)	_
(2) prefix+suffix	0.311	(0.291, 0.33)	_
(3) outermost	0.320	(0.3, 0.339)	_
(4) baseline	0.438	(0.417, 0.459)	(1)-(3)
(5) suffix	0.452	(0.431, 0.473)	(1)-(3)
(6) last	0.452	(0.431, 0.473)	(1)-(3)
(7) phonology+prefix	0.819	(0.803, 0.835)	(1)-(5)
(8) phonology+outermost	0.830	(0.814, 0.846)	(1)-(5)
(9) prefix+family	0.842	(0.826, 0.857)	(1)-(5)
(10) phonology+last	0.866	(0.851, 0.88)	(1)-(8)
(11) phonology+suffix	0.866	(0.851, 0.88)	(1)-(8)
(12) phonology	0.866	(0.852, 0.881)	(1)-(8)
(13) family	0.867	(0.853, 0.881)	(1)-(8)
(14) prefix+family+suffix	0.874	(0.86, 0.888)	(1)-(9)
(15) family+suffix	0.895	(0.882, 0.908)	(1)-(13)
(16) phonology+prefix+family+suffix	0.935	(0.924, 0.945)	(1)-(15)
(17) phonology+family+suffix	0.938	(0.928, 0.948)	(1)-(15)
(18) phonology+family	0.938	(0.928, 0.948)	(1)-(15)

Table A1: Accuracy of prediction of conjugation classes for various combinations of predictors: only prefixed verbs

Predictor	Accuracy	95% CI	Outperformed predictors
(1) prefix	0.575	(0.515, 0.634)	_
(2) baseline	0.653	(0.596, 0.71)	_
(3) prefix+family	0.713	(0.659, 0.767)	(1)
(4) family	0.769	(0.718, 0.819)	(1)-(2)
(5) outermost	0.847	(0.804, 0.89)	(1)-(3)
(6) family+suffix	0.993	(0.982, 1.003)	(1)-(5)
(7) phonology+family	0.993	(0.982, 1.003)	(1)-(5)
(8) prefix+family+suffix	0.993	(0.982, 1.003)	(1)-(5)
(9) phonology+prefix+family+suffix	0.996	(0.989, 1.004)	(1)-(5)
(10) phonology+family+suffix	0.996	(0.989, 1.004)	(1)-(5)
(11) phonology+outermost	1.0	(1.0. 1.0)	(1)-(5)
(12) phonology+prefix	1.0	(1.0. 1.0)	(1)-(5)
(13) phonology+suffix	1.0	(1.0. 1.0)	(1)-(5)
(14) suffix	1.0	(1.0. 1.0)	(1)-(5)
(15) last	1.0	(1.0. 1.0)	(1)-(5)
(16) prefix+suffix	1.0	(1.0. 1.0)	(1)-(5)
(17) phonology+last	1.0	(1.0. 1.0)	(1)-(5)
(18) phonology	1.0	(1.0. 1.0)	(1)-(5)

Table A2: Accuracy of prediction of conjugation classes for various combinations of predictors: only suffixed verbs

Predictor	Accuracy	95% CI	Outperformed predictors
(1) prefix	0.234	(0.201, 0.266)	_
(2) prefix+family	0.344	(0.307, 0.38)	(1)
(3) baseline	0.472	(0.434, 0.51)	(1)-(2)
(4) family	0.505	(0.467, 0.544)	(1)-(2)
(5) prefix+gender	0.602	(0.564, 0.639)	(1)-(4)
(6) prefix+family+gender	0.722	(0.688, 0.756)	(1)-(5)
(7) prefix+suffix	0.728	(0.694, 0.762)	(1)-(5)
(8) gender	0.771	(0.739, 0.803)	(1)-(5)
(9) family+gender	0.777	(0.745, 0.809)	(1)-(5)
(10) phonology+prefix	0.792	(0.761, 0.823)	(1)-(6)
(11) phonology	0.794	(0.763, 0.825)	(1)-(7)
(12) suffix	0.806	(0.776, 0.836)	(1)-(7)
(13) last	0.806	(0.776, 0.836)	(1)-(7)
(14) phonology+family	0.821	(0.792, 0.851)	(1)-(7)
(15) prefix+family+suffix	0.824	(0.795, 0.854)	(1)-(7)
(16) outermost	0.838	(0.81, 0.866)	(1)-(9)
(17) family+suffix	0.841	(0.813, 0.869)	(1)-(9)
(18) phonology+suffix	0.843	(0.815, 0.871)	(1)-(9)
(19) phonology+last	0.843	(0.815, 0.871)	(1)-(9)
(20) phonology+outermost	0.853	(0.826, 0.881)	(1)-(11)
(21) phonology+family+suffix	0.858	(0.831, 0.885)	(1)-(11)
(22) phonology+prefix+family+suffix	0.860	(0.833, 0.886)	(1)-(11)
(23) prefix+suffix+gender	0.922	(0.902, 0.943)	(1)-(22)
(24) outermost+gender	0.934	(0.915, 0.953)	(1)-(22)
(25) last+gender	0.942	(0.924, 0.96)	(1)-(22)
(26) suffix+gender	0.942	(0.924, 0.96)	(1)-(22)
(27) phonology+last+gender	0.942	(0.924, 0.96)	(1)-(22)
(28) phonology+suffix+gender	0.942	(0.924, 0.96)	(1)-(22)
(29) prefix+family+suffix+gender	0.948	(0.931, 0.965)	(1)-(22)
(30) family+suffix+gender	0.948	(0.931, 0.965)	(1)-(22)
(31) phonology+gender	0.950	(0.933, 0.966)	(1)-(22)
(32) phonology+family+suffix+gender	0.950	(0.933, 0.966)	(1)-(22)
(33) phonology+prefix+gender	0.953	(0.936, 0.969)	(1)-(22)
(34) phonology+prefix+family+suffix+gender	0.953	(0.936, 0.969)	(1)-(22)
(35) phonology+family+gender	0.954	(0.938, 0.97)	(1)-(22)
(36) phonology+outermost+gender	0.956	(0.94, 0.971)	(1)-(22)

Table A3: Accuracy of prediction of declension classes for various combinations of predictors: only prefixed nouns

Predictors	Accuracy	95% CI	Outperformed predictors
(1) prefix+family	0.496	(0.469, 0.523)	_
(2) prefix	0.527	(0.5, 0.554)	_
(3) family	0.575	(0.548, 0.602)	(1)
(4) baseline	0.613	(0.586, 0.639)	(1)-(2)
(5) prefix+family+gender	0.668	(0.643, 0.694)	(1)-(4)
(6) prefix+gender	0.683	(0.658, 0.708)	(1)-(4)
(7) family+gender	0.714	(0.689, 0.738)	(1)-(4)
(8) gender	0.727	(0.702, 0.751)	(1)-(5)
(9) phonology+family	0.866	(0.848, 0.885)	(1)-(8)
(10) phonology	0.875	(0.857, 0.893)	(1)-(8)
(11) phonology+prefix	0.878	(0.861, 0.896)	(1)-(8)
(12) prefix+suffix	0.935	(0.921, 0.948)	(1)-(11)
(13) outermost	0.935	(0.921, 0.948)	(1)-(11)
(14) prefix+family+suffix	0.935	(0.922, 0.949)	(1)-(11)
(15) suffix	0.937	(0.924, 0.95)	(1)-(11)
(16) last	0.937	(0.924, 0.95)	(1)-(11)
(17) family+suffix	0.939	(0.926, 0.952)	(1)-(11)
(18) phonology+prefix+family+suffix	0.945	(0.932, 0.957)	(1)-(11)
(19) phonology+family+suffix	0.946	(0.934, 0.958)	(1)-(11)
(20) phonology+outermost	0.951	(0.939, 0.963)	(1)-(11)
(21) phonology+suffix	0.952	(0.941, 0.964)	(1)-(11)
(22) phonology+last	0.952	(0.941, 0.964)	(1)-(11)
(23) phonology+family+gender	0.965	(0.955, 0.975)	(1)-(17)
(24) phonology+prefix+gender	0.971	(0.962, 0.98)	(1)-(19)
(25) phonology+gender	0.972	(0.963, 0.981)	(1)-(19)
(26) family+suffix+gender	0.986	(0.98, 0.993)	(1)-(23)
(27) outermost+gender	0.986	(0.98, 0.993)	(1)-(23)
(28) prefix+family+suffix+gender	0.988	(0.982, 0.994)	(1)-(25)
(29) phonology+family+suffix+gender	0.988	(0.982, 0.994)	(1)-(25)
(30) last+gender	0.988	(0.982, 0.994)	(1)-(25)
(31) suffix+gender	0.988	(0.982, 0.994)	(1)-(25)
(32) phonology+prefix+family+suffix+gender	0.988	(0.983, 0.994)	(1)-(25)
(33) phonology+outermost+gender	0.989	(0.984, 0.995)	(1)-(25)
(34) prefix+suffix+gender	0.989	(0.984, 0.995)	(1)-(25)
(35) phonology+suffix+gender	0.990	(0.985, 0.995)	(1)-(25)
(36) phonology+last+gender	0.990	(0.985, 0.995)	(1)-(25)

Table A4: Accuracy of prediction of declension classes for various combinations of predictors: only suffixed nouns

Postal and email addresses of authors

Olivier Bonami
LLF, CNRS – UMR 7110
Université de Paris
Case 7031 – 5, rue Thomas Mann,
75205 Paris cedex 13
olivier.bonami@linguist.univ-paris-diderot.fr

Matteo Pellegrini
CIRCSE Research Centre
Università Cattolica del Sacro Cuore
Largo Gemelli, 1,
20123 Milan, Italy
matteo.pellegrini@unicatt.it

Conj.	Sample lexeme	Stem	PRS.ACT.INF	PRS.ACT.IND.3PL	Frequency
1st	LAUDO ‘praise’	/lawd/	*laudāre*	*laudant*	1,412
2nd	MONEO ‘warn’	/mon/	*monēre*	*monent*	313
3rd	LEGO ‘read’	/leg/	*legere*	*legunt*	1,204
4th	VENIO ‘come’	/wen/	*venīre*	*veniunt*	190
mix.	CAPIO ‘take’	/kap/	*capere*	*capiunt*	152
irr.	SUM ‘be’	/es/	*esse*	*sunt*	77

Table 1: Type frequency of conjugation classes in the LatInfLexi dataset

Lexeme	Prefix	Suffix	Family	Last morph	Outermost operation
CREO 'create'	—	—	CREO_V	CREO_V	CREO_V
CRESCO 'grow'	—	*-sc-*	CREO_V	*-sc-*	*-sc-*
RECREO 'recreate'	*re-*	—	CREO_V	CREO_V	*re-*
CONDOLESCO 'suffer much'	*con-*	*-sc-*	DOLEO_V	*-sc-*	*-sc-*
INARDESCO 'take fire'	*in-*	*-sc-*	ARDEO_V	*-sc-*	*in-*
CORONO 'crown'	—	—	CORONA_N	CORONA_N	conversion (N→V)
DIGNO 'worthy'	—	—	DIGNUS_A	DIGNUS_A	conversion (A→V)
RAPTO 'seize'	—	—	RAPIO_V	RAPIO_V	conversion (V→V)
DEPILO 'pull out the hair'	*de-*	—	PILUS_N	PILUS_N	*de-* (N→V)
DEPRAVO 'distort'	*de-*	—	PRAVUS_A	PRAVUS_A	*de-* (A→V)
BENEDICO 'commend'	—	—	DICO_V	DICO_V	compounding

Table 2: Sample annotation of derivational information on verbs

Decl.	Sample lexeme	Stem	NOM.SG	GEN.SG	Frequency
1st	ROSA 'rose'	/ros/	*rosa*	*rosae*	758
2nd	LUPUS 'wolf'	/lup/	*lupus*	*lupī*	833
3rd	URBS 'city'	/urb/	*urbs*	*urbis*	1,102
4th	ARCUS 'bow'	/ark/	*arcus*	*arcūs*	272
5th	SPES 'hope'	/sp/	*spes*	*speī*	17

Table 3: Type frequency of nominal declension classes in our dataset

Lexeme	Gender	Prefix	Suffix	Family	Last morph	Outermost operation
ALIMENTUM 'nourishment'	N	_	*-ment-*	ALO$_{\mathrm{V}}$	*-ment-*	*-ment-*
PROAVUS 'greatgrandfather'	M	*pro-*	_	AVUS$_{\mathrm{N}}$	AVUS$_{\mathrm{N}}$	*pro-*
DEA 'goddess'	F	_	_	DEUS$_{\mathrm{N}}$	DEUS$_{\mathrm{N}}$	conversion (N→N)
COLLEGA 'colleague'	M	*con-*	_	LEGO$_{\mathrm{V}}$	LEGO$_{\mathrm{V}}$	*con-* (V→N)
AGRICOLA 'farmer'	M	_	_	COLO$_{\mathrm{V}}$	COLO$_{\mathrm{V}}$	compounding

Table 4: Sample annotation of derivational information on nouns

Predictor	Accuracy	95% CI	Outperformed predictors
(1) prefix	0.335	(0.319, 0.351)	_
(2) outermost	0.367	(0.35, 0.383)	_
(3) prefix+suffix	0.373	(0.357, 0.39)	(1)
(4) baseline	0.431	(0.414, 0.448)	(1)-(3)
(5) last	0.485	(0.468, 0.502)	(1)-(4)
(6) suffix	0.485	(0.468, 0.502)	(1)-(4)
(7) phonology+prefix	0.657	(0.641, 0.674)	(1)-(6)
(8) phonology+outermost	0.816	(0.803, 0.829)	(1)-(7)
(9) prefix+family	0.819	(0.806, 0.832)	(1)-(7)
(10) family	0.821	(0.808, 0.834)	(1)-(7)
(11) phonology	0.823	(0.809, 0.836)	(1)-(7)
(12) phonology+last	0.824	(0.811, 0.837)	(1)-(7)
(13) phonology+suffix	0.824	(0.811, 0.837)	(1)-(7)
(14) prefix+family+suffix	0.857	(0.845, 0.869)	(1)-(13)
(15) family+suffix	0.866	(0.854, 0.877)	(1)-(13)
(16) phonology+family	0.920	(0.911, 0.93)	(1)-(15)
(17) phonology+family+suffix	0.923	(0.914, 0.932)	(1)-(15)
(18) phonology+prefix+family+suffix	0.925	(0.916, 0.934)	(1)-(15)

Table 5: Accuracy of prediction of conjugation classes for various combinations of predictors

	Conjugation					
Prefix	1st	2nd	3rd	4th	mix.	irr.
e(x)-	116	18	101	23	10	3
con-	104	19	107	12	13	4
in- (entering)	88	14	104	13	9	4
ad-	78	21	78	8	9	3
de-	78	16	80	8	11	4
re-	75	19	82	10	7	3
per-	50	17	58	5	7	3
ob-	38	6	45	7	5	4
dis-	29	7	48	6	7	2
prae-	26	16	36	8	6	3
other prefixes	102	38	211	22	37	24

Table 6: Distribution across conjugations of verbs displaying the same derivational prefixes

Outermost operation	Frequency
Prefixation	2,101
None (simple lexemes)	586
Conversion	407
Suffixation	165
Compounding	62
Parasynthesis	27

Table 7: Type frequency of outermost derivational operation in the verbs of our dataset

Suffix	Conjugation 1st	2nd	3rd	4th	mix.
-sc-	0	0	172	0	0
-it- (frequentative)	69	0	0	0	0
-it- (factitive)	7	0	0	0	0
-cin-	4	0	0	0	0
-ess-	0	0	4	0	0
-ill-	3	0	0	0	0
-ig-	3	0	0	0	0
-er-	2	0	0	0	0
-ic-	2	0	0	0	0
-uri-	0	0	0	2	0
-il-	1	0	0	0	0

Table 8: Distribution across conjugations of verbs displaying the same derivational suffixes

Predictor	Accuracy	95% CI	Outperformed predictors
(1) prefix+family	0.309	(0.292, 0.325)	_
(2) prefix	0.332	(0.315, 0.349)	_
(3) family	0.347	(0.33, 0.364)	(1)
(4) baseline	0.370	(0.352, 0.387)	(1)-(2)
(5) gender	0.522	(0.505, 0.54)	(1)-(4)
(6) prefix+family+gender	0.572	(0.554, 0.59)	(1)-(5)
(7) family+gender	0.578	(0.56, 0.596)	(1)-(5)
(8) prefix+gender	0.593	(0.575, 0.611)	(1)-(5)
(9) suffix	0.610	(0.593, 0.628)	(1)-(6)
(10) last	0.610	(0.593, 0.628)	(1)-(6)
(11) prefix+suffix	0.618	(0.6, 0.635)	(1)-(7)
(12) prefix+family+suffix	0.631	(0.614, 0.648)	(1)-(8)
(13) family+suffix	0.643	(0.626, 0.66)	(1)-(8)
(14) phonology+prefix	0.657	(0.64, 0.674)	(1)-(11)
(15) phonology+family	0.659	(0.642, 0.676)	(1)-(11)
(16) phonology	0.664	(0.647, 0.681)	(1)-(11)
(17) outermost	0.667	(0.65, 0.684)	(1)-(12)
(18) phonology+last	0.695	(0.679, 0.712)	(1)-(15)
(19) phonology+suffix	0.695	(0.679, 0.712)	(1)-(15)
(20) phonology+prefix+family+suffix	0.699	(0.683, 0.716)	(1)-(16)
(21) phonology+family+suffix	0.701	(0.684, 0.717)	(1)-(16)
(22) phonology+outermost	0.717	(0.7, 0.733)	(1)-(17)
(23) last+gender	0.792	(0.778, 0.807)	(1)-(22)
(24) suffix+gender	0.792	(0.778, 0.807)	(1)-(22)
(25) prefix+family+suffix+gender	0.819	(0.805, 0.833)	(1)-(22)
(26) prefix+suffix+gender	0.825	(0.811, 0.838)	(1)-(24)
(27) family+suffix+gender	0.831	(0.817, 0.844)	(1)-(24)
(28) outermost+gender	0.865	(0.853, 0.877)	(1)-(27)
(29) phonology+prefix+gender	0.879	(0.867, 0.891)	(1)-(27)
(30) phonology+gender	0.881	(0.87, 0.893)	(1)-(27)
(31) phonology+family+gender	0.884	(0.873, 0.896)	(1)-(27)
(32) phonology+suffix+gender	0.886	(0.875, 0.898)	(1)-(27)
(33) phonology+last+gender	0.886	(0.875, 0.898)	(1)-(27)
(34) phonology+family+suffix+gender	0.895	(0.884, 0.906)	(1)-(28)
(35) phonology+prefix+family+suffix+gender	0.896	(0.885, 0.907)	(1)-(28)
(36) phonology+outermost+gender	0.900	(0.89, 0.911)	(1)-(28)

Table 9: Accuracy of prediction of nominal declension classes for various combinations of predictors

Declension	M	F	N
1st	24	734	0
2nd	327	9	497
3rd	298	699	105
4th	266	5	1
5th	0	17	0

Table 10: Distribution of nominal declensions across genders

	Declension				
Prefix	1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}
con-	12	25	67	26	0
e(x)-	6	11	30	10	0
ad-	5	14	22	15	0
de-	9	9	28	8	0
in- (negation)	22	6	17	3	0
other prefixes	40	67	145	57	1

Table 11: Distribution across declensions of nouns displaying the same derivational prefixes

	Declension				
Suffix	1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}
-(t)ion-	0	0	390	0	0
-tat-	0	0	157	0	0
-(t)or-	0	0	124	0	0
-i(a)-	119	0	0	0	0
-i(u)-	0	85	0	0	0
-cul-	11	21	0	0	0
-ll-	10	10	0	0	0
other suffixes	413	115	125	6	12

Table 12: Distribution across declensions of nouns displaying the same derivational suffixes

Outermost operation	Frequency
Suffixation	1,291
None (simple lexemes)	968
Conversion	653
Prefixation	28
Compounding	24
Parasynthesis	18

Table 13: Type frequency of outermost derivational operation in the nouns of our dataset

Conj.	Sample lexeme	PRS.ACT.INF	PRF.ACT.IND.1SG	SUP.ACC
1st	LAUDO 'praise'	*laudāre*	*laudāvī*	*laudātum*
2nd	MONEO 'warn'	*monēre*	*monuī*	*monitum*
3rd	CURRO 'run'	*currere*	*cucurrī*	*cursum*
4th	VENIO 'come'	*venīre*	*vēnī*	*ventum*
mix.	CAPIO 'take'	*capere*	*cēpī*	*captum*
irr.	FERO 'bring'	*ferre*	*tulī*	*lātum*

Table 14: Allomorphy in wordforms based on the Perfect and Third Stem

Lexeme	Family	Conj.	PRS.ACT.INF	PRF.ACT.IND.1SG
FACIO 'make'	FACIO	3^{rd}	*facere*	*fēcī*
CALEFACIO 'make warm'	FACIO	3^{rd}	*calefacere*	*calefēcī*
INFICIO 'put into'	FACIO	3^{rd}	*inficere*	*infēcī*

Table 15: Variation in the inflectional behaviour of verbs that derive from FACIO

Lexeme	PRS.ACT.IND.1SG Form	Type	PRS.ACT.INF Form	Type
LAUDO	lawdoː	-oː	lawdaːre	-aːre
MONEO	moneoː	-eoː	moneːre	-eːre
CURRO	kurroː	-oː	kurrere	-ere
UENIO	wenioː	-ioː	weniːre	-iːre
CAPIO	kapioː	-ioː	kapere	-ere

Table 16: Toy example for conditional entropy computations following the methodology of Ackerman, Blevins & Malouf (2009)

Lexeme	PRS.ACT.IND.1SG		PRS.ACT.INF	
	Form	Type	Form	Type
LAUDO	lawdoː	-oː	lawdaːre	-aːre
MONEO	moneoː	-eoː	moneːre	-eːre
CURRO	kurroː	-oː	kurrere	-ere
UENIO	wenioː	-ioː	weniːre	-iːre
CAPIO	kapioː	-ioː	kapere	-ere
COMMEO	kommeoː	-oː or -eoː?	kommeaːre	-aːre
GLACIO	glakioː	-oː or -ioː?	glakiaːre	-aːre
IMMEIO	immeːioː	-oː or -ioː?	immeːiere	-ere

Table 17: Toy example of uncertainty applying the methodology of Ackerman, Blevins & Malouf (2009)

Lexeme	PRS.ACT.IND 1SG form	PRS.ACT.INF form	Alternation	Class of PRS. ACT.IND.1SG
LAUDO	lawdoː	lawdaːre	$a_1 : X$oː $\sim X$aːre	$\{a_1, a_3\}$
MONEO	moneoː	moneːre	$a_2 : X$eoː $\sim X$eːre	$\{a_1, a_2, a_3\}$
CURRO	kurroː	kurrere	$a_3 : X$oː $\sim X$ere	$\{a_1, a_3\}$
VENIO	wenioː	weniːre	$a_4 : X$ioː $\sim X$iːre	$\{a_1, a_3, a_4, a_5\}$
CAPIO	kapioː	kapere	$a_5 : X$ioː $\sim X$ere	$\{a_1, a_3, a_4, a_5\}$
COMMEO	kommeoː	kommeaːre	$a_1 : X$oː $\sim X$aːre	$\{a_1, a_2, a_3\}$
GLACIO	glakioː	glakiaːre	$a_1 : X$oː $\sim X$aːre	$\{a_1, a_3, a_4, a_5\}$
IMMEIO	immeːioː	immeːiere	$a_3 : X$oː $\sim X$ere	$\{a_1, a_3, a_4, a_5\}$

Table 18: Toy example for conditional entropy computations following the methodology of Bonami & Boyé (2014)

Predictors	Accuracy
shape+last_morph	0.968
shape+family+suffix	0.968
shape+family	0.962
shape+suffix	0.875
shape	0.868

Table 19: Word-based classifiers: average results

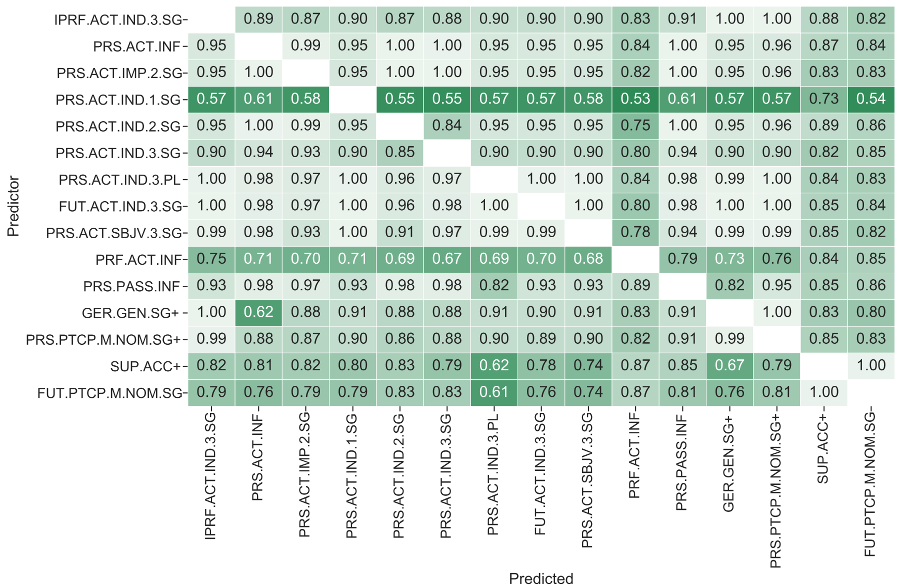

Figure 1: Accuracy of the 'vanilla' classifiers: prediction of alternation between two cells on the basis of the shape of the predictor cell. Darker color corresponds to lower accuracy.

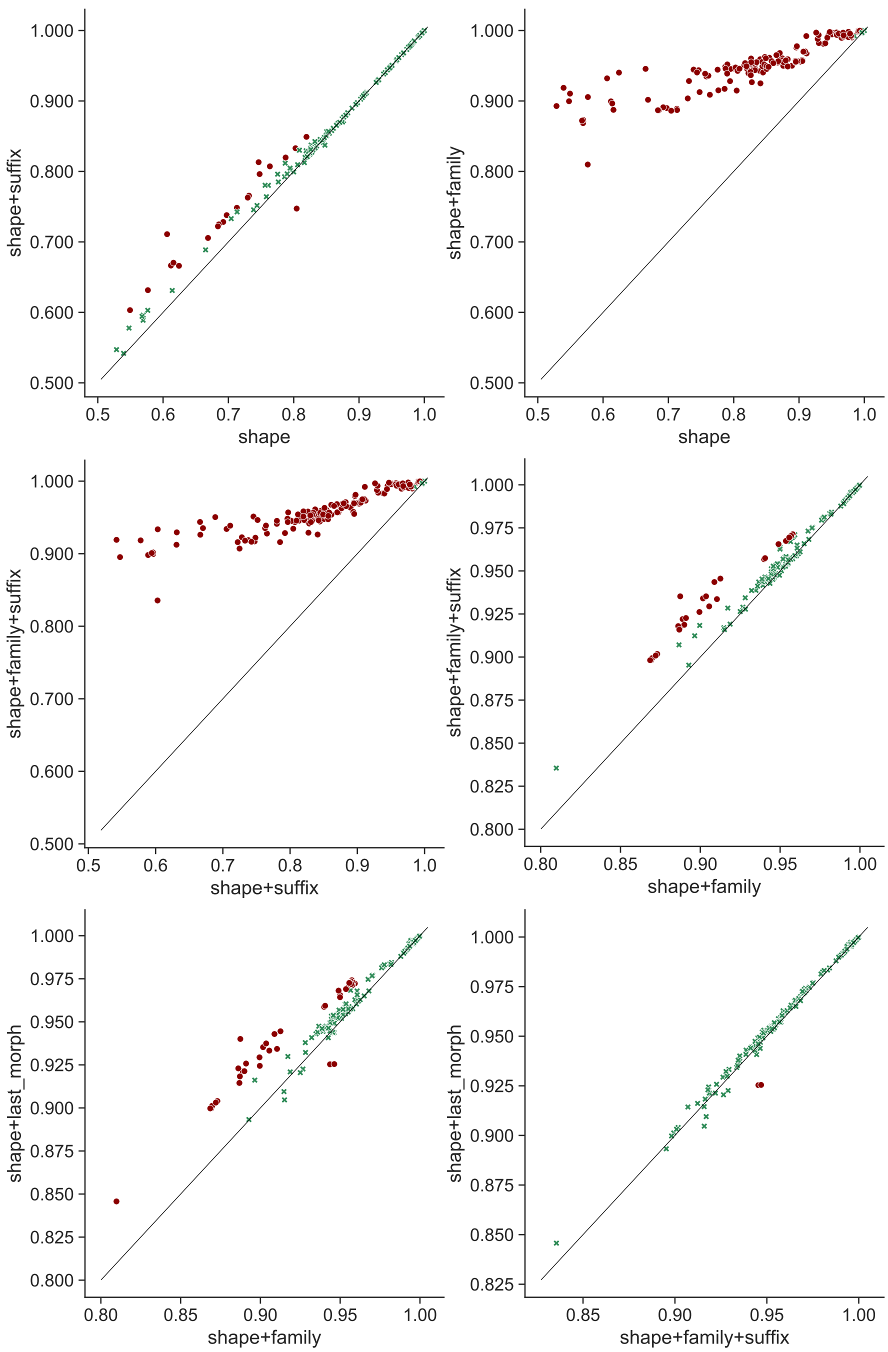

Figure 2: Comparisons of the accuracy of classifier types.