The role of conditional entropy in determining the base for leveling in Old High German inflectional paradigm

JUNYU RUAN
The Ohio State University
ruan.118@osu.edu

ABSTRACT

In the development from Old High German (OHG) to Modern High German (NHG), analogical changes occurred in the inflectional paradigm, in that the older past singular stem and past plural stem are leveled to a single past stem in the later language. However, different bases for leveling are attested in different inflectional classes. This paper takes an information-theoretic approach to the explanation of the multi-directional leveling found in the history of German. We assume that speakers use the most informative cell in the paradigm as the base for deriving other forms in the paradigm, and when analogical changes happen in the language, other forms in the paradigm might be leveled to the base but the base form would not be affected by leveling. Under this assumption, we studied the inflectional paradigms of verbs collected from an OHG corpus, and measured informativeness of a cell by its conditional entropy in predicting other cells in the paradigm. Results show that, in classes where the conditional entropy analysis makes strong predictions, the predictions are consistent with the attested historical changes, though other factors could also play a role in determining the base in classes where no strong predictions can be made based on informativeness.

KEYWORDS
leveling, German, paradigm, predictiveness, analogy

RÉSUMÉ

Des changements d'ordre flexionnel ont eu lieu dans l'évolution de le vieux haut allemand (VHA) vers le haut allemand moderne (HAM). Ainsi, les versions antérieures des radicaux du passé au singulier et au pluriel ont été nivelées vers un radical unique du passé. Différentes bases à partir desquelles ce nivellement s'est effectué ont par contre été observées selon les différentes classes flexionnelles. Le présent article adopte une approche théorique informative pour expliquer ce nivellement multidirectionnel observé dans l'histoire de l'allemand. Il est assumé que les locuteurs utilisent la plupart des cellules informatives du paradigme comme base à partir desquelles les autres formes sont dérivées. Dans cette approche, lorsqu'un changement analogue se produit dans la langue, les autres formes du paradigme peuvent être nivelées, mais la base n'est pas affectée par ce nivellement. À partir de cette prémisse, nous avons étudié les paradigmes flexionnels de verbes provenant d'un corpus d'OHG et mesuré le caractère informatif d'une cellule en fonction de son entropie conditionnelle dans la prévision des autres cellules du même paradigme. Les résultats obtenus démontrent que lorsque l'analyse de l'entropie conditionnelle génère de fortes prédictions, ces prédictions sont cohérentes avec les changements historiques rapportés. D'autres facteurs peuvent cependant jouer un rôle dans la définition des bases pour les classes ou aucune prédiction forte ne peut être émise en fonction du caractère informatif.

MOTS-CLÉS
nivellement, allemande, paradigme, prévisibilité, analogie

1. Introduction

Paradigmatic leveling refers to the diachronic change in which morpho-phonemic alternations in a paradigm are partially or completely eliminated (Hock 1991: 168). It has been widely attested cross-linguistically, and various approaches have been used in an effort to identify the base for leveling in paradigms. For example, Tiersma (1982) suggests that local markedness, which is relevant to related frequency of different paradigmatic forms of lexical items in a language, can play a role in the identification of the base for leveling. Besides, Albright (2002a, 2002b) considers the predictiveness of a single surface form relative to other forms to be crucial for the selection of the base for leveling, as speakers might tend to use the most predictive form as the base for deriving the remaining ones in the paradigm in language acquisition. He suggests that the predictiveness of a form can be indicated by its preservation of morphological alternation and accuracy in predicting other forms. For example, compare the inflected forms of Yiddish libən [to love] and zipən [to sift]: lib (1sg) ~ lipst (2sg) ~ lipt (3sg) vs. zip (1sg) ~ zipst (2sg) ~ zipt (3sg). Since the stem-final /b/ is assimilated to the personal suffix in terms of voicing, the first person singular form is more reliable in deriving other forms in the paradigm as the base than the second and the third person singular; it is thus considered more predictive (Albright 2002a). Two underlying hypotheses are proposed for his model:

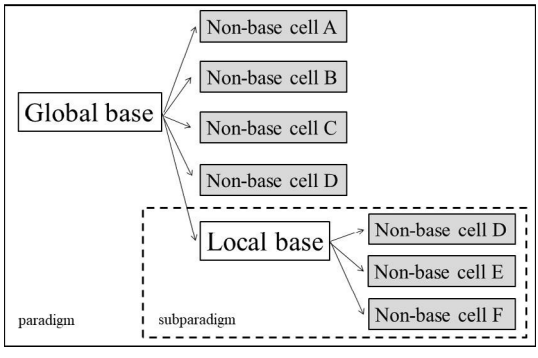
> The first hypothesis is that learners are limited to selecting a single form as the base, and that the base form must be a surface form from somewhere within the paradigm. Furthermore, the choice of base is global, meaning that the same part of the paradigm must serve as the base for all lexical items. ... The second hypothesis is that learners select the base form that is maximally informative, in the sense that it preserves the most contrasts, and permits accurate and productive generation of as many forms of as many words as possible. (Albright 2002b: 129)

There are often cases where certain forms in the paradigm have relatively high mutual predictiveness. For example, the stem of the Spanish present indicative 1sg form *conosko* is consistent with that of present subjunctive forms (e.g. *conoska*, *conoskas*), but differs from the stem of other forms (e.g. *conoses*, *conose*) (Albright 2002b: 119).[1] To cope with these, the model also supposes that subparadigms and local bases might exist, so that forms that

have low predictability from the global base could be formed with certain local bases with higher predictability, which, then, could be built from the global base. The relationship among global base, local base, and non-base forms is presented in the following figure:

Figure 1

The organization of a paradigm with local base and subparadigm structure proposed in Albright (2002b)



Note: local or global bases are in white, and non-base cells are in gray

The crucial point, however, is that the relations between the base form and other forms in the paradigm are asymmetric. Given these hypotheses about the synchronic organization of paradigms, Albright (2002b: 120) suggests that, diachronically, only non-base forms could be leveled by analogical changes, and hence re-formed by applying productive rules to base forms. The base form, however, along with the phonological contrasts it preserves, would not be affected by leveling.

It is questionable, however, whether there has to be only one single global base for paradigmatic leveling across all lexical items. As can be seen below, different cells in the paradigm could serve as the base in the merger of past[2] indicative stems from Old High German (OHG) to Modern High German (NHG) for different verb inflectional class. If feeding data are not pooled across all verbs, but consist of only one inflectional class, it is possible that different cells in the paradigm could be selected as the base for different classes, which might match the attested diachronic change better.

That being said, the ways of measuring the "informativeness" of a paradigm cell are not unique as well. Synchronic informativeness of a given cell in a paradigm can be quantified in a different way other than Albright's

definition as preservation of phonological contrasts and accuracy in predicting other forms. Another possible interpretation of informativeness is the uncertainty or predictability of a cell in the paradigm when another cell is known. This uncertainty can be measured by information entropy (Shannon 1948). The notion of entropy has been introduced as a quantification method for the study of the synchronic complexity of paradigms (Ackerman, Blevins, 2009; Ackerman and Malouf 2013). Similar to Albright's (2002b) approach, the theoretical work in Ackerman and Malouf (2013) emphasizes speakers' acquisition of paradigms and their productive use of rules. In other words, for non-suppletive paradigms, speakers might choose a small set of forms in the paradigm as principal parts and predict the remaining forms from those parts. Under this assumption, each cell in the paradigm is considered a random variable with a set of possible forms, and conditional entropy serves as an indicator of the implicative relations among cells within a paradigm. The lower the conditional entropy of Cell B given Cell A, the more certain Cell B is when Cell A is known, and, thus, the more accurately speakers can predict the form occurring in Cell B. However, different from Albright (2002b), the focus in Ackerman and Malouf (2013) is not the comparison of the predictiveness of individual cells or the selection of a base for analogical changes, but the average conditional entropy of all cells in a paradigm as a measure of that paradigm's complexity as a whole. Also, Ackerman and Malouf (2013) do not make any specific claims about tendencies in diachronic changes.

This paper will first give an overview of how the distinctive OHG past indicative stems are leveled into a single one in NHG in each strong verb class, and then introduce the method of calculating conditional entropy applied in this study. Finally, values for conditional entropy based on data for each strong verb class where leveling occurred in history (i.e. Strong Classes I, II, III & V) will be presented and compared with the values based on data across all verbs, in order to better explain the distinctive base for paradigmatic leveling in the development of NHG from OHG.

2. Analogical changes from Old High German to Modern High German

We are interested here in the patterns of leveling attested in the history of German. The majority of OHG verbs are traditionally divided into strong and weak verbs. Strong verbs build their tense stems through vowel alternation, whereas weak verbs do so through affixation. In the transition from OHG to NHG, unusual patterns of analogical change are attested. In

OHG, on the one hand, all inflected forms of strong verbs in the paradigm can maximally have five different stems: (a) the infinitive stem, deriving the infinitive and present forms other than indicative singular and imperative; (b) the pres.1.sg. stem, deriving indicative present singular forms and the imperative 2nd singular; (c) the past.1.sg. stem, deriving indicative past 1st and 3rd singular; (d) the past.1.pl. stem, deriving other past forms; (e) the past participle stem, deriving the past participle only (Braune 2012: 235). In NHG, on the other hand, all past indicative forms share the same stem.[3] The change is partially driven by regular sound change and partially by leveling, as shown in the following examples. Interestingly, for inflectional classes in which leveling has taken place, different stems are chosen as the base for leveling in different strong classes (Hill 2013).

Table 1
Classes VI & VII: preservation of OHG stems

Class	OHG		NHG		Gloss
	past.1.sg.	past.1.pl.	past.1.sg.	past.1.pl.	
VI	sluog	sluogum	schlug	schlugen	'struck'
VII	hialt	hialtum	hielt	hielten	'held'

There are several subclasses in Class VII, but the stem vowels in their past indicative forms are the same. Past forms in Classes VI and VII originally share the same stem vowel in the OHG paradigm. Hence, no leveling occurred from OHG to NHG.

Table 2
Class IV: merger of past indicative forms driven by regular sound change

Class	OHG		NHG		Gloss
	past.1.sg.	past.1.pl.	past.1.sg.	past.1.pl.	
IV	nam	nâmum	nam	namen	'took'

Class IV verbs are characterized by their root-final liquids or nasals, and the two past stems in Class IV differ only in the stem vowel's length. As OHG short vowels in monosyllabic words were regularly lengthened before a sonorant coda in NHG (Mettke 1967: 70; Paul 1929: 22), the vowel in the past.1.sg. stem became long, and merged with other past indicative stems as a result. Note that in NHG orthography, <a> in nam and namen represents a long vowel.

Table 3
Classes I & V: leveling of the past 1st/3rd singular stem to other past forms

Class	OHG			NHG			Gloss
	past.1.sg.	past.1.pl.	past.part.	past.1.sg.	past.1.pl.	past.part.	
Ia	steig	stigum	gistigan	stieg	stiegen	gestiegen	'climbed'
Ib	zêh	zigum	gizigan	zieh	ziehen	geziehen	'accused'
V	maz	mâzum	gimëzzan	maß	maßen	gemessen	'measured'

The lengthening rule mentioned above did not apply for Class V verbs like maz. Nevertheless, the OHG past.1.sg. stem is leveled by the past.1.pl. stem in Classes I and V, so that the vowels of NHG past indicative stems stieg, zieh (for Class I) and maß (for Class V) originate from OHG past plural stems. Note that in NHG orthography, <a> in maß and maßen also represents a long vowel.

Class I has two subclasses, with distinction in the stem vowel of the past.1.sg. form. Verbs in Subclass Ib usually show alternation in the stem-final consonant between past singular and plural forms, conventionally termed Grammatischer Wechsel, which is eliminated in NHG. However, its direction of leveling is sometimes inconsistent with the direction of leveling of stem vowels; it will not be further discussed in this paper for the sake of simplicity.

Table 4
Class III: spreading of the past 1st/3rd singular stem into other past indicative forms

Class	OHG			NHG			Gloss
	past.1.sg.	past.1.pl.	past.part.	past.1.sg.	past.1.pl.	past.part.	
IIIa	trank	trunkum	gitrunkan	trank	tranken	getrunken	'drank'
IIIb	warf	wurfum	giworfan	warf	warfen	geworfen	'threw'

In contrast to Classes I & V, in Class III, it was the OHG past plural stem that was leveled in NHG by analogical change. However, all three classes have a feature in common: the past participle was independent of the leveling occurring among past forms. It was not affected by the leveling. Nor did it serve as a base for analogical change, unless the past plural forms happened to have the same stem, as in Class I.

Table 5
Class II: Adoption of the past participle stem

Class	OHG			NHG			Gloss
	past.1.sg.	past.1.pl.	past.part.	past.1.sg.	past.1.pl.	past.part.	
IIa	bôt	butum	gibotan	bot	boten	geboten	'offered'
IIb	boug	bugum	gibogan	bog	bogen	gebogen	'bended'

As for Class II, it is worth noting that the NHG past indicative stem can be regarded as the regular outcome of the OHG past participle stem or the past.1.sg. stem of Subclass IIa. However, since analogy across different classes or subclasses is rare and unsystematic, it might be safer to consider Class II as a whole and assume that the past participle stem should be the base for leveling in this class.

This overview of leveling directions in each strong class is rather simplified, as exceptions are attested sporadically. For example, OHG skeran [to shear] belongs to Class IV and no morphological leveling should occur in NHG, but the past indicative stem schor-, whose stem vowel was introduced from the past participle, is attested. The past indicative forms of NHG werden [to become] chose the past plural stem as the base for leveling (wurde [I became] ~ wurden [we became]), whereas among other Class III verbs it was the singular stem that won out. Nevertheless, the number of exceptions is very small and they mostly result from non-morphological factors (Mettke 1967). This paper focuses on morphological motivation and pays more attention to the general tendency.

3. Methods and materials

3.1. Methods: calculation of conditional entropy

The method of calculating conditional entropy is proposed in Ackerman, Blevins, *et al.* (2009) and Ackerman and Malouf (2013); it is used as a way of measuring cell informativeness in this paper. In this theoretical model, cell *X* in the paradigm might have several potential forms – x_1, x_2, ..., x_n –, which occur with certain probabilities $p(x_1)$, $p(x_2)$, ..., $p(x_n)$ in the language. The amount of uncertainty in X, therefore, can be indicated by its entropy $H(X)$:

$$(1)\ H(X) = -\sum_i p(x_i) \log_2 p(x_i)$$

The unit for entropy is bit. A larger entropy indicates either that there are more potential forms in cell *X* or that the probabilities for all potential forms of *X* are more evenly distributed, suggesting that we are more uncertain of which form might occur in *X*. The following toy data show how entropy is used for comparison between the predictiveness of cells in a paradigm:

Table 6
Toy examples of OHG inflection paradigm

Past Participle (ppp)	Infinitive (inf)	Gloss
gi-bit-an	bîṭ-an	'to wait, to expect'
gi-rit-an	rîṭ-an	'to ride'
gi-lit-an	lîd-an	'to go'
gi-mit-an	mîd-an	'to avoid'
gi-lib-an	lîb-an	'to take care of'
gi-klib-an	klîb-an	'to attach'

Note: hyphen (-) marks morpheme boundary

All forms in the data belong to OHG Strong Class I, and the stem vowels are thus the same within a cell in the paradigm. However, there is a morphological change called Grammatischer Wechsel [grammatical change] with regards to the stem-final consonant (Braune 2012: 75-77): stem-final dentals /t/ and /d/ in infinitive form merge into /t/ in the past participle. Therefore, there are two possible values for the past participle stem in the data: one with stem-final /t/ and one with stem-final /b/, with probabilities of 4/6 and 2/6 respectively. The entropy of the past participle stem can be calculated using Equation (1):

$$(2)\ H(ppp) = -\left(\frac{4}{6}\log_2\frac{4}{6} + \frac{2}{6}\log_2\frac{2}{6}\right) = .918 \text{ bit}$$

The infinitive stem has three possible values: one with stem-final /t/, one with stem-final /d/, and one with stem-final /b/. The probability for each value is 2/6. Thus, the entropy of infinitive is:

$$(3)\ H(inf) = -\left(\frac{2}{6}\log_2\frac{2}{6} + \frac{2}{6}\log_2\frac{2}{6} + \frac{2}{6}\log_2\frac{2}{6}\right) = 1.585 \text{ bit}$$

When the past participle stem ends with /b/, the corresponding infinitive stem must have stem-final /b/ as well. The conditional probability of infinitive stem with /b/ given that its past participle has stem-final /b/ is therefore 2/2 = 1; the corresponding conditional entropy is calculated as follows:

(4) $H(inf|ppp = b) = -(\log_2 1) = 0$ bit

By contrast, lexemes whose past participle forms have stem-final /t/ have two possible infinitive stems. The conditional probability of each possible value for infinitive given that its past participle has stem-final /t/ is 2/4 = .5. The corresponding conditional entropy is:

(5) $H(inf|ppp = t) = -\left(\frac{2}{4}\log_2\frac{2}{4} + \frac{2}{4}\log_2\frac{2}{4}\right) = 1$ bit

With frequencies of each value for the past participle considered, the conditional entropy of infinitive given a past participle can be averaged as:

(6) $H(inf|ppp) = -\frac{4}{6}\left(\frac{2}{4}\log_2\frac{2}{4} + \frac{2}{4}\log_2\frac{2}{4}\right) + \left(-\frac{2}{6}\log_2\frac{2}{2}\right) = .667$ bits

Similarly, the conditional entropy of the past participle given an infinitive is:

(7) $H(ppp|inf) = \left(-\frac{2}{6}\log_2 1\right) + \left(-\frac{2}{6}\log_2 1\right) + \left(-\frac{2}{6}\log_2 1\right) = 0$ bit

The fact that H(*ppp*|*inf*) is smaller than H(*inf*|*ppp*) indicates that the paradigm presented in Table 6 is less uncertain when the infinitive is known than when the past participle is known. In other words, the infinitive is more informative and more predictive than the past participle in the data. We therefore expect that speakers tend to use the infinitive for the formation of the past participle, and that the contrast between /t/ and /d/ preserved in the infinitive would not be affected by leveling in diachronic changes.

The same algorithm is applicable for the comparison of predictiveness and base identification in data with more words and more complex paradigms.

Generally speaking, given a set of inflectional paradigms of verbs belonging to a single inflectional class or all inflectional classes in a language, suppose that c_1 and c_2 are two cells in the paradigm (in the case of OHG, they are any two of the five tense stems), and that r_1, r_2 are two sets of forms that could possibly occur in c_1, c_2 across the data, respectively. When c_2 is used as the base to predict c_1, the conditional entropy $H(c_1|c_2)$, which serves as a quantifiable measure of predictiveness of this projecting direction, can be obtained using the formula:

$$(8)\ H(c_1|c_2) = \sum_{r_1}\sum_{r_2} P_{c_1}(r_1)P_{c_2}(r_2)\log_2 P_{c_1}(r_1|c_2 = r_2)$$

The lower the conditional entropy, the more predictive the projecting direction is. The cell that shows the highest predictiveness as the base is considered the most informative. This it might be chosen by speakers as the base from which they derive forms for the remaining cells in the paradigm. Diachronically, it might also serve as the base for leveling and spread into other cells in the paradigm, while remaining unaffected by leveling.

3.2 Materials

The OHG data were derived from Deutsch Diachron Digital (ANNIS 2019; Krause and Zeldes 2016), which contains 707,835 tokens in 1621 texts. Since speakers are supposed to build formation rules between cells based on their memorization of complete paradigms for lexemes with high frequency (Ackerman and Malouf 2013:437), verbs with token frequency of 10 or more were selected. After removing the prefixes, 386 lemmata from all inflectional classes (including strong classes, weak classes, irregular classes, etc.) were extracted from the corpus. Their tense stems were constructed based on a standard available descriptive grammar (Braune 2012), so as to avoid the problem of dialectal variation, and all forms were phonetically transcribed for the calculation of conditional entropy. Words that had alternative paradigms were divided into as many lemmata as there were alternative paradigms. For example, wizzan [to know] (note that it is not a strong verb in OHG) was treated as four independent lemmata (Braune 2012: 261):

Table 7
Alternative paradigms of the lemma wizzan [to know]

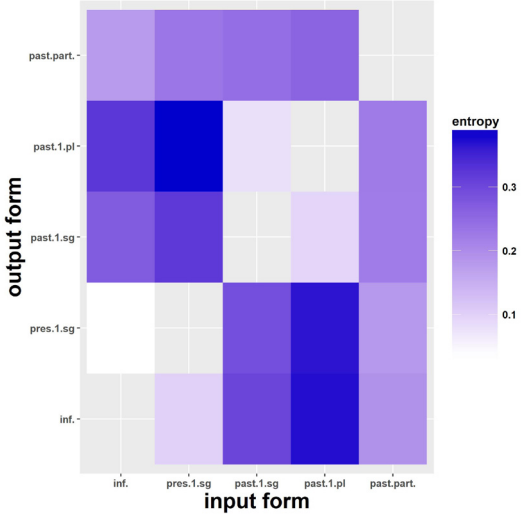
infinitive	pres.1.sg.	past.1.sg.	past.1.pl.	past.part.
wizzan	weiz	wissa	wissum	giwizzan
wizzan	weiz	wista	wistum	giwizzan
wizzan	weiz	wëssa	wëssum	giwizzan
wizzan	weiz	wësta	wëstum	giwizzan

Nevertheless, many trivial variants, such as regional or rarely attested forms, were not included in the data. To identify the base, each stem was used for the prediction of other stems in the paradigm, yielding 20 projecting directions in total.

4. Results

As conditional entropy, by definition, refers to the uncertainty of a paradigm cell when another cell is known, information about which inflection class the paradigm belongs to should not be considered in principle, when values for conditional entropy are calculated. Nevertheless, it is necessary to feed the model with the entire set of data and with different subsets of data collected from each verb inflectional class. This allows us to compare results with inflectional class information and results without class information, as well as to explore the best explanation for the attested historical change.

Figure 2
Conditional entropy of each projecting direction in the paradigm among all verbs in the data

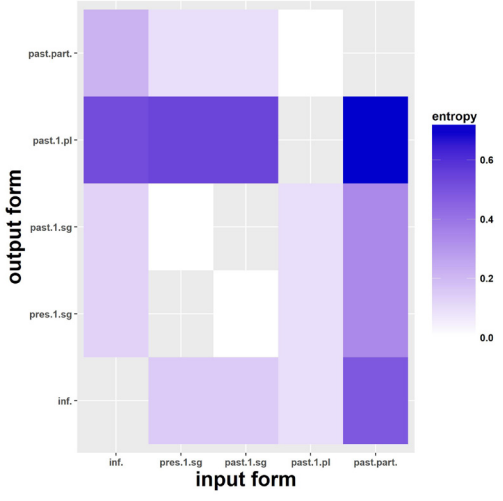


Note: the brighter the color, the lower the conditional entropy, and thus the higher the predictiveness in this direction

The results across all verb inflectional classes (including strong, weak, and other classes) suggest that subparadigms might exist in OHG verb inflection. Stems of the same tense (infinitive and pres.1.sg.; past.1.sg. and past.1.pl) can be mutually predicted with low conditional entropies, whereas predictions across different tenses are much more uncertain. This should be mostly influenced by the inflectional pattern of the weak classes, to which the majority of verbs in the data belong.

However, when conditional entropies are calculated for individual verb classes, the patterns are completely different; and some of them are consistent with the directions of leveling attested in the development from OHG to NHG. Only the results of Classes I, II, III, and V are presented in the following figures, because systematic analogical changes are not involved in other strong classes (see Section 2).

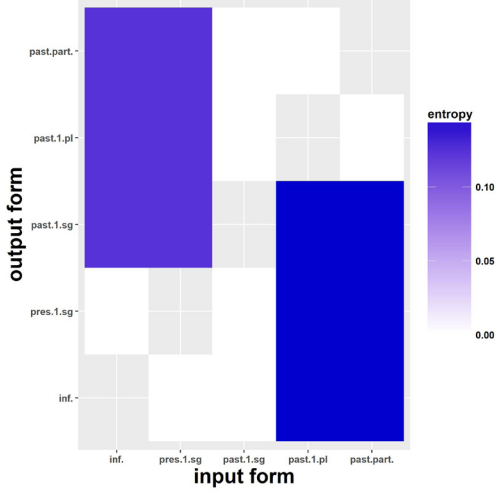
Figure 3
Conditional entropy of each projecting direction in the paradigm among Strong Class I verbs



Note: the brighter the color, the lower the conditional entropy, and thus the higher the predictiveness in this direction

The values for conditional entropy suggest that the past.1.pl. stem is the most predictive one for Strong Class I in general and that it should serve as the base for deriving other stems in the paradigm in Class I. As Table 3 shows, in Class I, it is also the OHG past plural stem that has become the stem for all past indicative forms in NHG Strong Class I (for example, NHG stieg, stieg-en from OHG past.1.pl. stig-um [we climbed]). The historical evidence thus supports conditional entropy as an indicator of the selection of the base for leveling.

Figure 4
Conditional entropy of each projecting direction in the paradigm among Strong Class III verbs

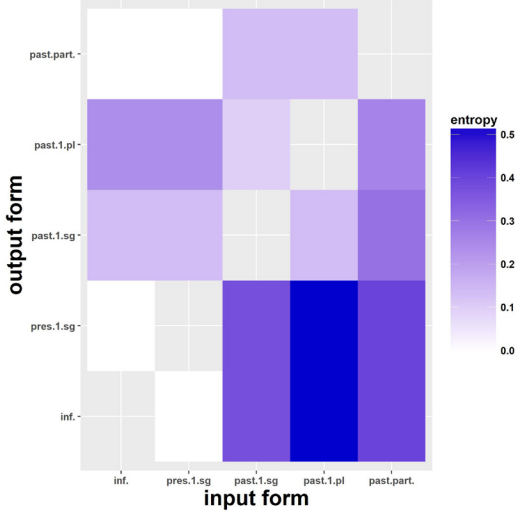


Note: the brighter the color, the lower the conditional entropy, and thus the higher the predictiveness in this direction

The results for Strong Class III also correspond to the attested diachronic change. Table 4 shows that all NHG past indicative forms are inflected based on the OHG past.1.sg. stem (for example, NHG trank, trank-en from OHG past.1.sg. trank [I drank]), which shows the lowest conditional entropies in predicting other forms in the paradigm, as indicated in Figure 4.

Nevertheless, there are inflectional classes in which historical changes that cannot be satisfactorily explained by cell informativeness occur.

Figure 5
Conditional entropy of each projecting direction in the paradigm among Strong Class V verbs

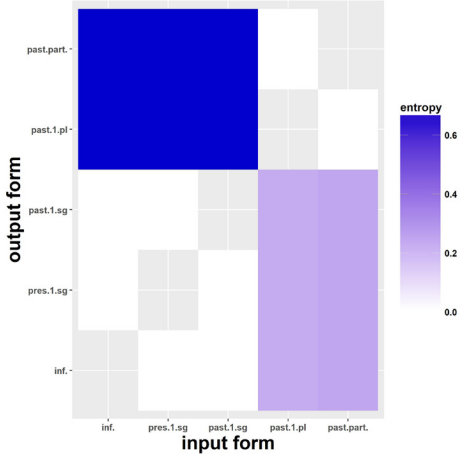


Note: the brighter the color, the lower the conditional entropy, and thus the higher the predictiveness in this direction

Similar to Strong Class I, NHG Strong Class V past indicative stems are also leveled to the OHG past plural stem (for example, NHG maß, maß-en from OHG past.1.pl. mâz-um [we climbed]; see Table 3). However, the past.1.pl. form is not the most informative one in Figure 5. Compared to Strong Classes I and III, cells for present forms (inf. and pres.1.sg.) in Strong Class V appear to be more predictive than cells for past forms in general and, remarkably, the differences in conditional entropies between the two past stems are smaller in Class V than in Classes I and III.

Figure 6

Conditional entropy of each projecting direction in the paradigm among Strong Class II verbs



Note: the brighter the color, the lower the conditional entropy, and thus the higher the predictiveness in this direction

The pattern of leveling attested in Strong Class II is unusual. The past indicative forms in NHG gave up their original OHG stems and adopted the past participle stem (for example, NHG bog, bog-en from OHG gi-bog-an [bent]; see Table 5). Although the past participle stem enjoys remarkably high predictiveness relative to other forms in Figure 6, the conditional entropies for the past.1.pl. stem are even slightly lower in general.

5. Discussion

5.1. Informativeness of cells as a factor in determining the base for leveling

Figures 3-6 provide evidence that informativeness of cells in a paradigm can partially explain the multiple directions of leveling attested in the development from OHG to NHG, though patterns found in some classes cannot be perfectly explained. However, the differences in values for conditional entropy between cells (especially cells for past tense and past participle) vary enormously across classes. For Classes I and III, in which the most informative cell in the OHG inflection paradigm happens to serve as the base for leveling in diachronic change (see Figure 3 and 4), the values for conditional entropy of the base are remarkably lower than other forms in the paradigm.

By contrast, the patterns found in Classes II and V are more complicated, in that there is no cell in the paradigm whose conditional entropies are much lower than those of other cells; it is thus not easy to identify the unique base for the paradigm (see Figure 5 and 6). For Class V, analysis of conditional entropy suggests that cells for present forms are more predictive than cells for past forms in Strong Class V and hence they are expected to be the base for the formation of NHG past indicative forms. However, the spread of present stems into past forms might be blocked for other reasons. On the one hand, present and past tenses are two distinctive semantic categories; semantically, past tense is also more closely related to the past participle than to the present tense. Though the past participle stem can spread into past forms in NHG without affecting the present forms, as we see in Strong Class II (see Table 5), systematic change in which present stems spread into past forms without affecting the past participle is not attested. On the other hand, analysis of conditional entropy based on the entire data set (see Figure 2) shows that, with all inflectional classes collapsed, the predictiveness of present forms relative to past forms is considerably lower than the predictiveness of the past participle relative to past forms. The remote relationship not only in semantics but also in overall predictiveness between present forms and past forms might explain why present stems fail to be the base for leveling among past forms in NHG (see also Section 5.2).

No reasonable explanation could be provided for the discrepancy in Strong Class II between the expected outcome of leveling indicated by conditional entropy and the attested diachronic change. However, it is worth noting that the differences in conditional entropy between the past.1.pl. form, the most informative form with lowest conditional entropies, and the past participle, the actual base for leveling in historical change, are not as large in Strong Class II as the differences found in other strong classes. If the selection of the past participle as the base for leveling is motivated by some unknown factors, the choice would hardly be restricted by cell informativeness, since the past participle in Strong Class II enjoys considerably low conditional entropies in general.

The exceptional changes in Strong Classes II and V suggest that cell informativeness should not be the only factor that determines the base for leveling. However, it could still be one of the most crucial. Comparing the results of Classes I, II, III, and V (see Figure 3-6), we can conclude that, in classes where conditional entropy makes a strong prediction, the prediction is consistent with the attested diachronic change. Exceptions occur only in classes where differences in conditional entropy for the most informative

cells are not large and thus strong predictions cannot be made. In these cases, other factors, like semantics, might play the decisive role in determining the base for leveling. These potential factors, however, cannot be identified based on the present data.

5.2. Evidence for the existence of subparadigms and local bases

While maintaining his hypothesis of single global base, Albright (2002b: 118-126) improves his model by postulating the existence of subparadigms and local bases. In a subparadigm, the local base is produced from the global base of the entire paradigm, whereas the remaining forms in the subparadigm are not directly produced from the global base, but instead from the local base in the subparadigm. When the local base is irregular and has to be stored in the lexicon rather than being directly derived from the global base, the remaining forms should share the irregularities of the local base, since they have a tighter relationship with the local base than with the global base. Two types of leveling are predicted in this model: non-base forms in a subparadigm might be leveled either to their local base or to the global base (via their local base). However, it is impossible for some forms in the subparadigm to be leveled to the global base, while the local base maintains its irregularity and remains a base for the formation of other forms in the subparadigm (see Figure 1).

Evidence of subparadigms in OHG inflection can also be found in the comparison of informativeness among cells as well as in diachronic changes. On the one hand, as shown in Figure 2, mutual predictiveness between cells of the same tense is higher than between cells of different tenses. The exact values from Figure 2 are presented in Table 8.

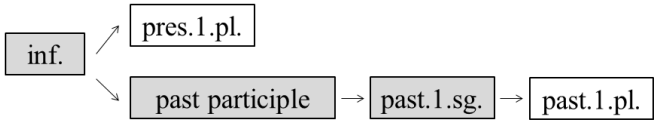
Table 8
Values for conditional entropy of each projecting direction in the paradigm among all verbs in the data

↓ output/input →	inf.	pres.1.sg.	past.1.sg.	past.1.pl.	past participle
inf		.1	.306	.374	.192
pres.1.sg.	.036		.293	.369	.183
past.1.sg.	.272	.322		.094	.224
past.1.pl.	.326	.385	.08		.226
past.part	.179	.232	.244	.26	

Note: values that could indicate the global or local bases are in blue

Though the algorithm used for quantifying informativeness is different from that used in Albright (2002b), the way to determine the local base can remain consistent. Table 8 suggests a paradigm organization with subparadigms and local bases. The infinitive generally has the lowest entropies as input and thus should be the global base. As for the past participle, it can not only be reliably predicted by the infinitive, but also has higher predictiveness relative to cells for past forms than do cells for present forms. Therefore, cells for the past participle and for past forms should belong to a subparadigm whose local base is the past participle. Finally, the conditional entropy for the direction from the past.1.sg. stem to the past.1.pl. stem is higher than the conditional entropy for the reverse direction. A hierarchical structure of local bases can thus be built:

Figure 7
Supposed organization of the OHG inflectional paradigm



Note: forms that serve as global or local bases are in grey

This paradigm organization is mostly compatible with attested diachronic changes. Some OHG strong verbs were regularized and became weak verbs.[4] In these changes, their past forms and past participle are rebuilt by applying derivational rules of weak classes to their present forms (the infinitive, in particular). On the contrary, systematic leveling of past forms to the present forms without affecting the past participle is not attested. Also, with regards to leveling within strong classes, the leveling of past forms to the past participle stem is found in Strong Class II (see Table 5); the leveling of past forms to the past.1.sg. stem is found in Strong Class III (see Table 4). Notably, the past participle stem was not subject to the leveling based on the past.1.sg. stem in Class III.

However, the situation is more complicated in Strong Classes I and V, in which the base for leveling was the past.1.pl. stem. In Class I, the past participle happens to share the same stem with the past.1.pl. form; we could hence suppose that the past participle stem was in fact the base for leveling. Nevertheless, since the past participle has a stem different from the past.1.pl. stem in Class V, the selection of base for leveling cannot be explained by the paradigm organization in Figure 7.

Moreover, although the model of subparadigms and local bases excludes some logically possible directions of leveling, it does not explain why leveling to a particular base is attested in one inflectional class, while leveling to another base occurred in another class. This is found in OHG Strong Classes I, II, and III, as well as in some sporadic cases where some strong verbs changed into weak verbs (see endnote 4).

To sum up, the results actually support the assumption of subparadigms and local bases in the OHG inflection paradigm. However, new problems might arise when attempts to build a more sophisticated paradigm with hierarchical structures are made. Since the present work focuses on the application of conditional entropy in explaining leveling, the tasks of defining and identifying local bases and subparadigms have to remain for future work.

6. Conclusion

This paper discusses conditional entropy as a measure of informativeness for explaining the multiple directions of leveling attested in the development from OHG to NHG. Quantitative analysis of the corpus data shows that the predictiveness of cells relative to other cells in the paradigm, quantified by their conditional entropy in paradigms within a single verb inflectional class, might play a crucial role in determining the base for leveling, though it is not the only factor. The base for paradigmatic leveling does not have to be the same across all lexical items, but might vary across different inflectional classes; effects of subparadigms and local bases can be found in the results as well. However, a more sophisticated model of the relationships among global base, local base, and non-base forms in the paradigm still needs to be developed.

NOTES

1. The Spanish forms are given in their pronunciation instead of their conventional spelling in Albright (2002b).

2. The tense is termed Präteritum in conventional German grammar; the most accurate English translation would be *preterite*. However, to facilitate understanding for readers who do not have background in German grammar, the term *past* is used for the tense that contrasts with the present tense in this paper.

3. In NHG, past subjunctive forms might use a stem that is different from the past indicative stem; this results from regular sound change and is thus not inherited from OHG. The new past subjunctive stem in NHG is irrelevant to the problem and thus excluded from our discussion.

4. The change from OHG strong verb to NHG weak verb occurred only sporadically. For example, OHG bliuwan [beat] is a Strong Class II verb, but its descendant in NHG, bläuen, is a weak verb.

RÉFÉRENCES

ACKERMAN, Farrell, BLEVINS, James P., and MALOUF, Robert (2009): Parts and wholes: Implicative patterns in inflectional paradigms. *In*: BLEVINS, James P. and BLEVINS, Juliette, eds. *Analogy in Grammar: Form and Acquisition*. Oxford: Oxford University Press, 54-82.

ACKERMAN, Farrell and MALOUF, Robert (2013): Morphological organization: the low conditional entropy conjecture. *Language*. 89(3):429-464.

ALBRIGHT, Adam (2002a): Base selection in analogical change in Yiddish. *In*: LARSON, Julie and PASTER, Mary, eds. *Proceedings of the twenty-eighth annual meeting of the Berkeley Linguistics Society*. Berkeley: The Berkeley Linguistics Society, 1-13. https://doi.org/10.3765/bls.v28i1.3814.

ALBRIGHT, Adam (2002b): *The Identification of bases in morphological paradigms*. Doctoral dissertation. Los Angeles: University of California, Los Angeles.

ANNIS (2019): Deutsch Diachron Digital (ANNIS Corpus Search). *ANNIS*. https://korpling.german.hu-berlin.de/annis3/ddd.

BRAUNE, Wilhelm (2012): *Althochdeutsche Grammatik* [Old high German grammar]. Munich: LINCOM Europa.

HILL, Eugen (2013): *Einführung in die historische Sprachwissenschaft des Deutschen* [Introduction to the historical linguistics of German]. Darmstadt: Wissenschaftliche Buchgesellschaft.

HOCK, Hans Henrich (1991): *Principles of Historical Linguistics*. Berlin: Mouton de Gruyter.

KRAUSE, Thomas and ZELDES, Amir (2016): ANNIS3: A new architecture for generic corpus query and visualization. *Digital Scholarship in the Humanities*. 31(1):118-139. http://dsh.oxfordjournals.org/content/31/1/118.

METTKE, Heinz (1967): *Mittelhochdeutsche Grammatik: Laut- und Formenlehre* [Middle high German grammar: sounds and morphology]. Halle: Max Niemeyer Verlag.

PAUL, Hermann (1929): *Mittelhochdeutsche Grammatik* [Middle high German grammar]. Halle: Max Niemeyer Verlag.

R CORE TEAM (2017): *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. https://www.R-project.org.

SHANNON, Claude (1948): A mathematical theory of communication. *The Bell System Technical Journal*. 27:379-423, 623-656.

TIERSMA, Peter Meijes (1982): Local and general markedness. *Language*. 58(4):832-849. https://www.jstor.org/stable/413959.

WICKHAM, Hadley (2016): *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag. https://ggplot2.tidyverse.org.