

---

# The direction of analogical extensions in the verbal roots of Old French and Old Florentine Italian: a corpus study

Sascha Gaglia

Georg-August-Universität Göttingen

---

## 1 Introduction

Old French shows a great amount of analogical extensions of verbal root allomorphs (Fouché 1967; Walker, 1987; among others). Some of the involved paradigms were levelled in later stages, for example O. Fr. *trover* (Mod. Fr. *trouver*) ‘find’, *paroler* (*parler*) ‘talk’, *amer* (*aimer*) ‘love’, *plorer* (*pleurer*) ‘cry’ and *lever* ‘lift/elevate’. Other verbs like *venir* ‘come’ and *tenir* ‘have/hold’ still show some significant allomorphic patterns in Modern French. Not all of these paradigms were extended in the same way. Where some paradigms generalized the unstressed root (e.g. *trover*, *paroler*), others generalized a stressed one (e.g. *amer*, *plorer*, *lever*). Hence, we know the starting point for analogical extensions and the situation in Modern French as well as the diachronic developments in general. However, a thorough quantitative analysis of analogical extensions based on Old French corpus data is still missing, which is the reason why it is still not possible to unveil the ‘path’ of each extension in the sense of paradigmatic patterns, subparadigms and paradigm cells. The same also holds true for Old Italian. Having this in mind, I will present some important findings from my ongoing project on the paradigmatic extension of verbal roots in French and Italian funded by the German Research Council (DFG). My first goal here is proposing a quantitative analysis based on corpus data of the direction of the main analogical extensions for the morphomic patterns N, L and U (Maiden 2018). This will be conducted for each century of the Old French period and paradigm cell by paradigm cell, if possible. For a better understanding and to strengthen our observations, I shall include also the Old Florentine Italian verbs *venire* ‘come’ and *tenere* ‘have/hold’. On the basis of these data I will then test some of the factors discussed in the literature. They comprise phonological markedness, morphomic attraction and stability as well as frequency. Moreover, I will discuss one additional factor, namely semantic-syntactic feature specification. I shall show that our data indicate that this factor is relevant to the direction of analogical extensions.

## 2 Influencing factors (previous research)

The following three factors have been often mentioned while trying to account for the direction of the analogical extension of Romance verbal roots, especially with respect to Old French:

*Phonological markedness:* In most cases, it was the phonologically unmarked root of a paradigm which was extended, e.g. Mod. Fr. *trouver* exhibits the Old French unstressed roots *trouv-* throughout the paradigm, while the segmentally more marked stressed roots got lost (e.g. *truev-/treuv-/truis(s)-/troiss-*). Mod. Fr. *aimer* ‘love’ and *pleurer* ‘cry’ are exceptions due to possible interparadigmatic analogies on the grounds of *\*esmer* ‘esteem’ and *pleuvoir* ‘rain’ (Esher, 2016). In the case of *aimer* and *pleurer*, the stressed roots served as basis for analogical extension.

*Frequency*: Verbs like O. Fr. *amer* lost their N-pattern distribution of root allomorphs (i.e. 1SG, 2SG, 3SG/3PL vs. 1PL/2PL present indicative/subjunctive) since this pattern had a low type frequency in the first conjugation class of Old French (Esher, 2016). I shall discuss the possible effects of type frequency and of token frequency, which have been proposed in more than one occasion to serve as factor for analogical extension or for being responsible for the resistance of a form against analogical extension (see Bybee 1985, among others).

*Morphomic stability* is a major factor constraining the direction of the extension of a root in the sense of determining the paradigm cells in which an extension may or may not take place (see ‘coherence’ in Maiden 2018).

### 3 Research questions

In line with the aims of the present paper and building upon the factors addressed in previous research, there are two research question which shall be discussed here:

- 1) Which of the above-mentioned factors can be shown to influence the direction of analogical extensions in Old French and Old Florentine Italian verbal roots by means of a quantitative corpus analysis?
- 2) Does semantic-syntactic feature specification matter?

## 4 Methodology

### 4.1 Data

With these goals in mind, all variants of verbal root allomorphs of the aforementioned Old French verbs *trover*, *amer*, *plorer*, *venir* and *tenir* were extracted from the *Nouveau Corpus d’Amsterdam* (= NCA), with a total number of 24,702 tokens. (NCA covers the entire Old French period, i.e. from 1100 to 1350, and includes 3,184,834 words from 299 texts.) The Old Florentine Italian data was taken from the *Corpus OVI dell’italiano antico* (= OVI), including the period from 1100 to Giovanni Boccaccio’s death in the year 1375, which conventionally marks the end of the Old Italian period (23,176,174 words from 2,335 texts). Due to practical reasons, only two verbal paradigms were investigated with respect to Old Florentine Italian, i.e. *venire* and *tenere* (5,758 tokens). While data from the NCA could be easily extracted for each paradigm cell together with relevant manuscript metadata via the tool TigerSearch, this was not the case for our Old Florentine Italian data, which I transferred token by token to a Filemaker database and then annotated each token with the relevant information. Hence, the amount of our Old Florentine Italian is significantly smaller than compared to our investigated data from Old French.

To compare similar phenomena for the verbs in question, I divided the lexemes in three groups (1, 2, 3), with respect to the main morphomic patterns they adhere to (i.e. N, L, U; L’ corresponds to the DARK-L pattern in Esher, 2016). Moreover, the groups 1 and 2 contain two subgroups (A, B) regarding the direction of analogical extension in prosodic terms (i.e. extension of the stressed or unstressed root). The differences between groups 1 and 2 on the one hand and 3 on the other hand are to be traced back to the overall differences between Old French and Old Florentine Italian. Table 1 (see on next page) represents mainly patterns that concern the present indicative and the present subjunctive and which are the domain of the patterns in question. Analogical extensions that go beyond the present tenses are not included in the table and are instead discussed below. (I use the following abbreviations in Table 1: R<sub>M</sub> = root including a monophthong, e.g. Fr. *trouv-* [truv-]; R<sub>D</sub> = root including a

diphthong, e.g. Fr. *vien-* [vjē(n)-];  $R_{\text{VEL}}$  = root including a final velar consonant, e.g. It. *veng-* [veŋg-];  $R_{\text{3PL PRS.IND}}$ : extension of a 3PL PRS.IND root, e.g. denasalized Fr. [vjən-].)

## 4.2 Statistical analysis

The orthography of medieval texts is far from being standardised. Moreover, these can include diatopic and stylistic variation. Hence, each cell of the paradigm of a given lexeme can display so-called ‘overabundance’, i.e. it can have more than one surface representation; see Thornton (2012), who reasonably uses the notion ‘cell-mate’ instead of ‘doublet’ for each overabundant form.

Table 1. *Analysed verbal groups*

Group	Patterns	Extended Root	Paradigms (Tokens)	Variety
1A	L >> levelled	$R_{\text{M}}$ (unstressed)	<i>trover</i> (4,250)	Old French
1B	L >> L'	$R_{\text{D}}$ (stressed) $R_{\text{3PL PRS.IND}}$ (stressed)	<i>venir</i> (9,728) <i>tenir</i> (2,968)	
2	N >> levelled	$R_{\text{D}}$ (stressed)	<i>amer</i> (913) <i>plorer</i> (1,085)	Old French
3	U >> U	$R_{\text{VEL}}$ (stressed)	<i>venire</i> (3,289) <i>tenere</i> (2,469)	Old Fl. Italian

I, therefore, summarized all cell-mates of a paradigm cell that show the same morpho-phonological phenomenon as in Table 1; e.g. *truev-*, *troev-*, *trueu-* are variants of a phonologically diphthongized root ( $R_{\text{D}}$ ), while *truv-* and *trouv-* phonologically display a monophthong ( $R_{\text{M}}$ ) in the case of *trover*. Our statistical analysis takes into consideration (a) percentages representing the proportion of analogically extended roots vs. non-analogically extended roots in each relevant cell of the paradigm for a given verb (e.g.  $R_{\text{M}}$  vs.  $R_{\text{D}}$  in the case of *trover*; our method is comparable to the count of cell-mate ratios in Thornton, 2012), and (b) a normalized count (or *normalized frequency*) of each analogically extended root vs. non-analogically extended roots in relation to corpus size. Every count was limited to time intervals corresponding to the 12<sup>th</sup>, 13<sup>th</sup> and 14<sup>th</sup> century, respectively.

## 5 Main Results

The following results shall be discussed in order to answer our research questions (see § 2 above):

*First*, each of the three assumptions found in the literature on the direction of analogical extension can be confirmed by our data: By comparing the verbs of groups 1A and 1B, I will show that in the case of 1SG PRS.IND the root [trʊis] (paradigm of O. Fr. *trover*) behaves more stably towards its extinction than [vēŋ] (sometimes [vīŋ]) and [tēŋ], the 1SG forms of O. Fr. *venir* and *tenir*, respectively, which means that the analogical extension of stressed [vjē(n)-] and [tjē(n)-] was faster than the extension of unstressed [truv-] (*trover*) towards this cell of the paradigm, as the normalized frequencies in Table 2 show:

Table 2. *Analogical extension to 1SG PRS.IND (normalized frequencies)*

Century	[truv-]	[vjẽ(n)-]	[tjẽ(n)-]
12th	0.00	1.23	0.41
13th	0.00	1.68	1.97
14th	0.08	2.00	2.88

In line with Bybee’s assumption of morpho-phonemic irregularity (Bybee 1985), I claim that this has to do with the fact that [truʝis] was phonologically more marked than [vẽn] and [tjẽ]. Moreover, I claim that the morphomic stability of the L-pattern contributed simultaneously to the relative stability of [truʝis], although due to their low type frequency in Old French, most of the morphomic patterns got lost very early from first conjugation class verbs like *trover* (see also Esher, 2016 with respect to the N-pattern). In the paper I will also argue against token frequency as being responsible for the stability of [truʝis] and, in general, discuss it with respect to 1SG PRS.IND (against Bybee 1985; see more evidence below in this abstract). Morphomic stability is also at issue when comparing the verbs of 1B and 2. It can be shown that wherever a subparadigm was not shaped by a morphomic pattern (i.e. future, conditional, weak preterite forms etc.), analogical extension occurred in these subparadigms first. From there, analogical extension spread to the morphomically shaped present indicative and present subjunctive of *amer* and *plorer*. This confirms, in principle, the findings in Fennell (1975) with respect to the analogical extension of verbal roots in the future tense and conditional. However, our data show that this extension occurred much earlier than assumed in Fennell’s corpus analysis of Middle French.

*Second*, a comparison of the developments affecting the verbs of group 1B and 3 reveals that the Old French denasalised root [vjɛn-] extended first from 3PL PRS.IND to 3PRS.SBJ, which also holds true for O. Fl. It. [vejŋ-] (paradigm of *venire*), as shown by its including a final velar consonant (the predecessor was [vejɛn-] as a result from sound change). O. Fl. It. *tenere* appears to deviate slightly from this ‘path’ in that the 3SG PRS.SBJ seems to undergo analogical extension slightly prior to 3PL PRS.SBJ with respect to [teŋg-]. I claim that these results serve as a further piece of evidence against the assumption that token frequency determines the direction of extension, which has already been questioned by Fertig (2013). If token frequency would be a determinant, we would first expect 3SG PRS.SBJ to undergo analogical extension due its higher frequency compared to 3PL PRS.SBJ. (See also the results for token frequency in spoken and written modern Spanish in Bybee 1985 with respect to Person and Number, even if her data concern only the present indicative and preterite.) I propose in line with our research question 2 that the path 3PL PRS.IND to 3PL PRS.SBJ can only be explained by taking into consideration the semantic-syntactic generalization of a form, which means that O. Fr. [vjɛn-] and O. Fl. It. [vejŋ-] were extended from 3PL PRS.IND to 3PL PRS.SBJ due to their syntactic feature specification being PERS = 3, NUM = PL (or semantically *reference = bystander*, *cardinality = x > 1*) and TENSE = PRS. In the case of *tenere*, the generalization could have worked for PERS = 3 and TENSE = PRS. All analysed groups (see Table 1) exhibit a third person cell being the first affected by analogy. This can especially be shown for group 1A and 1B: These are the groups in our sample where analogical extension could also affect 3SG PRS.IND, being the cell which has the highest frequency.

## Bibliography

- Bybee, Joan L. 1985. *Morphology. A study of the relation between meaning and form*. Amsterdam: Benjamins.
- Esher, Louise. 2016. Morphome death and transfiguration in the history of French. *Journal of Linguistics* 53. 51-84.
- Fennell, T. G. 1975. *Morphologie du futur en moyen français*. Genève: Droz.
- Fertig, David. 2013. *Analogy and morphological change*. Edinburgh: Edinburgh University Press.
- Fouché, Pierre. 1967. *Morphologie historique du français. Le verbe*. Paris: Klincksieck.
- Maiden, Martin. 2018. *The Romance verb*. Oxford: Oxford University Press.
- NCA = Stein, Achim. et al. 2011. *Nouveau Corpus d'Amsterdam. Corpus informatique de textes littéraires d'ancien français (ca 1150-1350), établi par Anthonij Dees (Amsterdam 1987), remanié par Achim Stein, Pierre Kunstmann et Martin-D. Gleßgen*. Stuttgart: Institut für Romanistik/Linguistik.
- OVI = Larson, Pär et al. 2005. *Corpus OVI dell'Italiano antico*. Istituto Opera del Vocabolario Italiano/CNR.  
([http://gattoweb.ovi.cnr.it/\(S\(gpccezaigip4ix55l2qo4lqx\)\)/CatForm01.aspx](http://gattoweb.ovi.cnr.it/(S(gpccezaigip4ix55l2qo4lqx))/CatForm01.aspx))
- Thornton, Anna M. 2012. Reduction and maintenance of overabundance. A case study on Italian verb paradigms. *Word Structure* 5/2. 183-207.
- Walker, Douglas C. 1987. Patterns of analogy in the Old French verb system. *Lingua* 72. 109-131.