

Chapter 4

The Latin *honor* analogy

In older stages of Latin, many nouns exhibited paradigmatic $s \sim r$ alternations created by rhotacism in suffixed forms, as in (44a). In the period immediately before Classical Latin, these $s \sim r$ alternations were eliminated by extending the r to the nominative form. At approximately the same time, an independent change shortened long vowels before word-final sonorants, resulting in the paradigm in (44b):¹

(44) Elimination of $s \sim r$ alternations in Pre-classical Latin

a. *Pre-leveling*

| | <i>sg.</i> | <i>pl.</i> |
|-------------|------------|------------|
| <i>nom.</i> | hono:s | hono:re:s |
| <i>gen.</i> | hono:ris | hono:rum |
| <i>dat.</i> | hono:ri: | hono:ribus |
| <i>acc.</i> | hono:rem | hono:re:s |
| <i>abl.</i> | hono:re | hono:ribus |

b. *Post-leveling*

| | <i>sg.</i> | <i>pl.</i> |
|-------------|------------|------------|
| <i>nom.</i> | honor | hono:re:s |
| <i>gen.</i> | hono:ris | hono:rum |
| <i>dat.</i> | hono:ri: | hono:ribus |
| <i>acc.</i> | hono:rem | hono:re:s |
| <i>abl.</i> | hono:re | hono:ribus |

As mentioned in chapter 1, the switch from [hono:s] to [honor] has attracted the attention of numerous linguists over the years (Hoenigswald 1960, pp. 107-111; Hooper 1976, 95-96;

¹There is some evidence for an intermediate stage in the late Pre-Classical stage in which [hono:r] with long [o:] was at least one possible variant of the nominative form. Early poets such as Plautus and Ennius still allowed final long [o:r], using it to satisfy metrical requirements for heavy syllables, and also sometimes used [-o(:)r] instead of [-o:s], e.g., Ennius 545 (Skutsch 1985) *Clám[ō:]r ād cáelūm vólvēndūs pēr áethĕrǎ vágĭt*; also Ennius 409, 428. As far as I know, there is no evidence concerning how “clean” this intermediate stage was – the shortened variant [honor] may also have been used from the very beginning of the [s] > [r] change. The analysis that I propose here is compatible with the existence of an intermediate [hono:r] stage, but does not rely on it. It would, however, make a crucial difference for some other analyses, such as the Uniform Exponence analysis sketched below.

Wetzels 1984; Hock 1991, pp. 179-190; Barr 1994; Kiparsky 1997; Kenstowicz 2002; Hale, Kisseck and Reiss 1998; and others), because the direction of leveling runs counter to so many other changes. Why was the nominative singular rebuilt on the basis of oblique forms in Latin,² while in so many other languages (including the Polish diminutives example in chapter 1), entire paradigms have been rebuilt on the basis of nominative singular forms?

There have been numerous attempts to explain why Latin speakers might have rebuilt the nominative singular instead of the oblique forms; indeed the problem has never been in coming up with hypotheses about what would have made [honor] an appealing form. Looking at the paradigms in (44), various possibilities come to mind: perhaps it was the predominance of forms in [r], or perhaps one of the oblique forms (such as the accusative) was more common than the nominative singular, or perhaps it was the existence of many other words with *-or* nominatives. The real challenge is to explain why one of these factors was able to cause a leveling of the nominative in Latin, but not in Polish or other cases. In this chapter, I will propose that what distinguishes Latin from many other cases is not some difference in the frequency or markedness of forms, but rather in the form that serves as the base. Employing the base identification algorithm from the previous chapter, I will show that neutralizations and exceptions make the nominative singular an uninformative form in Latin, and favor the choice of an oblique form as base (it does not matter which one). This explains how learners could have set up a different base in Latin than in other languages (a problem noted by, among others, Wetzels 1984, pp. 582-583), and helps to account for the unusual direction of the analogical change. Then, in section 4.3, I will show that when an oblique form is chosen as base, the grammar that the model learns makes the correct predictions for the *honor* analogy. Finally, in sections 4.4.1-4.4.3, I will show that this analysis provides a more accurate explanation for the details of the change—in particular, by differentiating words that changed from those that did not—than many of the factors that have previously been proposed.

4.1 Statement of the problem

As Hock points out, the *honor* analogy can be described in at least two different ways. The first is as a *four-part analogy*, in which [hono:s] was influenced by words that already had an [r] in the nominative, such as [soror] ‘sister’:

- (45) Extension of *r* by four-part analogy:
 [soro:ris] : [soror] :: [hono:ris] : X (X = [honor])

Although the four-part analogy notation expresses the change as the influence of one particular lexical item (in this case, [soror]), it is generally recognized that such changes are actually due to the collective influence of many words, such as [soror], [kruor] ‘blood’, and also the numerous agentive nouns ending in *-or* ([o:ra:tor] ‘speaker’, [gladia:tor] ‘gladiator’, etc.). However, even when we recognize that the four-part analogy notation is just a shorthand for the influence of a larger pattern, it is far from an explanation of the change. How many words does it take for speakers to construct such an analogy? How similar do they have to be? More importantly, it has often been noted that the four-part analogy notation cannot tell us why

²I will use the term *oblique* to denote all cases other than the accusative, which in the case of Latin, includes genitive, dative, accusative, ablative, and dative.

the influence was not from [hono:s] to [soror] ([hono:ris]:[hono:s] :: [soro:ris]:X, X=[soro:s]) (Hock 1991; Barr 1994; Kiparsky 1998), or why the change was not in the opposite direction, undoing rhotacism (something like [soror]:[soro:ris] :: [hono:s]:X, X = [hono:sis]). This last question is especially puzzling, because the actual change in (45) violates the tendency for analogical change to extend from more “basic” or unaffixed forms to less basic, or affixed forms (Kuryłowicz 1947).


The four-part analogy notation can equally well capture the leveling of alternations ([hono:s] > [honor]) or the extension of alternations ([soror] > [soro:s]). An alternative description of the spread of [r] to the nominative in [honor] is as paradigm leveling, with the nominative form changing to match the remainder of the paradigm ([hono:ris], [hono:ri:], [hono:rem], etc.). The pressure to level paradigms has been formalized in Optimality Theory (OT, Prince and Smolensky 1993) as constraints on paradigm uniformity or uniform exponence (Kenstowicz 1997b; Kenstowicz 2002; Steriade 2000); for example, Kenstowicz (2002) schematizes the [hono:s] > [honor] change as the promotion of a constraint demanding uniformity in noun paradigms (UE). In the first stage, shown in the tableau in (46), the ban on intervocalic [s] (*VsV) outranks Faithfulness for /s/ (Faith(s)), yielding rhotacism in suffixed forms. Faith(s) in turn outranks UE, meaning that rhotacism does not overapply in the nominative form. The result is a paradigm with *s* ~ *r* alternations, as in (46a). Note that candidate (c) contains a vowel length alternation ([or] ~ [o:ris] – Kiparsky 1997; Hale et al. 1998; Baldi 1999, p. 323), for which Kenstowicz does not record a UE violation. It is entirely possible that there are separate UE constraints for different alternations, with UE for *s* ~ *r* alternations ranked higher than UE for vowel length alternations. If that is the case, then UE(*s* ~ *r*) is the relevant constraint here, and this what I will assume.

(46) Stage 1: *VsV ≫ Faith(s) ≫ UE(*s* ~ *r*)

| /hono:s/ (nom.), /hono:s-is/ (gen.), /hono:s-em/ (acc.), ... | *VsV | Faith(s) | UE(<i>s</i> ~ <i>r</i>) |
|--|-----------------|----------|---------------------------|
| a. ☞ [hono:s], [hono:r-is], [hono:r-em], ... | ✓ | ** | * |
| b. [hono:s], [hono:s-is], [hono:s-em], ... | *(gen.)!*(acc.) | ✓ | ✓ |
| c. [honor], [hono:r-is], [hono:r-em], ... | ✓ | ***! | ✓ (?) |

Under this analysis, the change consists of promoting UE over Faith(s), so that rhotacism overapplies in the nominative, as in (47). This leaves two possible candidates for the nominative: [hono:r] (candidate (c)) and [honor] ((d) and (e)). The first of these violates a high-ranking phonotactic constraint against final [o:r] in Latin, favoring a paradigm with shortening in the nominative ((d) or (e)); of these, the paradigm that preserves long [o:] in the oblique forms is more faithful to the long [o:] of the input (Faith(V:)), and thus candidate (d) is selected.

(47) Stage 2: *VsV, UE(*s~r*) \gg Faith(*s*), Faith(*V*)

| | /hono:s/ (nom.), /hono:s-is/ (gen.), /hono:s-em/ (acc.) | *o:r# | *VsV | UE(<i>s~r</i>) | Faith(<i>s</i>) | Faith(<i>V</i>) |
|--|---|-------|---------------------|------------------|-------------------|-------------------|
| a. | [hono:s], [hono:r-is], [hono:r-em] | ✓ | ✓ | *(nom.)! | ** | ✓ |
| b. | [hono:s], [hono:s-is], [hono:s-em] | ✓ | *(gen.)! *(acc.) | ✓ | ✓ | ✓ |
| c. | [hono:r], [hono:r-is], [hono:r-em] | *! | ✓ | ✓ | *** | ✓ |
| d.  | [honor], [hono:r-is], [hono:r-em] | ✓ | ✓ | ✓ | *** | * |
| e. | [honor], [honor-is], [honor-em] | ✓ | ✓ | ✓ | *** | **!* |

This analysis avoids many of the problems pointed out by Hale et al. (1998) by limiting UE to *s~r* alternations. An alternative, however, is that the Uniform Exponence analysis actually captures the creation of the intermediate variant [hono:r] (see fn. 1), with perfect UE satisfaction, and that the underlying form of this word had already been reanalyzed as /hono:r/ (as Hale et al. claim) by the time final vowel shortening occurred, so Uniform Exponence was irrelevant by that stage.

The uniform exponence account formalizes the intuition that paradigm leveling is due to a pressure for nonalternating paradigms, and that the resulting paradigm is one which (in this case) satisfies both paradigmatic constraints and also general phonotactic constraints of the language (such as *VsV, and *o:r#). However, it leaves many details unaccounted for. Why, for example, was the old *s~r* alternation suddenly intolerable, at the same time that a new *o~o:* alternation was being created? We might have expected the UE constraint to move above both the *o:r# and *VsV constraints, since there was no crucial ranking between them before the change. This should have yielded a uniform paradigm with overapplication of both rhotacism and shortening: [honor], [honoris], [honorem], etc. Furthermore, we might expect an increased drive for paradigm uniformity to level other alternations in noun paradigms, but in fact these remained by and large intact:

(48) Alternations preserved in Latin noun paradigms

| <i>gloss</i> | 'honor' | 'city' | 'art' |
|--------------------|------------|--------------------|-------------------|
| <i>alternation</i> | [o] ~ [o:] | [p] ~ [b] | ∅ ~ [t] |
| <i>nom.</i> | [honor] | [urps] | [ars] |
| <i>gen.</i> | [hono:ris] | [urbis] (*[urpis]) | [artis] (*[aris]) |
| <i>dat.</i> | [hono:ri:] | [urbi:] (*[urpi:]) | [arti:] (*[ari:]) |

Another unexplained mystery is why UE was promoted over Faith(s) to extend rhotacism, rather than Faith(s) being promoted over *VsV to eliminate rhotacism. As mentioned above, the rhotacism constraint (*VsV) plays a crucial role in Kenstowicz's analysis, ensuring that the resulting paradigm will have uniform [r] and not uniform [s]. However, we could just as easily have used the same analytical device (historical reranking of two constraints) to predict the opposite change, promoting the IO Faith(s) constraint to yield uniform [s]. Thus, the mere existence of an active rhotacism constraint at one stage in the grammar is not sufficient to explain why it should continue to be true at the next stage in the grammar. A proposal by McCarthy (1998) that output-to-output constraints start at the top of the grammar in the initial state might help to explain the tendency for uniform exponence to move up in grammars over time. If this is correct, then we need only assume that learners sometimes fail to demote UE below the relevant markedness constraints – but why would one generation of Latin learners suddenly fail to apprehend the correct ranking of Faith(s) \gg UE? If uniform exponence is to have any explanatory force in accounting for paradigm leveling, we would ideally like to be able to predict when such UE $\ll = \gg$ Faithfulness flips are likely to occur, and which faithfulness constraints will be demoted.³

A uniform exponence analysis also fails to capture various other details about the [hono:s] > [honor] change. Most notably, the spread of [r] to the nominative form was complicated by the fact that it was restricted primarily to polysyllabic, non-neuter nouns such as [hono:s] 'honor' (masc.) and [arbo:s] 'tree' (fem.), shown in (49a) (Hock 1991; Barr 1994; Kiparsky 1998). Monosyllabic nouns, such as [flo:s] 'flower' (masc.) were not affected (49b), nor were polysyllabic neuter nouns, such as [korpus] 'body' (neut.) (49c).

(49) Leveling restricted to masc. and fem. polysyllables

- | | | | | |
|----|------------------------|----------|---------------------------|--|
| a. | [hono:s] \Rightarrow | [honor] | 'honor' (<i>masc.</i>) | |
| | [arbo:s] \Rightarrow | [arbor] | 'tree' (<i>fem.</i>) | |
| | [odo:s] \Rightarrow | [odor] | 'odor' (<i>masc.</i>) | |
| | [augus] \Rightarrow | [augur] | 'omen' (<i>masc.</i>) | |
| b. | [flo:s] \Rightarrow | [flo:s] | 'flower' (<i>masc.</i>) | (\nRightarrow [flor]) |
| | [o:s] \Rightarrow | [o:s] | 'mouth' (<i>neut.</i>) | (\nRightarrow [or]) |
| | [mo:s] \Rightarrow | [mo:s] | 'custom' (<i>masc.</i>) | (\nRightarrow [mor]) |
| c. | [korpus] \Rightarrow | [korpus] | 'body' (<i>neut.</i>) | (\nRightarrow [korpor], [korpor]) |
| | [tempus] \Rightarrow | [tempus] | 'time' (<i>neut.</i>) | (\nRightarrow [tempor], [tempur]) |
| | [onus] \Rightarrow | [onus] | 'burden' (<i>neut.</i>) | (\nRightarrow [oner], [onur] ⁴) |

Furthermore, the new [r] forms appear to have replaced the older [s] forms relatively slowly; in the historical period we find both [hono:s] and [honor], [odo:s] and [odor], [arbo:s] and

³Dresher (2000, p. 60) makes this same point, and likewise concludes that UE-type constraints do not provide an adequate explanation for paradigm leveling.

⁴We might expect an [e] on the basis of the oblique forms of this word, which have [oner-]. A [u] might also be possible, on analogy with words like [femur] 'thigh'.

[arbor], [lepo:s] and [lepor] ‘charm’, [labo:s] and [labor] ‘labor’ (Leumann 1977, p.179).⁵ Both of these problems could be handled by various means – UE could be restricted to non-neuter polysyllables in some way, for example, and the free variation could be accomplished by the gradual promotion of stochastically ranked constraints (Boersma and Hayes 2001). As above, the real problem is not in finding theoretical machinery that can describe the change; it is in understanding why the change should have occurred in this direction, to these particular words, and why Latin behaved differently from so many other languages.

In sum, there are a variety of questions that must be answered if our understanding of the Latin *honor* analogy is to move beyond description to actual explanation:

- (50) a. Why was a basic, “unmarked” isolation form (the nominative) rebuilt on the basis of more marked suffixed forms, contrary to the usual direction of analogical change? (Lahiri and Drescher 1984; Bybee 1985, chap. 3)
- b. What role (if any) did similar words, like [soror], play in the change? Is there a minimum number of such words that are necessary to effect such a change?
- c. Why did [hono:s] change to [honor], and not [soror] to [soro:s]?⁶
- d. Why were monosyllables and neuters generally not affected?
- e. Why might both [o:s] and [or] variants have persist for so long?

In this chapter, I propose that the change of [hono:s] to [honor] was driven by more than just a phonological change involving paradigm uniformity constraints. I suggest that it was actually the result of the way that the morphological grammar of Latin noun paradigms was organized, with an oblique form serving as the base, and rules deriving the remaining forms (including the nominative singular). When the base identification of the previous chapter is applied to Latin, it emerges that the preferred base is an oblique form, not the nominative form. Furthermore, when an oblique form is used to project nominative forms, the system makes essentially the right predictions for the *honor* analogy: [-o:s] nominatives are strongly favored for monosyllabic and neuter nouns, [-or] nominatives are preferred for polysyllabic masculine and feminine nouns with [-o:s] remaining a strong second choice, and [-or] is strongly favored for agentive nouns.

This analysis is similar in spirit to the four-part analogy explanation of the change, but the gradient nature of the rules in this system gives us a quantitative expression of the influence of other lexical items (50b), and also helps predict which pattern should win out for each class of words (50cd). Finally, the close competition between [-o:s] and [-or] for many forms suggests an intriguing interpretation for the persistence of [-o:s] nominatives into the attested period: I conjecture that perhaps they were not merely a conservative retention of memorized archaic

⁵It is difficult to know in many cases whether the occurrence of forms like *hono:s* alongside *honor* reflects free variation, or simply literary archaism. Cicero, for example, systematically used the form [hono:s] instead of [honor], in both philosophy and oratory texts (elevated styles), as well as in letters (potentially less elevated/archaic); at the same time, he used [labor] instead of [labo:s] in all three contexts.

⁶One possible answer is that the paradigm of [hono:s] already had [r] forms in it, while the paradigm of [soror] never had [s] forms. Following Steriade (1994) we might call this the “lexical conservatism” analysis, in which speakers may only use or extend allomorphs that are already attested. It is possible that a lexical conservatism analysis could explain this part of the asymmetry ([honor] but not *[soro:s]), but it would tell us nothing about the effect of syllable count or gender, since monosyllables and neuters also had an available [r] allomorph which could have been extended to the nominative singular.

forms, but that they may also have supported to a certain extent by the synchronic grammar of Classical Latin. In other words, when Latin speakers heard an archaism like [hono:s], even if it was not the synchronically preferred form, it may have struck them as moderately grammatical, allowing [-o:s] forms to persist as an archaism much longer than some other archaic features. Clearly more philological work is needed to support this hypothesis, but it is a good example of how the current model that makes predictions not only about how forms can be innovated, but also about how they may be retained or lost.

4.2 Applying the model to Latin noun paradigms

In the small, hypothetical examples discussed in chapter 3, the neutralizations affected just one form in the paradigm, making the mappings in one direction clearly easier than those in the opposite direction. In real languages, however, the situation is rarely so clear. Neutralizations typically affect only a subset of the segments in the language, so the uncertainty that they cause may only affect a small number of words. Furthermore, neutralizations often affect different parts of the paradigm for different words. Thus, it is not always easy to intuit whether a mapping is easier in one direction than the other, or the magnitude of the asymmetry.

The question of interest for this chapter is whether Latin nouns were easier to project in the oblique → nominative direction than vice versa. Latin nouns are traditionally divided into five classes or declensions, each of which was inflected for five major cases: the nominative, genitive, dative, accusative, and ablative. (Two additional cases, the vocative and locative, were almost always identical to other cases.) A full description of all of the declensions and their subclasses is clearly beyond the scope of this chapter – see Leumann (1977), Kühner (1912), or Allen (1903) for in-depth discussions. What is important here is that the distinctions among many of these classes were neutralized or nearly neutralized in various parts of the paradigm.

As with Yiddish, it is possible to consider what neutralizations affected each slot in the paradigm. In the nominative singular, several morphological classes were neutralized by having the same case endings. For example, at least four classes of nouns all had the ending [-us] in the nominative singular: 2nd declension masculines like [populus] ‘people’; 3rd declension neuters, some with [-oris] ([korpus], [korporis] ‘body’) and some with [-eris] ([genus], [generis] ‘kind’); and 4th declension masculines, like [manus]. These are shown in the shaded row in (51a). Another ambiguous ending was the nominative singular [-er] (51b), which included both second declension (genitive [-i:]) and third declension ([-is]) nouns. Nouns with [-er] also exhibited a neutralization between [e]’s that alternated with ∅ in the oblique forms ([ag_ri:]) and those that did not ([generi:]). Finally, the noun [iter] exhibited a relatively idiosyncratic alternation between ∅ and [in].

(51) Morphological neutralizations in the nominative

a. Nouns ending in -us in the nominative

| | ‘people’ | ‘body’ | ‘kind’ | ‘hand’ |
|----------------|-----------|------------|-----------|----------|
| <i>nom.sg.</i> | [populus] | [korpus] | [genus] | [manus] |
| <i>gen.sg.</i> | [populi:] | [korporis] | [generis] | [manu:s] |
| <i>dat.sg.</i> | [populo:] | [korpori:] | [generi:] | [manui:] |
| <i>acc.sg.</i> | [populum] | [korpus] | [genus] | [manum] |
| <i>abl.sg.</i> | [populo:] | [korpore] | [genere] | [manu:] |

b. Nouns ending in *-er* in the nominative

| | 'field' | 'father- in-law' | 'brother' | 'prison' | 'journey' |
|----------------|---------|---------------------|------------|------------|------------|
| <i>nom.sg.</i> | [ager] | [soker] | [fra:ter] | [karker] | [iter] |
| <i>gen.sg.</i> | [agri:] | [sokeri:] | [fra:tris] | [karkeris] | [itineris] |
| <i>dat.sg.</i> | [agro:] | [sokero:] | [fra:tri:] | [karkeri:] | [itineri:] |
| <i>acc.sg.</i> | [agrum] | [sokerum] | [fra:trem] | [karkerem] | [iter] |
| <i>abl.sg.</i> | [agro:] | [sokero:] | [fra:tre] | [karkere] | [itinere] |

In addition to these (and other) morphological neutralizations, the nominative singular was also affected by several phonological neutralizations induced by the *-s* suffix. For example, the voiced stops [b] and [g] became devoiced in this environment, merging with [p] and [k], as in (52a,b). The coronal obstruents [t], [d], and [s] were deleted altogether in this environment, as in (52c).

(52) Devoicing before *-s*a. Neutralization of *g* with *k*

| | 'murder' | 'flock' |
|----------------|----------|----------|
| <i>nom.sg.</i> | [neks] | [greks] |
| <i>gen.sg.</i> | [nekis] | [gregis] |
| <i>dat.sg.</i> | [neki:] | [gregi:] |
| <i>acc.sg.</i> | [nekem] | [gregem] |
| <i>abl.sg.</i> | [neke] | [grege] |

b. Neutralization of *b* with *p*

| | 'feast' | 'city' |
|----------------|---------|---------|
| <i>nom.sg.</i> | [daps] | [urps] |
| <i>gen.sg.</i> | [dapis] | [urbis] |
| <i>dat.sg.</i> | [dapi:] | [urbi:] |
| <i>acc.sg.</i> | [dapem] | [urbem] |
| <i>abl.sg.</i> | [dape] | [urbe] |

c. Deletion of *t*, *d*, and *s*

| | 'foot' | 'lawsuit' | 'penny' |
|----------------|---------|-----------|---------|
| <i>nom.sg.</i> | [pe:s] | [li:s] | [a:s] |
| <i>gen.sg.</i> | [pedis] | [li:tis] | [assis] |
| <i>dat.sg.</i> | [pedi:] | [li:ti:] | [assi:] |
| <i>acc.sg.</i> | [pedem] | [li:tem] | [assem] |
| <i>abl.sg.</i> | [pede] | [li:te] | [asse] |

Another phonological neutralization was caused by restrictions on word-final clusters, which resulted in the deletion of final coronals in clusters: *[kord] > [kor] 'heart', *[lakt] > [lak] 'milk':⁷

⁷There appear to be no neutralizations in noun paradigms caused by *[kt] > [k], since nouns ending in simple [k] end in *-s* in the nominative: [paks]/[pa:kis] 'peace-nom./gen.'.

(53) Neutralizations caused by deletion of coronals

| | 'heart' | 'sister' |
|----------------|----------|------------|
| <i>nom.sg.</i> | [kor] | [soror] |
| <i>gen.sg.</i> | [kordis] | [soro:ris] |
| <i>dat.sg.</i> | [kordi:] | [soro:ri:] |
| <i>acc.sg.</i> | [kor] | [soro:rem] |
| <i>abl.sg.</i> | [korde] | [soro:re] |

The nominative was not the only form with neutralizations, however. Oblique forms also suffered from both morphological and phonological neutralizations.

The (morpho)phonological rhotacism process that yielded the [s] ~ [r] alternation in [hono:s] ~ [hono:ris], for example, created a neutralization in oblique forms, between alternating [s] ~ [r] and non-alternating [r]:

(54) Neutralization of /s/ and /r/ caused by rhotacism

| | 'honor' | 'sister' |
|----------------|------------|------------|
| <i>nom.sg.</i> | [hono:s] | [soror] |
| <i>gen.sg.</i> | [hono:ris] | [soro:ris] |
| <i>dat.sg.</i> | [hono:ri:] | [soro:ri:] |
| <i>acc.sg.</i> | [hono:rem] | [soro:rem] |
| <i>abl.sg.</i> | [hono:re] | [soro:re] |

In addition, a process of vowel reduction in non-final unstressed syllables caused several vowels to surface as [i] in oblique forms:

(55) Neutralization of /e/, /u/, /i/ to [i] in oblique forms

| | 'soldier' | 'head' | 'pyramid' |
|----------------|------------|-----------|-------------|
| <i>nom.sg.</i> | [mi:les] | [kaput] | [pyramis] |
| <i>gen.sg.</i> | [mi:litis] | [kapitis] | [pyramidis] |
| <i>dat.sg.</i> | [mi:liti:] | [kapiti:] | [pyramidi:] |
| <i>acc.sg.</i> | [mi:litem] | [kaput] | [pyramidem] |
| <i>abl.sg.</i> | [mi:lite] | [kapite] | [pyramide] |

These are just a few of the more serious neutralizations that would have occurred in the nominative and oblique forms in Latin. Unlike Yiddish verbs, which involved just a few neutralizations, it would be difficult in the case of Latin nouns to give a comprehensive list of all of the possible neutralizations, because there were many more inflectional classes (including nouns of three genders), and more irregular morphophonological processes involved. This sampling of the complexity of the system should at least serve to show that the choice of a base form is not nearly as straightforward as it was in the synthetic languages of chapter 3, or in Yiddish verbs. What we must take into account, therefore, is how serious these neutralizations were, by considering the number of lexical items involved.

Some of these neutralizations affect relatively large numbers of words. The neutralization between masculine ([-us]) and neuter ([-um]) second declension nouns in the genitive (both [-i:]) involves two very large (and productive) classes of nouns. The neutralizations caused by voicing agreement in final obstruent clusters in (52a,b), on the other hand, affected relatively fewer words. An additional complication is that other factors, such as grammatical gender,

could help the speaker know which suffix to use in a potentially ambiguous situation – so, for example, the use of [-us] or [-um] in the nominative of a word with [-i:] in the genitive is almost completely predictable given the gender of the word.⁸ Thus, if there is an asymmetry in predictability between nominative and other forms, it would be because of differences in the “severity” of the neutralizations involved, and the ability to predict the correct form using gender. Were the neutralizations in the nominative in fact more severe than in other cases?

In order to answer this question, I started with a database of fully inflected classical Latin nouns, prepared in 1997-1998 by a group working under the supervision of Bruce Hayes at UCLA. This database contained all of the nouns with five or more tokens in a lemmatized frequency count from classical texts (Delatte, Evrard, Govaerts, and Denooz 1981), based on a corpus of approximately 800,000 words (582,000 from prose, 212,000 from poetry). Nouns beginning with the letters R through Z were omitted from the simulations because the database was found to have incomplete information for many paradigms in this section of the alphabet. Nominative forms were listed in their forms prior to the [hono:s] > [honor] change; in cases of uncertainty, words were listed with a final [s]. The rationale for this was that we are interested in seeing if the model will favor [r] forms in spite of numerous [s] forms in the training data, and we do not want this to be the result of the influence of spurious [r] forms.

The model of base selection being tested here is that learners evaluate the usefulness of prospective bases early in the learning process. Therefore, the only input data which would be available to the learner for comparisons would be the most common words. As an idealization, words with 50 or more tokens in Delatte, Evrard, Govaerts, and Denooz (1981) were selected, for a total of 494 input nouns. Six forms were considered as possible bases: the nominative, genitive, dative, accusative, and ablative singular, and the nominative plural.⁹ For each possible base form, training data files were then constructed to project each of the remaining forms (*nom.* → *gen.*, *nom.* → *dat.*, *nom.* → *acc.*, etc.), yielding 30 (=6×5) training sets in total.¹⁰

Nouns in the input files were listed in phonemic transcription. In order to take into account the effect of phonological processes, a list of illegal sequences was also prepared, including final clusters disagreeing in voicing (*bs#, *gs#, *ds#), final geminates (*ll#, *dd#, *ss#), the clusters *rts, *lts, and *nts, and a few other illegal sequences whose repair caused alternations in the nominative (*o:r#, *kt, *ii:). Because some of these illegal sequences refer crucially to word boundaries, word boundaries were also marked explicitly in the input files with brackets. Since rhotacism is not surface-true in this stage of Latin (cf: [ka:sus] ‘fall’, [rosa] ‘rose’), intervocalic

⁸There were a few isolated second declension neuters ending in [us] in the nominative, including [vulgus] ‘people’, [pelagus] ‘sea’, and [virus] ‘poison’.

⁹The remaining plural oblique cases were not considered because they never preserved distinctions that were neutralized elsewhere, and in fact some cases (especially the dative and ablative) involved massive neutralizations between different declension classes.

¹⁰All of the training input sets and results files for the simulations discussed here, as well as the original database of nouns, can be downloaded from <http://www.linguistics.ucla.edu/people/grads/aalbrigh/papers/latin.html>.

[s] was not included as an illegal sequence (Hoenigswald 1960, pp. 106-107).¹¹ Finally, each noun was provided with a numeric code indicating the grammatical gender and the number of syllables (monosyllabic vs. polysyllabic), since the current implementation of the minimal generalization learner does not have an independent capacity for considering general prosodic properties of words. These codes allowed words to be categorized as masculine, feminine, or neuter, and as monosyllabic or polysyllabic. Token frequencies were also included in the input files, but they were not employed in the simulations reported here.

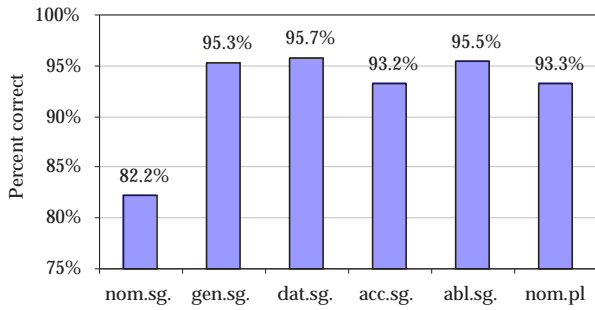
The input files were submitted to the minimal generalization learner, yielding subgrammars of rules with confidence values. The word-specific rules were then eliminated, and the resulting subgrammars were tested on the input forms. The metrics proposed in section 3.3 were calculated for each subgrammar, to obtain an estimate of the usefulness of each slot in the paradigm for predicting the remainder of the paradigm. The results, given in Appendix B, show that the predictability between *all* forms is quite high (over 80%); it is not the case that any part of the paradigm suffers from neutralizations that affect the majority of nouns in the input set. Nevertheless, the oblique forms tend to be substantially better than the nominative form on almost all of the metrics considered. In Figure 4.1, the candidates for base status are compared in terms of their mean effectiveness in projecting the five other forms in the paradigm.

As can be seen, the criteria proposed in section 3.3 generally agree on the relative effectiveness of the various forms as possible bases. The combined results from all five remaining criteria (excluding number of rules) are shown in Figure 4.2.

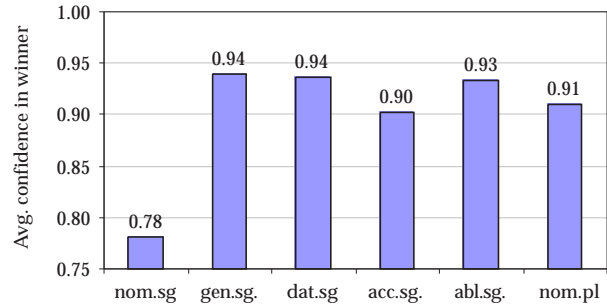
The nominative is the worst choice of base under all criteria, and thus receives the lowest rank for all metrics. This reflects the fact that the nominative suffers from more neutralizations, affecting both more words and more segments, than the oblique forms. Interestingly, the accusative also fares relatively poorly, because it is the same as the nominative for all neuter nouns, and thus shares many of the same neutralizations. Among the remaining forms, the dative comes out slightly ahead of the genitive and ablative.

An additional factor that has not been discussed here is the relative frequency of the different forms in the paradigm. As an idealization, I have assumed that learners have access to the six candidate forms in equal proportions for all nouns. Clearly this is not true in real life however;

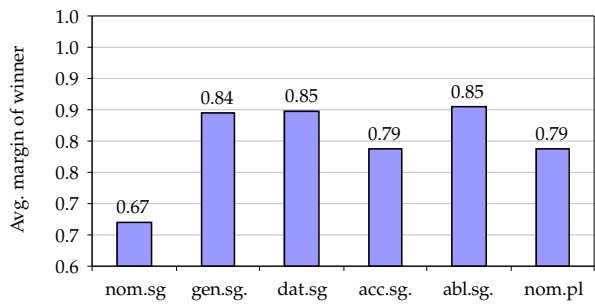
¹¹The presence of intervocalic [s] does not necessarily preclude the possibility that rhotacism continued to be a synchronically active process in Latin, possibly restricted to a particular morphological environment, such as /V__+V, where '+' indicates a morpheme boundary. For the purposes of the current model, the synchronic status of rhotacism actually makes very little difference. Including a *Vs+V constraint would improve the reliability of nominative → oblique grammars slightly, because the model could learn to apply rhotacism in mappings like [hono:s] → [hono:ris] (instead of [hono:sis]); however, -o:s nominatives make up only a small fraction of the language as a whole, so improving the model's predictions for this subset of the vocabulary does not make a substantial difference in the calculations reported below. Note also that including a rhotacism constraint does not help the model at all in the oblique → nominative direction, since an oblique form with -o:ris could come from either underlying /s/ or underlying /r/. More generally, assuming that rhotacism was synchronically active in Latin can help to explain why the paradigm of 'honor' was not leveled to [hono:s], [hono:sis], etc., but it cannot explain other facts, like why speakers did not assume that forms like [soro:ris] were also the result of rhotacism (predicting the incorrect nominative [soro:s]), or why speakers did not continue to tolerate the rhotacism alternation. It could also be added that many authors have tried to make use of the exact opposite intuition, arguing that since rhotacism had been obscured by numerous exceptions caused by borrowings, dissimilations, and degemination of *ss > s, it was no longer synchronically active in Latin. If we assume that rhotacism was no longer productive, then we can understand why [hono:s] and [hono:ris] could no longer be related to one another by an automatic phonological process, and why the alternation was then open to leveling (Hoenigswald 1960, pp. 108-109; Klausenburger 1979; Wetzels 1984; Barr 1994, pp. 519-524; and others).



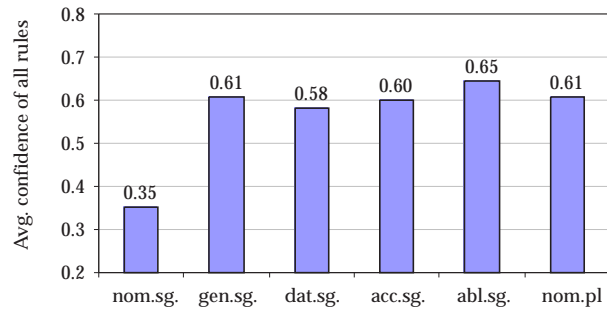
(a) Percent correct



(b) Avg. confidence in winner



(c) Avg. margin between winner and nearest competitor



(d) Avg. confidence of rules

Figure 4.1: Comparison of potential bases according to the criteria in §3.3

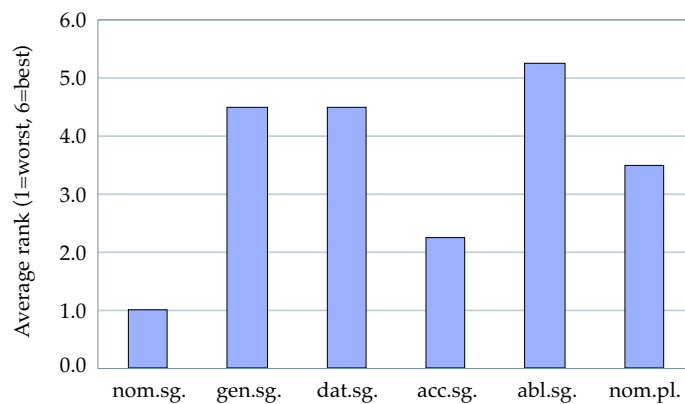


Figure 4.2: Average rank of forms as potential bases

some cases are more frequent than others, and the frequency of cases may differ from word to word. As mentioned on page 40, the Albright and Hayes implementation of minimal generalization uses confidence statistics to estimate the effectiveness of rules, so that rules covering a few forms are penalized more than rules covering many forms. Therefore, with more realistic input data, including different amounts of data about different cases, subgrammars involving less frequent cases would be penalized because their rules would be based on fewer forms. I will return to this issue section 6.2.1.

A simulation taking this into account would require more detailed frequency information about Latin noun paradigms than is currently available to me. Nevertheless, intuitively, it seems that there are substantial differences in the frequency of the oblique cases, and this could be the decisive factor in choosing a base from among the oblique forms that are more or less equivalent by all other criteria. (In section 4.4.2, I will show that at least for a small sample of words in a small text corpus, the ablative and genitive are both relatively frequent forms.) For the purposes of the [hono:s] > [honor] change, it is sufficient that the model proposed here select something other than the nominative as the base form; in the discussion that follows, I will use the genitive singular as the base for deriving the nominative, but the same result could be achieved using the ablative or dative singular.¹² This proposal is in line with Hooper's claim (Hooper 1976, p. 95) that an oblique form was the basic form in Latin, although Hooper claims (without much discussion) that it was in fact the accusative, not the genitive or dative. A possible solution that is compatible with both of these claims is that the global base of Latin noun paradigms was the dative or accusative, while the nominative was derived more locally from the accusative; I will discuss this possibility in more detail in section 6.3.

4.3 Projecting nominatives from the genitive

Choosing an oblique form as the base in Latin noun paradigms gives us only half of the explanation for the [hono:s] > [honor] change. In particular, it explains the “backwards” direction of the change (oblique forms affecting nominatives). This answers the question in (50a), of why it was the nominative that changed in Latin. What remains to be shown, then, is that once an oblique form has been chosen as the base, the model makes the right predictions for nominative forms: namely, that polysyllabic non-neuter *-o:s* nouns changed to *-or*.

Recall that an assumption of the current model is that bases are selected early in the learning process, but learners continue to fine-tune their grammars to derive the remainder of the paradigm. Therefore, in order to test the predictions of the model for nominatives using an oblique form as the base, the model was trained on the full set of 1,687 nouns in the gen.→nom. direction. The resulting grammar was then used to generate possible nominatives for all genitive forms ending with sequences that could potentially arise from rhotacism: [-o:ris], [-oris], [-uris], [-eris] (157 in all). The grammar derived several possible nominatives for each noun, each with its own confidence value. For example, for [-o:ris] and [-oris] genitives, the possible nominatives typically included an [-o:s] nominative, an [-or] nominative, and various other possibilities, such as [-o:ris] (on the basis of words like [kanis] ‘dog-nom./gen.sg.’, which were

¹²A common intuition is that the [hono:s] > [honor] change may be due to the collective influence of *all* of the oblique forms combined, and not the effect of a single oblique form on the nominative singular. I will discuss this possibility further in section 4.4.1.

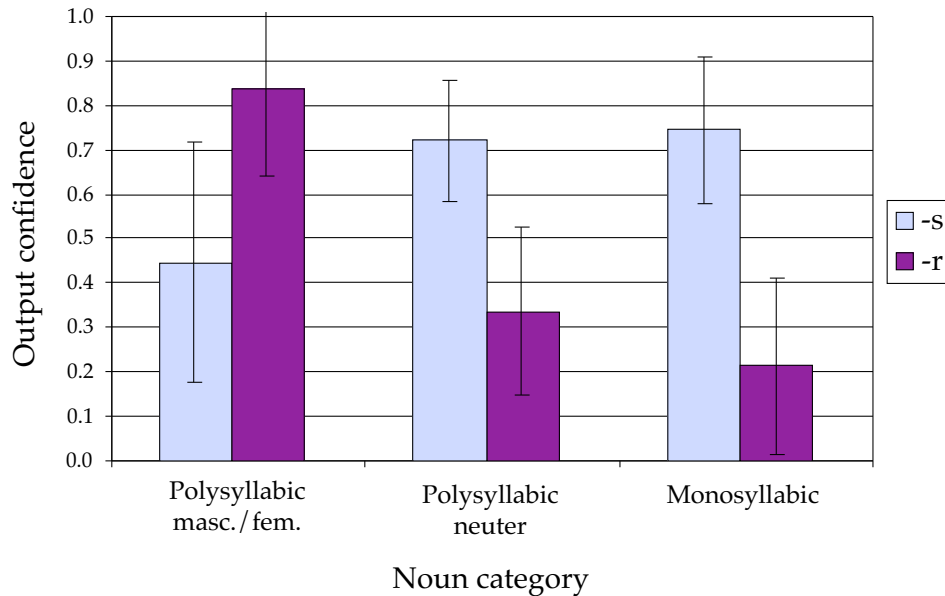


Figure 4.3: Preference for *-r* or *-s* in different categories of nouns

identical in the nominative and genitive), [-us] (like [korpus] ‘body-nom.sg.’), etc. For each noun, the best possible *-r* nominative was compared against the best possible *-s* nominative, in order to gauge the model’s preference for *-r* nominatives. As expected, the preference for *-r* or *-s* varied substantially from word to word. For masculine and feminine polysyllabic nouns, *-r* was generally favored, while for neuters and monosyllables, *s* was preferred (Figure 4.3), mirroring the observed outcome of the *honor* analogy. (Note that in the graph, bars indicate the size of a standard deviation, not the standard error.)

Why does this pattern emerge? The differing strength of *-r* and *-s* for different words is due to the fact that the system employs multiple rules, with different confidence values in different contexts. Among polysyllabic non-neuter nouns, genitives in [-o:ris] frequently have nominatives in *-or*. Thus, the rule of [o:ris] → [or] / [X]_{polysyll,-neut} # has a relatively high confidence (.727), correctly deriving words like [soror] and [cruor], and all agentives, but failing for words like [hono:s]. Among these forms, then, there is a slight preference for *-or* in the nominative.

Furthermore, within the masculine and feminine polysyllabic nouns, two groups can be distinguished (Figure 4.4). The first are the agentive nouns, which strongly favor *r* in all cases. In fact, these words contained *-r* etymologically, and continued to have *-r* in the nominative with no variation or hypercorrections. The second are the non-agentive masculine and feminines, which show a slight tendency to favor *-r*, but with strong competition from *-s*. This is the *honor* class of words, which were etymologically *-s* but changed to *-r*, with some attested variation and occasional hypercorrections of etymological *-r* to *-s* (Neue-Wagener 1902, p. 265).

The model can capture this difference because it is able to posit a more specific rule that covers just the agentives; these are not only polysyllabic and non-neuter, but they also all have

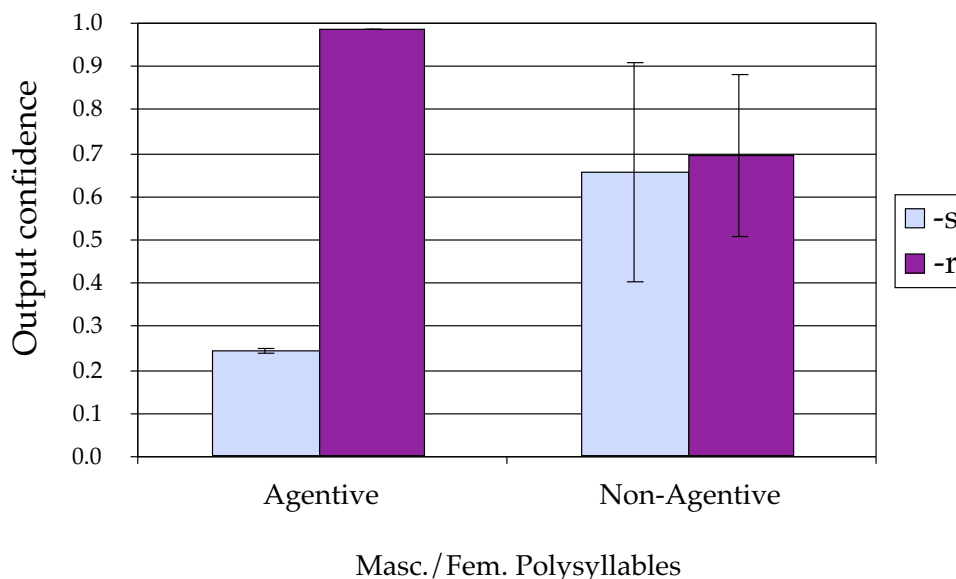


Figure 4.4: Comparison of agentive vs. non-agentive nouns

a stem-final [s] or [t]: *doctor*, *audītor*, *cēnsor*, etc.¹³ Thus, the more specific rule [o:ris] → [or] / [X {s,t}]_{polysyl,-neut}___# is able to describe the agentives quite narrowly, and has an extremely high reliability (.980).

Outside the class of polysyllabic non-neuter nouns, the reliability of *-r* is much lower. Among polysyllabic neuters, very few nominatives end in *-r*, so the general rule [ris] → [r] / [X]_{polysyl,+neut}___# has a rather low confidence (.196). Among this group of words, the competing rule [ris] → [s] / [X]_{polysyl,+neut}___# has a high reliability, meaning that for these forms, non-uniformity (anti-correspondence) prevails, and there is no change. There are, however, a few local pockets of *-r* nominatives among the neuters, especially among those with *-aris* and *-eris* in the genitive ([kalkar]/[kalka:ris] ‘spur-nom./gen.’, [nektar]/[nektaris] ‘nectar-nom./gen.’, [aker]/[akeris] ‘maple-nom./gen.’) This is the reason why the model disfavors innovative *-r* nominatives among neuters, and why there is also a good deal of item-by-item variation among them.¹⁴ Finally, among monosyllables, almost all *-ris* genitives had nominatives ending in *-s* – a notable exception being [fu:r] ‘thief’. Thus, the model correctly learns that in this environment, the *s* ~ *r* alternation is extremely robust, and final *-r* cannot compete with it.

It should be reiterated that the slight preference of the model for *-r* in words like [honor] emerges in spite of the fact that they were listed with *-s* in the training data. In other words, the grammar produces an output which is different from the existing form. Therefore, under this analysis, pre-change forms like [hono:s] would have been considered irregular, and would have

¹³This common phonological trait is not a coincidence: agentives were formed from the perfect passive participle (4th stem), which was generally formed by adding a [t], or in some phonological contexts, by changing a stem consonant to [s] (e.g., [kad-] → [ka:s-] ‘fall’).

¹⁴In fact, Kieckers (1960) points to one, possibly quite isolated example of a neuter noun with etymological *-r* being written with an *-s*: *femus* ‘femur’ (vol 2, §22).

had to have been listed as exceptions to the *-ris* ~ *-r* pattern. Of course, if learners had perfect memories and access to all forms of all words, then they could perfectly well have memorized [hono:s] and continued to produce it, and the language would not have changed. However, in real life this is not the case, and speakers must sometimes synthesize new forms. The model is intended to predict what forms a speaker would produce in such situations, and in this case it correctly predicts errors, or overregularizations, like *honor*. It cannot, however, predict when existing forms would be unavailable—either because they have never been encountered, or because lexical access has failed—forcing speakers to use their grammars. What is missing from this model, then, is a production mechanism which uses both the lexicon and grammar to produce forms. Even so, the result in (4.3) is still significant, because we can assume that speakers do sometimes make overregularization errors (Marcus, Pinker, Ullman, Hollander, Rosen, and Xu 1992; Pinker 1999), and the errors that the model makes correctly mirror the attested historical change.

It is worth mentioning that the [hono:s] > [honor] change was just one of many changes that affected nominatives in the history of Latin. Numerous nouns with highly irregular nominative forms were regularized; for example, the nominative of [juppiter]/[jowis] ‘Jupiter-nom./gen.’ was eventually replaced by [jowis] (cf. [kanis]/[kanis] ‘dog-nom./gen.’), and the nominative of [bo:s]/[bowis] ‘cow-nom./gen.’ is attested as [bowis] (Kieckers 1960, vol. 2, §I.21.3; Kühner §63.2). The converse change, of regularizing nouns by fixing the oblique forms, generally did not occur (*[juppitris] or *[juppitri:], *[bo:ris], etc.).¹⁵ Furthermore, the form of nouns in modern Romance languages can generally be traced back to oblique forms in Latin. For example, Latin [pe:s]/[pedis] ‘foot-nom./gen.’ has yielded Italian [pjede], instead of the expected *[pe] (cf. Latin [tre:s] > Italian [tre] ‘three’); similarly [ars]/[artis] ‘skill’ > Ital. [arte], [flo:s]/[flo:ris] ‘flower’ > Ital. [fjore], [niks]/[nivis] ‘snow’ > Ital. [neve], and so on. The analysis of Latin presented here helps to explain why nominatives were subject to many changes in the later history of Latin, and not just leveling of *s* ~ *r* alternations. The Uniform Exponence account of the Latin change presented above, on the other hand, does not predict leveling of alternations other than *s* ~ *r*.

4.4 Discussion

This analysis captures two common intuitions about Latin nouns, and about the [hono:s] to [honor] change. The first is that oblique forms are “more revealing” about the declension of a noun than the nominative – seen, for example, in the common practice of listing both the nominative and genitive forms in dictionary entries, as the nominative alone is not considered

¹⁵There seems to be at least one case in which a property of the nominative was extended to the remainder of the paradigm: (1) the paradigm of words like [vo:ks] ‘voice’ originally had a long [o:] and [k] in the nominative, and a short [o] and [k^w] elsewhere ([vo:ks], [vok^wis], etc.; see Meiser 1998, p. 141 regarding vowel length, Leumann 1977, p. 148 and Kieckers 1931, p. II.13 regarding [k^w]). The long [o:] and simple [k] of the nominative were subsequently extended to all forms: [vo:ks], [vo:kis], etc. It would be interesting to compare the relative chronology of these changes, since these nominative-driven changes may have occurred at an older stage of the language in which nominatives suffered from fewer phonological reductions, while changes driven by obliques occurred throughout the Classical and Late Latin period.

“informative enough” to predict the entire paradigm.¹⁶ This intuition is reflected in the current analysis by the fact that an oblique form is chosen as the base, and the remainder of the paradigm is derived from an oblique form.

The second intuition is that the change from [hono:s] to [honor] involved replacing a small, irregular, morphologized alternation with the more general default pattern of non-alternation. This analysis shares with Barr (1994) the idea that this can be captured with competing rules, but differs with respect to why different classes of words were treated differently. Barr claims that monosyllabic nouns and neuter nouns retained *s* for fundamentally different reasons. The monosyllabic nouns, she argues, retained *s* because in a shorter word, the alternating $s \sim r$ segment constituted a larger proportion of the word, and was therefore more salient and more easily retained. The neuters, on the other hand, retained *s* because it occurred in two slots in the paradigm (both the nominative and the accusative) instead of just one, and was more salient for this reason. Both of these points seem to involve arbitrary thresholds. Even if we grant that the final segment of *flo:s* is “a larger portion of the word” than the final segment of *odo:s* (which both have four phonemes, but do differ in their syllable count), is the extra vowel in *odo:s* really enough to distract learners from reliably noticing the final $s \sim r$? And in the case of neuters, why was occurring in two slots in the paradigm sufficient to guarantee that the *s* would be preserved, when occurring in two slots was not enough to guarantee that [ɔ] would be preserved in Polish feminine diminutives (5b) (among many other cases)?

In the current system, the difference between different genders and word lengths is attributed to the existence of multiple versions of the rules in question, at varying levels of generality, and with differing reliability in different contexts. The use of multiple overlapping rules might be seen as unwanted redundancy in the model, but in fact cases like Latin are taken as evidence that speakers, too, have at least a certain amount of detailed knowledge about the reliability of different processes in different environments. Furthermore, an ability to assess the reliability of rules in different environments is required in any event in order for learners to locate the best morphological and phonological rules to describe the patterns of their language.

Although this analysis makes use of several intuitions about the factors that are thought to drive paradigm leveling, it ignores certain other factors that have been proposed in the literature. Some notable factors that do not play a role in this analysis are the frequency of an allomorph within the paradigm, the token frequency of various surface forms, or the semantic naturalness of different nouns in different cases. It is useful to consider, therefore, the extent to which these other factors could provide an alternative explanation of the [honor] analogy, and whether the current model would benefit from incorporating any of these factors.

4.4.1 Frequency of occurrence within the paradigm

It is often suggested that the [hono:s] to [honor] change was encouraged by the fact that every form in the paradigm except the nominative singular contained [r]. I will refer to this as the “majority rule” hypothesis. Under the model proposed here, each paradigm has a single unique base, and forms are derived by grammars relating individual pairs of forms. If the most

¹⁶The relative uninformative-ness of the nominative in Latin is due, in part, to the fact that the nominative suffix for one large class of nouns lacked a vowel (-s), creating coda clusters that resulted in phonological simplifications (e.g., **arts* > *ars*). The oblique forms always provided a prevocalic context for the stem, resulting in far fewer neutralizations; rhotacism is a rare exception.

informative form had turned out to be the nominative singular, the prediction of this model is that all of the remaining forms could have been rebuilt on the basis of a single form. Thus, this model has no way to capture the majority rule intuition. It is not clear to me, however, that there is evidence that paradigm leveling is truly driven by majority rule. There are numerous cases in which a single form seems to have driven a paradigmatic change—for example, the leveling in Polish masculine diminutives in chapter 1 and the Yiddish change discussed in chapter 2 both involve the extension of a form that is vastly outnumbered in the paradigm. Conversely, it is difficult to prove that a leveling like [hono:s] > [honor] would not have happened if [s] had occurred in more slots in the paradigm. Barr (1994, p. 543) points out there was a difference between masculines and feminines (like [hono:s]) and neuters (like [tempus]): masculines and feminines had [s] in just one slot in the paradigm (the nominative), while neuters had [s] in two slots (the nominative and accusative). She suggests that this alone was enough to produce the difference between these classes of words. However, it seems unlikely that this threshold would work in general. Furthermore, this majority rule hypothesis does not actually explain the Latin facts. First, even if we accept that having two slots with [s] was enough to protect the neuter nouns from leveling, there is still the problem that monosyllabic masculine and feminine nouns had only one [s] in the paradigm (the nominative singular), but they did not change either. Furthermore, a few neuter nouns did change, or at least acquired [r] variants (e.g., [kinus]/[kiner], ‘ash’ (neuter)), in spite of the fact that they had more than one [s] variant. From the point of view of paradigm-internal pressures, there is no reason why these words should have behaved differently. In addition, there are many other noun paradigms in which the nominative had a different form from the rest of the paradigm, but was not leveled (e.g., [iter] ~ [itineris] ‘road-nom./gen.sg.’). Therefore, frequency of occurrence within the paradigm does not seem to add anything to the account of the change.

4.4.2 Token frequency of different paradigm members

A natural hypothesis, pursued by Mańczak (1958) and others, is that less frequent forms are often rebuilt on the basis of more frequent forms within the paradigm. Could it be the case that the nominative was significantly less frequent than the oblique forms in Latin? This is especially relevant in Latin because many or most of the words affected by the [honor] analogy were inanimate or abstract nouns, which are perhaps more frequent in oblique forms than in the nominative.

In order to get a rough (and very informal) estimate of the relative frequency of case forms for different nouns, I performed some counts on the complete works of Cicero, as found in the Perseus Digital Library (<http://www.perseus.tufts.edu/>). As (4.1) shows, it is true that among singular forms, nominatives rarely constitute the majority of tokens for any noun. This might possibly help to explain why nominative forms were open to rebuilding in Latin – perhaps they were not frequent enough to be reliably memorized and retrieved.¹⁷

¹⁷This explanation is not really convincing without a more explicit theory of how frequent a form must be before it can be reliably memorized and retrieved. In fact, 20% of the tokens for a relatively frequent noun seems like sufficient exposure to remember and maintain the form. I am simply granting for the sake of argument that perhaps the lower token frequency of nominatives in Latin could have made them susceptible to leveling.

¹⁸A problem arises in counting frequencies for neuters, since the nom. and acc. forms are identical. The hypothesis being tested here is that the frequency of [s] forms in the paradigm determines their susceptibility to leveling, so I have counted all *s* forms in the nom. column, to facilitate comparison with the masc. and fem. nouns.

Table 4.1: Distribution of singular tokens for some Latin nouns

| Noun | Total Sg. | Nom. | Gen. | Acc. | Abl. |
|-------------------------------|-----------|------|------|------|------|
| Polysyllabic, non-neuter | | | | | |
| <i>hono:s/ honor</i> ‘honor’ | 285 | 19% | 25% | 18% | 32% |
| <i>labo:s/ labor</i> ‘work’ | 163 | 17% | 21% | 37% | 35% |
| <i>odo:s/ odor</i> ‘odor’ | 4 | 50% | 0% | 0% | 50% |
| Monosyllabic | | | | | |
| <i>flo:s</i> ‘flower’ (masc.) | 16 | 25% | 0% | 56% | 19% |
| <i>mo:s</i> ‘custom’ (masc.) | 146 | 18% | 2% | 17% | 63% |
| <i>o:s</i> ‘mouth’ (neut.) | 65 | 29% | 9% | 6% | 55% |
| Neuter ¹⁸ | | | | | |
| <i>corpus</i> ‘body’ | 174 | 21% | 47% | – | 28% |
| <i>onus</i> ‘burden’ | 40 | 45% | 40% | – | 15% |
| <i>tempus</i> ‘time’ | 935 | 32% | 15% | – | 51% |
| Masculine, agentive | | | | | |
| <i>re:x</i> ‘king’ | 207 | 23% | 21% | 27% | 18% |
| <i>homo:</i> ‘man’ | 1049 | 19% | 23% | 35% | 12% |
| <i>senator</i> ‘senator’ | 43 | 33% | 23% | 28% | 14% |

What these counts cannot explain, however, is why the change should have been restricted only to the non-neuter polysyllabic nouns. The nominative does not seem to be less frequent in this class of nouns than in any other class. Furthermore, there is apparently not even a difference between masculine agentive nouns like ‘king’, ‘man’ and ‘senator’, and inanimate, abstract nouns like ‘honor’ or ‘custom’. Thus, a frequency-based account can explain only the direction, but not the details of the [hono:s] > [honor] change.

4.4.3 Semantics and local markedness

Another intuition, related to token frequency, but logically distinct from it, is that the semantics of particular lexical items make them more “natural” in some case forms than in others. Tiersma (1982), for example, shows that singular forms in Frisian have been rebuilt on the basis of plural forms, but just for those nouns which occur more naturally in the plural than in the singular (such as ‘teeth’ or ‘geese’). He refers to this phenomenon as *local markedness*. On the whole, we would expect local markedness to be reflected in token frequency, which is much easier to measure, but does not provide an adequate explanation of the Latin change (see above). However, pursuing the Jakobsonian view of markedness, one might attempt to come up with a definition of case markedness as distinct from token frequency; I assume that among the least marked nominatives, in this case, would be things like agentive nouns, since they would be most likely to act as (nominative-marked) agents. I have no estimate of the naturalness of the nominative forms which changed from [-o:s] to [-or], but I see no reason why this would fare any better than token frequency as an explanation of the change. Nouns like [onus] ‘burden’, [korpus] ‘body’, and [flo:s] ‘flower’ seem to me to be just as “non-agentive” as [hono:s] ‘duty’,

[odo:s] ‘odor’, or [arbo:s] ‘tree’. It appears that the class of nouns that changed is best defined by prosodic and morphological properties, and adding a sensitivity to frequency or semantics would not improve the model’s predictions in this case.

4.4.4 Leveling vs. extending alternations

The analysis of paradigm leveling proposed here relies on a strong pre-existing pattern of non-alternation in the lexicon – in this case, the non-alternation of [r]. This proposal immediately raises two related questions: first, if paradigm uniformity is really just the extension of an existing pattern of non-alternation, then what happens when the dominant pattern is alternation? Why does there seem to be a universal tendency towards leveling?

As an example of a language with a dominant pattern of alternation, consider a previous stage of Korean (Martin 1992):

(56)

| / __ # | Example | / __ V | (ACC -i) | gloss |
|--------|---------|--------|----------|------------|
| [tʰ] | [patʰ] | [tʰ] | [patʰi] | ‘field’ |
| [t] | [tʃʌt] | [d̥] | [tʃʌd̥i] | ‘milk’ |
| [tʰ] | [kʰotʰ] | [tʰ] | [kʰotʰi] | ‘flower’ |
| [t] | [ot] | [s] | [osi] | ‘clothing’ |

As (56) shows, all stem-final coronal obstruents alternate with [tʰ] word-finally. As with Latin, this alternation could be expressed as the result of a markedness constraint against manner and laryngeal specifications in coda position (favoring [tʰ]) outranking faithfulness constraints (which preserve underlying contrasts). If there was a universal pressure for uniform exponence constraints to move above IO-Faithfulness constraints, then we would expect that paradigmatic changes in Korean should bring Korean closer to non-alternating paradigms, perhaps as in (57). (The intervocalic voicing of /t/ → [d] is a completely predictable process in Korean.) Note that although the phonotactics of Korean rule out a completely non-alternating paradigm ([natʰ] ~ *[nati], or *[nad] ~ [nadi]), we may assume that the relatively minor, predictable allophonic alternation between [tʰ] and [d] better satisfies Uniform Exponence than a [tʰ] ~ [s] or [tʰ] ~ [tʰʰ] alternation, just as the shortening of final /o:r/ → [or] in Latin is assumed to be a less serious violation of Uniform Exponence than a [s] ~ [r] alternation is.

(57) Expected Korean paradigm leveling:

| / __ # | | / __ V | | |
|--------|---------|--------|---------|------------|
| [tʰ] | [patʰ] | [d] | [padi] | ‘field’ |
| [t] | [tʃʌt] | [d] | [tʃʌdi] | ‘milk’ |
| [tʰ] | [kʰotʰ] | [d] | [kʰodi] | ‘flower’ |
| [t] | [ot] | [d] | [odi] | ‘clothing’ |

In fact, the attested change in Korean noun paradigms is quite different. As it turns out, the majority of coronal obstruent-final stems contained [s] or [tʰ] etymologically (i.e., most were like [kʰotʰ]/[kʰotʰi] or [otʰ]/[osi]), and many Korean noun paradigms are being rebuilt to contain [s] or [tʰ]:

(58) Actual change in Korean Paradigms: (Martin 1992; Hayes 1995, 1998)

| / __ # | | / __ V | | |
|--------|---------|------------------|---------------------------|------------|
| [tʰ] | [patʰ] | [tʰ], [tʰʰ], [s] | [patʰi], [patʰʰi], [pasi] | 'field' |
| [t] | [tʰʌtʰ] | [s], [dʒ] | [tʰʌsi], [tʰʌdʒi] | 'milk' |
| [tʰ] | [kʰotʰ] | [tʰʰ], [s] | [kʰotʰʰi], [kʰosi] | 'flower' |
| [t] | [ot] | [s] | [osi] | 'clothing' |

Although there is a considerable amount of word-by-word and speaker-by-speaker variation, it is clear that the restructuring underway in Korean is introducing, not eliminating, alternations. For the most part, the dominant alternations of [tʰ] ~ [s] and [tʰ] ~ [tʰʰ] are coming to replace other, arguably less drastic alternations like [t] ~ [d].

The explanatory challenge, therefore, is to explain why in some cases a pattern of alternation is extended (as in Korean), while in other cases, alternations are eliminated (as in Latin). The model of paradigm learning advocated in this chapter always extends the strongest pattern, regardless of whether it is alternating or uniform. The reranking of paradigm uniformity constraints, on the other hand, can explain only leveling; the spread of alternations would have to be handled by other means, such as anti-correspondence constraints (Hayes 1999), leaving us with no explanation for why sometimes paradigm uniformity wins out, and sometimes anti-correspondence wins out. I will discuss another example in which an alternation was extended in the next chapter. In chapter 6, I will also return to the question of why leveling may be somewhat more common than anti-correspondence typologically.

4.4.5 Local summary

In this chapter, I have argued that the Latin [hono:s] > [honor] change was caused by more than simply a sporadic pressure for paradigm uniformity or uniform exponence constraints to assert themselves over IO-Faithfulness constraints. I have shown that the spread of [r] to nominative forms did more than just create uniform paradigms; it also extended a pattern of non-alternation that was already dominant in the lexicon. Details of the change, such as its restriction to polysyllabic nouns and non-neuters reflect the fact that these were especially strong contexts for [r] stems. Furthermore, the “backwards” direction of the leveling, which is the most puzzling aspect of the change, can be explained by the model of base identification that was proposed in chapter 3. More generally, this result provides further evidence for a model of paradigm learning in which learners choose the base form that is “the most informative” – i.e., that preserves the most distinctions between classes of words, and allows the remainder of the paradigm to be predicted with the greatest accuracy and confidence. This echoes a proposal by Lahiri and Dresher (1984) that certain forms in the paradigm “matter more than others” to learners when they are determining what class a word belongs to. However, what we see from examples like the Latin *honor* analogy is that the most important part of the paradigm is not universal; learners can pay attention to which part of the paradigm would make the best base, and use that as a base to derive the remainder of the paradigm. The prediction, then, is that distinctions that are preserved in the base form will be easily learned and maintained, whereas distinctions that are neutralized in the base form may be lost by leveling or regularization.

The question of maintaining lexical distinctions has implications for how underlying forms are discovered. In traditional models of phonology (e.g., Chomsky and Halle 1968) as many

surface contrasts as possible are given unique underlying representations, so that ideally all forms of a word can be derived unambiguously from a single UR. In the case of Latin, this would lead us to posit some underlying difference between words like *hono:s* and words like *soror* – perhaps *hono:s* ends in underlying /s/ and is marked with a diacritic to take a rhotacism rule in its suffixed forms, or perhaps *hono:s* has some underspecified archiphoneme (/Z/), or some other difference. We would need to compare various surface forms of each word (e.g., *hono:s* and *hono:ris*) in order to learn that the word has an alternation, and set up the appropriate UR. If learners are paying more attention to some surface forms than others, however, then this constitutes a restriction on how underlying forms can be inferred. In the next chapter, I will pursue this hypothesis even further, arguing that underlying forms must be established on the basis of just one single surface form.