Language Acquisition Guiding Theory and Diachrony: A Case Study from Latin Morphology

Jordan Kodner – University of Pennsylvania

The complex relationship between Latin stem forms and the (ir)regularity of the Latin past participle have generated significant theoretical discussion. I employ a model of productivity learning, the Tolerance Principle as a new avenue for analysis of the Latin verbal system consistent with progress in child language acquisition and the system's eventual diachronic development.

Background: Classical Latin (CL) verbs are classified into four conjugations by theme vowel (Th), and each verb is expressed with up to three stems: a present (pres), perfect (perf), and past participle (pptc) which are not necessarily predictable from one another. Table 1 summarizes common perf and pptc formations. The lack of predictability of stem derivation, the relative roles of form-based and meaning-based similarity, and related issues have generated significant theoretical discussion on the regularity of past participle derivation, and the relative role of shared semantics between stem categories and form-based generalization (e.g., Aronoff [2], Embick [4], Steriade [10]). Theoretical work seeks to explain empirically valid generalizations in the data. Part of what makes a generalization valid is that it can plausibly be acquired by a native learner.

Conjugation	Th	Common perfs	Common pptcs
1st	*ā*	*-āv-*	*-āt-*
2nd	*ē*	*-u-*	*-it-*, *-t-*
3rd	*e*	*-u-*, *-s-*, bare	*-it-*, *C-t-*
3rd-*iō*	*i*	bare	*C-t-*
4th	*ī*	*-īv-*	*-īt-*, *-it-*, *-Ct-*

Table 1: *The most common "regular" perfect and pptc types by conjugation/Th. Pptcs and bare perfects often undergo stem vowel mutations. Th-less pptcs are sometimes in -s- instead of -t-, which is sometimes synchronically phonologically conditioned and sometimes not.*

Learnability and morphological generalizations: I adopt a model of productivity learning to more rigorously assess the regularity of pptc stem derivation. To do so, I extract and lemmatize all verbs from Perseus Online (250 BCE - 200 CE) [9] and apply the Tolerance Principle (TP), which provides a measure of which plausible productive generalizations learners can discover from input data as their vocabularies grow (Formula 1). Children seek out which potential mappings are productive without pre-specifying which mappings should or should not hold. Since the TP is independently motivated, we can use it as a metric to evaluate whether and when generalizations can be made between Latin stem types in the input and inform which such patterns then should be encoded in a theoretical treatment and which should be listed instead.

I consider the productivity of pptc patterns on the top $N = 100$, 500, and 1000 most frequent lemmas (approximating knowledge of learners as they mature).These learning generalizations are independent from theoretical implementation: i.e., they can be interpreted as surface phonotactic patterns ("stem ends in /a:/"), or as the presence of certain morphemes ("$\sqrt{}$-*v* selects Th *ā*"). Table 2 lists all generalizations which achieve productivity for some N and plausible ones which fail because there are too many exceptions in CL.

The Tolerance Principle:

$$\textit{tolerable} \text{ if: } e < \theta;\ \ \theta = \frac{N}{\ln N}$$

Figure 1: *If the number of exceptional types e out of N word types that would follow the pattern is below threshold θ, that rule is productive, and the exceptions are memorized. Otherwise the rule fails and a narrower gen. or memorization of all N is needed. [12]*

Synchronic Implications: By the TP (Tab. 2.4), most pptcs for 2nd and 3rd verbs should listed since neither common pptc formation is productive even though both are high frequency. This happens because neither has a clear advantage over the other and provides too many exceptions to the other's generalization. The few that may be represented productively with rules (*-ētus*, *-ūtus*, *-tus* for 3rd-*iō* only, -[*l,r*]*sus*) are all low type frequency patterns that apply only to verbs with specific phonological forms. Other plausible generalizations do not quite hold quantitatively, for example, while Steriade's rhythmic correspondence (e.g., 2016 (16b)) works for many verbs and

so is worth considering, there are actually too many exceptions between the perf and pptc for its broadest expressions to achieve productivity by the TP (e.g., *valuī* ∼ *valitus*, but *saluī* ∼ *saltus*, **sal*[*i,ī*]*tus*). It is possible that related narrower generalizations would hold.

For theories which relate the form of pptc directly to another stem (e.g., Aronoff, Steriade), many more verbs could productively form pptcs on the basis of their perfects than presents (Tab. 2.2-3), particularly 4th verbs and 2nd verbs with *-ēv-* perfs, which means fewer verbs have to be listed if pptcs were formed with reference to the perfect. There is reason for a Latin speaker to relate the perf and pptc on the basis of form alone which can be accomplished with reference to semantic relatedness between stems. Form-based learnability provides motivation for the otherwise arbitrary morphomic relationships of Aronoff, and unlike Steriade, no active process of phonological correspondence is needed to maintain it since it is the result of learning historically contingent patterns and their exceptions.

If productivity-driven correspondence patterns are form-based, nothing prevents a speaker from adopting a pres-pptc correspondence for the cases where that works out better perf-pptc (e.g., for *-ūtus* verbs), since the semantic relationship between the perf and pptc is not critical. However, this result agrees with Steriade over Aronoff in usually preferring a perf-pptc correspondence over pres-pptc. Additional evidence for a perf-pptc relationship comes from pptcs reworked on the basis of the perf (eg *crēvī* ∼ *crētus* ← *certus*). Nasal-infix spread is ambiguous evidence because all but one example (*pungō* ∼ *pepugī* ∼ *punctus*) attests a perf nasal as well, and so cannot be direct evidence for pres-pptc correspondence. This one example can be learned as an exception by the TP.

Evidence from diachrony: Evidence for the proposed productivity pattern comes from the fate of the pptcs in Late Latin since productivity is expected to have an impact on morphological change [1].

Three pptc forms spread at the expense of the others in most regions: **-atu*, **-itu*, **-utu* < *-ātus*, *-ītus* (not *-itus*), *-ūtus* [7,11]. **-utu* spread among 2nd and 3rd verbs, especially those with *-u-* perfects and replaced much higher frequency *-tus*/*-itus*. However, *-tus*/*-itus* were not productive in general in CL, and there were no productive pptc formations for any *-u-* perf verbs except for the subset with *-u-* roots (e.g., *soluō*), so even though the analogical basis for **-utu* was slim, it had no competitors for productivity.

Theme V	**PPtc**	**Example**	**at 100?**	**at 500?**	**at 1000?**
(1st) *ā*	*-ātus*	*vocāre*∼*vocātus*	**yes**	**yes**	**yes**
(3rd *-iō*) *i*	*-tus*	*capiō*∼*captus*	**yes**	**yes**	**yes**
(4th) *ī*	*-ītus*	*audīre*∼*audītus*	**yes**	marg.	no
(4th) *ī*	*-tus*	*venīre*∼*ventus*	**yes**	no	no
(2nd) *ē*	either	*docēre*∼*doctus*	no	no	no
(3rd) *e*	either	*reddere*∼*redditus*	no	no	no
Pres stem	**PPtc**	**Example**	**at 100?**	**at 500?**	**at 1000?**
-veō	*-au/ōtus*	*faveō*∼*fautus*	N/A	**yes**	**yes**
-[¬velar]*eō*	*-itus*	*debeō*∼*debitus*	marg.	no	no
-vere	*-ūtus*	*solvere*∼*solūtus*	**yes**	marg.	marg.
-[velar]*eō*	*-tus*	*doceō*∼*doctus*	N/A	no	no
Perf stem	**PPtc**	**Example**	**at 100?**	**at 500?**	**at 1000?**
-āv-	*-ātus*	*amāvī*∼*amātus*	**yes**	**yes**	**yes**
-ēv-	*-ētus*	*flēvī*∼*flētus*	**yes**	**yes**	marg.
-īv-	*-ītus*	*dormīvī*∼*dormītus*	**yes**	**yes**	**yes**
-Cs-	*-tus*	*iūnxī*∼*iūnctus*	**yes**	**yes**	**yes**
-u-	*-itus*	*valuī*∼*valitus*	no	no	no
-u-	*-tus*	*tenuī*∼*tentus*	no	no	no
bare	*-itus*	*lēgī*∼*lēctus*	no	no	no
-vere & -u-	*-ūtus*	*volvere*∼*volūtus*	**yes**	**yes**	**yes**

Table 2: *All productive and selected failed pptc generalizations given Th, narrower present stem, and perf stem patterns suggested in the literature for* $N = 100, 500, 1000$. ***1)*** *Pptcs that surface with -ātus endings are productive for 1st verbs regardless of whether form is cued by the pres or perf stem,* ***2)*** *There is no productive pptc for 4th verbs as a whole, but there is for the large subset with -ī-v- perfs since most 4th verbs with -tus or -itus pptcs have other perf formations instead, e.g., aperīre ∼ aperuī ∼ apertus.* ***3)*** *There are a few pptcs that barely achieve productivity among narrow sets of 2nd and 3rd verbs, including -autus/-ōtus, -ētus, -ūtus.* ***4)*** *The two most common pptc formations for 2nd and 3rd are both non-productive because they provide too many exceptions for one another. (marginal=1 verb away)*

[1] Anderson, S. R. (2015). Morphological change. *The Routledge handbook of historical linguistics*, pages 264–285.

[2] Aronoff, M. (1994). *Morphology by itself: Stems and inflectional classes*. Number 22. MIT press.

[3] Chan, E. (2008). *Structures and distributions in morphological learning*. PhD thesis, Ph. D. dissertation, Dept. of Computer and Information Science, UPenn.

[4] Embick, D. (2000). Features, syntax, and categories in the latin perfect. *Linguistic inquiry*, 31(2):185–230.

[5] Kodner, J. (2019). Estimating child linguistic experience from historical corpora. *Glossa: A Journal of General Linguistics*, 4(1):122.

[6] Labov, W. (1972). Some principles of linguistic methodology. *Language in society*, 1(1):97–120.

[7] Laurent, R. (1999). *Past participles from Latin to Romance*, volume 133. Univ of California Press.

[8] Nivre, J. e. a. (2018). Universal dependencies 2.3. LINDAT/CLARIN digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University.

[9] Smith, D. A., Rydberg-Cox, J. A., and Crane, G. R. (2000). The perseus project: A digital library for the humanities. *Literary and Linguistic Computing*, 15(1):15–25.

[10] Steriade, D. (2016). The morphome vs. similarity-based syncretism. *The morphome debate*.

[11] Weiss, M. L. (2009). *Outline of historical and comparative grammar of Latin*. Beech Stave Press.

[12] Yang, C. (2016). *The price of linguistic productivity: How children learn to break the rules of language*. MIT Press.