Conjoined quadruplets in West Slavic

Guy Tabachnick[1*]

[1*]Department of Czech Language, Masaryk University Faculty of Arts, Arne Novaka 1, Brno, 60200, Czech Republic.

Corresponding author(s). E-mail(s): guy@phil.muni.cz;

Abstract

Polish and Czech both have "conjoined quadruplets": sets of four paradigms that differ in exactly two cases. In this paper, I compare how Distributed Morphology and Nanosyntax account for these patterns, which include multiple configurations identified as problematic for Nanosyntax by Janků (2022) and others. Since both theories must relegate some of the patterns to unproductive morphology, mere ability to account for all of the attested patterns cannot distinguish them. Instead, I compare the theories' ability to account for measures of productivity like frequency, acquisition, and behavioral patterns. Accounting for the full range of attested phenomena requires speakers to store generalizations over their lexicon in a pattern-matching module that enforces the shape of lexical entries and can be applied productively but is not active in formal derivations. The nature of lexical representations in Nanosyntax requires a much more powerful pattern-matching module to account for the attested phenomena than does Distributed Morphology. These gradient patterns, often ignored in theoretical analyses, thus provide valuable sources of data for evaluating morphological theories.

Keywords: Polish, Czech, inflectional morphology, gradience, productivity

1 Introduction

Distributed Morphology and Nanosyntax have different ways of handling allomorphy in inflectional paradigms. In Distributed Morphology, suffixes expressing morphosyntactic features like number, gender, and case are spelled out through rules of realization (known as

Vocabulary Items) that take properties of their adjacent base into account. Stems that differ in their inflectional paradigm must differ in some way in their featural content, whether this be phonological (rules of realization may be conditioned on phonological properties of their context), morphosyntactic (e.g. gender), or purely formal (these rules can be conditioned by diacritic features that index inflectional class or, roughly equivalently, contain lists of roots in whose context they apply). In Nanosyntax, on the other hand, functional sequences of morphosyntactic features are spelled out through matches between lexically stored trees representing roots and suffixes; thus, stems with different inflectional paradigms may have different morphosyntactic features (like gender) but crucially differ in the size and shape of the syntactic tree in their lexical entry. These then interact with the lexical entries of suffixes to determine how each paradigm cell is spelled out.

Nanosyntax imposes greater restrictions upon patterns of allomorphy than Distributed Morphology—in particular, its inability to derive certain typologically rare patterns has long been considered a calling card for the framework. However, such patterns are attested, requiring Nanosyntax to resort to some form of unproductive morphology to account for them. Janků (2022) points out one such configuration that is problematic in Nanosyntax (that is, the version of the theory she used, which I call "classical" Nanosyntax): in Czech, there are two inflectional paradigms identical except for one case, which she calls "conjoined twins". In fact, in Czech and its neighbor in the West Slavic group, Polish, the problem is even worse: there are sets of four paradigms that are identical except for two cases; these paradigms contain all four logically possible combinations of two realizations for each of the two cases. I thus call these patterns conjoined quadruplets.

The mere *existence* of these patterns is not necessarily problematic for either theory: there will always be some morphological irregularities that speakers have to store for individual lexical items. Thus, what is more important is evidence about how the patterns are integrated into the morphological system. This evidence can include factors such as relative frequency, gradient generalizations, and evidence of productivity from loanwords and nonce word studies.

In this paper, I provide a quantitative perspective on the conjoined quadruplet patterns in Czech and Polish. Following e.g. Gouskova, Newlin-Łukowicz, and Kasyanenko (2015), I conclude that there are demonstrable aspects of speakers' knowledge of their language that cannot be encapsulated in the grammar as it is typically understood in generative theory—that is, the module(s) that derive forms online from syntactic structure to phonological form. Instead, this knowledge is best placed in a separate module that tracks and enforces statistical patterns across a language's lexicon. While such a module can be paired with any formal theory of morphology, the nature of this module depends on the form of entries stored in the lexicon, which is in turn dependent on the shape of the morphological grammar. I show that, with Distributed Morphology, this pattern-matching module essentially comprises morpheme structure constraints over phonological underlying representations (see Booij, 2011), while Nanosyntax requires something much more powerful and representationally cumbersome. This work will thus be a demonstration that quantitative data like the sort I discuss, which is too often ignored by theoretical morphologists, is crucial for getting a full understanding of the morphological systems of speakers of a given language, and can thus give us a broader understanding of the implications of different theoretical frameworks.

Section 2 provides an overview of the conjoined quadruplet paradigms in Czech and Polish. Section 3 presents the architecture of Distributed Morphology and Nanosyntax and presents possible analyses of the paradigms in question in the two theories. Section 4 brings quantitative data about the paradigms into the picture and discusses their implications for the theories. Section 5 concludes.

2 Basic paradigms

In this section, I will introduce the nominal paradigms that are the subject of this study in two closely related West Slavic languages, Polish and Czech.

The main inflectional macroclass of masculine nouns in Polish has seven subclasses in the singular, which are shown in Table 1 (Cameron-Faulkner & Carstairs-McCarthy, 2000;

Halle & Marantz, 2008; Saloni, Woliński, Wołosz, Gruszczyński, & Skowrońska, 2015). The

Table 1 Full Polish singular masculine noun paradigms (Cameron-Faulkner & Carstairs-McCarthy, 2000; Halle & Marantz, 2008; Saloni et al., 2015)

	'country'	'leaf'	'gentleman'	'brother'	'money-grubber'	'store'	'column'
NOM	kraj	liɕtɕ	pan	brat	xtɕivʲɛts	sklɛp	swup
GEN	kraju	liɕtɕa	pana	brata	xtɕivtsa	sklɛpu	swupa
LOC	kraju	liɕtɕu	panu	bratɕɛ	xtɕivtsu	sklɛpʲɛ	swupʲɛ
DAT	krajɔvʲi	liɕtɕɔvʲi	panu	bratu	xtɕivtsɔvʲi	sklɛpɔvʲi	swupɔvʲi
INS	krajɛm	liɕtɕɛm	panɛm	bratɛm	xtɕivtsɛm	sklɛpɛm	swupɛm
VOC	kraju	liɕtɕu	panʲɛ	bratɕɛ	xtɕivtʂɛ	sklɛpʲɛ	swupʲɛ

accusative case is omitted from this table because it is identical to the nominative for inanimate nouns like [kraj] 'country' and to the genitive for animate nouns like [brat] 'brother'. In these paradigms, there are four cases that vary: genitive, locative, dative, and vocative. Each has two possible suffixes: the profligate suffix [u] alternates with [a] in the genitive, [ɔvʲi] in the dative, and [ɛ] in the dative and vocative (this last suffix also triggers alternations in some stem-final consonants).

The analogous masculine paradigms in Czech, which are often called "hard-stem" because they only appear after nouns ending in phonologically "hard" consonants, are shown in Table 2. In Czech, the same suffixes are distributed across eight paradigms (as in Polish,

Table 2 Full Czech singular masculine hard-stem noun paradigms (Křen et al., 2020; Ústav pro jazyk český AV ČR, 2008–2026)

	'age'	'today'	'team'	'evening'	'time'	'forest'	'doctor'	'grandson'
NOM	vjɛk	dnɛʃɛk	tiːm	vɛtʃɛr	tʃas	lɛs	doktor	vnuk
GEN	vjɛku	dnɛʃka	tiːmu	vɛtʃɛra	tʃasu	lɛsa	doktora	vnuka
LOC	vjɛku	dnɛʃku	tiːmu	vɛtʃɛru	tʃasɛ	lɛsɛ	doktorovɪ∼u	vnukovɪ∼u
DAT	vjɛku	dnɛʃku	tiːmu	vɛtʃɛru	tʃasu	lɛsu	doktorovɪ∼u	vnukovɪ∼u
INS	vjɛkɛm	dnɛʃkɛm	tiːmɛm	vɛtʃɛrɛm	tʃasɛm	lɛsɛm	doktorɛm	vnukɛm
VOC	vjɛku	dnɛʃku	tiːmɛ	vɛtʃɛrɛ	tʃasɛ	lɛsɛ	doktorɛ	vnuku

the accusative is syncretic with either the nominative or the genitive depending on animacy).

Almost all animate nouns can appear with either [u] or [ovɪ] in the dative and locative; the choice between them is determined in large part by syntactic context (Rusínová & Nekula, 1995, 247).

In this paper, I largely ignore allomorphy in the dative and vocative to focus on the patterns in the genitive and locative. For the sake of clarity, the four paradigms from the previous tables that differ only in their genitive and locative—the namesake "conjoined quadruplets" in this paper—are shown again in Table 3 for Polish and Table 4 for Czech. All of the paradigms in question are inanimate, so these tables also include the accusative, which is identical to the nominative. Since the vocative is often considered a different sort of case than the others and ignored in morphological analyses (e.g. Caha, 2009), I remove it in these paradigms.

Table 3 Conjoined quadruplets in Polish

	'country'	'leaf'	'store'	'column'
NOM	kraj	liɕtɕ	sklɛp	swup
ACC	kraj	liɕtɕ	sklɛp	swup
GEN	kraju	liɕtɕa	sklɛpu	swupa
LOC	kraju	liɕtɕu	sklɛpʲɛ	swupʲɛ
DAT	krajɔvʲi	liɕtɕɔvʲi	sklɛpɔvʲi	swupɔvʲi
INS	krajɛm	liɕtɕɛm	sklɛpɛm	swupɛm

Table 4 Conjoined quadruplets in Czech

	'team'	'evening'	'time'	'forest'
NOM	tiːm	vɛtʃɛr	tʃas	lɛs
ACC	tiːm	vɛtʃɛr	tʃas	lɛs
GEN	tiːmu	vɛtʃɛra	tʃasu	lɛsa
LOC	tiːmu	vɛtʃɛru	tʃasɛ	lɛsɛ
DAT	tiːmu	vɛtʃɛru	tʃasu	lɛsu
INS	tiːmɛm	vɛtʃɛrɛm	tʃasɛm	lɛsɛm

As we will see in greater detail in Section 3.2, the conjoined quadruplet paradigms have two features that are problematic for the version of Nanosyntax used by Janků (2022). First, when reading the case forms as an ordered list (as Nanosyntax does), these paradigms diverge

(in the genitive) and then converge again (in the dative). Second, the Czech paradigm for [tʃas] 'time' includes a configuration known as ABA: the suffix [u] is used in two non-adjacent cases (genitive and dative), with the case in between them (locative) having a different form. The inability of Nanosyntax to derive ABA patterns has been an important aspect of the theory since its early days (cf. Baunaz & Lander, 2018; Caha, 2009), although recent advances in the theory have allowed for ABA patterns to be captured in certain circumstances (van Craenenbroeck, 2025).

3 Theoretical analyses

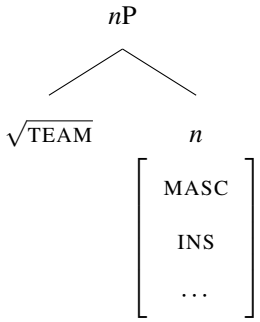
In this section, I present theoretical analyses of the Polish and Czech conjoined quadruplet paradigms in Distributed Morphology and Nanosyntax, including a brief overview of relevant architectural properties of these theories.

3.1 Distributed Morphology

3.1.1 Basics

In Distributed Morphology, individual nodes of morphosyntactic trees are spelled out through a process called Vocabulary Insertion, in which rules of realization (Vocabulary Items) compete to realize these nodes. For the sake of illustration, I assume that inflected nouns have the structure in (1): a root combines with a nominal categorizing head *n*, which contains (or acquires during the derivation) morphosyntactic features associated with the resulting noun, including but not necessarily limited to case and gender. I represent these features here as atomic, though I decompose case below (gender presumably decomposes as well, though this is not relevant for our purposes).

(1) *Distributed Morphology tree for the instrumental of Czech 'team'*



Following Bobaljik (2000), I assume that Vocabulary Insertion precedes from the root outwards: first the root is spelled out, then the *n* head, meaning that the phonological form of the former is available in the context on which spellout of the latter may depend. That is, spellout of the *n* head (the inflectional suffix) may be sensitive to the phonological properties of the root. Like Bobaljik (2000), but unlike Privizentseva (2024), I assume that phonological forms may contain diacritic features that index their inflectional properties. For example, if we place [tiːm] 'team' into Czech inflection class H, its underlying form would contain a [H] feature, as shown by the Vocabulary Item spelling out the relevant root:

(2) *Vocabulary Item for Czech 'team'*

$\sqrt{\text{TEAM}} \leftrightarrow /\text{tiːm}_{[\text{H}]}/$

Diacritic features are typically used as a last resort indicating when a noun's inflectional properties must be memorized and cannot be derived from (contextually determined by) other properties such as its gender or its final consonant. However, as we will see, Distributed Morphology gives us the analytic freedom to rely as little or as much on these diacritic features as we see fit.

After Vocabulary Insertion of the root, the instrumental case marker can be spelled out using the following Vocabulary Item:

(3) *Vocabulary Item for the instrumental of Czech 'team'*

INS ↔ /ɛm/ / [H]___

This Vocabulary Item will only apply when the [H] feature is in the context at the point of application; that is, when attaching to a root marked with [H].

Though I often represent cases as atomic units, many linguists (e.g. Jakobson, 1984; Müller, 2004) decompose them into binary features. I assume that cases decompose into features as shown in Table 5, which Müller (2004) uses for Russian. Thus, in (3), /ɛm/ is really

Table 5 Assumed featural decomposition of cases (cf. Müller, 2004)

	subj(ect)	gov(erned)	obl(ique)
NOM	+	−	−
ACC	−	+	−
GEN	+	+	+
LOC	−	−	+
DAT	−	+	+
INS	+	−	+

spelling out the feature set [+subj, −gov, +obl].

This case decomposition allows us to demonstrate two closely related features of Distributed Morphology: the Subset Principle and the Elsewhere Principle. The noun [tiːm] 'team' has [u] in the genitive, locative, and dative. These cases all share the feature [+obl], which appears in the Vocabulary Item in (4):

(4) *Vocabulary Item for the genitive, locative, and dative of Czech 'team'*

[+obl] ↔ /u/ / [H]___

This Vocabulary Item is not an exact match for spelling out an *n* head, because it does not match all of its case features ([±subj] and [±gov]), not to mention any other features on that head like gender. Nonetheless, it is considered a match because the features it does spell out are a *subset* of the features on the *n* head. This is known as the Subset Principle.

Since the instrumental case is also [+obl], both (3) and (4) are viable candidates to spell out the instrumental of [tiːm], and thus compete to spell it out. This competition is adjudicated by the Elsewhere Principle, which states that the Vocabulary Item spelling out the most specific set of features wins (similarly, a Vocabulary Item correctly specifying a context will win over one spelling out the same set of features with no context or a less specific context). Thus, (3) beats out (4) for the instrumental because it is an exact match for the case features, whereas the more general (4) only matches one feature. However, in other oblique cases, (3) will not be a viable competitor because it spells out features ([+subj] and/or [−gov]) that are not included in, say, the dative.

3.1.2 Analyses of conjoined quadruplets

Since only one suffix for a given set of morphosyntactic features can be the unmarked default, at least some words must have their case suffix selection marked with diacritic features. There are three categories of analysis depending on which suffixes are unmarked.

First, the [u] suffix may be default. This means that words with genitive [a] and/or locative [ɛ] must have diacritic features ([Ga] and [Le], respectively) which are indexed in corresponding Vocabulary Items:

(5) *Vocabulary Items for [u] default*

a. [+obl] ↔ /u/

b. GEN ↔ /a/ / [Ga]___

c. LOC ↔ /ɛ/ / [Le]___

When, for example, the diacritic feature [Ga] is present when spelling out the genitive, the more specific rule in (5-b) wins out over (5-a) by the Subset Principle; however, when it is absent, the conditions for (5-b) are not satisfied and (5-a) is the only viable option.

Under this system, Czech [tiːm] ‘team’ would have the unmarked underlying form /tiːm/, because it has [u] in both genitive and locative, while Polish [sklɛp] ‘store’ would be underlyingly /sklɛp$_{[Le]}$/, marking the fact that its locative is [sklɛp^{j}ɛ]. Czech [lɛs] ‘forest’, which has [a] in the genitive and [ɛ] in the locative, would be marked with both features: /lɛs$_{[Ga,\ Le]}$/.

Halle and Marantz (2008) take the opposite approach in their analysis of Polish inflection: namely, words that take [u] in the genitive and locative are marked with diacritic features [Gu] and [Lu], respectively, while those with [a] in the genitive and [ɛ] in the locative are unmarked. Thus, the rules spelling out genitive [a] and locative [ɛ] must be context-free, as in (6):

(6) *Vocabulary Items for non-[u] default*

a. [+obl] ↔ /u/

b. GEN ↔ /a/

c. LOC ↔ /ɛ/

Since (6-b) and (6-c) are more specific than (6-a), the former will always win over the latter by the Subset Principle. Thus, for case forms to ever be spelled out as [u], we need some way to remove (6-b) and (6-c) from competition. Halle and Marantz (2008) achieve this by deleting the relevant case features through rules of Impoverishment—post-syntactic operations that remove morphosyntactic features from syntactic structures before they are sent to spellout. In this case, the rules are conditioned by the diacritic features [Gu] and [Lu], as shown in (7):

(7) *Rules of Impoverishment for non-[u] default*

a. [+subj, +gov] → Ø / [Gu]___

b. [−subj, −gov] → Ø / [Lu]___

These rules delete the case features that distinguish genitive and locative from other oblique cases. Thus, after their application, all that remains to be spelled out is [+obl], meaning that the rule in (5-a) will be the only applicable rule and the case will be spelled out as [u].

In this analysis, Czech [tiːm] 'team' has the underlying form /tiːm$_{[Gu,\ Lu]}$/, while Polish [sklɛp] 'store' is underlyingly /sklɛp$_{[Gu]}$/ and Czech [lɛs] 'forest' is unmarked: /lɛs/. That is, the patterns of marking are exactly the opposite as in the previous analysis.

Halle and Marantz (2008) choose this analysis because it allows for more convenient generalizations about the full set of Polish inflection classes shown in Table 1 in the introduction. However, it is more complicated than the previous one: it requires not just Vocabulary Items but post-syntactic operations, the use of which has been criticized both within and outside of Distributed Morphology (see e.g. Haugen & Siddiqi, 2016). It also requires a specific order of operations: if diacritic features are inserted at spellout, as I assume here, then the root must be spelled out before the Impoverishment rules in (7) apply. That is, post-syntactic rules must be interleaved with root-outward spellout.

The third possible analysis is that *all* case forms are marked—there are no default inflection classes, and all nouns have diacritics in their underlying form for both cases: /tiːm$_{[Gu,\ Lu]}$/, /sklɛp$_{[Gu,\ Le]}$/, /lɛs$_{[Ga,\ Le]}$/. These could then be referenced in similarly fully marked Vocabulary Items:

(8) *Vocabulary Items for no default*

a. GEN ↔ /u/ / [Gu]___

b. LOC ↔ /u/ / [Lu]___

c. GEN ↔ /a/ / [Ga]___

d. LOC ↔ /ɛ/ / [Le]___

However, in this analysis the syncretism between genitive and locative [u] becomes accidental: formally, they are two homophonous suffixes. Alternately, the Vocabulary Items (and Impoverishment rules) from the previous analyses yield the same results when all nouns have

markers in both cases. For example, (5-a) spells out the genitive and locative of nouns with [Gu] and [Lu] as [u], even though these features are not referenced in this (or any) Vocabulary Item. In this case, [Gu] and [Lu] would be used strictly in the organization of the lexicon and not in the course of spelling out case suffixes. I discuss this point further in Section 4.1.2.

Finally, it is possible to mix and match the first two analyses such that [u] is default in one case but not the other. For example, the grammar could have default [a] in the genitive—by including the context-free Vocabulary Item for the genitive in (6-b) paired with the contextually specified genitive Impoverishment rule in (7-a)—and [u] in the locative, with the contextually specified Vocabulary Item for the locative in (5-c). In this case, nouns with genitive [u] would be marked with [Gu], while nouns with locative [ɛ] would be marked with [Le].

To summarize, in Distributed Morphology, at least one suffix for each case must be indexed by diacritic markers in underlying forms of nouns that take them. The grammar fragment is somewhat simpler if [u] is the default, but otherwise, there are no limitations on which allomorph of each suffix can be the default: all combinations of genitive and locative defaults are possible.

3.2 Nanosyntax

I begin this section with a theoretical note. This work expands on issues raised by Janků (2022) in her analysis of Czech inflection classes. Accordingly, I assume the same version of Nanosyntax as she does. Recent work in Nanosyntax (e.g. Cortiula, 2023; Kasenov, 2025) has adopted a new version of the spellout algorithm (see De Clercq, Caha, Starke, & Vanden Wyngaerd, 2025) whose additional flexibility allows it to overcome many of the limitations discussed in Section 3.2.2 (cf. van Craenenbroeck, 2025). However, that does not mean that this work is outdated: analyses of the previous, well-established spellout algorithm can facilitate and direct exploration of the new algorithm, the implications of which are still

not fully understood. Indeed, Caha (2026) proposes an analysis of Czech inflection classes using the new algorithm in direct response to an earlier version of this work.

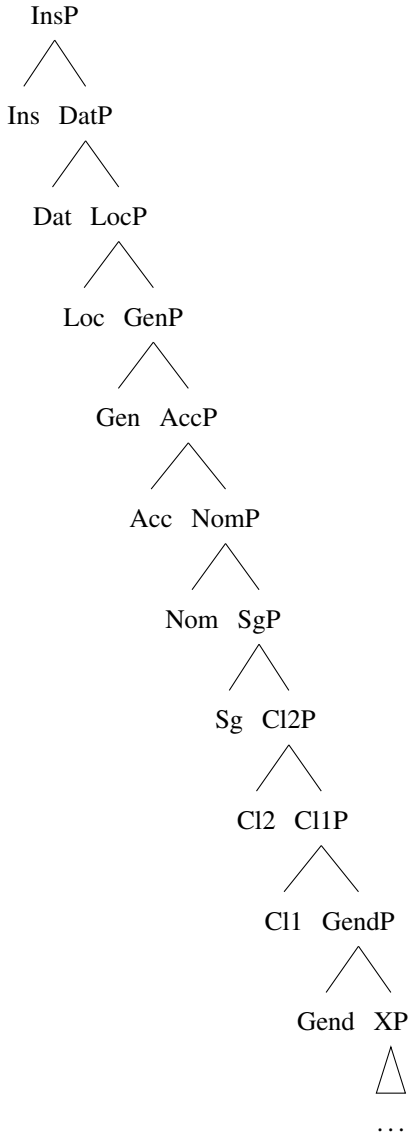
3.2.1 Basics

In this section, I briefly explain the architectural properties of Nanosyntax that are relevant for this paper. For a more thorough demonstration of how Nanosyntax works, see Janků (2022) or the introduction to this volume.

In Nanosyntax, syntactic structures are spelled out by matching lexical items of particular shapes. Whereas spellout in Distributed Morphology targets one node at a time, Nanosyntax spellout targets entire trees. Likewise, Distributed Morphology allows nodes to comprise bundles of morphosyntactic features, whereas in Nanosyntax, each feature sits in a different node (and projection) of the functional sequence.

Thus, in Nanosyntax, the feature geometry of case and other morphosyntactic features is one of containment: for example, singular number is expressed by a Sg head in the nominal functional sequence, while plural number is expressed by both Sg and a Pl head on top of it. Similarly, "bigger" cases, like the instrumental, contain the functional projections for "smaller" cases like the accusative and genitive. Janků (2022) assumes the following functional sequence for Czech masculine singular nouns:

(9) *Functional sequence for Czech masculine singular nouns*

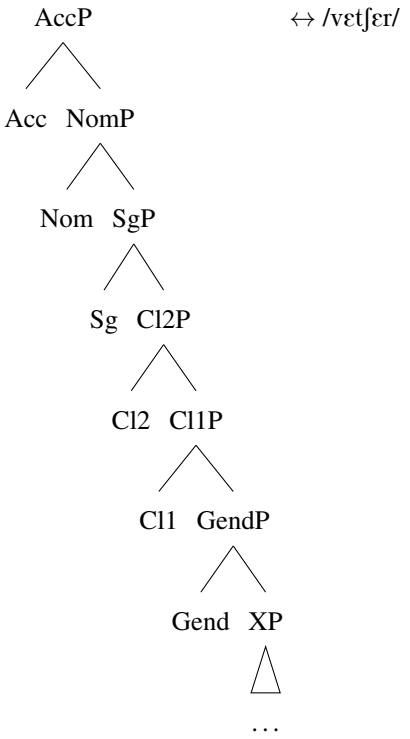


Different cases will include more or less of this sequence: the locative contains four case projections (Loc, Gen, Acc, and Nom), while the nominative only includes the lowest one. The order of the case projections is derived from cross-linguistic study of syncretisms, which tend to occupy contiguous segments of the above hierarchy. For example, it is very common

for nouns to have the same ending in the nominative and accusative or the accusative and genitive, but extremely unlikely for a noun to have the same ending in the nominative and instrumental to the exclusion of all other cases.

Nanosyntax derives these syncretism patterns through the Superset Principle. Suppose our lexicon contains the following lexical entry for [vɛtʃɛr] 'evening', taken from the analysis of Czech nominal declension in Janků (2022):

(10) *Lexical entry for Czech 'evening'*



The Superset Principle states that a lexical entry can spell out not just trees that exactly match its structure, but also subtrees contained within its structure. Thus, the lexical entry in (10) exactly matches accusative masculine singular nouns, which have the structure in (9) up to AccP. However, (10) also matches nominative masculine singular nouns, which comprise the sequence up to NomP. This is because the NomP sequence is a subtree of the syntactic

structure in (10). This is how Nanosyntax derives the nominative–accusative syncretism that [vɛtʃɛr] has no suffix in the nominative or accusative.

If we wish to derive the genitive, we must add a GenP to our functional sequence. Here, (10) is no longer a match (since it doesn't include GenP), so we must do something else: rearrange the tree through what is called spellout-driven movement in an attempt to exhaustively spell out the entire tree.

Janků (2022) uses the spellout algorithm of "classical" Nanosyntax:

(11) *"Classical" Nanosyntax spellout algorithm (Caha, 2021; Janků, 2022; Starke, 2018)*

a. Merge F and spell out.

b. If (11-a) fails, move the specifier of F's complement and spell out.

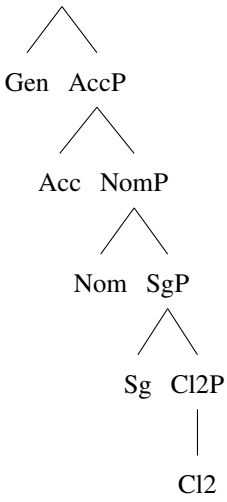
c. If (11-b) fails, move the complement of F and spell out.

d. If all else fails, go back to the previous cycle and try the next option for that cycle.

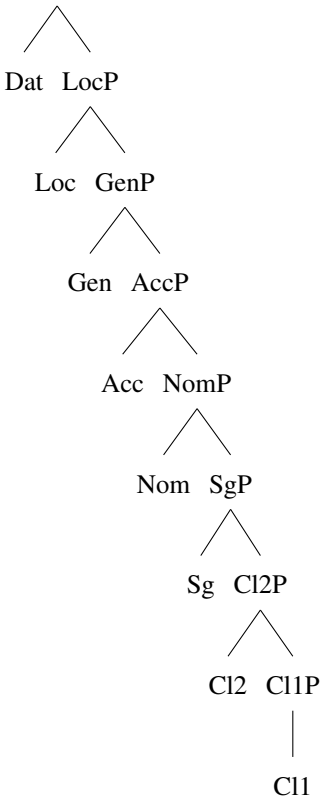
Her analysis also includes the following lexical entries for case suffixes:

(12) *Lexical entries for Czech case suffixes*

a. GenP ↔ /a/

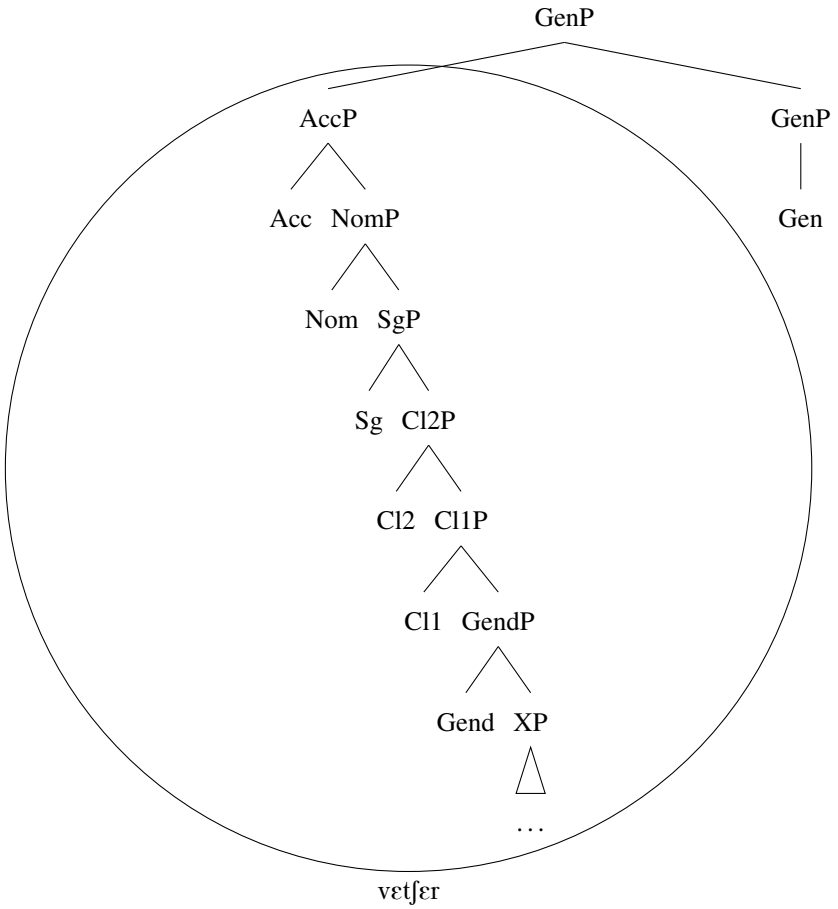


b. DatP ↔ /u/



Our structure, which exactly matches that in (10) currently has no specifiers, so we move the complement of Gen—that is, the AccP— and attempt spellout again. This yields the following tree:

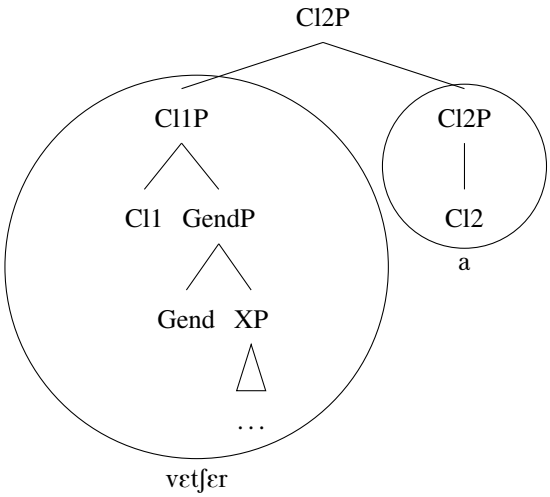
(13) *Attempted spellout of the genitive of 'evening': first try*



While we can successfully spell out the left half as /vɛtʃɛr/, as before, the isolated GenP is not a match for either of the lexical entries in (12). Thus, we must backtrack: undo the derivation until we reach a point where things will work out again. The lexical entry for /a/ in (12-a) is rooted in Cl2 (i.e., its lowest projection is Cl2), so in order for it to match part of our structure, we must go back to the point of the derivation where Cl2 is added and pursue an alternate

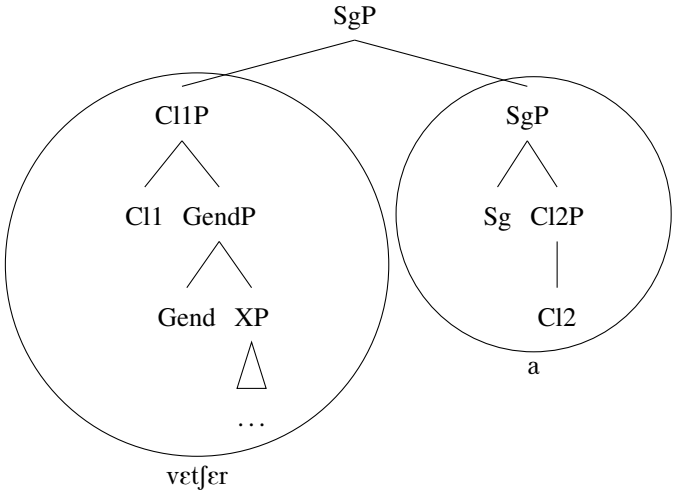
derivational pathway where Cl2 has no complement. Now, instead of lexicalizing the entire tree with Cl2 at the top and moving on to the next step (adding Sg), we try the next applicable option: movement of the complement of Cl2.

(14) *Spellout of the genitive of 'evening': intermediate step after backtracking*



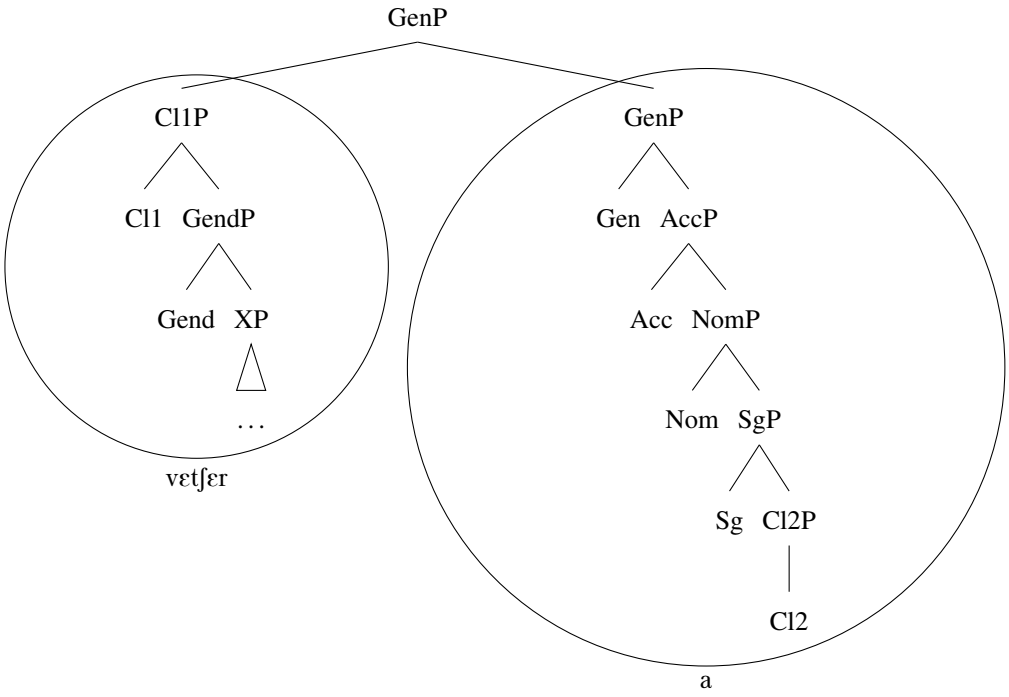
This can be successfully lexicalized, the left branch as /vɛtʃɛr/ and the right branch as /a/ (both by the Superset Principle, as they are subtrees of the full lexical entries that are matched). Now we add SgP and try the first movement option in the lexicalization algorithm: moving the specifier of the complement of Sg (that is, the left branch in (14)):

(15) *Spellout of the genitive of 'evening': intermediate step after attaching SgP*



In this manner, we pass the left branch all the way up to the top, moving it above each newly merged projection until we reach the genitive:

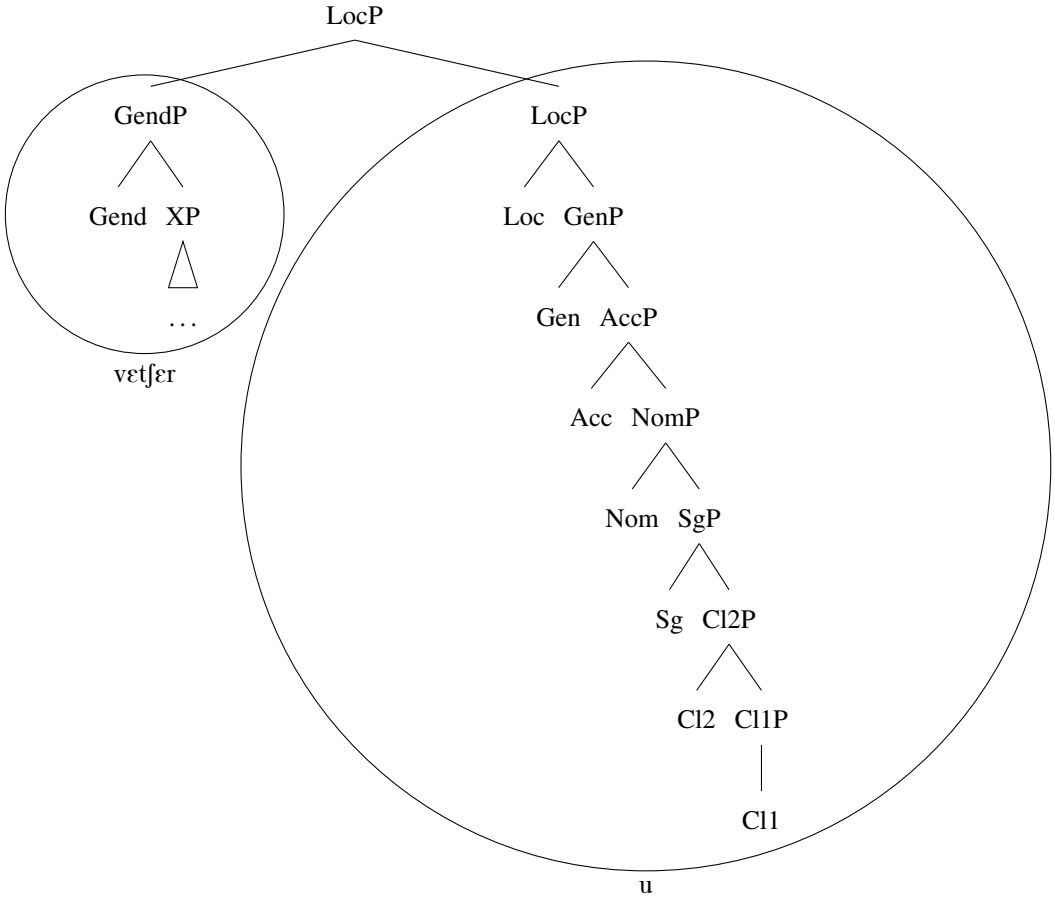
(16) *Spellout of the genitive of 'evening': final step*



As before, the left branch is spelled out as /vɛtʃɛr/ and the right branch as /a/. This is how we get the genitive form. Note that this derivation also spells out the nominative and accusative as /a/, but since we only go through this pathway as an intermediate step to reach the genitive after backtracking, this is not a problem.

When we merge the locative, we run into the same problem as before: the only lexical entry that contains LocP, (12-b), is rooted in Cl1 and cannot spell out Loc on its own. Thus, we must backtrack again, this time all the way to Cl1, then pass the left branch along as before until we get back to LocP:

(17) *Spellout of the locative of 'evening': final step*

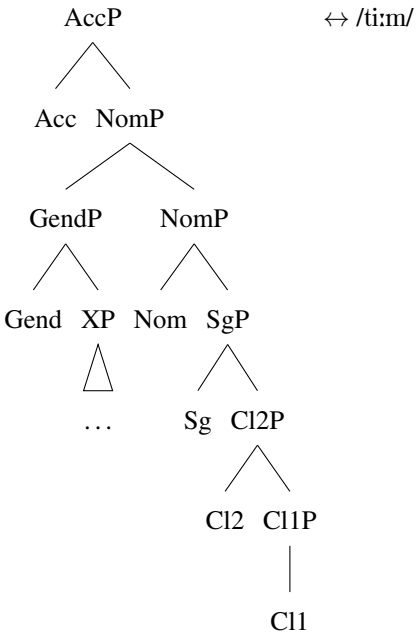


Here, the left branch is spelled out as /vɛtʃɛr/ and the right branch as /u/. Since the lexical entry for [u] in (12-b) also contains a DatP projection, we can merge this projection and pass the left branch up one more time to spell out the dative as /u/ as well, deriving the dative–locative syncretism in masculine inanimate paradigms (these two cases are syncretic elsewhere in Czech as well).

Nanosyntax explicitly rejects the diacritic features common in Distributed Morphology (Caha, 2021). Instead, different inflectional classes for a given set of morphosyntactic features (in this case, masculine singular) are distinguished by root shape. Thus, Janků (2022) gives nouns like [tiːm] 'team', which have [u] in the genitive, locative, and dative, the shape in (18).

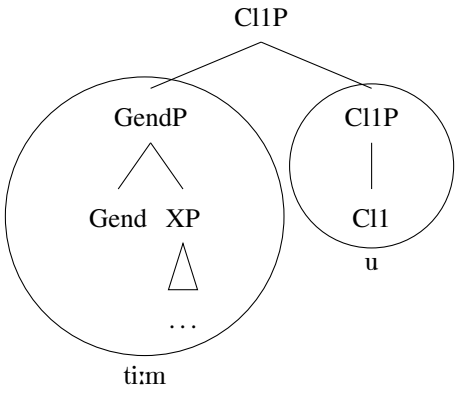
This root shape, like that of [vɛtʃɛr] ‘evening’ in (10), has AccP as its maximal projection. While (10) is a straight functional sequence, though, this root has a complex left branch (see Blix, 2021):

(18) *Lexical entry for Czech ‘team’*



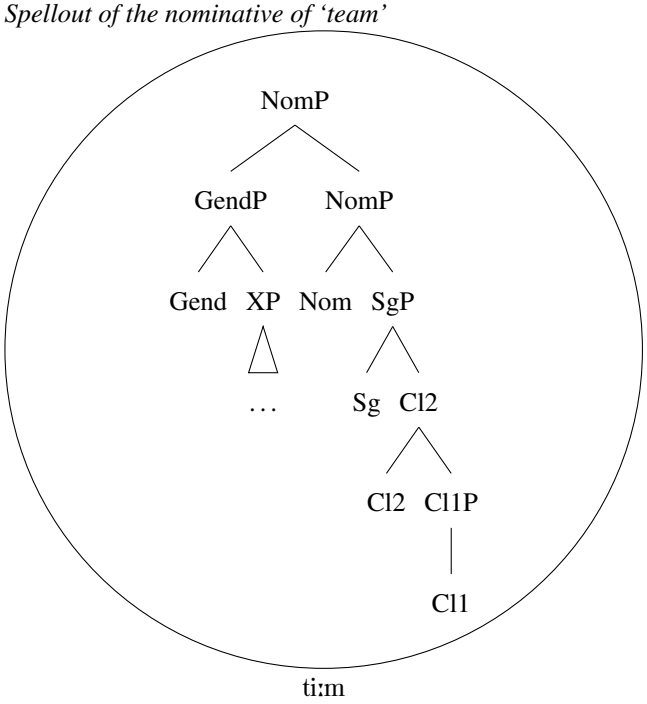
The derivation for this word follows a similar path as that of [vɛtʃɛr] after the last step of backtracking. Here we are forced into the movement after attaching Cl1P the first time, since (18) cannot match a tree including a complement to Cl1:

(19) *Spellout of Cl1 of 'team'*



Here the left branch is lexicalized as /tiːm/, matching a subtree of (18), and the right branch as /u/. Of course, since trees ending in Cl1P never appear on their own, this is not a problem. We pass the left branch up all the way to NomP. Here we face a potential issue: in the previous steps, the right branch was spelled out as /u/, but we need the nominative to be suffixless. Let us look at the tree at this point:

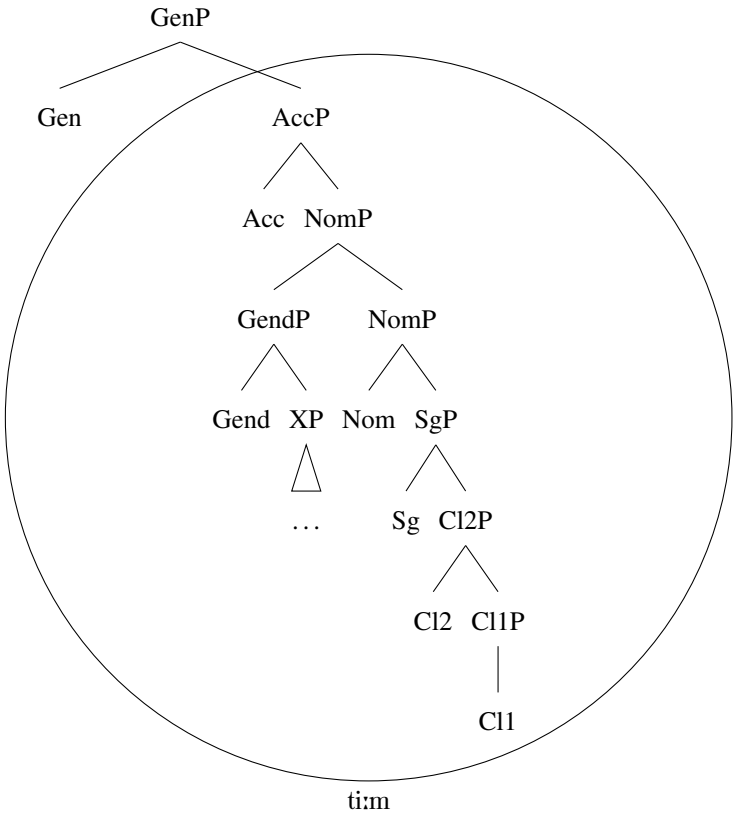
(20) *Spellout of the nominative of 'team'*



In fact, this entire tree is a subtree of the lexical entry of [tiːm], so it is spelled out in its entirety as /tiːm/ (given the option, it is preferable to spell out a tree using one lexical entry instead of two). When we add AccP on top, we have exactly matched the lexical entry in (18), so the accusative does not have a suffix either.

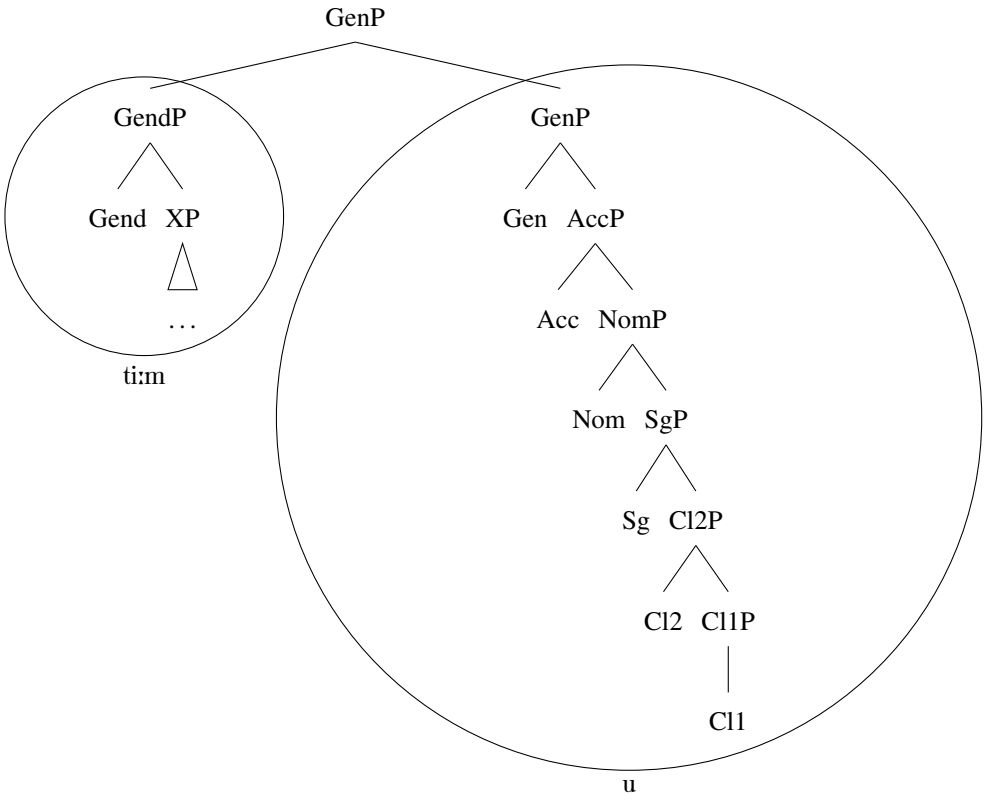
When we merge the genitive on top, we get the following structure:

(21) *Spellout of the genitive of 'team': intermediate step after attaching GenP*



Here we cannot spell out Gen, and the available movements, of Acc alone or of the entire complement, do not help either. Accordingly, we must backtrack one step to the attachment of AccP and try the next option: moving the left branch of its complement, the GendP. After this, we can reattach GenP and pass this left branch up once more, yielding the following structure:

(22) *Spellout of the genitive of 'team': final step*



As was the case earlier in the derivation before we attached NomP, for example in (19), the root now spells out GendP and the right branch is spelled out by (12-b) as /u/—thus, the genitive is /tiːmu/. From here, the derivation proceeds as with [vɛtʃɛr]: Loc and Dat are attached and the left branch gets passed up, so all three cases have [u].

This example shows how two nouns with the same set of features can have different inflectional patterns: the shape of the root in their lexical entry forces different derivational pathways that make different suffix lexical entries available at different points in the derivation.

3.2.2 Analyses of conjoined quadruplets

Given that no noun's inflection class must be encoded in the size and shape of its lexical item (or the morphosyntactic features in its functional sequence), the four patterns of the conjoined quadruplets (which have identical morphosyntactic features) must be associated with four different root shapes. However, the spellout algorithm only allows paradigms to differ in particular ways. For example, as shown in the previous section, two nouns can start out divergent in their patterns, and then merge through backtracking. Thus, the derivation for [vɛtʃɛr] 'evening' backtracks to Cl2 to match the lexical entry for the suffix [a] in the genitive, then backtracks further to Cl1 to match the lexical entry for the suffix [u] in the locative; by contrast, the derivation for [tiːm] already has its right branch rooted in Cl1 by the genitive, so it matches the lexical entry for [u] at that point. As Janků (2022, 176–7) discusses (somewhat indirectly), we cannot productively encode the opposite pattern, where two nouns' inflections start out the same and then diverge, unless they differ in some morphosyntactic feature. She also points out that it is totally impossible to capture the "conjoined twin" pattern, in which two nouns start out with the same overt suffix (for example, share a genitive), then diverge (have different locatives), then converge again (share a dative).

Table 6 and Table 7 show all possible pairs of paradigms in Polish and Czech, respectively. In each language, the (b) and (e) pairs are conjoined twins and cannot both be productively

Table 6 All possible pairs of paradigms in Polish, with pairs that can be captured in Nanosyntax highlighted

	(a)		(b)		(c)		(d)		(e)		(f)	
	⋮		⋮		⋮		⋮		⋮		⋮	
GEN	u	a	u	u	u	a	a	u	a	a	u	a
LOC	u	u	u	ɛ	u	ɛ	u	ɛ	u	ɛ	ɛ	ɛ
DAT	ɔvʲi	ɔvʲi	ɔvʲi	ɔvʲi	ɔvʲi	ɔvʲi	ɔvʲi	ɔvʲi	ɔvʲi	ɔvʲi	ɔvʲi	ɔvʲi
	⋮		⋮		⋮		⋮		⋮		⋮	

captured through different root shapes. This means that we can only capture two of the four

Table 7 All possible pairs of paradigms in Czech, with pairs that can be captured in Nanosyntax highlighted

	(a)	(b)	(c)	(d)	(e)	(f)
	⋮	⋮	⋮	⋮	⋮	⋮
GEN	u a	u u	u a	a u	a a	u a
LOC	u u	u ɛ	u ɛ	u ɛ	u ɛ	ɛ ɛ
DAT	u u	u u	u u	u u	u u	u u
	⋮	⋮	⋮	⋮	⋮	⋮

patterns productively in total, since any combination of three paradigms must include either both of the patterns in (b) or both of the patterns in (e).

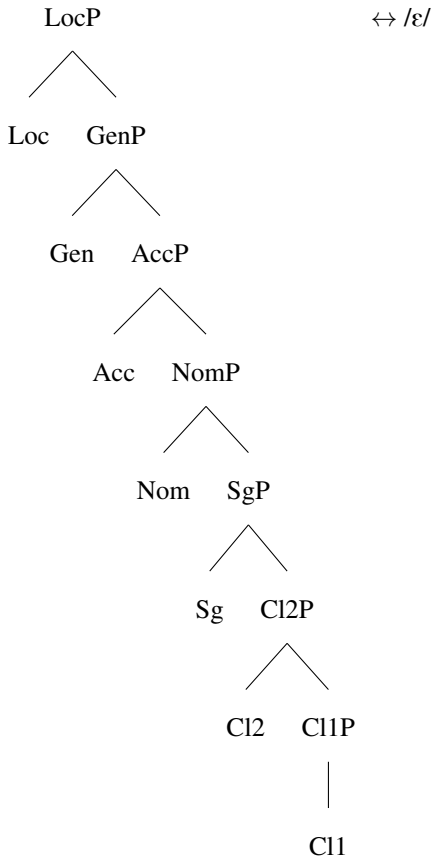
However, there are also additional limitations. The pairs in (d) cannot coexist either, because they have a single element, [u], being used in mutually exclusive cases for the two paradigms. Let us see what happens if we try to modify the previous analysis to account for this "cross pattern", exemplified by Polish [liɕtɕ] 'leaf' (which has genitive [a] and locative [u]) and [sklɛp] 'store' (which has genitive [u] and locative [ɛ]). Since [u] is capable of spelling out both genitive and locative, its lexical entry must contain both Gen and Loc heads, as it does in (12-b). Let us assume, as before, that this entry is rooted in Cl1. In the genitive, this lexical entry is preempted by that of [a], which must be rooted higher—in the previous section, at Cl2. Thus, when the genitive is added in the derivation of [liɕtɕ] 'leaf', the right branch must be rooted at Cl2. Since [a] does not contain Loc, adding Loc forces backtracking further to Cl1, where [u] becomes available. By the same logic, the lexical entry for locative [ɛ] must be rooted even further back: at Gend (the projection below Cl1), say. Now let us see what happens in the derivation of [sklɛp] 'store'. Since its genitive is [u], the right branch must be rooted at Cl1 at that point in its derivation. When Loc is attached, we can pass the left branch up and are left with a suffix containing Loc and rooted at Cl1. But this is an acceptable match for [u] as well—so there is nothing to trigger further backtracking which would make [ɛ] available. Thus, [ɛ] is impossible to reach, and we cannot capture the "cross pattern".

In addition, the [tʃas] paradigm (genitive [u], locative [ɛ], dative [u]) in Czech constitutes an ABA pattern: [u] appears in two non-contiguous sections of the case hierarchy. This is famously impossible to capture in classical Nanosyntax, as demonstrated by Janků (2022, 206–7), and its typological rarity has traditionally been considered one of Nanosyntax's main selling points. Nonetheless, such patterns do exist, and Czech inflection is one example.

Putting all this together, a Nanosyntax analysis of Polish must choose to productively represent any one of the following pairs of paradigms: (a), (c), and (f). Czech is even more limited: (f) includes the ABA pattern, so the only options are (a) and (c). These pairs are highlighted in Table 6 and Table 7, respectively. Of course, any actual analysis of these languages is much more constrained, as it must take into account all the other inflection classes (including in other genders) as well. As demonstrated in Section 3.2.1, Janků (2022) chooses pair (a) as the productive ones. She uses a third root shape for the main class of Czech masculine animate nouns, which, as mentioned in Section 1, have [a] in the genitive and accusative and vary between [u] and [ovɪ] in the dative and locative.

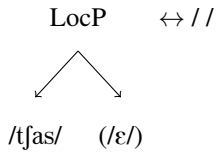
The question now is how to represent the other two inflectional patterns, which have [ɛ] in the locative. Janků (2022) does not address these patterns directly. To account for other (unrelated) patterns that cannot be represented productively in her analysis, she makes use of private lexical entries (De Clercq & Vanden Wyngaerd, 2019). These have a similar structure to normal lexical entries—thus, the private lexical entry for [ɛ] in (23) is the same as that for [u] in (12-b) except that the latter also has a dative projection on top.

(23) *Private lexical entry for Czech locative [ɛ]*



The difference between the two lies in their accessiblity. While the lexical entry in (12-b) is always visible, the private lexical entry in (23) can only be accessed through special lexical entries with pointers that require their constituent parts to be spelled out through other lexical entries. Words with [ɛ] in the locative, like [tʃas] 'time', are thus referenced in a pointer lexical entry along with the private lexical entry in (23):

(24) *Pointer lexical entry for the locative of Czech 'time'*



Since [tʃas] takes [u] in the genitive, its lexical entry has the same shape as that of [tiːm] 'team' in (18), which yields [u] in both genitive and locative. Thus, (24) will apply to a tree with LocP at the top whose left branch can be spelled out as /tʃas/ and whose right branch can be spelled out as /ɛ/ by (23) (which could, of course, always be spelled out as /u/ by (12-b) as well).

Thus, in this analysis, a Czech noun's inflection class may be distributed across multiple lexical entries: its genitive is determined by the shape of its lexical entry, and each noun with [ɛ] in the locative is additionally referenced in a separate pointer lexical entry alongside the private lexical entry in (23). The former are productive, while the latter—and, accordingly, the patterns that require them—are unproductive.

3.3 Comparison of analyses

Neither Distributed Morphology nor Nanosyntax allow for all of the patterns to be captured as productive. However, the possibilities of which patterns can be productive differ between the theories. In Distributed Morphology, any one of the four patterns (or none) can be the unmarked default, while the others must be marked with diacritic features, one for each suffix. Moreover, the choice of default suffix for each case is independent. Nanosyntax, which avoids diacritic features, instead aims to encode inflectional patterns in the shape of a root's lexical entry; it allows for up to two of the four patterns to be captured as productive. The choice, though, is not free: paradigms must be considered in relation to one another and to the surrounding case markers. Thus, only three pairs of paradigms can even potentially be productive classes in Polish, and only two in Czech, thanks to the ABA pattern caused by dative [u] in the latter. Words in the unproductive patterns have lexical entries of the same shapes

as those in the productive patterns, but additionally appear in pointer lexical entries referencing private lexical entries for unproductive case suffixes. Accordingly, the representation of inflection class is more distributed in Nanosyntax than in Distributed Morphology, where a noun's inflectional patterns can be determined entirely by examining the diacritic features (or lack thereof) in its underlying form.

4 Empirical data and its implications

In this section, I turn to empirical patterns of inflection classes in Polish and Czech (in particular, their relative type frequency and behavioral measures of productivity) and discuss their implications for the analyses in Distributed Morphology and Nanosyntax. I show that many of these empirical patterns cannot be explained by the formal grammar and propose that they are instead handled in a pattern-matching module which stores and productively applies generalizations over properties of the lexicon. The properties of this module, in turn, are dependent on the properties of the lexical entries assumed in a given theory: generalizations over the sort of lexical entries used in Distributed Morphology can be captured with a much more restrictive pattern-matching module than generalizations over those used in Nanosyntax.

4.1 Polish

4.1.1 Distribution

Table 8 lists the type frequencies of the various conjoined quadruplet paradigms among Polish masculine inanimate nouns according to Saloni et al. (2015), with nouns marked as rare or outdated removed. Also removed are a small number of nouns that are listed as variably or categorically taking [u] in the dative instead of [ɔvʲi]. Words most frequently have [u] in *exactly one case*: 12,521 words have [u] in the genitive or locative but not both, while only 3,614 have [u] in both cases or neither case.

The distribution of genitive and locative allomorphs is very different. The choice of locative suffix is almost entirely conditioned by the last consonant, which Saloni et al. (2015),

Table 8 Type frequencies of genitive and locative for Polish masculine inanimate nouns in Saloni et al. (2015)

		locative		
		u	u∼ɛ	ɛ
genitive	u	2570	0	8095
	u∼a	363	4	352
	a	4426	0	1044

following the traditional phonological division between "hard" and "soft" consonants, divide into four groups: soft (palatals, palato-alveolars, and palatalized labials), hard (non-affricate dentals and non-palatalized labials), velar (which are mostly hard but interact with the vowel system in a different way from other hard consonants), and hardened (dental affricates and retroflex sibilants, which were historically soft but now mostly act phonologically as hard). Non-velar hard consonants take locative [ɛ], which triggers a palatalization alternation in them: the locatives of [sklɛp] 'store' and [lɔt] 'flight' are [sklɛpʲɛ] and [lɔtɕɛ], respectively. Nouns ending in all other consonants take [u] in the locative. Underlying palatalization in labials is neutralized in the unsuffixed base form but inferrable from other inflected forms: the genitive of [jɛdwab] 'silk' is [jedwabʲiu], not [jɛdwabu], indicating that the underlying form ends in a palatalized labial (/jɛdwabʲ/); thus, its locative is [jɛdwabʲu], not [jɛdwabʲɛ]. The only exception in the data set is [dɔm] 'house, home', whose locative is [dɔmu] instead of the expected [dɔmʲɛ].

A noun's genitive is not fully predictable from its phonology (or any other property), and under any circumstances, speakers must memorize the genitive of large numbers of nouns. Indeed, Dąbrowska (2001) shows that during acquisition, children treat neither [u] nor [a] as a productive default, instead learning the genitive for each word separately.

As mentioned in Section 1, animate nouns show slightly different patterns from inanimate ones (their accusative is syncretic with the genitive rather than the nominative); in addition, animate nouns (with the variable exception of one noun and related forms) exclusively take [a] in the genitive. However, even when animates are added, the overall pattern remains the

same: most nouns take [u] in exactly one of the genitive and locative. This is shown in Table 9. The animate [sɨn] 'son' also exceptionally takes [u] in the locative instead of the expected

Table 9 Type frequencies of genitive and locative for Polish masculine inanimate and animate nouns in Saloni et al. (2015)

		locative		
		u	u∼ɛ	ɛ
genitive	u	2570	0	8095
	u∼a	363	4	355
	a	10 302	2	5763

[ɛ], as do (variably) forms derived from it like [skurvɨsɨn] 'son of a bitch'.

4.1.2 Distributed Morphology

The distributional acquisitional data for the genitive suggests a no default analysis: all inanimate nouns are marked with a tag for their genitive, [Gu] for words that take [u] and [Ga] for words that take [a]. There are two options for the phonological conditioning of the locative (nouns ending in non-velar consonants take [ɛ], others take [u]). First, this phonological conditioning can be built into the descriptions of the Vocabulary Items themselves. Recall the Vocabulary Items relevant for the locative in the [u] default analysis:

(5) *Vocabulary Items for [u] default locative*

a. [+obl] ↔ /u/

c. LOC ↔ /ɛ/ / [Le]___

The rule in (25) makes the context phonological, eliminating the need for the [Le] diacritic feature (here I ignore the issue of how this class is defined in terms of phonological features):

(25) *Phonologically defined Vocabulary Item for the Polish locative*

LOC ↔ /ɛ/ / [hard, non-velar]___

Exceptions like [dɔm] 'house' must avoid (25) in order to receive the locative [u] despite their phonology. One way to achieve this would be to give these words the [Lu] tag that triggers Impoverishment of the locative case features, as in the non-[u] default analysis repeated below from Section 3.1.2:

(7-b) *Rule of Impoverishment for non-[u] default locative*

[−subj, −gov] → Ø / [Lu]___

The second possibility is that all nouns are marked with [Le] or [Lu], as in the no default analysis. In this case, the near-categorical phonological conditioning would be represented in a *pattern-matching module* that derives generalizations over stems that share a diacritic feature (cf. Gouskova et al., 2015; Tabachnick, 2024). This module probabilistically assigns features (in this case, [Le] or [Lu]) to underlying forms of novel lexical items by evaluating candidate forms using a constraint-based Maximum Entropy grammar (see Hayes & Wilson, 2008). Each constraint has a weight that determines the penalty it assesses to forms that violate it; a form's score is the sum of its weighted violations. These scores are then converted to probabilities: forms with better scores (less bad violations) are more likely to be selected as the underlying form and placed into the lexicon, accordingly. In Polish, this module would thus contain very strong morpheme structure constraints (cf. Booij, 2011) indicating that underlying forms marked with the [Le] feature always end with non-velar hard consonants, while underlying forms marked with the [Lu] feature almost never do. To form the locative of a new word (a loan or nonce word), Polish speakers must assign either [Le] or [Lu] to its underlying form, and do so by invoking the pattern-matching module with the constraints which enforce the correlation between features and stem-final consonants.

The pattern-matching module of Gouskova et al. (2015) does not comprise constraints on underlying forms, but rather on base (nominative singular) and inflected surface forms for stems with a given feature. The Polish pattern cannot be fully described in terms of either the base or inflected form alone. On the one hand, the unsuffixed nominative singular neutralizes

the distinction, described above, between stem-final unpalatalized labials (which take locative [ɛ]) and palatalized labials (which take locative [u]). On the other hand, since locative [ɛ] palatalizes consonants, the two suffixes do not make a clear cut in the consonants that precede them on the surface: for example, [tɕ] is followed by [ɛ] in [ɔtstɕɛ], the locative of [ɔtsɛt] 'vinegar', but by [u] in [kɔptɕu], the locative of [kɔpʲɛtɕ] 'soot'. While these problems may be surmountable, generalizations over surface forms are also problematic for the Czech data discussed in Section 4.2, so I will not consider them further.

The pattern-matching account of conditioned allomorphy has several advantages. First, as Gouskova et al. (2015) note, it allows exceptions like [dɔm] 'house' without the need for Impoverishment rules like (7-b).

Second, it provides a mechanism for no-default behavior, as described for the Polish genitive, without the accidental homophony we saw in (8):

(8) *Vocabulary Items for no default*

a. GEN ↔ /u/ / [Gu]___

b. LOC ↔ /u/ / [Lu]___

c. GEN ↔ /a/ / [Ga]___

d. LOC ↔ /ɛ/ / [Le]___

That is, as described in Section 3.1.2, we can have the equivalent of rules like (5) for the genitive, which invoke [Ga] but not [Gu]; nouns with genitive [u] are tagged with [Gu] as a bookkeeping mechanism, but this feature does not get used in the course of derivation. On its own, this pattern would still yield [u] as a productive default genitive: new nouns, which do not have a feature, would go through the default rule in (5-a). The lack of default, then is enforced in the pattern-matching module: a morpheme structure constraint requires that lexical items contain some genitive feature, either [Ga] or [Gu]. To make the genitive of a new word, speakers must first provide it with a licit underlying form, which means assigning it

one of the features. Thus, they would not have one suffix that is used by default in the absence of marking; a decision must be made one way or the other.

Third, the pattern-matching module can capture the gradient generalization that speakers tend to have [u] in either the genitive or locative but not both: this would be expressed in a weaker morpheme structure constraint that moderately, but not overwhelmingly, penalizes [Gu] and [Lu] appearing on the same lexical items (or, alternately, correlating the presence of [Gu] and [Ga] with certain stem-final consonants determining the locative). Speakers who have learned these constraints would apply them stochastically when forming the genitive of new words, for example in nonce word studies. Dąbrowska (2008) found that a minority of Polish 18-year-olds used phonological regularities to assign genitives to masculine inanimate nonce words; other speakers used other strategies, including default [a], default [u], using both interchangeably, and generalizing a semantic regularity to assign [a] to objects and [u] to substances (see also Dąbrowska, 2019). However, she does not elaborate on which phonological regularities her participants applied. As discussed in Section 4.2, Czech speakers do show stochastic sensitivity to gradient generalizations of this sort in nonce word studies.

The pattern-matching module is not an inherent or necessary component of Distributed Morphology, and represents an alternative source of productivity located in the assignment of diacritic inflectional features to new words before they are derived through the modules described by Distributed Morphology. However, the use of diacritic features in Distributed Morphology makes the theory a suitable complement to the pattern-matching module described here, which takes advantage of the presence of such features in underlying forms.

4.1.3 Nanosyntax

As discussed in Section 3.2.2, a Nanosyntax analysis of Polish must choose one of the three pairs of paradigms highlighted in Table 10 to capture productively (without use of private lexical entries). This table replicates Table 6 and adds the number of nouns covered by each pair of paradigms; that is, the number of nouns falling into either of the two paradigms in a given pair.

Table 10 All possible pairs of paradigms in Polish and the number of inanimate nouns they include, with pairs that can be captured in Nanosyntax highlighted

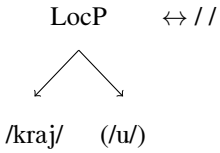
	(a)	(b)	(c)	(d)	(e)	(f)
	⋮	⋮	⋮	⋮	⋮	⋮
GEN	u a	u u	u a	a u	a a	u a
LOC	u u	u ɛ	u ɛ	u ɛ	u ɛ	ɛ ɛ
DAT	ɔvʲi ɔvʲi	ɔvʲi ɔvʲi	ɔvʲi ɔvʲi	ɔvʲi ɔvʲi	ɔvʲi ɔvʲi	ɔvʲi ɔvʲi
	⋮	⋮	⋮	⋮	⋮	⋮
	6996	10665	3614	12521	5470	9139

Table 10 restates the finding from Table 8 that most nouns have [u] in the genitive or the locative but not both—that is, pair (d) is the most frequent. However, this "cross-pattern" cannot be captured in Nanosyntax. The best we can do is pattern (f), which has two productive genitive suffixes and a fixed [ɛ] in the locative. On the one hand, the symmetrical variation in the genitive matches the claim of Dąbrowska (2001) that neither genitive suffix is productive—speakers must place words into either of two viable options. However, the locative is problematic for this analysis (or the alternative, (a), which uses [u] as the productive locative instead of [ɛ]). Although a noun's locative is almost entirely predictable from its phonology, Nanosyntax has no way to encode this, so all nouns ending in certain classes of consonants would have to have their locative specially encoded in private lexical entries. This is not a particularly appealing way to encode phonologically conditioned suppletive allomorphy.

The only way to resolve this issue in Nanosyntax would be to say that the locative allomorphy is not suppletive; that is, that locative [u] and [ɛ] share a single underlying form whose realization in a given word is determined by the phonology. This analysis faces multiple challenges: /u/ and /ʲɛ/ (where /ʲ/ indicates whatever property palatalizes preceding consonants) do not share phonological features, and the allomorphy does not seem to be phonologically optimizing (cf. Paster, 2015). An approach like that of Myler (2026), in which unproductive alternations are captured through underspecified underlying segments, may be adaptable to this sort of analysis, though I do not pursue one further at present.

If the two locatives cannot be collapsed to a single underlying form, one of them must be unproductive. For example, if we take the pair of paradigms in (f) in Table 10 to be productive, we have two shapes of productive lexical entries that differ in whether they have [a] or [u] in the genitive; both types have [ɛ] in the locative. The locative is only spelled out by [u] as a private lexical entry, as described in Section 3.2.2 (the exact shape of this private lexical entry would depend on the rest of the analysis, so I do not show it here). All nouns that have [u] in the locative also appear in a pointer lexical entry referencing this private lexical entry, as shown in (26) for [kraj] 'country':

(26) *Pointer lexical entry for the locative of Polish 'country'*



Since these classes are (by assumption of the analysis) unproductive, we need a back route to add new words to them—even though they together contain every word ending in a consonant other than a non-velar hard consonant. One possibility would be for a pattern-matching module similar to that described in Section 4.1.2 to enforce a strong restriction on lexical entries. Looking at all lexical entries, this module will learn a strong constraint that, in pointer lexical entries like (26), the underlying form spelled out on the left branch (i.e. the root) very rarely ends in a non-velar hard consonant. This constraint will correctly—if gratuitously—prevent the creation of new pointer lexical entries for nouns ending in these consonants. However, it cannot force the creation of such pointer lexical entries for nouns ending in other consonants. In order to do so, the pattern-matching module must take an even broader scope to be able to make the following generalization: if a lexical entry spells out a root shape for the two paradigms in pair (f) in Table 10 with a phonological form ending in one of a certain set of consonants (all but non-velar hard consonants), the lexicon must also contain a pointer lexical entry like (26) containing that phonological form on its left branch. That is, in

considering additions to the lexicon, this pattern-matching module must be able to evaluate how it affects the lexicon as a whole. Thus, the way that Nanosyntax handles non-productive morphology—by distributing a root's inflectional properties between productive root shapes and unproductive pointer lexical entries—requires a much more powerful pattern-matching module than does Distributed Morphology with diacritic markers, which captures the same productive generalizations by looking at the properties of individual underlying forms and adding to the lexicon through local evaluation of candidates for a new underlying form.

4.2 Czech

4.2.1 Distribution

Table 11 shows the type frequencies of the genitive and locative of Czech hard-stem masculine inanimate nouns from the SYNv11 corpus of the Czech National Corpus (Křen et al., 2022), a very large corpus mostly comprising texts (especially from newspapers) from the past two decades. For reasons unrelated to the present data, I limited my search to noun tokens appearing as prepositional objects. I also removed proper nouns, nouns with other irregularities (mostly unassimilated foreign words), and very rare words (with fewer than five tokens in the data set). Nouns with greater than 1% of tokens with each suffix allomorph for a given case were considered variable. For more details about the data set, see [self-citation].

Table 11 Type frequencies of genitive and locative for Czech hard-stem masculine inanimate nouns in Křen et al. (2022)

		locative		
		u	u∼ɛ	ɛ
genitive	u	9686	523	21
	u∼a	145	18	3
	a	32	19	31

The distribution of suffixes in Czech is very different from that of Polish. The vast majority of nouns in the data set have [u] in both cases, and nouns that never appear with [u] in

either case are exceedingly rare. In fact, the distribution of suffixes is not quite as extreme as shown in Table 11. First, my decision to remove proper nouns eliminated a large class of place names suffixed with [ov] or [iːn] that typically take [a] in the genitive and very high rates of [ɛ] in the locative (on the other hand, it is possible that speakers have learned this as a property of the suffixes [ov] and [iːn], in which case their influence on generalizations made over type counts stored in the lexicon is minimal). Second, as mentioned in Section 2, masculine animate nouns in this class always have [a] in the genitive and almost always vary between [u] and [ovɪ] in the locative. That being said, the results discussed in this section make sense under the assumption that speakers are keeping track of masculine inanimate nouns as a separate class from masculine animate nouns. This assumption is reasonable given that animacy is represented morphosyntactically: as mentioned previously, for example, in inanimate nouns the accusative is syncretic with the nominative, while in animate nouns it is typically syncretic with the genitive, and this pattern also appears in agreeing adjectives. While the inflection of animates may have an influence on analytical choices in Nanosyntax—as they do for Janků (2022)—I do not discuss them further.

Table 11 shows a correlation between the two cases: of the 10,230 nouns that take [u] in the genitive, 9,686 (94.7%) also take [u] in the locative; of the 248 nouns that variably or always take [a] in the genitive, only 177 (71.4%) always take [u] in the locative. For the even smaller group of nouns that always take [a] in the genitive, locative [u] is the minority pattern (32/82, or 39.0%). In addition, there are some gradient phonological generalizations over the choice of locative (see [self-citation]): in particular, locative [ɛ] is relatively rare among nouns ending in velars, in which it triggers a salient consonant alternation (for example, the locative of [jazɪk] ‘language’ is [jazɪtsɛ]).

In [self-citation], I conducted a pair of nonce word studies showing that speakers are sensitive to both the phonological and morphological generalizations that are partially predictive of locative allomorphy. First, participants assigned locative [ɛ] less often to nonce words that ended in velars than those that ended in consonants with a higher rate of [ɛ] in the lexicon, like

alveolar stops. Second, they assigned [ɛ] in the locative more often to words which had been presented as having [a] in the genitive. Thus, these studies showed that speakers have learned the gradient correlation between genitive and locative in their lexicon and apply it (alongside other factors) in stochastically selecting the locative of new words (cf. Hayes, Zuraw, Siptár, & Londe, 2009).

4.2.2 Distributed Morphology

The lopsided distribution of forms suggests an analysis where [u] is default in both cases. That is, roots with [u] in both cases are unmarked, while nouns with genitive [a] have the [Ga] feature and nouns with locative [ɛ] have the [Le] feature—setting aside the pervasive variation, which I discuss at length in [self-citation]. These features are then referenced by the rules below, repeated from Section 3.1.2:

(5) *Vocabulary Items for [u] default*

a. [+obl] ↔ /u/

b. GEN ↔ /a/ / [Ga]___

c. LOC ↔ /ɛ/ / [Le]___

The gradient generalizations can be captured using the pattern-matching module discussed for Polish in Section 4.1.2. That is, speakers know that underlying forms with the [Ga] feature are relatively likely to also have the [Le] feature, so they have an active medium-strength constraint pushing newly formed lexical entries in this direction. For the phonology, they also know have constraints encoding the fact that certain final consonants (in particular, velars) are underrepresented on lexical entries with the [Le] feature.

In this case, the phonological effect has a functional explanation: [ɛ] is dispreferred after consonants in which it induces a salient alternation. The pattern-matching module does not capture or encode this causality, and could just as easily encode the opposite pattern, in which [ɛ] is preferred after consonants in which it induces a salient alternation (this is, in fact, what

happens in Polish). Thus, in this conception, the module does not have any particular bias towards learning phonologically or morphologically "natural" patterns, as has been proposed by Becker, Ketrez, and Nevins (2011) and others. Instead, in this case we can say that the lexicon has the form that it does due to functional pressures having an effect outside of the grammar, and the pattern-matching module faithfully encodes the patterns of the resulting lexicon rather than the functional pressures themselves.

4.2.3 Nanosyntax

While Polish has three possible pairs of paradigms that can be captured productively in Nanosyntax, Czech has only two. As discussed in Section 3.2.2, the additional limitation in Czech comes from the fact that the dative is also [u], creating the possibility for an impossible ABA pattern with the genitive and the locative. This is shown in Table 12, which extends Table 7 to include the number of nouns captured by each pair of paradigms. Variable nouns are counted under [a] in the genitive and [ɛ] in the locative, respectively. Unsurprisingly,

Table 12 All possible pairs of paradigms in Czech and the number of inanimate nouns they include, with pairs that can be captured in Nanosyntax highlighted

	(a)	(b)	(c)	(d)	(e)	(f)
	⋮	⋮	⋮	⋮	⋮	⋮
GEN	u a	u u	u a	a u	a a	u a
LOC	u u	u ɛ	u ɛ	u ɛ	u ɛ	ɛ ɛ
DAT	u u	u u	u u	u u	u u	u u
	⋮	⋮	⋮	⋮	⋮	⋮
	9863	10230	9757	721	248	615

given the lopsided distribution of suffix forms, the main split in Table 12 is between pairs that include nouns with [u] in both cases and those that do not. Thus, the choice is which of the other patterns to capture productively.

The largest other class has [u] in the genitive and [ɛ] in the locative. However, this class forms an ABA pattern and cannot be captured productively in Nanosyntax at all. As shown in Section 3.2.2, Janků (2022) chooses pair (a) based on considerations of the rest of the morphological system (in particular, masculine animates, which always have [a] in the genitive and variably have [u] in the locative and dative). In this case, nouns with locative [ɛ] would each appear in associated pointer lexical entries. As in Distributed Morphology, the productive gradient phonological and morphological generalizations influencing locative allomorphy would have to be encoded in a pattern-matching module. However, as discussed in Section 4.1.3 for Polish, this pattern-matching module would have to be quite powerful, capturing correlations across multiple lexical entries (as opposed to the pattern-matching module paired with Distributed Morphology, which only needs to capture correlations within a single phonological underlying form). To capture the phonological effect, it is not enough to say that pointer lexical entries rarely point to noun roots ending in velar consonants: since pointer lexical entries are unproductive within the morphological system, speakers would not be motivated to create new pointer lexical entries for nonce words unless they have learned a cross-derivational constraint pairing certain productive lexical entries with pointer lexical entries. This is easier to conceive for the productive morphological generalization that words with genitive [a] also have a higher likelihood of having locative [ɛ]. Here the relevant constraint is: if a lexical entry spells out a root shape common to nouns with genitive [a] and locative [u]—shown in (10) in Section 3.2.2—the phonological form it spells out is likely to be referenced in a locative pointer lexical entry. Thus, when speakers encounter a masculine inanimate noun with genitive [a], giving them evidence to create a lexical entry for it with the root shape in (10), they will also have some motivation to create a parallel pointer lexical entry for its locative.

Alternatively, we could make productive root shapes for pair (c) in Table 12—that is, make the class with genitive [a] and locative [ɛ] productive. This has the benefit of capturing the correlation between genitive [a] and locative [ɛ]: a noun can productively have [u] in either both cases or neither. However, this analysis encodes the correlation too strongly, predicting

categorical application of this correlation to new words—in fact, there are still a number of words with [u] in only one case, and the experimental effect of a nonce word's genitive to participants' choice of locative in [self-citation] was quite modest. This analysis is also somewhat awkward representationally: the private lexical entries used for the two classes not captured spell out their case forms as suffixes which are also used productively. For example, let us consider a noun like [vɛtʃɛr] 'evening', which has [a] in the genitive and [u] in the locative. If this noun has the root shape of the all-[u] class, it needs a pointer lexical entry for the genitive referencing a private lexical entry spelling out the genitive as [a]. However, other masculine inanimate nouns productively spell out the genitive as [a] using a non-private lexical entry with a slightly different shape. Similarly, if [vɛtʃɛr] instead has the root shape of the no-[u] class, it has a pointer lexical entry to get [u] in the locative—but [u] is the productive locative suffix for the other class. Such duplication is unavoidable in this case, even before we consider all of the other noun paradigms in the Czech inflectional system as Janků (2022) did.

4.3 Summary

Both Distributed Morphology and Nanosyntax run into issues when faced with the full empirical scope of genitive and locative allomorphy in Polish and Czech. In particular, speakers show sensitivity to gradient generalizations that are not easily captured in formal, categorical morphological theories. Nanosyntax has particular trouble with the near-categorical phonological conditioning of the Polish locative—unless this conditioning is analyzed as allophony of a single, highly abstract underlying form. Indeed, the restrictiveness of Nanosyntax encourages linguists to attempt this sort of analysis for cases of (so-called) phonologically conditioned suppletive allomorphy in general.

To fully capture the patterns of speakers' behavior, I suggest that speakers make use of a pattern-matching module that makes (often gradient) generalizations over the forms in their lexicon and applies them stochastically to novel forms. This module is, in theory, compatible

with either Distributed Morphology or Nanosyntax. However, because non-productive morphology is indexed in Nanosyntax by the existence of a separate pointer lexical entry, the theory requires a much more powerful pattern-matching module than Distributed Morphology, where the module is limited to phonological underlying forms that may contain diacritic features indexing non-productive morphology.

5 Conclusion

Polish and Czech nouns show a pattern of "conjoined quadruplets": given two masculine inanimate suffixes in the genitive ([u] and [a]) and locative ([ɛ] and [a]), all four logically possible combinations are attested among nouns whose paradigms are otherwise identical. At first glance, this poses a problem for analysts in Distributed Morphology and "classical" Nanosyntax, which must choose a subset of the four paradigms as productive defaults and relegate the others to some way of marking unproductive morphology. Thus, the common theoretical practice of comparing two analyses according to their ability to predict the attested patterns is insufficient; instead, we must also consider relations *between* the attested patterns, including relative frequency, variation, gradient generalizations, and data from acquisition and behavioral studies.

In particular, I propose that speakers keep track of phonological and morphological patterns across the shape of their lexicon in a separate module that is invoked when they need to fill out underdetermined lexical entries (typically, for the purpose of inflecting novel lexical items). In Distributed Morphology, it is common to mark inflectional paradigms with diacritic features. This symbolic representation of inflectional behavior allows for the relevant patterns to be captured entirely through generalizations over phonological underlying forms in the tradition of morpheme structure constraints (see Booij, 2011). In Nanosyntax, by contrast, encoding of inflectional class is distributed across multiple lexical entries, requiring a much more powerful pattern-matching module. Proponents of Nanosyntax (e.g. Caha, 2021) have argued that the use of diacritic features leads to a less restrictive theory of grammar, since they

are clearly language-specific and thus allow more degrees of freedom than a theory of grammar that only makes use of universal features. However, we see that the avoidance of diacritic features in Nanosyntax here has the consequence of requiring a much more powerful (that is, less restrictive) pattern-matching module. Thus, claims about restrictiveness of theories of grammar should broaden their range to consider modelling speakers' knowledge of patterns, which is an important, though often underappreciated, aspect of linguistic knowledge.

Nanosyntax is still a relatively young theory, and this comparative work comes at a point of transition to a new spellout algorithm (not discussed in this paper) that broadens the range of patterns that can be captured productively. Indeed, Caha (2026), in response to an earlier version of this work, shows that all four of the Czech "conjoined quadruplets" can be accounted for in the new version of the theory under certain assumptions—in this case, met—about the rest of the inflectional system (van Craenenbroeck, 2025). It would be useful for consideration of this and other recent analyses in Nanosyntax to look at the implications for the sorts of gradient behavioral and distributional data I discuss in this paper, and it is my hope that linguists use this work as an inspiration and guide for how to do so productively.

Declarations

- Funding

  The research, presentation, and writing of this paper were financially supported by National Science Foundation (NSF) Doctoral Dissertation Research Improvement Grant BCS-2214315, Slovenian Research Agency (ARIS) grants P6-0382 and J6-4614, and Grant Agency of Masaryk University grant MUNI/SC/1432/2024.

References

Baunaz, L., & Lander, E. (2018). Nanosyntax: The basics. L. Baunaz, K. De Clercq, L. Haegeman, & E. Lander (Eds.), *Exploring nanosyntax* (pp. 3–56). New York: Oxford University Press.

Becker, M., Ketrez, N., Nevins, A. (2011). The surfeit of the stimulus: Analytic biases filter lexical statistics in Turkish laryngeal alternations. *Language*, *87*(1), 84–125, https://doi.org/10.1353/lan.2011.0016

Blix, H. (2021). Phrasal Spellout and Partial Overwrite: On an alternative to backtracking. *Glossa: a journal of general linguistics*, *6*, , https://doi.org/10.5334/gjgl.1614

Bobaljik, J.D. (2000). The ins and outs of contextual allomorphy. K. Grohmann & C. Struijke (Eds.), *University of Maryland working papers in linguistics* (pp. 35–71). College Park.

Booij, G. (2011). Morpheme structure constraints. M. van Oostendorp, C.J. Ewen, E. Hume, & K. Rice (Eds.), *Phonological interfaces* (Vol. IV, pp. 2049–2069). Malden: Wiley-Blackwell.

Caha, P. (2009). *The nanosyntax of case* (Unpublished doctoral dissertation). University of Tromsø.

Caha, P. (2021). Modeling declensions without declension features: The case of Russian. *Acta Linguistica Academica*, *68*(4), 385–425, https://doi.org/10.1556/2062.2021.00433

Caha, P. (2026, June). *ABAs and subextractions, Nanoseminar.* Zenodo. Retrieved from https://doi.org/10.5281/zenodo.20545383

Cameron-Faulkner, T., & Carstairs-McCarthy, A. (2000). Stem alternants as morphological stigmata: Evidence from blur avoidance in Polish nouns. *Natural Language and Linguistic Theory*, *18*, 813–835, https://doi.org/10.1023/a:1006496821412

Cortiula, M. (2023). *The nanosyntax of Friulian verbs* (Unpublished doctoral dissertation). Masaryk University, Brno.

Dąbrowska, E. (2001, July). Learning a morphological system without a default: the Polish genitive. *Journal of Child Language*, *28*(3), 545–574, https://doi.org/10.1017/s0305000901004767

Dąbrowska, E. (2008). The later development of an early-emerging system: the curious case of the Polish genitive. *Linguistics*, *46*(3), 629–650, https://doi.org/10.1515/LING.2008.021

Dąbrowska, E. (2019). Individual differences in grammatical knowledge. E. Dąbrowska & D. Divjak (Eds.), *Cognitive linguistics: Key topics* (Vol. 3, pp. 231–250). Berlin, Boston: De Gruyter Mouton.

De Clercq, K., Caha, P., Starke, M., Vanden Wyngaerd, G. (2025). Nanosyntax: State of the art and recent developments. P. Caha, K. De Clercq, & G. Vanden Wyngaerd (Eds.), *Nanosyntax and the lexicalization algorithm* (pp. 1–44). Oxford: Oxford University Press.

De Clercq, K., & Vanden Wyngaerd, G. (2019, November). On the idiomatic nature of unproductive morphology. *Linguistics in the Netherlands*, *36*(1), 99–114, https://doi.org/10.1075/avt.00026.cle

Gouskova, M., Newlin-Łukowicz, L., Kasyanenko, S. (2015). Selectional restrictions as phonotactics over sublexicons. *Lingua*, *167*, 41–81, https://doi.org/10.1016/j.lingua.2015.08.014

Halle, M., & Marantz, A. (2008). Clarifying "Blur": Paradigms, defaults, and inflectional classes. A. Bachrach & A. Nevins (Eds.), *Inflectional identity* (pp. 55–72). Oxford: Oxford University Press.

Haugen, J.D., & Siddiqi, D. (2016). Towards a restricted realization theory: Multimorphemic monolistemicity, portmanteaux, and post-linearization spanning. D. Siddiqi & H. Harley (Eds.), *Morphological metatheory* (pp. 343–385). Amsterdam, Philadelphia: John Benjamins.

Hayes, B., & Wilson, C. (2008, Summer). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry*, *39*(3), 379–440, https://doi.org/10.1162/ling.2008.39.3.379

Hayes, B., Zuraw, K., Siptár, P., Londe, Z. (2009). Natural and unnatural constraints in Hungarian vowel harmony. *Language*, *85*(4), 822–863, https://doi.org/10.1353/lan.0.0169

Jakobson, R. (1984). Morphological observations on Slavic declension (the structure of Russian case forms). L.R. Waugh & M. Halle (Eds.), *Russian and Slavic grammar: Studies 1931–1981* (pp. 105–133). Berlin: Mouton.

Janků, L. (2022). *The nanosyntax of Czech nominal declensions* (Unpublished doctoral dissertation). Masaryk University, Brno.

Kasenov, D. (2025). ABA in Russian adjectives, subextraction, and Nanosyntax. B. Gehrke, D. Lenertová, R. Meyer, D. Seres, L. Szucsich, & J. Zaleska (Eds.), *Advances in formal Slavic linguistics 2022* (pp. 255–292). Berlin: Language Science Press.

Křen, M., Cvrček, V., Henyš, J., Hnátková, M., Jelínek, T., J., K., … Škrabal, M. (2020). *SYN2020: reprezentativní korpus psané češtiny.* Prague: Ústav Českého národního korpusu FF UK. Retrieved from www.korpus.cz

Křen, M., Cvrček, V., Hnátková, M., Jelínek, T., Kocek, J., Kovářík ová, D., … Škrabal, M. (2022). *Korpus SYN, verze 11 z 14.12.2022.* Prague. Retrieved from www.korpus.cz

Müller, G. (2004). On decomposing inflection class features: Syncretism in Russian noun inflection. G. Müller, L. Gunkel, & G. Zifonun (Eds.), *Explorations in nominal inflection* (pp. 189–227). Berlin: Mouton de Gruyter.

Myler, N. (2026). *The Spanish PYTA morphome dissolved.* (lingbuzz/009692)

Paster, M. (2015). Phonologically conditioned suppletive allomorphy: Cross-linguistic results and theoretical consequences. E. Bonet, M.-R. Lloret, & J. Mascaró (Eds.), *Understanding allomorphy: Perspectives from Optimality Theory* (pp. 218–253). Sheffield: Equinox.

Privizentseva, M. (2024, May). Semantic agreement in Russian: Gender, declension, and morphological ineffability. *Natural Language & Linguistic Theory*, *42*(2), 767–814, https://doi.org/10.1007/s11049-023-09587-0

Rusínová, Z., & Nekula, M. (1995). Morfologie. P. Karlík, M. Nekula, & Z. Rusínová (Eds.), *Příruční mluvnice češtiny* (pp. 227–367). Prague: Nakladatelství Lidové noviny.

Saloni, Z., Woliński, M., Wołosz, R., Gruszczyński, W., Skowrońska, D. (2015). *Słownik gramatyczny języka polskiego* (3rd ed.). Warsaw.

Starke, M. (2018). Complex left branches, spellout, and prefixes. L. Baunaz, K. De Clercq, L. Haegeman, & E. Lander (Eds.), *Exploring nanosyntax* (pp. 239–249). New York: Oxford University Press.

Tabachnick, G. (2024, August). *Czech speakers productively apply correlations between inflected forms.* (Poster presented at International Morphology Meeting 21, Vienna University of Economics and Business)

Ústav pro jazyk český AV ČR, v.v.i. (2008–2026). *Internetová jazyková příručka.* Prague. Retrieved from https://prirucka.ujc.cas.cz

van Craenenbroeck, J. (2025). *Blixemes, pointers, and *ABA.* Ms.