

Overabundance and the interface

Noah Diewald
Ohio State University

February 25, 2021

There are three parts to this talk.

1. Description of overabundance
2. Problems for the Paninian morphology-syntax interface
3. The interface as a relationship between two paradigmatic systems

How the linguistic examples relate to these goals.

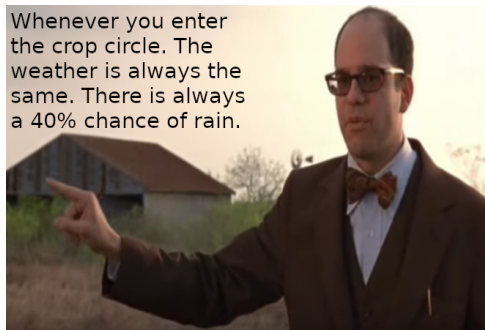
1. English patterns ground the discussion in a familiar language.
2. Quechua patterns provide simple, minimal conditions that break Paninian rule ordering.
3. Wao Tededo patterns demonstrate how a multi-paradigm approach can be leveraged to handle more complex cases.

What is overabundance?

Overabundance is when more than one inflected form of a lexeme is compatible with the same inflectional category. (Thornton, 2011)

I use the following definition of free variation.

Within the rules (or constraints) of a grammar, speakers have a non-deterministic choice for the realization of a category. Conditioned stochastic tendencies are compatible with this definition.



There is a famous example from English.

Quirk (1970) and Haber (1976) addressed the question of free variation in leap's past tense from different perspectives.

sleep/slept beep/beeped

keep/kept heap/heaped



leap/{leapt, leaped}

There is overabundance in Cochabamba Quechua plurals.¹

warmi-**s**

woman-PL

warmi-**kuna**

woman-PL

warmi-**s-kuna** warmi-**kuna-s**

woman-PL-PL

woman-PL-PL

‘women’

Lastra (1968), Muysken (2002), and Plaza Martínez (2010)

¹Thank you to my colleague Elvia Andina for reviewing this data.

The -s plural is more restricted based on phonological criteria.

In general -s cannot be used with stems ending in a consonant.

sipas-s **No** | sipas-kuna **OK**
young.woman-pl

There is some dialectal variation.

This meme was posted by a native speaker who doesn't like how Bolivian Quechua is taught at universities.



Misikuna

Wallpakuna

Wasikuna

Sach'akuna

Quwikuna



Misis

Wallpas

Wasis

Sach'as

Quwis

There may be meaning contrasts in some contexts.

The -kuna plural may be more appropriate for subjects with contrastive focus.

Warmikuna llaqtaman purinku.

women to.town they.walk

‘Son las mujeres las que van al pueblo.’

trans. ‘It is the women who walk to the town.’

Plaza Martínez (2010, p. 33 (English by me))

The forms overlap in distribution.

Consonant Stem

-s

+kuna

Focal Subject

-s

+kuna

Vowel Stem

+s

+kuna

Elsewhere

+s

+kuna

Why is overabundance 'too abundant'?

The term is deeply embedded in a Cartesian theory of the morphology-syntax interface. Thornton talks of 'cellmates' to describe the pairs in the plural column.

	Singular	Plural
Nominative	warmi	warmis warmikuna
Accusative	warmi	warmista warmikunata

This English pattern exemplifies an asymmetry between morphosyntactic features and inflected forms.

	Singular	Plural
1	walk	walk
2	walk	walk
3	walks	walk

This pattern has dominated paradigmatic theories (Anderson, 1992; Stump, 2001) and hybrid systems like Distributed Morphology (Halle and Marantz, 1993) since Robins (1959).

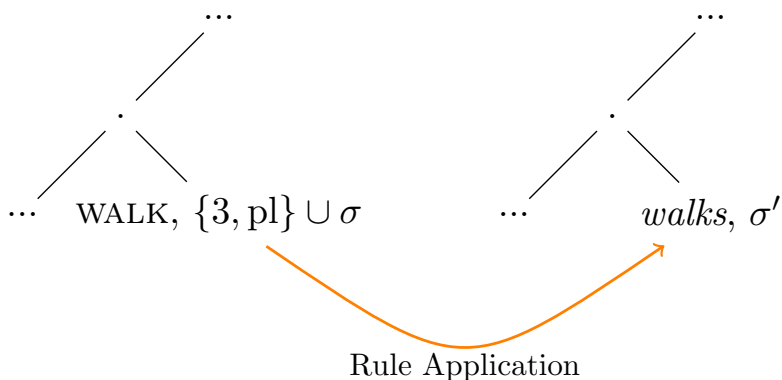
The asymmetric pattern is modeled using Paninian or elsewhere rule ordering.

Paninian rules are ordered by specificity. Rules are in competition and *only one* rule matches.

$$1: \langle X, \{\text{person: 3, number: singular}\} \in \sigma \rangle \rightarrow \langle X_s, \sigma \rangle$$

$$2: \langle X, \emptyset \in \sigma \rangle \rightarrow \langle X, \sigma \rangle$$

Often the features in question are said to exist at a syntactic node.



The Quechua plural pattern cannot be represented in a Paninian system.

1: $\langle X, \{\text{focal, plural, nominative}\} \in \sigma \rangle \rightarrow \langle X\text{kuna}, \sigma \rangle$

2: $\langle X, \{\text{plural}\} \in \sigma \rangle \rightarrow \langle X\text{s}, \sigma \rangle$

2: $\langle X, \{\text{plural}\} \in \sigma \rangle \rightarrow \langle X\text{kuna}, \sigma \rangle$

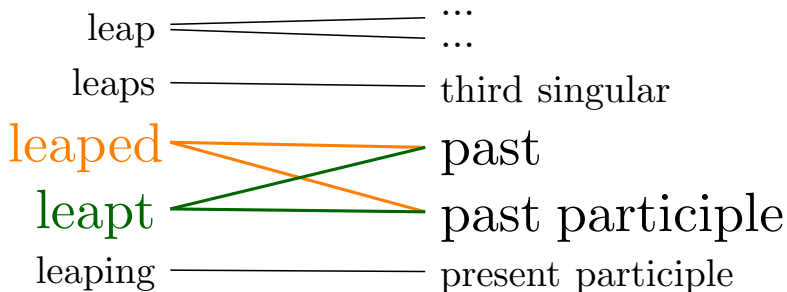
Non-deterministic specificity based ordering is possible.

Applying all rules that are compatible is not Paninian (Bonami and Boyé, 2005).

1: $\langle X, \{\text{focal, plural, nominative}\} \in \sigma \rangle \rightarrow \langle X\text{kuna}, \sigma \rangle$

2: $\langle X, \{\text{plural}\} \in \sigma \rangle \rightarrow \{\langle X\text{s}, \sigma \rangle, \langle X\text{kuna}, \sigma \rangle\}$

Logically, the relationship between word forms and inflectional categories is many-to-many.



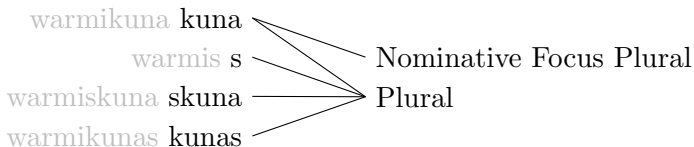
Purely morphological categories allow for abstraction and generalization.

Sadler and Spencer (2001) propose that there are morphological categories that are correlated to but not determined by syntactic (inflectional) categories. Similar tiered notions have gained popularity following general acknowledgment of overabundance (Stump, 2016).

leap	m-cat ₁	=====	...
leaps	m-cat ₂	—————	third singular
leaped	m-cat ₃	⌵	past
leapt	m-cat ₄	⌶	past participle
leaping	m-cat ₅	—————	present participle

For the Quechua data, we can propose multiple morphological "plural" categories.

I like to name these categories after characteristics of the form.



Using a hierarchy of categories makes generalizations possible.

**Morphological
Categories**

M-Plural

kunas skuna s kuna

**Syntactic(ish)
Categories**

Plural

Nominative Focus
Plural

What are the cells for?

The Cartesian cell is an artifact of the deterministic interface assumption. In a lexical theory, we can replace cells with a paradigm of lexical entries (Pollard and Sag, 1987; Koenig and Jurafsky, 1994; McConville, 2006).

$\left[\begin{array}{ll} \text{PHON} & \text{warmikuna} \\ \text{SYN} & \textit{N.pl} \\ \text{SEM} & \textit{women} \end{array} \right]$	$\left[\begin{array}{ll} \text{PHON} & \text{warmikuna} \\ \text{SYN} & \textit{N.nom.pl} \\ \text{SEM} & \textit{focus(women)} \end{array} \right]$
$\left[\begin{array}{ll} \text{PHON} & \text{warmis} \\ \text{SYN} & \textit{N.pl} \\ \text{SEM} & \textit{women} \end{array} \right]$	$\left[\begin{array}{ll} \text{PHON} & \text{warmiskuna} \\ \text{SYN} & \textit{N.pl} \\ \text{SEM} & \textit{women} \end{array} \right]$

...

The morphological paradigm is a paradigm of categorized forms with a lexeme identifier.

[PHON	warmikuna]	[PHON	warmikunas]
	MCAT	<i>kuna</i>			MCAT	<i>kunas</i>	
	LEX	WARMI			LEX	WARMI	

[PHON	warmis]	[PHON	warmiskuna]
	MCAT	<i>s</i>			MCAT	<i>skuna</i>	
	LEX	WARMI			LEX	WARMI	

...

Morphological patterns are mapped to lexical entries with mapping rules.

Mapping rules “generate” the lexicon from general patterns.

$$\frac{mcat(x) \leq \text{M-Plural}}{\left[\begin{array}{ll} \text{PHON} & \mathit{phon}(x) \\ \text{SYN} & \mathit{NP.pl} \\ \text{SEM} & \mathit{meaningOf}(\mathit{lex}(x)) \end{array} \right]}$$

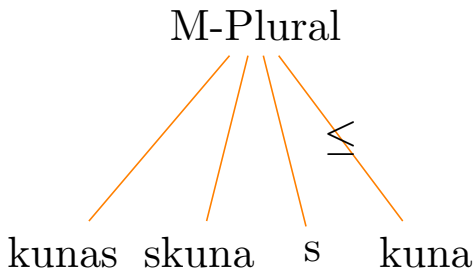
Think of this as a declarative natural deduction rule.

The variable x in $mcat(x)$ is for a morphological paradigm member.

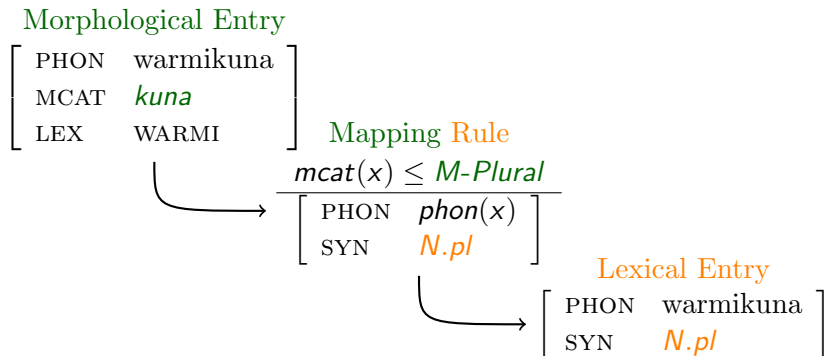
$$mcat \left(\begin{bmatrix} \text{PHON} & \text{warmikuna} \\ \text{MCAT} & \textit{kuna} \\ \text{LEX} & \text{WARMI} \end{bmatrix} \right) = \textit{kuna}$$

The $mcat(x) \leq M\text{-Plural}$ is making reference to the hierarchy of morphological categories.

$kuna \leq M\text{-Plural} = True$ given:



The result of rule application is demonstrated below.



Wao Tededo (Wao Terero) is spoken the Ecuadorian Amazon.



Zoom Group Chat

From **Abraham** to **Everyone**: 10:25 AM
Orellana tike mamente gotapa petróleo
contamina kepa

From **Abraham** to **Everyone**: 10:40 AM
Epene nagi atime impa
en el rio hay mucho lodo
Epe kidame kidame impa

The analytic periphrastic future tense is in free variation with a synthetic² form.

Be-ke-bopa. Be-ke ke-bopa.

drink-FUTURE-1

drink-FUTURE do-1

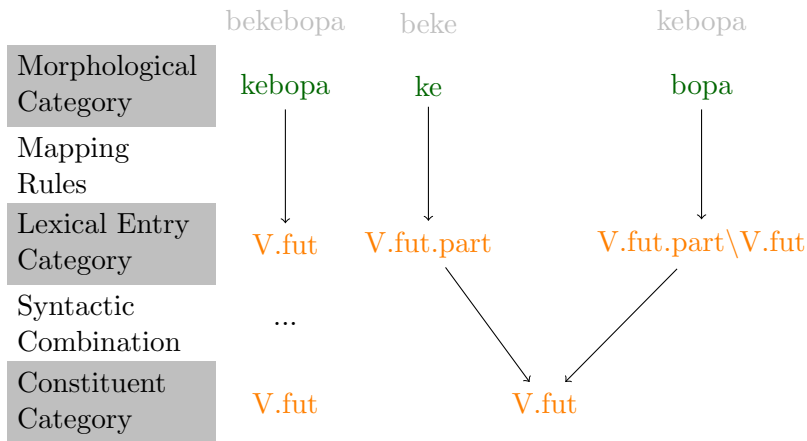
‘I will drink.’

²Peeke (1968) may indicate that the ‘merged’ form is the older form.

Both of these need to have the same category as syntactic constituents allowing them to share a distribution.

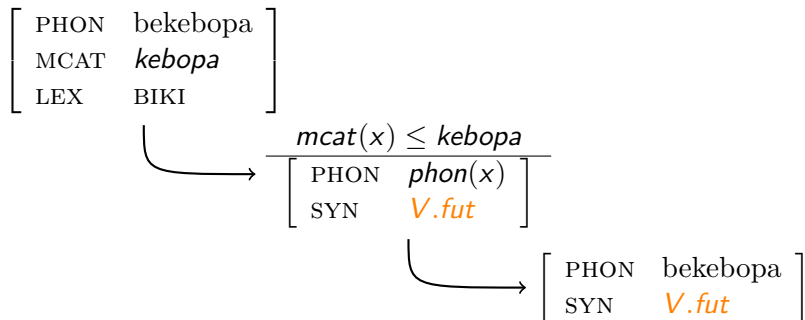
Be**ke**bopa. Be**ke** kebopa.
V.FUT V.FUT

The interface allows for a clear distinction between the analytic and synthetic construction.

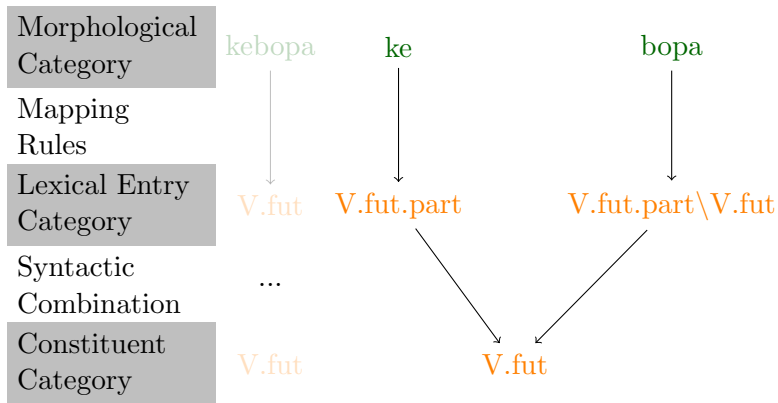


In categorial grammars *V.fut.part \ V.fut* means that if a future participle is supplied, the resulting category is *V.fut*.

The synthetic case is routine.

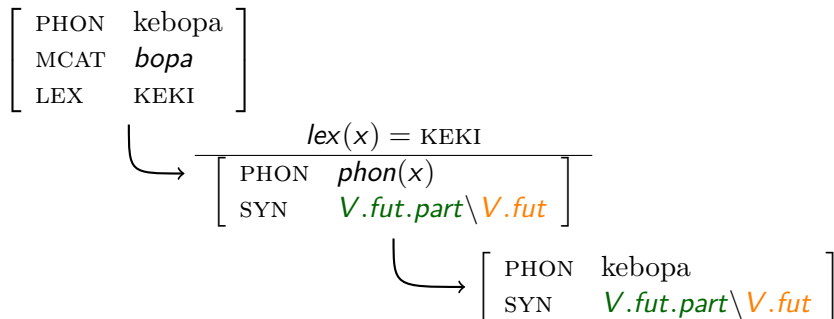


Now only the more complicated pattern remains.

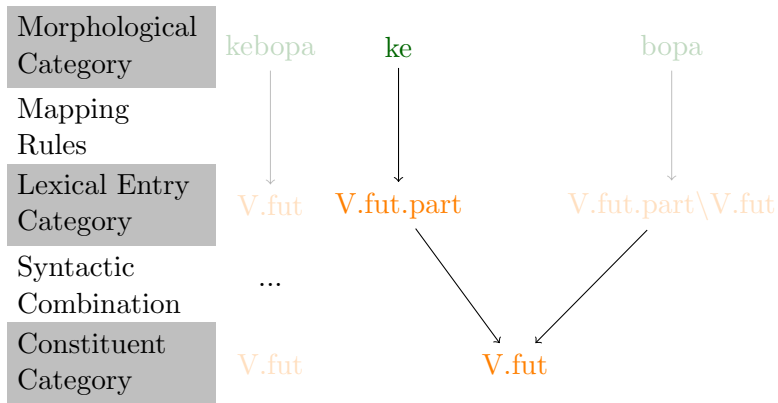


The first step is providing a rule specific to the auxiliary verb.

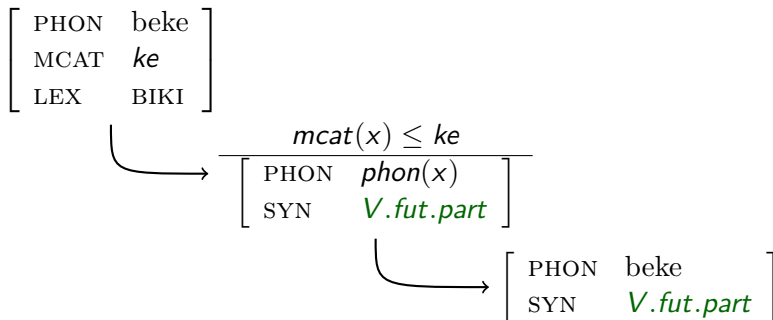
What is notable about the mapping rule is that it is specific to a lexeme identifier, rather than a morphological category.



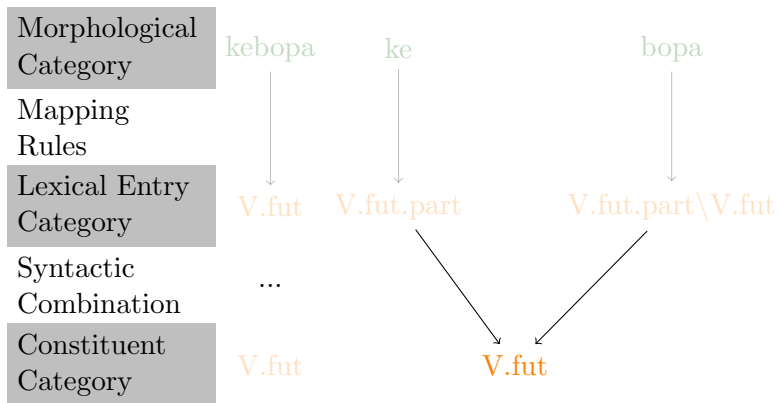
Now only a rule for future participles is needed.



This future participle rule is standard.



The morphological interface has completed its contribution.



Syntactic combination derives the correct category.

In categorial grammar, an elimination rule allows the two lexical entries to combine.

$$\left[\begin{array}{ll} \text{PHON} & \text{beke} \\ \text{SYN} & V.fut.part \end{array} \right] \left[\begin{array}{ll} \text{PHON} & \text{kebopa} \\ \text{SYN} & V.fut.part \setminus V.fut \end{array} \right]$$

$$\left[\begin{array}{ll} \text{PHON} & \text{beke kebopa} \\ \text{SYN} & V.fut \end{array} \right]$$

The takeaway is the following:

1. Morphological forms overlap in distribution.
2. Standard, popular models cannot accommodate it due to Paninian rules.
3. Tiered, multi-paradigm approaches, provide the necessary interface for overabundance.

This talk focused on high level concepts but the theory is completely formalized.

Please visit my GIT repositories if you'd like to see fragments implemented in a theorem prover:

<https://git.diewald.me/morphexamples>