Towards an axiomatic approach in Linguistics

with a special focus on segmental grammar

Artemij Keidan (Sapienza University of Rome) – September 21, 2023

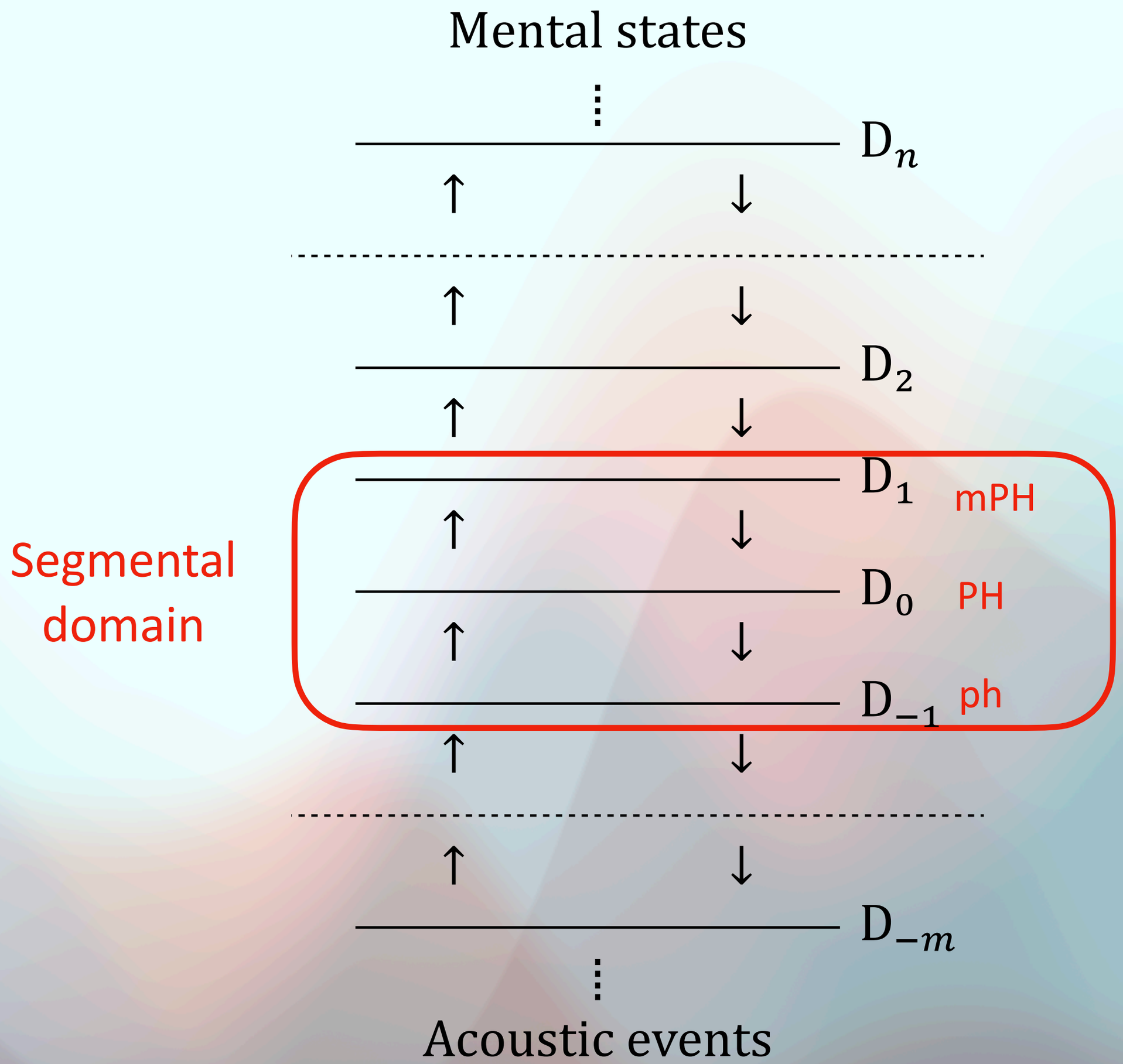
Main goals

- How do we know whether a statement is true or false, in linguistics?
- How can we compare two grammars?
- How do we uncover the logic on which a grammar is based?
  - with a special focus on the segmental domain

Some premises

- Linguistics = grammar writing
- Language = finite corpus of sentences
- Grammar = a cascade of representations/levels + mapping rules between them
  - grammar is constructed rather than “extracted”
- Representations (levels) form a hierarchy, from phonetics to semantics
- Examples of rewriting rules (in segmental gr.):

  ph→PH, PH→ph, mPH→PH


Mental states
D_n
D_2
D_1
mPH
Segmental domain
D_0
PH
D_{-1}
ph
D_{-m}
Acoustic events

Intuitive theories

- Traditional/intuitive theory (roughly) = a set of propositions about reality
  - verification is obtained from empirical data
- Intuitive theory in linguistics = traditional descriptive grammar
- data based verification has two weaknesses:
  - the data are vague
  - the data address the easiest questions only

Intuitive theories in linguistics

- E.g.: given a sound sample, which transcription is the correct one?

  [hju] ~ [hɪu] ~ [h^{j}u] ~ [çju] ~ [hy] ecc.

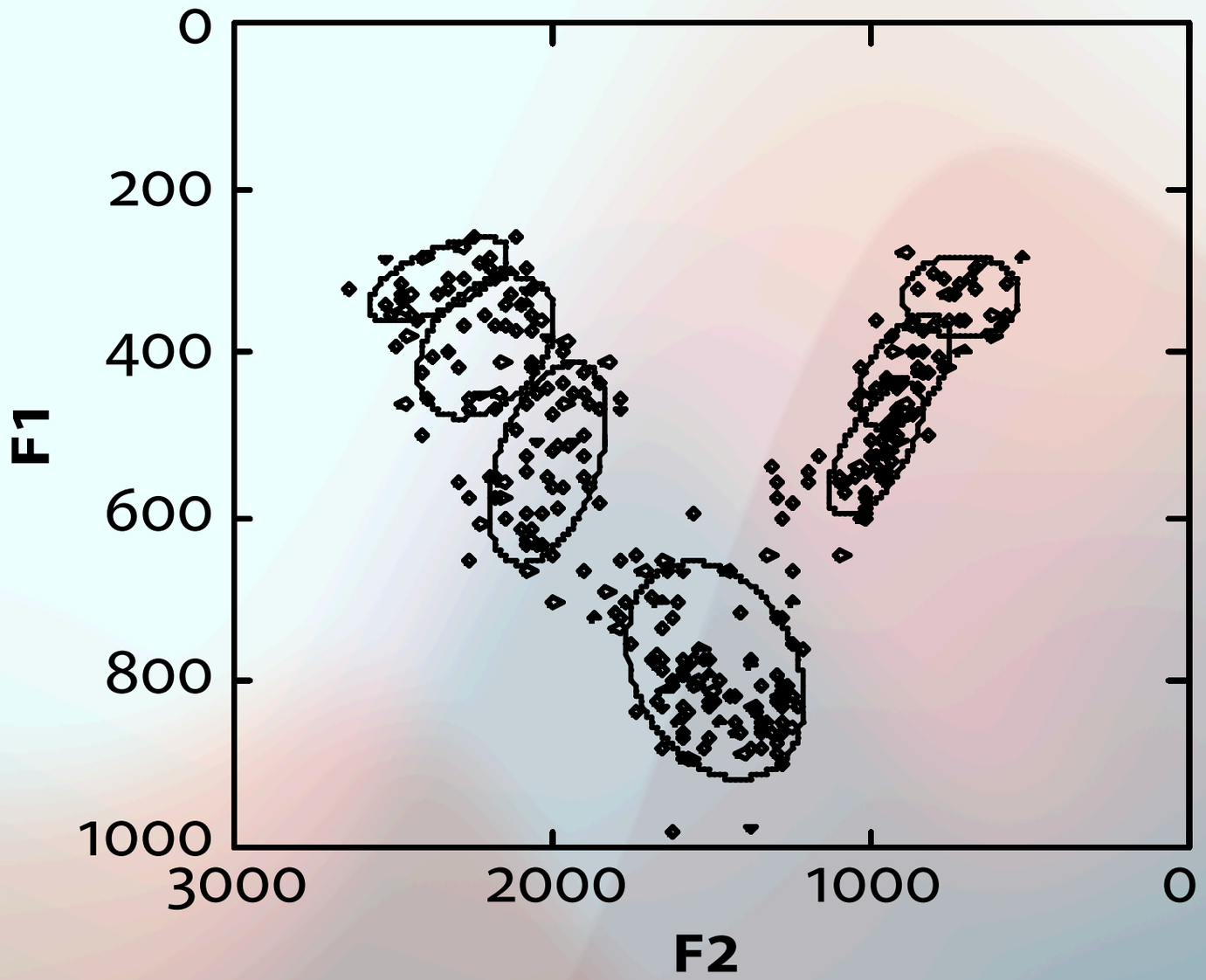
- E.g.: given a list of word-forms, how do we classify them?

  Eng. *love*: N, V or “N-V” ?

  Rus. *druga*: AccSg or GenSg of *drug* ‘friend’?

Intuitive theories in linguistics

- E.g.: formant dispersion (Italian vowels):



Intuitive theories in linguistics

- Intuitive/traditional grammars are, therefore:
  - falsely empirical: data are constructed anyway
    - empirical verification is a myth
  - too depending on grammarian's arbitrary decisions
  - impossible to evaluate or to compare
    - which grammar of Old Slavic is better?
    - how do two grammars of Old Slavic even differ?

Axiomatization

- Intuitive theories may be redefined in axiomatic form
- Axiomatic theory is detached from verification
  - logic and proof from data and truth
- Axiomatic theory = set of provable formulas
  - axioms = formulas assumed as proven
  - theorems = formulas inferred from axioms
  - rules of inference are purely formal

Axiomatization

- In order to obtain semantic values, a theory must be applied to a model
- Model = set of propositions + interpretation
  - variables of the theory take values from the objects in the model
  - formulas of the theory translate into propositions in the model
  - axioms of the theory correspond to true propositions in the model

Example: Groups

- Signature:
  - non-empty set G
  - binary operation "•"
- Axioms
  - associativity: for all a , b and c in G, $(a \bullet b) \bullet c = a \bullet (b \bullet c)$
  - identity: $\boldsymbol{e}$ exists in G such that, for all a in G, $a \bullet \boldsymbol{e} = \boldsymbol{e} \bullet a = a$
  - inverse: for each a in G, b exists, such that $a \bullet b = b \bullet a = \boldsymbol{e}$

Example: Groups

- Is $(\mathbb{Z}, +)$ a group?
  - for all a , b and c in $\mathbb{Z}$, $(a+b)+c = a+(b+c)$
  - 0 is in $\mathbb{Z}$, and, for all a in $\mathbb{Z}$, $a+0 = 0+a = a$

    e.g.: $3 + 0 = 0 + 3 = 3$
  - for each a in $\mathbb{Z}$, exists b, such that $a+b = b+a = 0$

    e.g.: $5 + (-5) = (-5) + 5 = 0$

Example: Groups

- Also groups:

  $(\mathbb{R}, +)$

  $(\mathbb{Q}, \cdot)$

- Not groups:

  $(\mathbb{N}, +)$

  $(\mathbb{Z}, \cdot)$

Axiomatization

- Models in Mathematics are invented
- Models in empirical sciences are retrieved from empirical data
- Advantages of axiomatization:
  - logical structure is detached from the truth
  - verification is formal, rather than empirical
  - theories become comparable
    - they may differ in signature, axioms or interpretation rules

Axiomatization in linguistics

- Axiomatic theory = **abstract grammar**, roughly:
  - variables = “empty tables”, “empty inventories”
  - axioms = correctness criteria limiting how levels are mapped, tables are filled
- Model = **specific grammar**
  - take a finite corpus of textual data
  - populate tables and inventories
  - …respecting the axioms

Axiomatization in linguistics

- E.g.: fragment of an abstract grammar of Latin (morphology)

	Singular	Plural
Nom		
Gen		
Dat		
Acc		
Abl		

Axiomatization in linguistics

- Consider a list of latin word-forms (in alphabetical order):

  …, lup=a, lup=ae, lup=am, lup=arum, lup=as, lup=i, lup=is, lup=o, lup=orum, lup=os, lup=us, …

- Let's use these forms to populate the empty tables of the abstract grammar
  - respecting the correctness criteria of morphology

Axiomatization in linguistics

- Consider the axioms of morphology (approx.):

  1. no word-form of a lexeme can be left outside
  2. no cell can remain empty in all lexemes
  3. any pair of cells must contain different word-forms in at least one lexeme
  4. no cell can contain multiple word-forms
  5. ecc.

Axiomatization in linguistics

- E.g.: fragment of a specific grammar of Latin

	Singular	Plural
Nom	lup=us	lup=i
Gen	lup=i	lup=orum
Dat	lup=o	lup=is
Acc	lup=um	lup=os
Abl	lup=o	lup=is

Axiomatization in linguistics

- The preceding table is correct because it respects the axioms
  - not because it is a “faithful picture of the Latin language”
    - (in fact, alternative grammars are possible)
- The following table is not correct because it does not respect the axioms
  - not because it contradicts “Latin language”
    - (corpus comes unlabelled)

Axiomatization in linguistics

- E.g.: fragment of specific grammar of Latin, different solution

	Singular	Plural
Nom	lup=us, lup=a	lup=i, lup=ae
Gen	lup=i, lup=ae	lup=orum, lup=arum
Dat	lup=o, lup=ae	lup=is
Acc	lup=um, lup=am	lup=os, lup=as
Abl	lup=o, lup=a	lup=is

Axiomatization in linguistics

- E.g.: fragment of an abstract grammar of Italian (phonological inventories)

	Front		Central		Back	
	+lab	-lab	+lab	-lab	+lab	-lab
High						
Mid						
Low						

Axiomatization in linguistics

- Consider the axioms of phonological inventories:

  1. no row or column must be empty
  2. no cell can contain more than one symbol
  3. no symbol can appear in more than one cell

- Therefore, the following table is not correct

Axiomatization in linguistics

- E.g.: fragment of a specific grammar of Italian (phonological inventories)

	Front		Central		Back	
	+lab	-lab	+lab	-lab	+lab	-lab
High		i			u	
Mid		e, ɛ			o, ɔ	
Low				a		

Axiomatization in linguistics

- In order to obtain a correct table, I can
  - either change the interpretation of data = arrange the vowels differently
    - no such arrangement is known to me
  - dismiss the abstract grammar as incorrect
    - then create a new abstract grammar
    - populate the table with the same date (see the following tables)

Axiomatization in linguistics

- E.g.: fragment of a specific grammar of Italian (phonological inventories)

	Front	**Central**	**Back**
High	i		u
Mid-high	e		o
Mid-low	ɛ		ɔ
Low		a	

Axiomatization in linguistics

- E.g.: fragment of a specific grammar of Italian (phonological inventories)

	Front	**Central**	**Back**
High	i		u
Mid	e		o
Low	ɛ	a	ɔ

Segmental levels

- mPH = morphophonological level
  - formatives maintain a constant segmental form
- PH = phonological level
  - minimal inventory of segments
- ph = phonetic level
  - non minimal inventory, more articulatory details

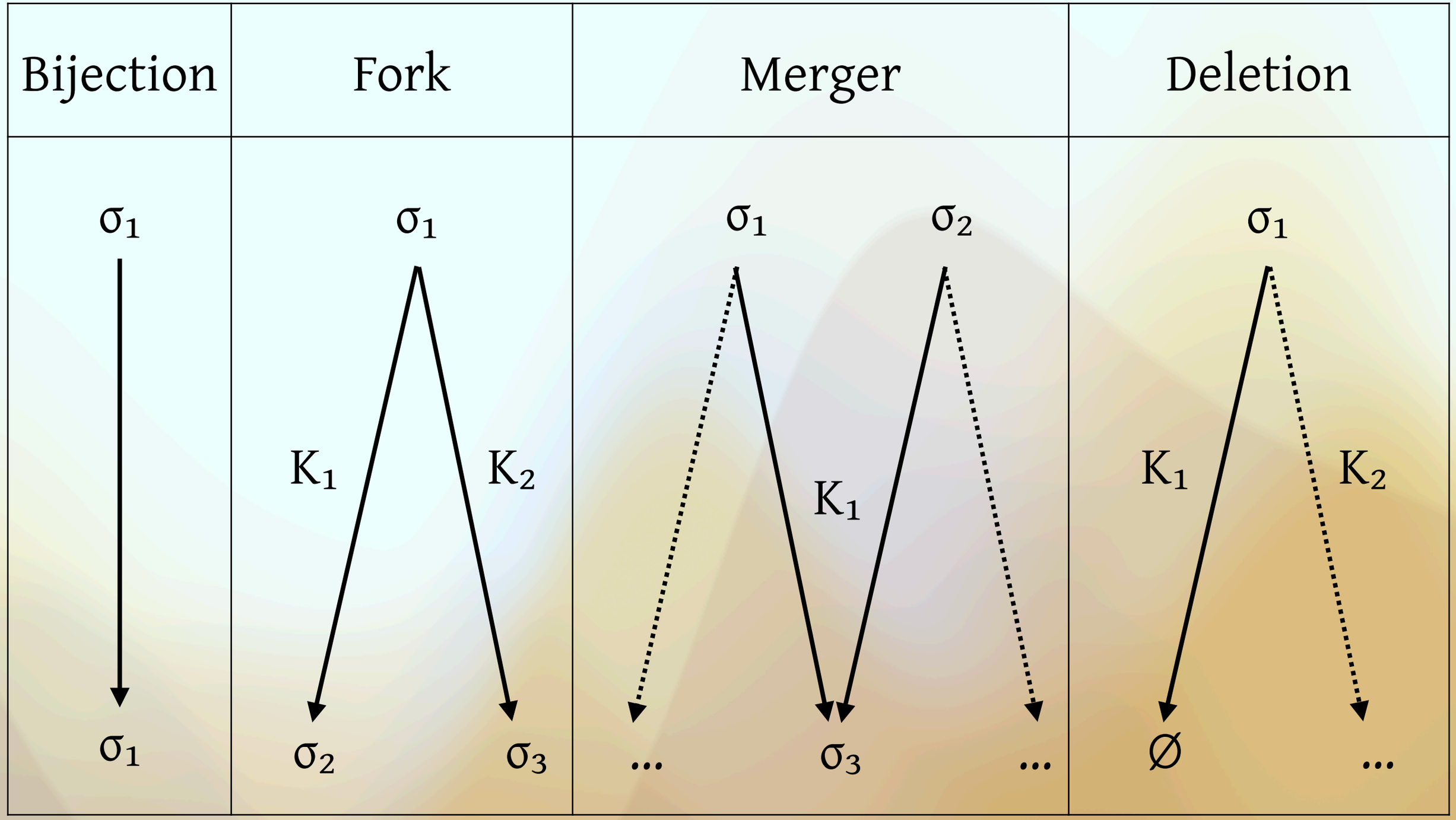
Segmental levels

- The input to mPH comes from the “inflectional request” ≈ interface with morphology
- The mapping rules within the segmental domain must be contextual
- The mapping rules in the preceding levels are non-contextual
  - rules InflR→mPH are non contextual
  - e.g.: InflR Pl(*cat*) → mPH cat.s

Segmental levels

- Possible types of rules:
  - Bijection: $\sigma_1 \rightarrow \sigma_1$ in all contexts
  - Fork: $\sigma_1 \rightarrow \sigma_2$ in context K_1 and $\sigma_1 \rightarrow \sigma_3$ in context K_2
  - Merger: in some context $\sigma_1 \rightarrow \sigma_2$ and $\sigma_3 \rightarrow \sigma_2$
  - Deletion: in some context $\sigma_1 \rightarrow \varnothing$

Types of rules



Axioms

- A_1 — Every utterance can be represented as a linear sequence of segments
  - Segmental representation must be possible, but A_1 says nothing on how do we get it

Axioms

A_2 — In segmental grammar we consider monolateral strings of segments.

- there is no meaning in segmental grammar
- if two forms correspond to the same string, it means they are the same form
- homonymy is not possible

Axioms

A_3 — Every word-form is given in three representations: mPH, PH and ph.

- Morphophonological (mPH) representations are stored in the dictionary
- the other two can be constructed by means of mapping rules
  - e.g. Polivanova's Slavic grammar

Axioms

A_4 — PH has the minimal distinctive inventory

- Minimal pairs must be in PH form
- It is always true that distinctivity ⇒ phonemes
- The opposite implication may be not true

Axioms

A_5 — Inventories of mPH and PH coincide.

- unlike in “traditional morphophonology”
- There is no limiting criterion for “morphophonemes”
- Mapping rules mPH→PH are a manipulation of phonemes

Excursus on “morphophonemes”

InflR	Sg(*parco*) ‘park’	Pl(*parco*) ‘parks’	Sg(*medico*) ‘doctor’	Pl(*medico*) ‘doctors’	Sg(*carne*) ‘meat’
mPH	park$_1$+o	park$_1$+i	mɛdik$_2$+o	mɛdik$_2$+i	kʔarn+e
PH	parko	parki	mɛdiko	mɛdiʧi	karne
ph	parko	parki	mɛːdiko	mɛːdiʧi	karne

Our morphophonemics

InflR	Sg(*parco*) ‘park’	Pl(*parco*) ‘parks’	Sg(*medico*) ‘doctor’	Pl(*medico*) ‘doctors’	Sg(*carne*) ‘meat’
mPH	park+o	park+i	mɛdik+o	mɛdiʧ+i	karn+e
PH	parko	parki	mɛdiko	mɛdiʧi	karne
ph	parko	parki	mɛːdiko	mɛːdiʧi	karne

Axioms

A_6 — All the mapping rules within the domain of segmental grammar must be contextual.

- These rules are: mPH→PH, PH→ph, ph→PH.
- All such rules must work in every occurrence of the relevant segment
- If a rule cannot be defined contextually, this rule does not exist
  - thus, there are no contextual rules from PH to mPH

Axioms

A_7 — Mergers are only allowed in the mapping rules mPH→PH

- mPH→PH rules are destructive: they destroy segmental information
  - consequence: there are no rules PH→mPH
- The other rules cannot be destructive

Example of a destructive rule

InflR	PartSgm(*piangere*) 'to cry'	Pres1Sg(*piantare*) 'to plant'
mPH	pjandʒ.t=o	pjant=o
	[deletion occurs]	
PH	pjanto	pjanto
ph	pjanto	pjanto

Example of a destructive rule

InflR	PartSgm(*piangere*) ‘to cry’	Pres1Sg(*piantare*) ‘to plant’
mPH	pjandʒ.t=o	pjant=o
	[merger of two forms]	
PH	pjanto	
ph	pjanto	

Some theorems

- Theorem 1. There are no PH→mPH rules.
  - A_6: all segmental rules must be contextual
  - A_7: only mPH→PH rules admit mergers and deletions
  - Once a segment has been deleted, you cannot restore it from the context
  - PH→mPH rules would require non-contextual information
  - Therefore, there PH→mPH rules do not exist

Some theorems

- Theorem 2. Reversibility of PH and ph.
  - A_7: only mPH→PH rules admit mergers
  - Rules PH→ph do not admit mergers
    - but admits forks!
  - Information is not destroyed
  - Therefore we can always restore PH from ph
    - there is no such thing as phonemes' neutralization

On shared allophones

- Our analysis of the flapping rule

	writer	*rider*
mPH	raɪtər	raɪdər
	[substitute /t/ → /d/ between vowels, after stress]	
PH	raɪdər	raɪdər
	[flap /d/ → [ɾ] between vowels]	
ph	raɪɾər	raɪɾər

Some theorems

- Theorem 3. Boundaries between formatives are only visible in mPH.
  - A_7: only mPH→PH rules admit mergers
  - Phonemes surrounding a boundary may be subject to a merger:

    mPH $\alpha.\beta \rightarrow$ PH μ
  - Therefore, formative boundaries are not warranted in PH and ph

Some Italian facts in our analysis

	infatti ‘in fact’	*impari* ‘uneven’	*tramvia* ‘tramway’
mPH	in.fatti	in.pari	tram.via
		n→m/_p	m→n/_v
PH	infatti	impari	tranvia
	n→ɱ/_f		n→ɱ/_f
ph	iɱfatti	impari	traɱviːa

Some Italian facts in our analysis

	tempo ‘time’	*temporale* ‘temporal’	*tono* ‘tone’	*tonale* ‘tonal’
mPH	tɛ́mp=o	tɛmp.or.ál=e	tɔ́n=o	tɔn.ál=e
		ɛ→e/unstressed		ɔ→o/unstressed
PH	tɛ́mpo	temporále	tɔ́no	tonále
		a→aː/stressed	ɔ→ɔː/stressed	a→aː/stressed
ph	tɛ́mpo	temporáːle	tɔ́ːno	tonáːle

Some Russian facts in our analysis

	NomSg(*kod*) 'the code'	GenSg(*kod*) 'of the code'	NomSg(*kot*) 'the cat'	GenSg(*kot*) 'of the cat'
mPH	kod	kod=a	kot	kot=a
	devoicing/__#			o→a/unstressed
PH	kot	koda	kot	kata
ph	kot	koda	kot	kata

Some Russian facts in our analysis

	LocSg(*los^{j}*) 'in the moose'	LocSg(*lisa*) 'in the fox'	NomSg(*urok*) 'in the lesson'
mPH	los^{j}=e	lis=e	urok=e
		palatalization/__e	
PH	las^{j}e	lis^{j}e	uroke
			palatalization/__e
ph	las^{j}e	lis^{j}e	urok^{j}e

In conclusion

- Let's answer some of the opening questions
- A grammar that respect the axioms is a good grammar
- There may be multiple grammars of the same language, all equally good
- No grammar can be considered a "better picture" of the language data
- The differences among grammars may be located in the axioms, signature or interpretation