Introducing Axiomatic Linguistics, with a special focus on segmental grammar

Artemij Keidan
artemij.keidan@uniroma1.it

Centro de Linguística da Universidade de Lisboa

Preliminary remarks

- Why yet another new framework?
  - we need verifiability
  - we need comparability
- Why a statement is true or false?
  - traditional answer: depends on its empirical “fidelity”
  - this is a mirage

Preliminary remarks

- Empirical data is never really decisive
  - it may suggest something very basic and simple
  - it is useless when abstract statements must be verified
    - NP vs. DP debate
- Important questions get no answer from the data
  - In practice, modern linguistics is often arbitrary

Preliminary remarks

- We suggest the following moves on such respect:
  1. From the infinite faculty of mind to a finite set of texts
  2. From “extraction” to “construction” of the theory
  3. From empiricism to axiomatization

Preliminary remarks

- From mind to texts
  - We *do not* investigate the “faculty of language”
  - We *cannot* investigate the “faculty of language”
    - Chomsky fooled us all with this seductive mirage
    - unsolvable problem
- We investigate one finite corpus of texts at a time

Preliminary remarks

- From “extraction” to “construction” of the theory
  - Traditionally the grammar is a discovery procedure
    - Trubeckoj’s rules are for discovering the phonological inventory of a language
- For us, the grammar is entirely “constructed”, from phonology to syntax

Preliminary remarks

- From empiricism to axiomatization
  - Traditionally, grammars describe “real languages”
  - But real linguistic data are vague, continuous, and highly interpretable
- For us, both the grammar and the data are constructed, then tested for consistency

Axiomatic theory

- Axiomatic approach comes from logic and algebra
  - see A. Tarsky's research from 40s and 50s
- A theory is a set of true statements
  - axioms are assumed as true
  - theorems are deduced from the axioms
- How do we know that a theory is consistent?

Theory and model

- A theory includes:
  - terms (variables), predicates, connectives
  - syntax (rules for constructing formulas)
  - axioms (formulas assumed as true)
  - rules of inference (rules for proving theorems)
- Theories have no semantics
  - they need a model

Theory and model

- A model is a structure such that
  - terms, predicates and formulas of the theory can be interpreted as terms, predicates and formulas of the model
  - true statements of the theory become true statements about the model
- The existence of a model proves the consistency of the theory

Theory and model

- Examples
  - given $a<b$, there exists c such that $a<c<b$
    - this is false in $\mathbb{N}$ (natural numbers)
    - this is true in $\mathbb{R}$ (real numbers)
    - therefore, $\mathbb{R}$ is a model for this theory

Theory and model

- Examples
  - every specific group is a model of the group theory
  - a plane is a model for the Euclidean geometry (notably, 1 parallel exists)
  - the surface of a sphere is a model for some non-Euclidean geometries (where no parallels exist)

Axiomatic theory in linguistics

- Theory is called abstract grammar
- Abstract grammar is a set of axioms and variables

  variables = empty tables
- The model is a corpus of linguistic data
  - sentences, lexemes, wordforms, morphemes, phonemes
  - the corpus is “constructed” rather than “extracted”

Axiomatic theory in linguistics

- The corpus is “constructed” rather than “extracted”
  - the theory cannot make appeal to intuition or analogy
  - the structures are not extracted from the data *by* the theory
  - every sentence comes in three already constructed forms: phonetic, phonological, morphophonological

Axiomatic theory in linguistics

- Abstract grammar is interpreted on the data
  - empty containers are filled with data from the corpus
- The result is a particular grammar
  - filled paradigms, tables, inventories
  - particular grammar must respect the axioms

Axiomatic theory in linguistics

- Example: a fragment of the abstract grammar of Latin

		Number	
		Singular	Plural
Case	Nom		
	Acc		
	Gen		
	Dat		
	Abl		

Axiomatic theory in linguistics

- Example: a fragment of the Latin corpus

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

* The wordforms are in the alphabetic order

** The symbol “=” separates the endings form the stems

Axiomatic theory in linguistics

- Example: a fragment of the particular grammar of Latin

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

Axiomatic theory in linguistics

- This table is correct because it respects the axioms
  - not because it is a “correct picture of the Latin language”
    - in fact, alternative pictures are also possible
- Not that the axioms are proven as consistent because a model has been found

Axiomatic theory in linguistics

- Axioms of morphology

1. lexemes must have no forms beyond the table
2. no cell can remain empty in all lexemes
3. each two cells must contain different forms in at least one lexeme
4. no cell can contain more than one form
5. ...

Axiomatic theory in linguistics

- The following is against axiom 4:

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

Abstract and particular grammars

- Traditional approach does not distinguish abstract and particular grammars
  - usually, the tables are considered self-evident
- The distinction is useful for comparing grammars
  - before deciding which one is “better”, let’s see where they differ

Abstract and particular grammars

- Instead of claiming intuitively that Grammar_1 is "better" than Grammar_2
- Two grammars may differ:
  - on the abstract side
  - by the model
  - by the interpretation of the theory in the model

Abstract and particular grammars

- Two grammars of English differing at the abstract level:
  - grammar with N ≠ V
    - each lexeme is either N, or V, with many homonyms
  - grammar with N = V
    - each word may occur as predicate or argument

Abstract and particular grammars

- Two grammars of English differing in the model:
  - grammars covering only modern British English
  - grammars covering all the main dialects of English
  - grammars covering English from Shakespeare to the present days
  - (the abstract level may be the same)

Abstract and particular grammars

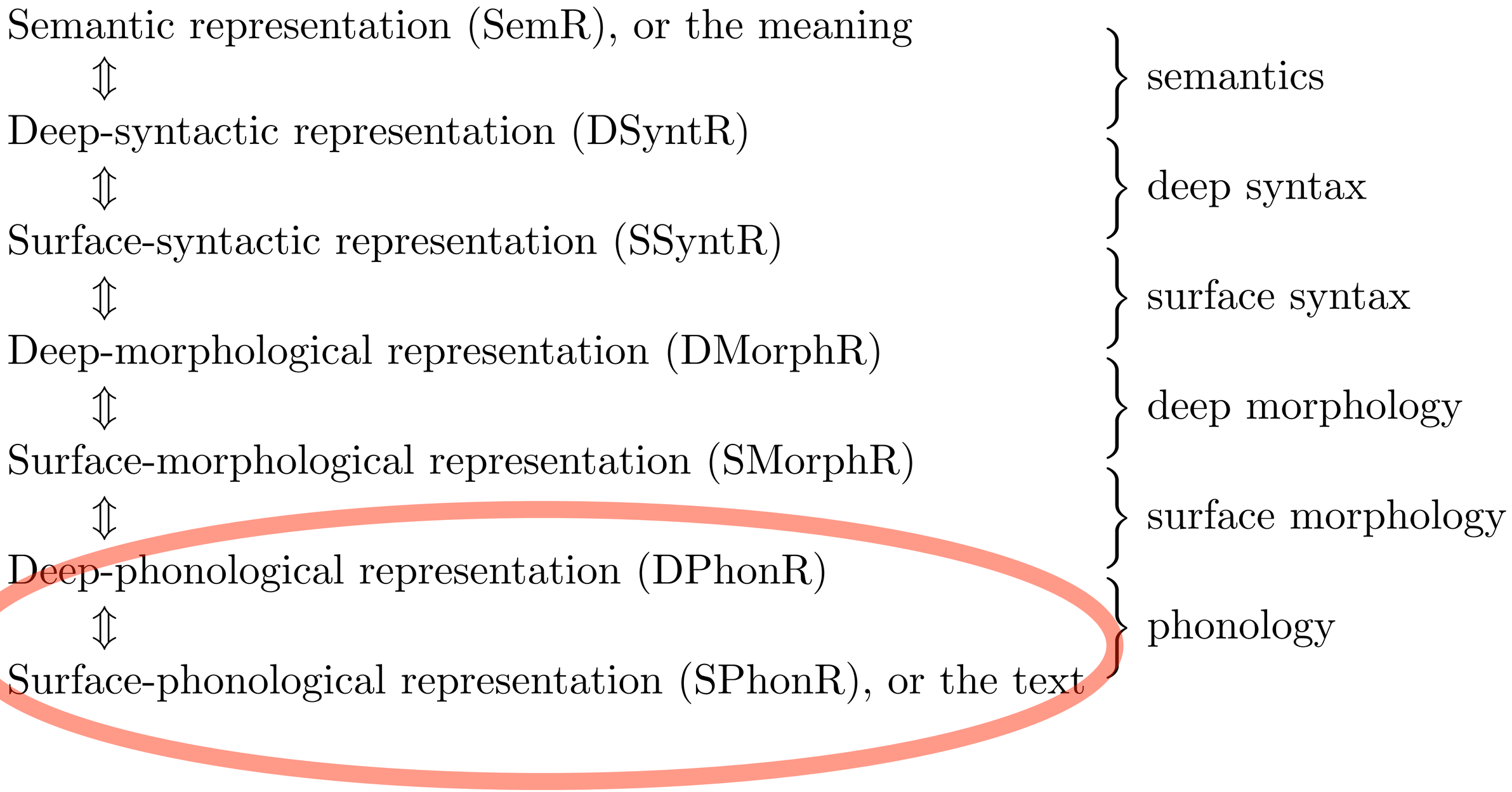
- Two grammars of English differing in the interpretation
  - (abstract grammar and model remain unchanged)
    - grammars with diphthongs
    - grammars with bisegmental sequences in place of diphthongs

Axioms of segmental grammar

- Let’s define the aim of segmental grammar
- Let’s expose the axioms of segmental grammar
- Let’s prove some theorems
- Let’s translate some statements about Italian into axiomatic form

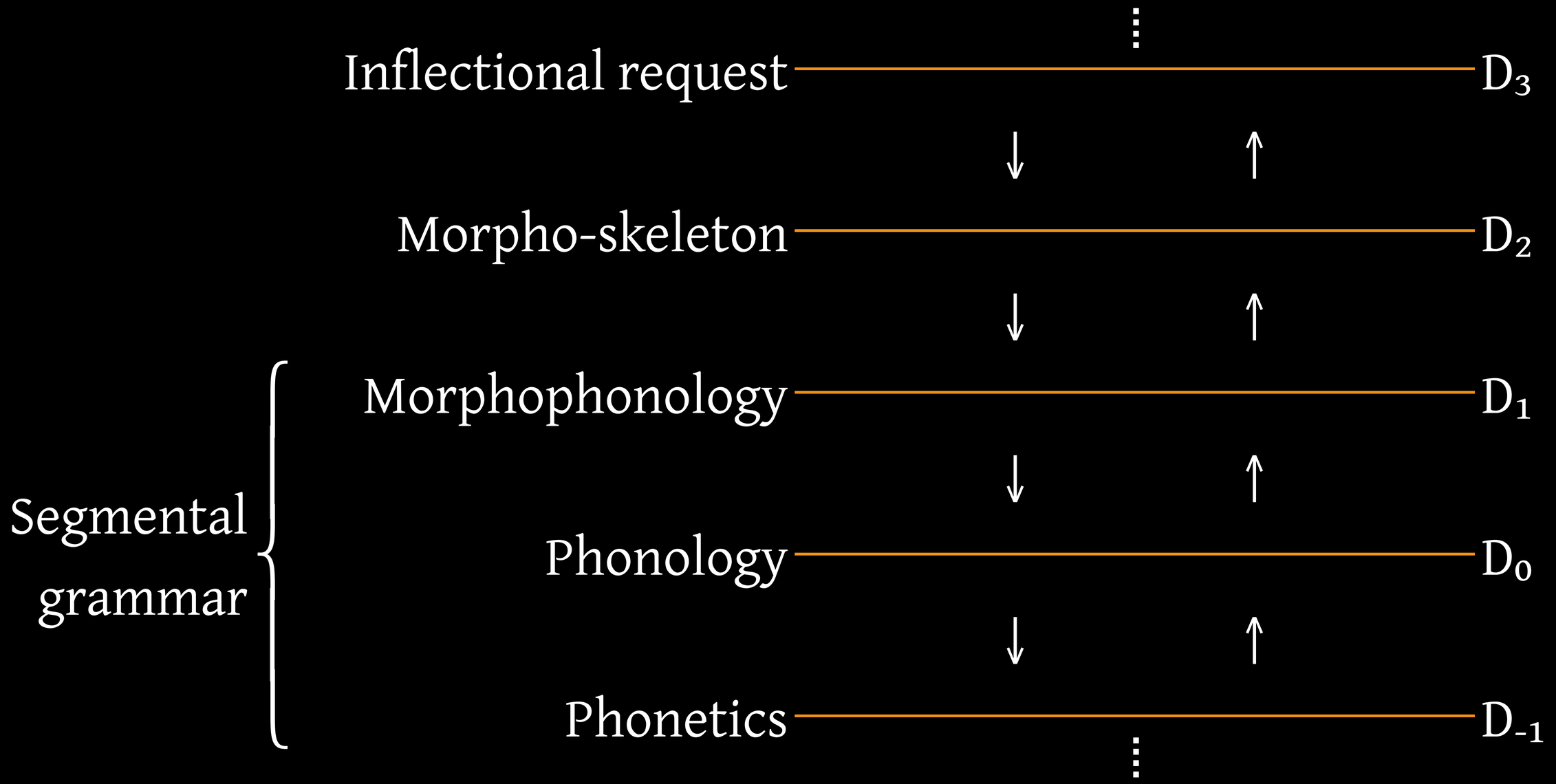
Origin of our approach

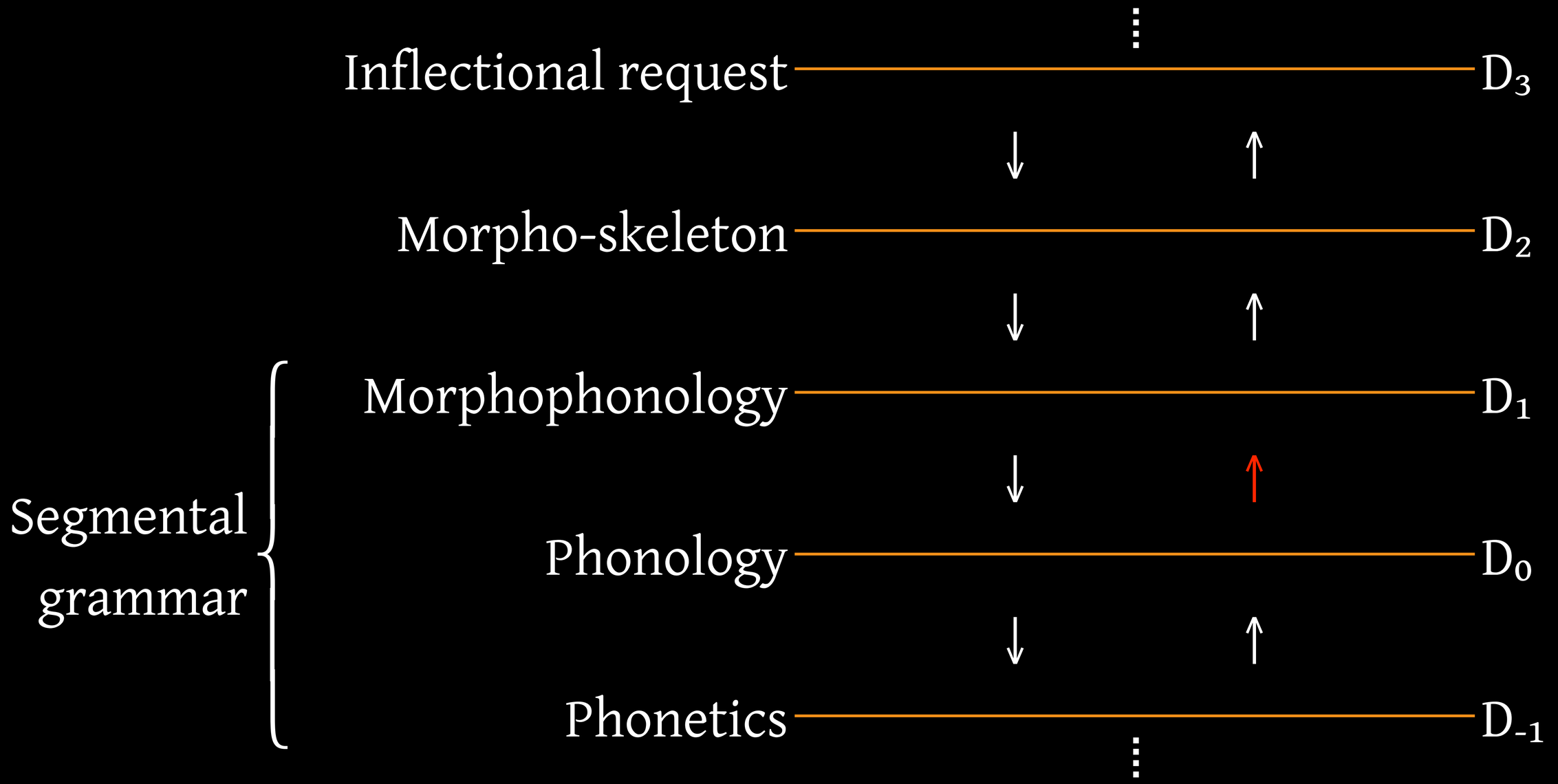
- Igor Melchuk's "Meaning ⇔ Text theory" (1960s):
  - The grammar describe how to get from a set of meanings to a text, and from a text to the meanings
  - The grammar is a stack of representations
    - from the deepest (SemR) to the most surface (PhonR)
    - mapping rules between each pair of adjacent representations


Semantic representation (SemR), or the meaning
⇕
Deep-syntactic representation (DSyntR)
⇕
Surface-syntactic representation (SSyntR)
⇕
Deep-morphological representation (DMorphR)
⇕
Surface-morphological representation (SMorphR)
⇕
Deep-phonological representation (DPhonR)
⇕
Surface-phonological representation (SPhonR), or the text
semantics
deep syntax
surface syntax
deep morphology
surface morphology
phonology

Definition of segmental grammar

- Main proponent of our theory: Anna Polivanova
  - different terminology
  - more levels of representation
    - especially in the segmental domain
  - plus, the axiomatization
- Let’s see our picture


Inflectional request
D3
Morpho-skeleton
D2
Morphophonology
D1
Phonology
D0
Phonetics
D-1
Segmental grammar


Inflectional request
D3
Morpho-skeleton
D2
Morphophonology
D1
Phonology
D0
Phonetics
D-1
Segmental grammar

Definition of segmental grammar

- Segmental domain includes those levels where:
  - the representation is made only of segments
    - thus, a syntactic tree is not segmental
  - the mapping rules use only segmental information
  - this corresponds to $\{D_{-1}, D_0, D_1\}$ = {ph, PH, mph}
    - observe that PH (=phonology, D_0) is central

Inflectional request (D_3)

- Inflectional request (D_3): the interface with syntax
  - select the wordform ω from the lexeme λ, in symbols: ω(λ)
  - example: select Present Indicative 3rd plural of PARLARE, in symbols: PresInd3Pl(PARLARE)
  - another example: Pl(MEDICO)
- This level is not segmental!

Morphological skeleton (D_2)

- Receives the request from D_3
  - Selects the default stem and default ending
  - Ex.: PresInd3Pl(PARLARE) → parl+ano
  - Ex.: Pl(MEDICO) → mɛdik+i
- D_2 is made of segments, but $D_2 \rightarrow D_1$ is not contextual
  - therefore, D_2 does not belong to the Segmental grammar

Morphophonology (mph, D_1)

- Substitutes the default formatives with the correct ones
- The rules $D_2 \rightarrow D_1$ require lexical information
  - it's where non-contextual allomorphs are selected
- From D_1 downwards the rules are only contextual
- Ex.: skeleton mɛdik+i → mph mɛdiʧ=i

Phonology (PH, D_0)

- Morphophonology shows the boundaries between formatives
- Mapping rules mph→PH delete the boundaries
  - in phonology, there are no boundaries!
- Sandhi are processed
- Ex.: mph mɛdiʧ=i → PH mɛdiʧi

Phonetics (ph, D_{-1})

- Phonetics adds articulatory details to the phonological representation
- Phonetics is not compulsory
  - if we stop at PH level, our speech remains comprehensible
- Ex.: PH mɛdiʧi → ph mɛːdiʧi
  - (accented vowel in open syllable is lengthened)

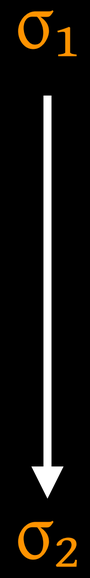
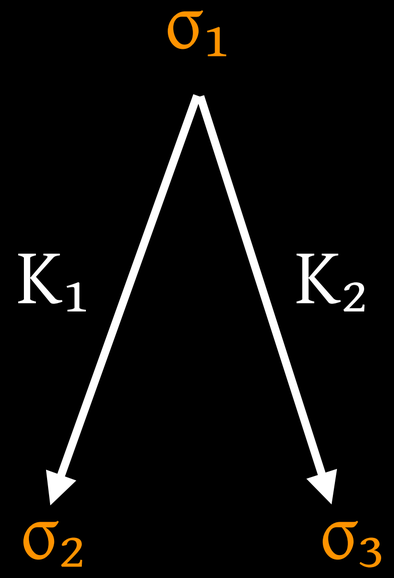
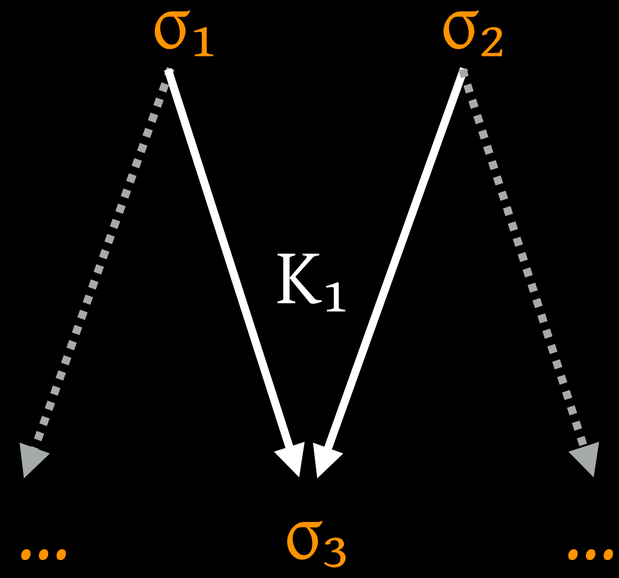
The whole picture

Infl. request	Pl(MEDICO)	PartSgm(PIANGERE)	Pres1Sg(PIANGERE)
Skeleton	mɛdik+i	pjandʒ+t+o	pjandʒ+o
mph	mɛditʃ=i	pjandʒ.t=o	pjang=o
PH	mɛditʃi	pjanto	pjango
ph	mɛːditʃi	pjanto	pjaŋgo

Mapping rules

- We map strings from one level to strings of another level by means of rules that map segments to segments
  1. $\sigma_1 \rightarrow \sigma_2$ in all contexts
  2. $\sigma_1 \rightarrow \sigma_2$ in context K_1 and $\sigma_1 \rightarrow \sigma_3$ in context K_2 FORK
  3. in some context $\sigma_1 \rightarrow \sigma_2$ and $\sigma_3 \rightarrow \sigma_2$ MERGER
  4. in some context $\sigma_1 \rightarrow \emptyset$ DELETION

Mapping rules

Simple rule	Fork	Merger
		

Inventories

- Each level of representation, in segmental domain, has its own
  - inventory of segments
    - alphabet of segments with phonetic classification
  - phonotactic rules
    - how segments can be combined

Axioms of segmental grammar

- A_1 — Every utterance can be represented as a linear sequence of segments defined by phonetic distinctive features.
  - Segmental representation must be possible, but A_1 says nothing on how do we get it
  - Phonetic features are both acoustic and articulatory

Axioms of segmental grammar

- 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 of segmental grammar

- A_3 — Every wordform has 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
    - see the example from Slavic

Axioms of segmental grammar

- A_4 — Segments of PH ≡ distinctive segments.
  - The distinctivity is on PH level
  - It is always true that distinctivity ⇒ phonemes
  - The opposite implication may be not true sometimes

Axioms of segmental grammar

- A_5 — Alphabets of mph and PH coincide.
  - Extremely important!
  - Otherwise there would be no limits to the alphabet of mph
  - Practically, there are no “morphophonemes”
  - Mapping rules mph→PH are a manipulation of phonemes

Excursus on morphophonemes

- Traditional morphophonology operates with morphophonemes
  - Because they try to solve an unsolvable problem:
  - explaining contextually something that is not contextual
    - For example, k_1 in $park_1o$ becomes k before i
    - but k_2 in $mɛdik_2o$ becomes ʧ before i

Excursus on morphophonemes

Infl. request	Pl(MEDICO)	Pl(PARCO)	Sg(CARNE)
Skeleton = mph	mɛdik$_2$+i	park$_1$+i	k$_ʔ$arn+e
PH	mɛditʃi	parki	karne
ph	mɛːditʃi	parki	karne

Axioms of segmental grammar

- A_6 — All the mapping rules within the domain of segmental grammar must be contextual.
  - These rules are: mph→PH, PH→mph, PH→ph, ph→PH.
  - If a rule cannot be defined contextually, this rule does not exist.

Axioms of segmental grammar

- A_7 — Mergers and deletions are only allowed in the mapping rules mph→PH
  - mph→PH rules are destructive: they destroy information
  - The other rules cannot be destructive
  - Particularly, PH→ph rules are not allowed to merge phonemes

Example of a destructive rule

Infl. request	PartSgm(PIANGERE)	Pres1Sg(PIANTARE)
Skeleton	pjandʒ+t+o	pjant+o
mph	pjandʒ.t=o	pjant=o
PH	pjanto	pjanto
ph	pjanto	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 are no PH→mph rules in segmental grammar

Some theorems

- Theorem 2. Biuniqueness of PH and ph.
  - A_7: only mph→PH rules admit mergers and deletions
  - Rules PH→ph do not admit mergers and deletions
    - but admits forks!
  - Information is not destroyed
  - There is always a contextual rule for restoring PH from ph
    - another consequence: there are no shared allophones

Some theorems

- Theorem 3. Boundaries between formatives are only visible in mph.
  - A_7: only mph→PH rules admit mergers and deletions
  - Rules mph→PH may involve phonemes around the boundary: α.β, which can therefore undergo merge or deletion
  - Therefore, boundaries are not warranted in PH (and ph).
    - In PH representation we simply don't have roots and endings anymore

Some Italian facts in our analysis

- These are some of the “difficulties” of the “phonetic transcription” of Italian
  - assimilation of nasals
  - lengthening of palatals
  - treatment of unstressed vowels

Some Italian facts in our analysis

	INFATTI	IMPARI	OGNI
mph	infatti	in.pari	oɲɲi
PH	infatti	impari	oɲɲi
ph	iɱfatti	impari	oɲɲi
	n→ɱ/_f (PH→ph)	n→m/_p (mph→PH)	short ɲ prohibited between vowels on all levels

Some Italian facts in our analysis

	TEMPO	TEMPORALE	TONO	TONALE
mph	tɛ́mp=o	tɛmp.or.ál=e	tɔ́n=o	tɔn.ál=e
PH	tɛ́mpo	temporále	tɔ́no	tonále
ph	tɛ́mpo	temporáːle	tɔ́ːno	tonáːle

ɛ→e/unstressed
(mph→PH)

ɔ→o/unstressed
(mph→PH)