Workshop

Phonological systems and complex adaptive systems

Complexity of phonological inventories: features & structures

Christophe Coupé, Egidio Marsico and François Pellegrino

Laboratoire Dynamique Du Langage

UMR 5596 CNRS – Université Lumière Lyon 2

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

Dynamique Du Langage

UNIVERSITÉ LUMIÈRE LYON 2

Objectives

Short term

- Find one or several measures allowing to compare the structure of phonological inventories both quantitatively and qualitatively

Mid term

- Be able to organize all system types according to various indices that may explicate frequencies of distribution

Long term

- Develop an evolutionary model for phonological inventories

Preliminaries: Proposal of indices

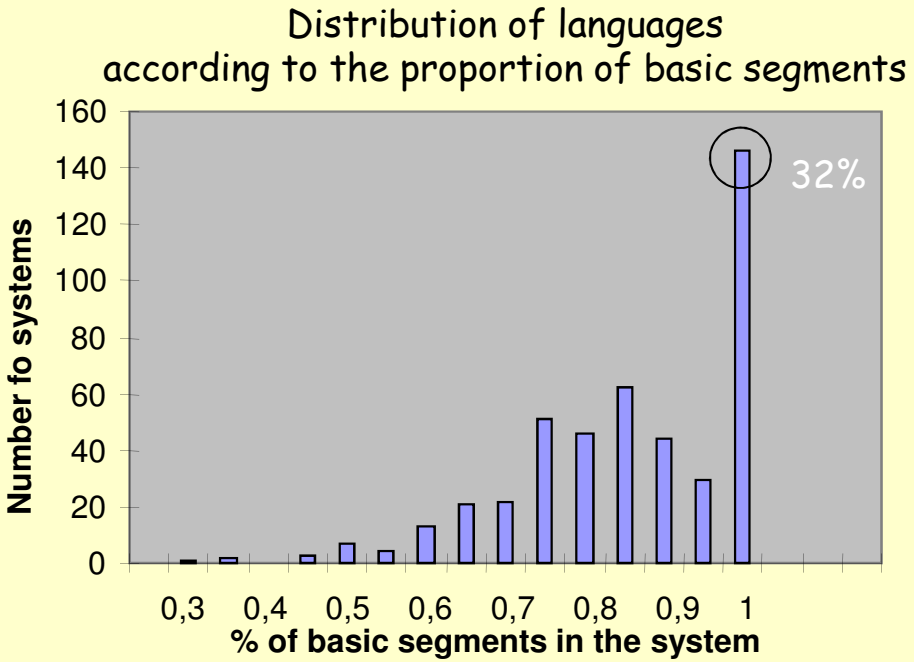
- Previous work
  - ✓ Attempt to define indices able to monitor the organization of phonological inventories (UPSID) (presented at BLS 2004)

- Selected indices:
  - ✓ Basicness
  - ✓ Redundancy
  - ✓ Generativity
  - ✓ Plasticity

Basicness

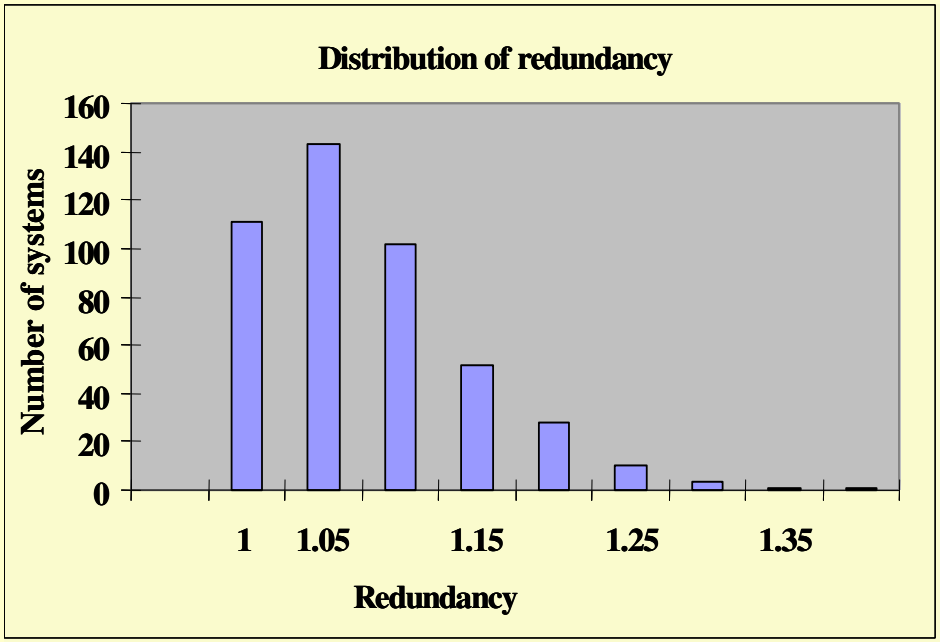
- A feature is basic if, when removed from the definition of a segment, the remaining set of features is not a segment

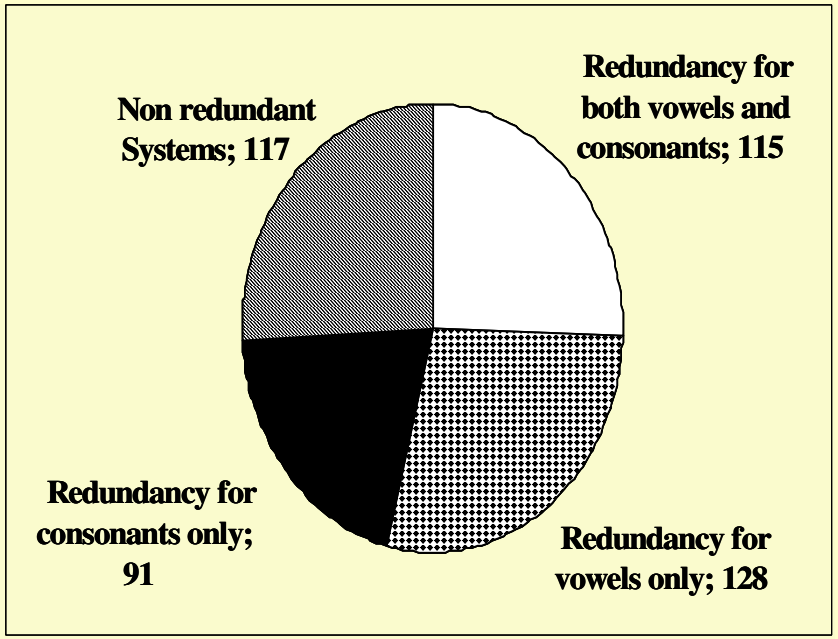
- Basicness is defined at the feature level and extended to segments and systems.



Redundancy

- Mean distance between each segment and its nearest neighbor (in terms of number of features)
- Calculated at the system level





Generativity for vowels

= Number of segments derived from the basic segment considered

i
a
o
u
e

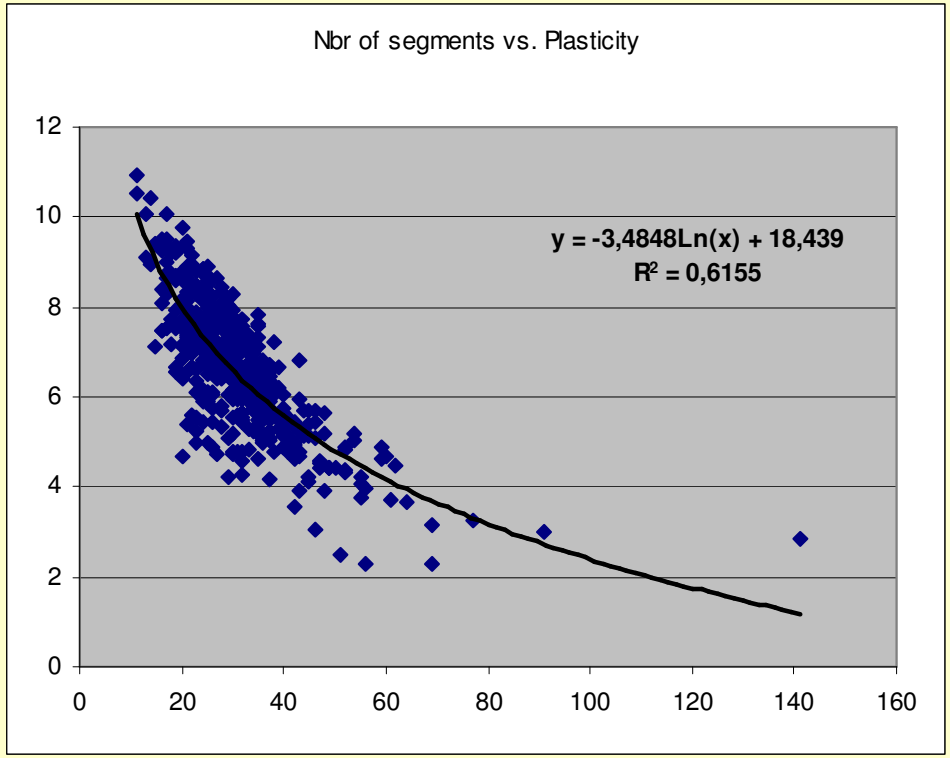
Segments	Generativity	Derivation degree	Frequency (in languages)
voiced high front unrounded	14	0	**0.87**
voiced low central unrounded	14	0	**0.87**
voiced higher-mid back rounded	12	0	**0.69**
voiced high back rounded	11	0	**0.82**
voiced higher-mid front unrounded	10	0	**0.65**
voiced lower-mid back rounded	8	0	0.36
voiced lower-mid front unrounded	6	0	0.41
voiced high central unrounded	5	0	0.15
voiced higher-mid front rounded	5	0	0.03
voiced higher-mid central unrounded	5	0	0.04
voiced higher-mid back unrounded	5	0	0.04
voiced mid central unrounded	5	0	0.17
voiced nasalized low central unrounded	5	1	0.18
voiced nasalized high front unrounded	4	1	0.18
voiced high back unrounded	4	0	0.09
voiced lowered-high back rounded	4	0	0.15
voiced low back rounded	4	0	0.04
voiced high front rounded	3	0	0.05
voiced lowered-high front unrounded	3	0	0.16

Generativity for consonants

Segments	Generativity	Derivation degree	Frequency (in languages)
voiceless velar stop	18	0	**0.89**
voiceless alveolar stop	14	0	**0.74**
voiceless postalveolar sibilant-affricate	14	0	0.42
voiceless uvular stop	13	0	**0.12**
voiceless bilabial stop	12	0	**0.83**
voiced bilabial stop	11	0	**0.64**
voiced alveolar stop	11	0	0.47
voiceless alveolar sibilant-affricate	11	0	0.24
voiced velar stop	10	0	**0.56**
voiced bilabial nasal	9	0	**0.94**
voiceless alveolar sibilant-fricative	9	0	**0.73**
voiceless uvular non-sibilant-fricative	9	0	**0.10**
voiceless dental stop	7	0	0.24
voiced velar nasal	7	0	**0.53**
voiced alveolar trill-or-unspecified	7	0	0.43
voiceless postalveolar sibilant-fricative	7	0	0.41
voiceless velar non-sibilant-fricative	7	0	0.21
voiced alveolar lateral-approximant	7	0	**0.69**
voiced alveolar nasal	6	0	**0.80**

Plasticity

- Extension of the notion of generativity to the system level
- Take a "lazy" way of evolution into account (reuse of already available segments…)

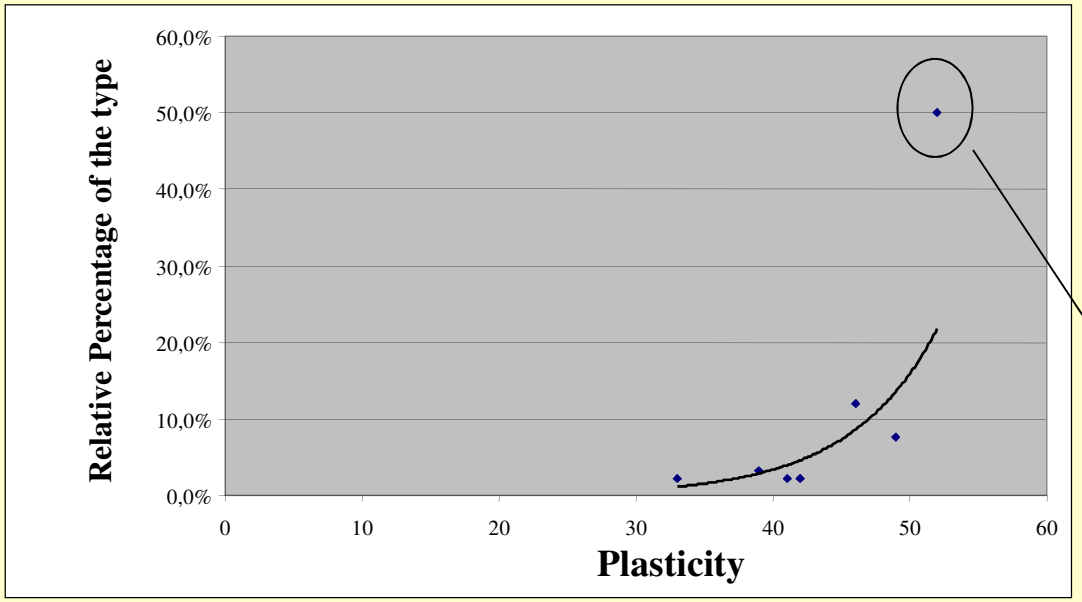


Plasticity (cont'd)

- Are "plastic" systems preferred?

- Example for 5-vowel systems:

% of each type x Plasticity



$R^2 = 0.7$

/i e a o u/

Conclusions about indices

- Basicness
  - ✓ Full basicness for 32% of the systems
  - ✓ If basicness < 1, regular distribution (though not normal)
- Redundancy
  - ✓ MUAF
  - ✓ Feature Economy
- Generativity
  - ✓ Linked to the frequency of occurrences for "best –seller" vowels
  - ✓ More complicate (= not understood yet) scheme for consonants
- Plasticity
  - ✓ Negatively correlated to the size of systems
  - ✓ Maybe correlated to the frequency of occurrence of systems (?)

How to go further?

- Complexity of Phonological Systems involves (at least):
  - ✓ Intrinsic complexity of the elements (primitives)
  - ✓ Complexity of interactions
  - ✓ Structural complexity

- How to characterize the structural complexity?
  - ↳ Networks of interactions

- How to take interactions into account ?
  - ↳ Weighting the structure according to the relationship between the constituents

- What are the correct primitives?
  - ✓ Features? Segments? Oppositions?
  - ↳ Discussion about the description of segments in phonological systems

A proposal to build phonological graphs

A graph built from a set of segments (and their relations in terms of features)

segments = nodes of the graph

Goal:

Build a network based on oppositions between segments, which translates the relations between basic and derived segments

Method:

Prune a fully-connected network to only retain relevant links between segments

Rely on a feature-based distance:

d(i,e)=1 ; d(i, i:) = 1 ; d(i,u) = 2 ; d(e,u) = 3 ; d(a:, ã) = 2

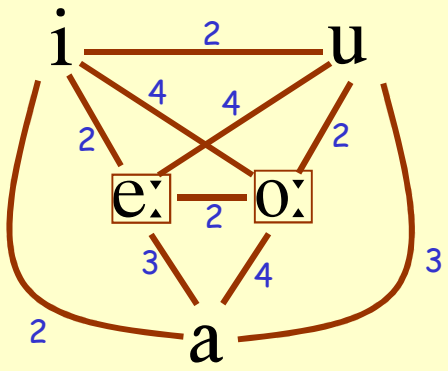
high – higher mid | long – ø | front – back
rounded - unrounded | front – back
rounded – unrounded
high – higher mid | long – ø
nasalized - ø

(secondary features do not get opposed to each others)

A proposal to build phonological graphs: Description of the algorithm

1. Compute the distances for all pairs of segments

	i	u	e:	o:	a
i	0	2	2	3	2
u		0	4	2	3
e:			0	2	3
o:				0	4
a					0



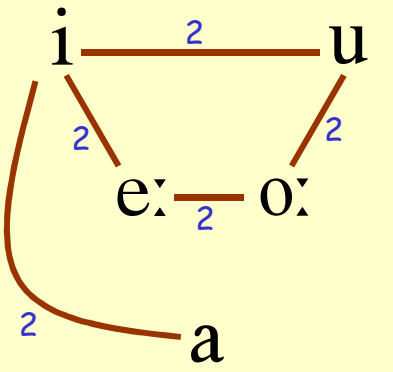
2. Compute shortest paths for all pairs of segments

	i	u	e:	o:	a
i	0	2	2	2	2
u		0	2	2	2
e:			0	2	2
o:				0	2
a					0

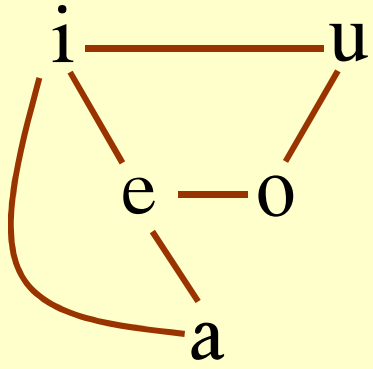
length of a path = maximum distance on this path

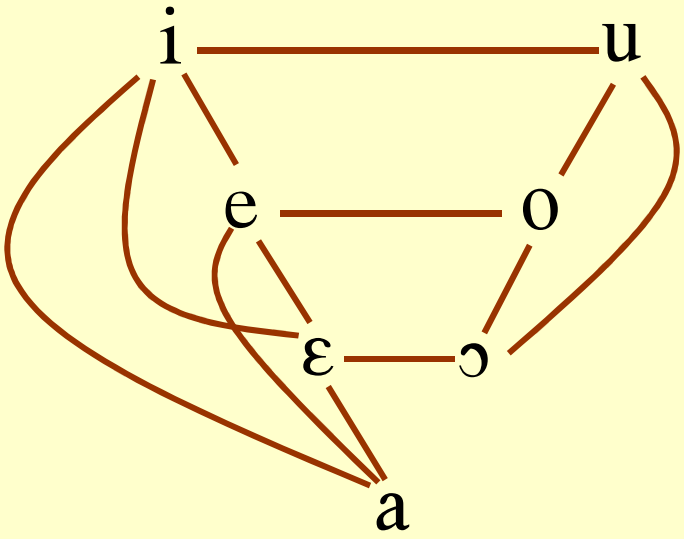
3. For all pairs of segments, remove direct link if it exceeds the length of the shortest path

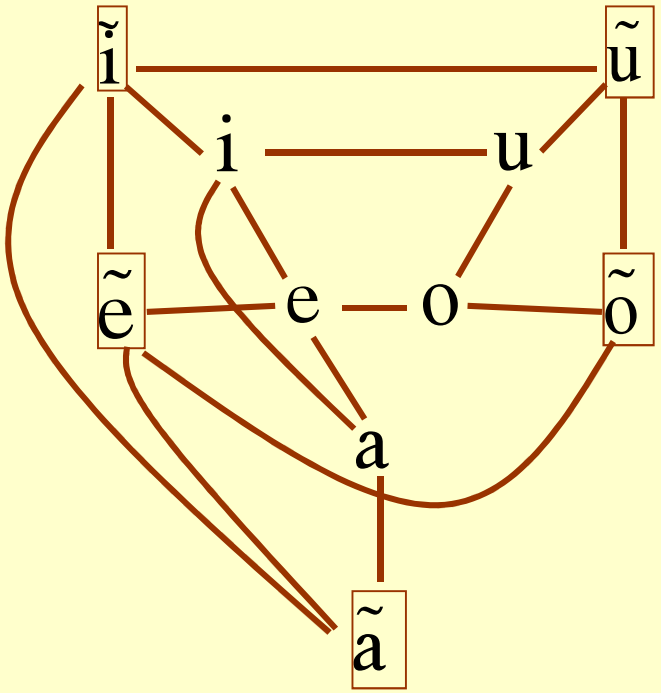
	i	u	e:	o:	a
i	0	2	2		2
u		0		2	
e:			0	2	
o:				0	
a					0

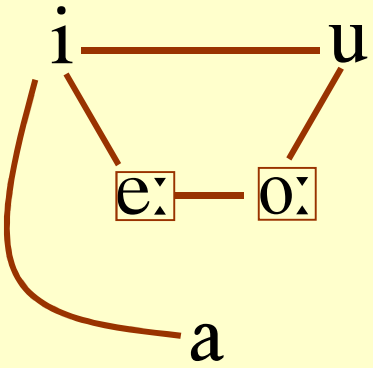


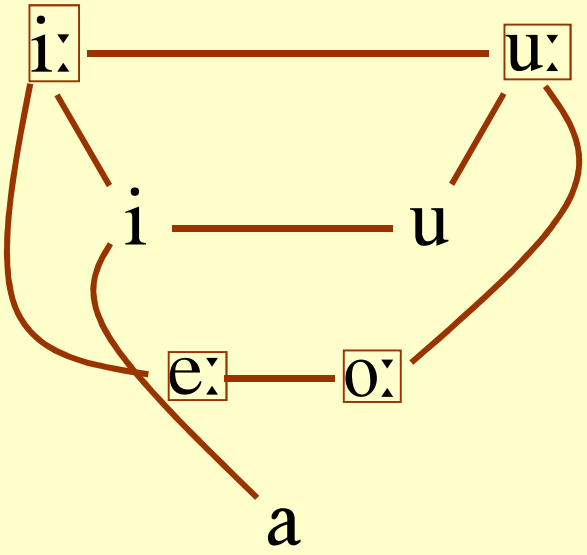
Examples of networks











Measuring graph complexity: Offdiagonal complexity

Claussen, J. C. (2004) Offdiagonal Complexity: A computationally quick complexity measure for graphs and networks. *q-bio.MN/0410024*.

Principle:

- Compute the degrees of the nodes of the graph (=number of connections)

- Fill a matrix M with M(k1,k2) = nb of links between nodes of d° k1 and nodes of d° k2

- Compute the entropy of this distribution (after summation on the minor diagonals):

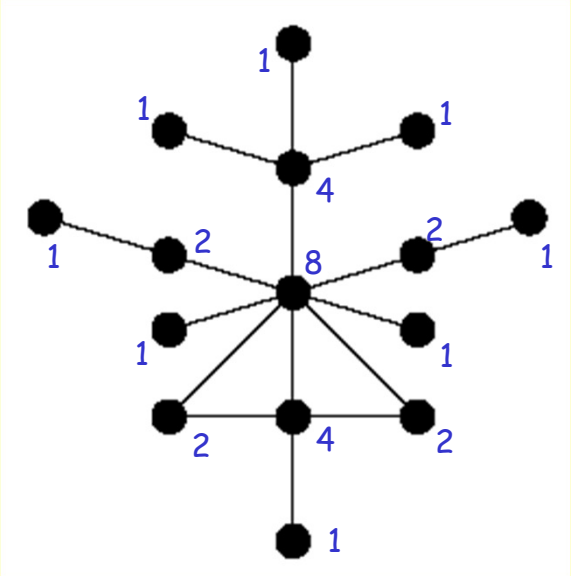
Properties:

- not related to graph size

- sensitive to hierarchical structures

- minimum value for regular graphs, maximum for scale-free networks

Offdiagonal complexity: Examples

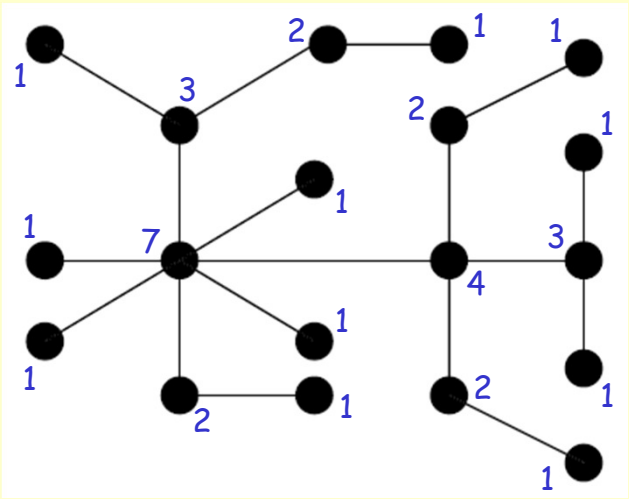


	1	2	4	8	k_2
1	0	2	4	2	2
2		0	2	4	8
4			0	2	6
8				0	0
k_1	C=0.974				

C = - [

2/16.log(2/16) +
8/16.log(8/16) +
6/16.log(6/16)]

C = 0.974



	1	2	3	4	5	6	7	k_2
1	0	5	3	0	0	0	3	3
2		0	1	2	0	0	1	1
3			0	1	0	0	1	1
4				0	0	0	1	1
					0	0	0	5
						0	0	7
k_1		C=1.503					0	

Applying offdiagonal complexity to phonological graphs

A measure of *structural* complexity (only)

. Different from the overall complexity of a system

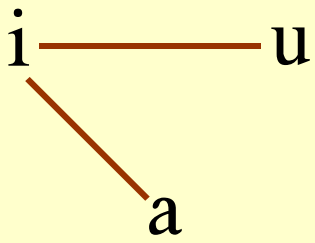
. No (a priori) relation with frequencies of types

Offdiagonal complexity works with non-valued graphs…

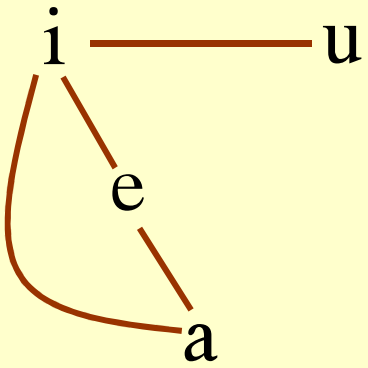
Discard the values of the links of the graph

(and think to something better later…)

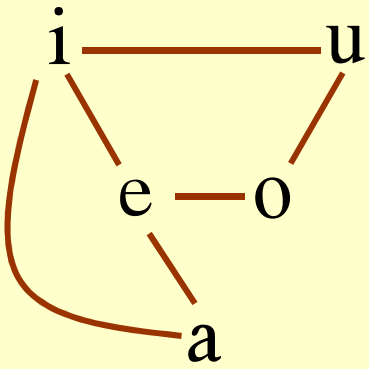
Examples of networks



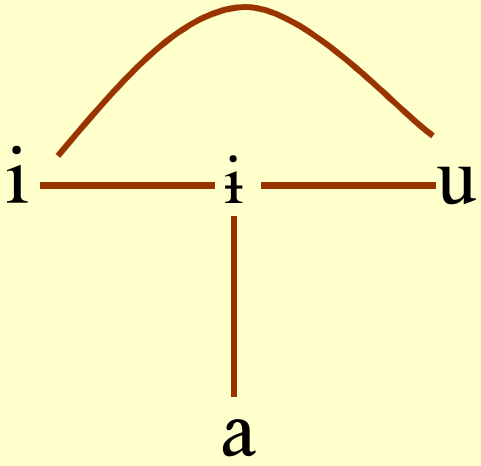
C = 0



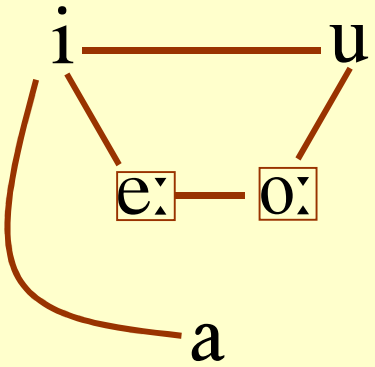
C = 1.04



C = 0.64

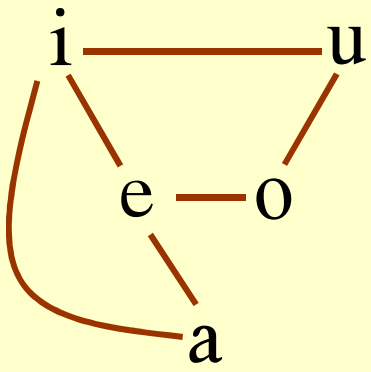


C = 1.04

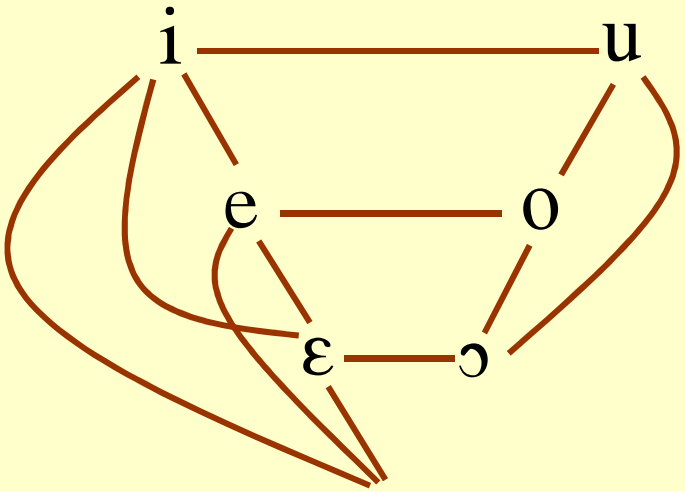


C = 0.69

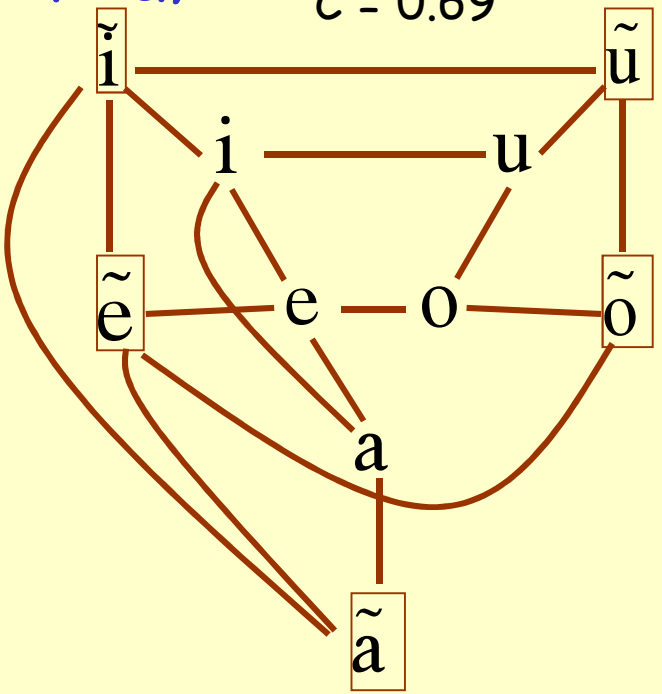
Examples of networks (cont'd)



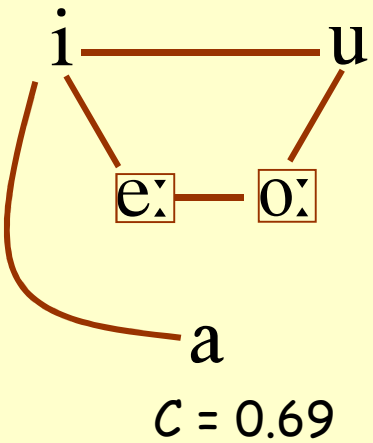
C = 0.64



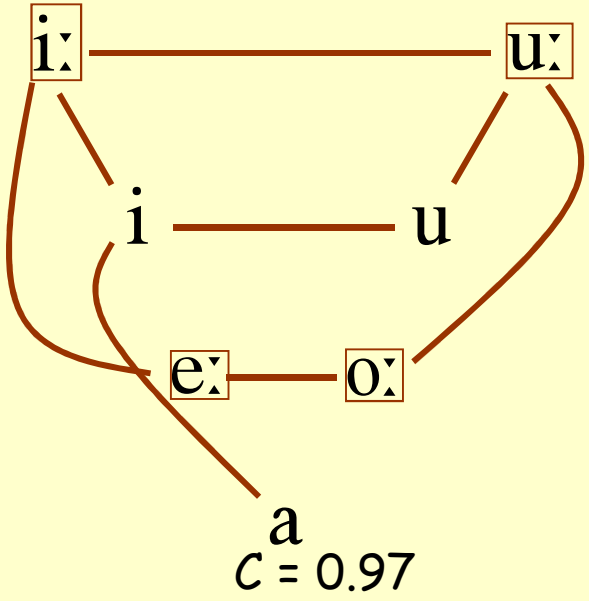
C = 0.69



C = 0.69



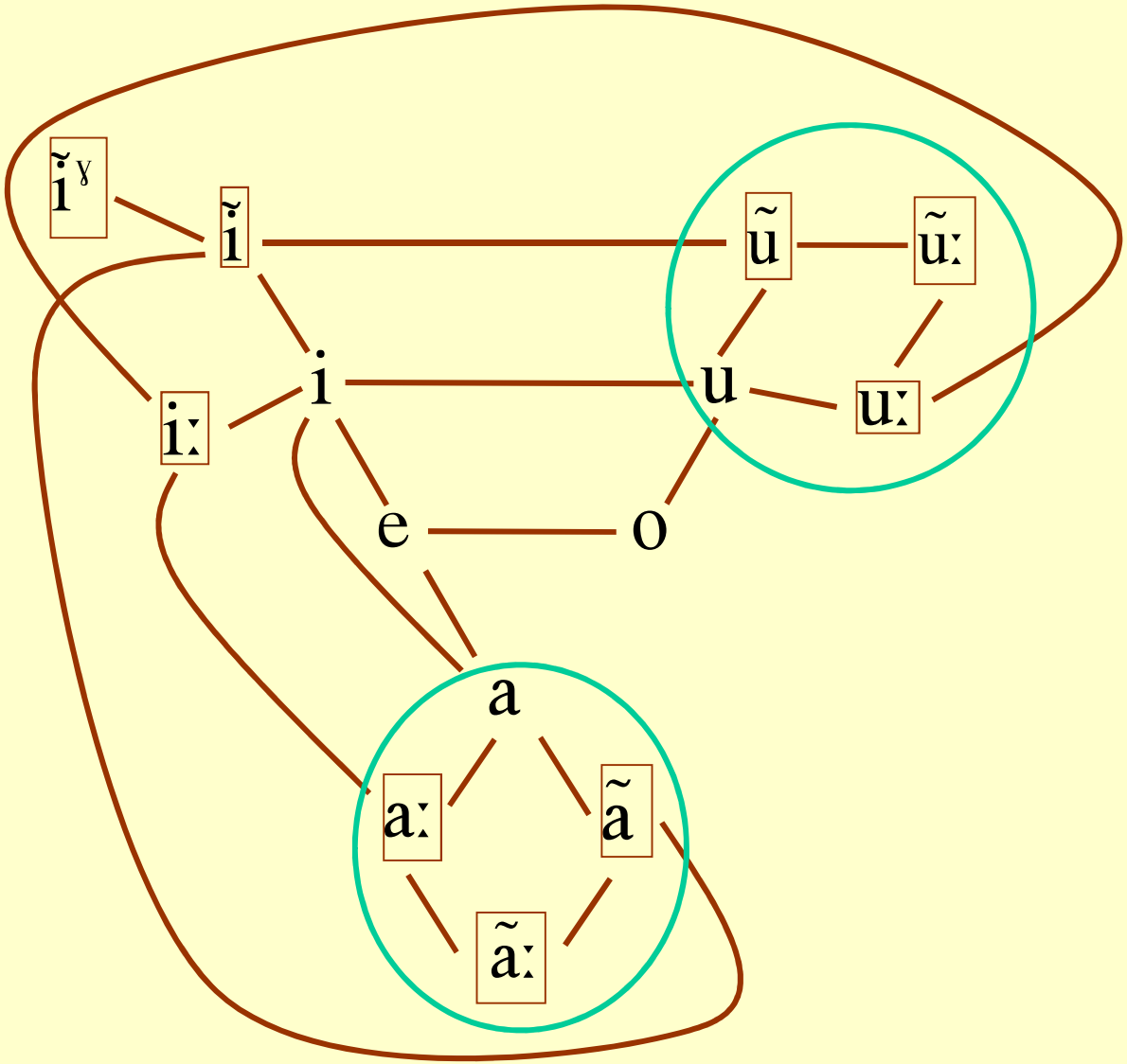
C = 0.69



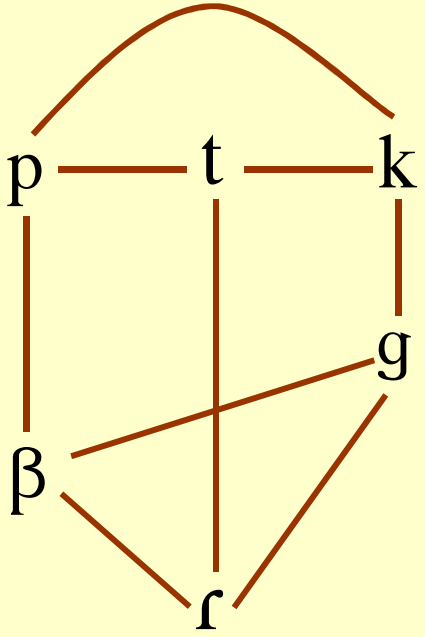
C = 0.97

Examples of networks (still cont'd)

Chipewyan
$C = 0.89$



Rotokas
$C = 0$



Estimating the phonological structural complexity of UPSID's sample of languages (1)

Average complexity for vocalic systems (diphtongs omitted) :

$C = 0.79$

$(\sigma = 0.31)$

Average complexity for random vocalic systems (diphtongs omitted) (similar size distribution)

$C = 1.06$ $(\sigma = 0.49)$

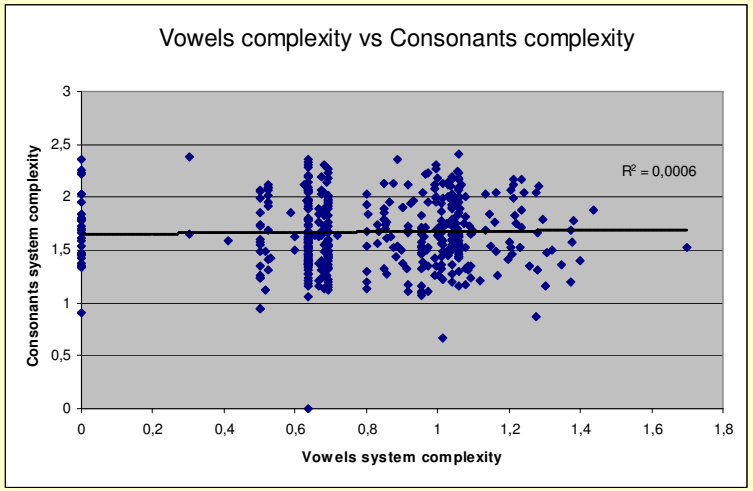
Significantly different: $t(450) = 9.85$, $p << 1$

Average complexity for consonantal systems

(clicks omitted for mental sanity reasons):

$C = 1.63$

Correlation between vocalic and consonantic structural complexity: No correlation at all…



Estimating the phonological structural complexity of UPSID's sample of languages (2)

Vocalic and consonantic structural complexities across groups:

	Europe, W. & S. Asia (71)	E. & S.E. Asia (108)	Africa (74)	N. Am. (68)	S. & Central Am. (66)	Australia & New Guinea (64)
V	0.90	0.87	0.79	0.73	0.81	0.65
C	1.83	1.61	1.84	1.67	1.51	1.45

First tests suggest significant differences between groups

Discussion
about the description of segments

1. Influence of the description

- ✓ A reduced and an extended feature sets have been defined and tested (with Ian Maddieson)
- ✓ Comparison of the indices (basicness, etc.) estimated with the 3 feature sets

2. Improving the description

- ✓ On the notion of segment, consonant and vowel
- ✓ On the nature of features

Standard set of features

Standard Set					
affricate	alveolar	front	high	ad-tg-root	raised
affricate-lateral	bilabial	front-back	higher-mid	advanced	raised-low
affricate-trill	dental	front-central	higher-mid-high	aspirated	raised-low-high
approximant	epiglottal	central	higher-mid-low	ejective	retracted
click	glottal	central-back	higher-mid-mid	glottalized	retroflexed
flap	labial-palatal	central-front	high-higher-mid	labialized	velar-fricated
fricative-flap	labial-velar	back	high-low	lateral-release	velarized
fricative-trill	labiodental	back-central	high-lower-mid	lip-compressed	with-breathy-release
implosive	palatal	back-front	high-mid	long	
lateral-approximant	pharyngeal		mid	nasalized	
lateral-flap	postalveolar	unrounded	mid-high	nasal-release	
lateral-fricative	retroflex	unrounding	mid-lower-mid	overshort	
nasal	uvular	rounded	low	palatalized	
non-sibilant-affricate	velar	rounding	lowered-high	pharyngealized	
non-sibilant-fricative			lowered-high-high	preaspirated	
sibilant-affricate			lowered-high-higher-mid	preglottalized	
sibilant-fricative	voiced		lower-mid	prenasalized	
stop	voiceless		lower-mid-high	prestopped	
tap	breathy-voiced		lower-mid-higher-mid		
trill-or-unspecified	creaky-voiced		low-high		
	narrow-voiceless		low-higher-mid		
			low-lower-mid		

Reduced set of features

Reduced Set			
affricate	alveolar	front	ad-tg-root
approximant	bilabial	central	advanced
click	dental	back	aspirated
flap	epiglottal		ejective
fricative	glottal	rounded	lateral-release
implosive	labiodental	unrounded	lip-compressed
lateral	palatal		long
nasal	pharyngeal	high	nasal-release
non-sibilant	postalveolar	higher-mid	overshort
sibilant	retroflex	mid	preaspirated
stop	uvular	low	preglottalized
tap	velar	lowered-high	prenasalized
trill-or-unspecified		lower-mid	prestopped
	voiced		raised
	voiceless		raised-low
	breathy-voiced		retracted
	creaky-voiced		velar-fricated
	narrow-voiceless		with-breathy-release

Expanded set of features

Expanded Set				
affricate-click	advanced-alveolar	palatal-velar-fricated-glottalized	unrounded	high
affricate-click-nasalized	alveolar	pharyngeal	unrounding	higher-mid
affricate-lateral	alveolar-glottalized	pharyngeal-glottal	rounded	higher-mid-high
affricate-lateral-click	alveolar-labialized	pharyngealized	rounding	higher-mid-low
affricate-lateral-click-nasalized	alveolar-palatalized	postalveolar		higher-mid-mid
affricate-trill	alveolar-pharyngealized	postalveolar-glottalized		high-higher-mid
approximant	alveolar-velar-fricated	postalveolar-labialized	front	high-low
approximant-nasalized	alveolar-velar-fricated-glottalized	postalveolar-palatalized	front-back	high-lower-mid
click	alveolar-velarized	postalveolar-velar-fricated	front-central	high-mid
click-nasalized	alveolar-velarized-glottalized	postalveolar-velarized	central	mid
flap	bilabial	preaspirated-voiceless	central-back	mid-high
flap-nasalized	bilabial-labialized	preglottalized-voiced	central-front	mid-lower-mid
fricative-flap	bilabial-labialized-velarized	retroflex	back	low
fricative-trill	bilabial-palatalized	uvular	back-central	lowered-high
implosive	bilabial-velarized	uvular-labialized	back-front	lowered-high-high
lateral-approximant	bilabial-velarized-labialized	uvular-labialized-pharyngealized		lowered-high-higher-mid
lateral-approximant-nasalized	dental	uvular-pharyngealized		lower-mid
lateral-flap	dental-glottalized	velar		lower-mid-high
lateral-fricative	dental-palatalized	velarized		lower-mid-higher-mid
nasal	dental-pharyngealized	velar-labialized		low-high
non-sibilant-affricate	dental-velar-fricated	velar-palatalized		low-higher-mid
non-sibilant-fricative	dental-velar-fricated-glottalized	velar-palatalized-labialized		low-lower-mid
prenansalized-sibilant-affricate	dental-velarized	velar-pharyngealized		raised-low
prenasalized-affricate-trill	epiglottal			raised-low-high
prenasalized-non-sibilant-affricate	glottal			
prenasalized-non-sibilant-fricative	glottal-labialized			
prenasalized-sibilant-affricate	glottal-palatalized			
prenasalized-sibilant-fricative	glottal-pharyngealized			
prenasalized-stop	labial-palatal	voiced		
prenasalized-trill-or-unspecified	labial-velar	voiced-aspirated	ad-tg-root	
prestopped-lateral-approximant	labial-velar-labialized	voiced-ejective	advanced	
sibilant-affricate	labiodental	voiceless	lip-compressed	
sibilant-fricative	labiodental-labialized	voiceless-aspirated	long	
stop	labiodental-palatalized	voiceless-ejective	nasalized	
stop-lateral-release	palatal	voiceless-with-breathy-release	overshort	
stop-nasal-release	palatal-glottalized	breathy-voiced	raised	
tap	palatal-labialized	creaky-voiced	retracted	
trill-or-unspecified	palatal-velar-fricated	narrow-voiceless	retroflexed	

Influence of the description: Conclusion

- Strong correlations between the indices estimated with the 3 feature sets

One step ahead: revisiting features and segments

- Should we use a unified frame to describe Consonants & Vowels?
- Features are not equivalent in the current description
  - ✓ Dynamical information (affricate, …)
  - ✓ Complex information (lateral,…)
- Should we use an explicitly dynamical description?

What is a segment?

- Intuitively, it seems reasonable to consider that vowels and consonants should be described with the same set of features. Even though they might play different roles in speech, yet they're produced with same vocal tract and under the same physical constraints.

- We could define any speech sound as the result of a laryngeal event (source) and a supra-laryngeal one (filter), along with a time dimension to instantiate these events.

- The laryngeal event only has to be characterized on its "how" part (mode), whereas the supra-laryngeal one also needs a "where" part (place).

Something like this…

New description settings

Basic elements: LS=Laryngeal Settings, FM=Filter Mode, FP=Filter place

	LS	FM	FP	
[p]	voiceless	closure	bilabial	
[i]	voiced	high open	palatal	
[y]	voiced	high open	labio-palatal	

Time dimension

		Onset	Steady State	Offset
[ts]	LS	voiceless	voiceless	voiceless
	FM	closure	closure	constriction
	FP	alveolar	alveolar	alveolar
[pʲ]	LS	voiceless	voiceless	voiced
	FM	closure	closure	near open
	FP	bilabial	bilabial	palatal

Revised features

- One unique set of features for both consonants and vowels => need to modify the current classical features

- Proposition
  - ✓ Constriction scale for mode of articulation (from closure to open)
  - ✓ Nasalization: nasalized for both vowels and consonants
  - ✓ Rounding is considered as a second place of articulation

(numerous) Problems

- Laterals, rhotics, nasal consonants, prenasalized stops, secondary articulations…

- All this is an attempt to characterize segments though their cognitive relevance is still at stakes
  - ✓ Segments as emergent properties (by-products)?

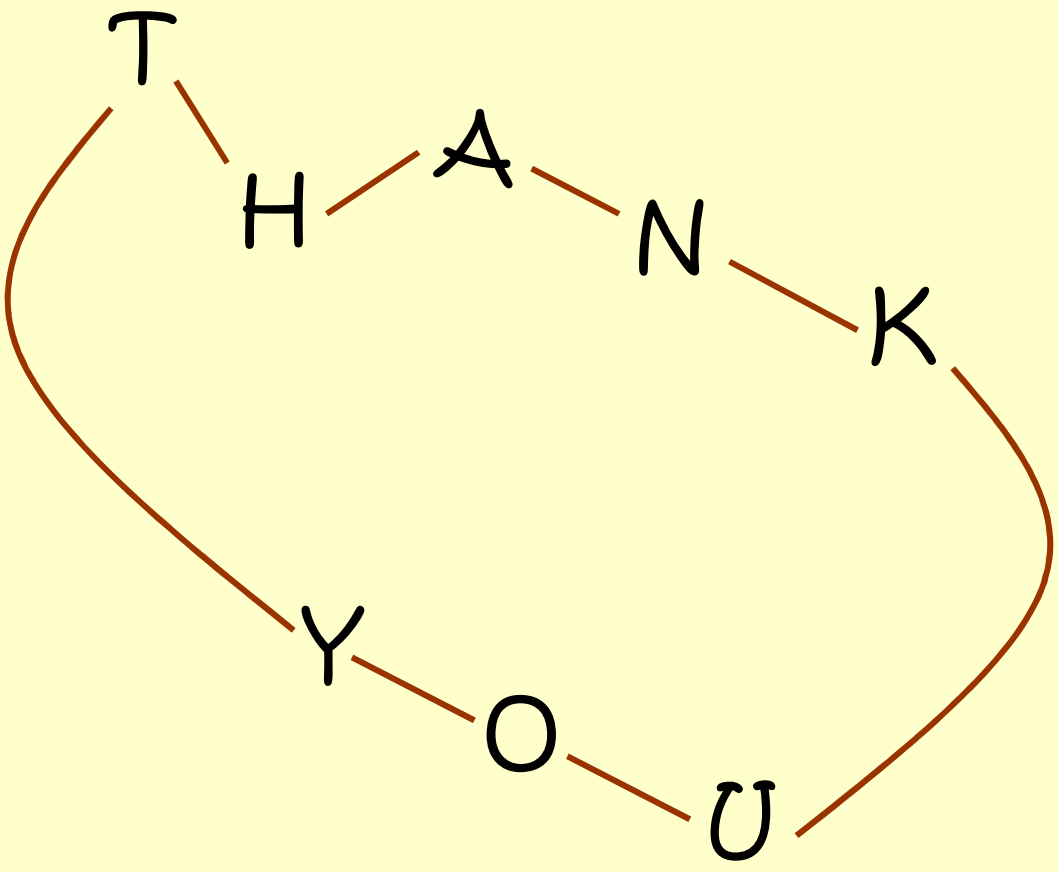
…If we get rid of the notion of segments then what is the meaning of phonological inventories?

Perspectives

- Structural complexity
  - ✓ Structural comparison of graphs (with ABSURDIST algorithm)
    → structural distances between systems
  - ✓ Approach based on graph theory and indices (feature, segment and system levels)
  - ✓ Necessity to evaluate this measurement

- Complexity of interactions
  - ✓ Possible transformation from segments to contrasts
  - ✓ From segment graphs to feature graphs or oppositions graphs

- Intrinsic complexity of the elements (primitives)
  - ✓ Proposition of a dynamic and unified descriptive set of features


T
H
A
N
K
Y
O
U