Approaches to Complexity in Language, Helsinki, August 24-26, 2005

Complex patterns in phonological systems

Christophe Coupé, Egidio Marsico & François Pellegrino

Laboratoire Dynamique Du Langage
UMR 5596 CNRS – Université Lumière Lyon 2

- ✓ Introduction
- ✓ Presentation of our approach
- ✓ First results
- ✓ Complexity measurement
- ✓ Conclusion and perspectives

- ✓ Introduction
- ✓ Presentation of our approach
- ✓ First results
- ✓ Complexity measurement
- ✓ Conclusion and perspectives

About complexity and complex adaptive systems

- Definition of a complex system
  - ✓ A set of elements (identical or not), interacting with each other, in a non-linear or non-hierarchical way
  - ✓ Characterized by emergent properties at the general level
- Complexity vs. understanding
  - ✓ Simple versus complicated systems ↔ understanding of the system
  - ✓ A complex system may be considered as complicated because
    - The relevant components have not been identified
    - Their interactions have not been understood
- Regarding phonological systems (and language in general)
  - ✓ Are phonological systems complex systems?
  - ✓ Can be considered as complicate until we manage to "break the complexity"
    - Identify relevant components and their *linear* or *non-linear* interactions

Objectives

Short term

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

Mid term

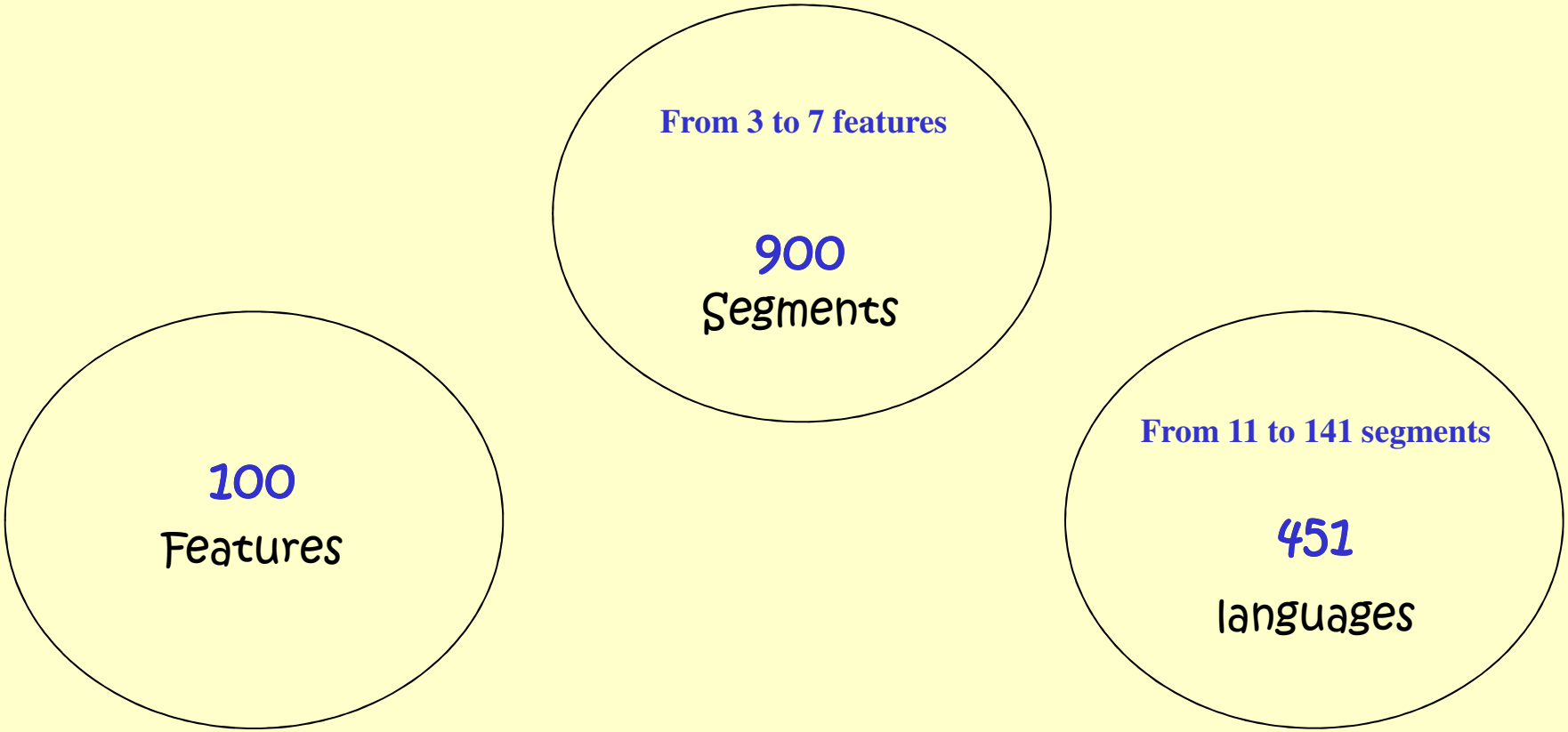
- Be able to organize all system types according to various indices that may explicate frequencies of distribution

Long term

- Develop an evolutionary model for phonological inventories

UPSID description

UCLA Phonological Segment Inventories Database, Maddieson, 1984; Maddieson & Precoda, 1990



… But theoretically

→ with 100 features we could generate around 10^{10} different segments (defined with 3, 4, 5, 6 or 7 features)

→ With 900 segments we could obtain around 10^{56} different systems with an average number of segments of 31

This difference shows that phonological systems are not randomly composed

But rather that some constraints are responsible for their shape

→ Identify and understand how these constraints weight on the content of phonological systems

Previous work

- Some principles are invoked to explain the structure of vowel systems

  (sufficient perceptual contrast, articulatory easiness (or economy), focalization, quantal effect…)

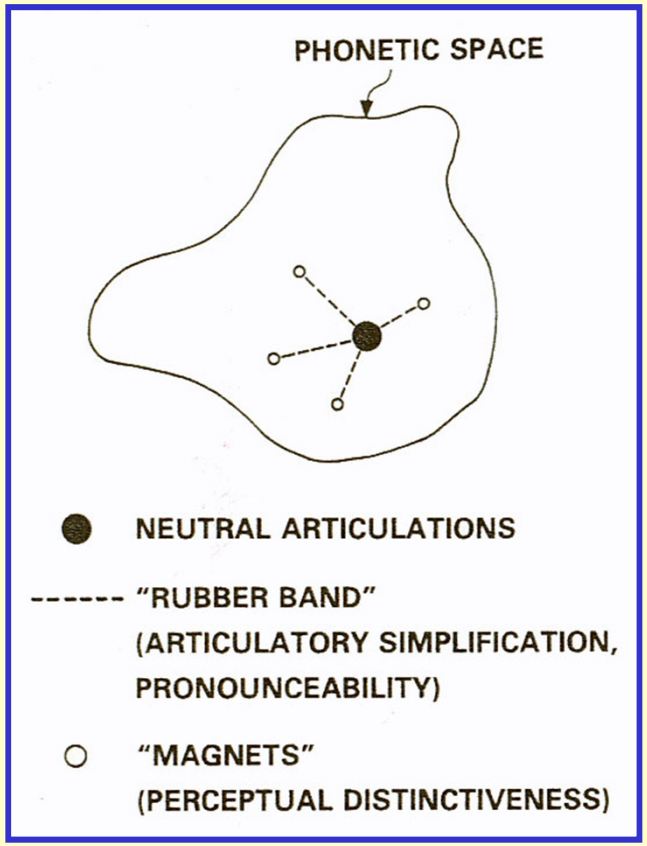
- Some cues regarding consonantal systems

  (simple, elaborated and complex consonants…)

- No global approach for phonological systems

  (The "size principle"…)

Characterize the "all inclusive universal phonetic space" (UPS), Lindblom and Maddieson 1988



- ✓ Introduction
- ✓ Presentation of our approach
- ✓ First results
- ✓ Complexity measurement
- ✓ Conclusion and perspectives

Usually...

* typological studies starts by considering the frequency of distribution of various patterns

But...

* we consider these frequencies as emergent properties: surface illustrations of the hidden structure

→ we thus keep them for the end as a validation

hypothesis

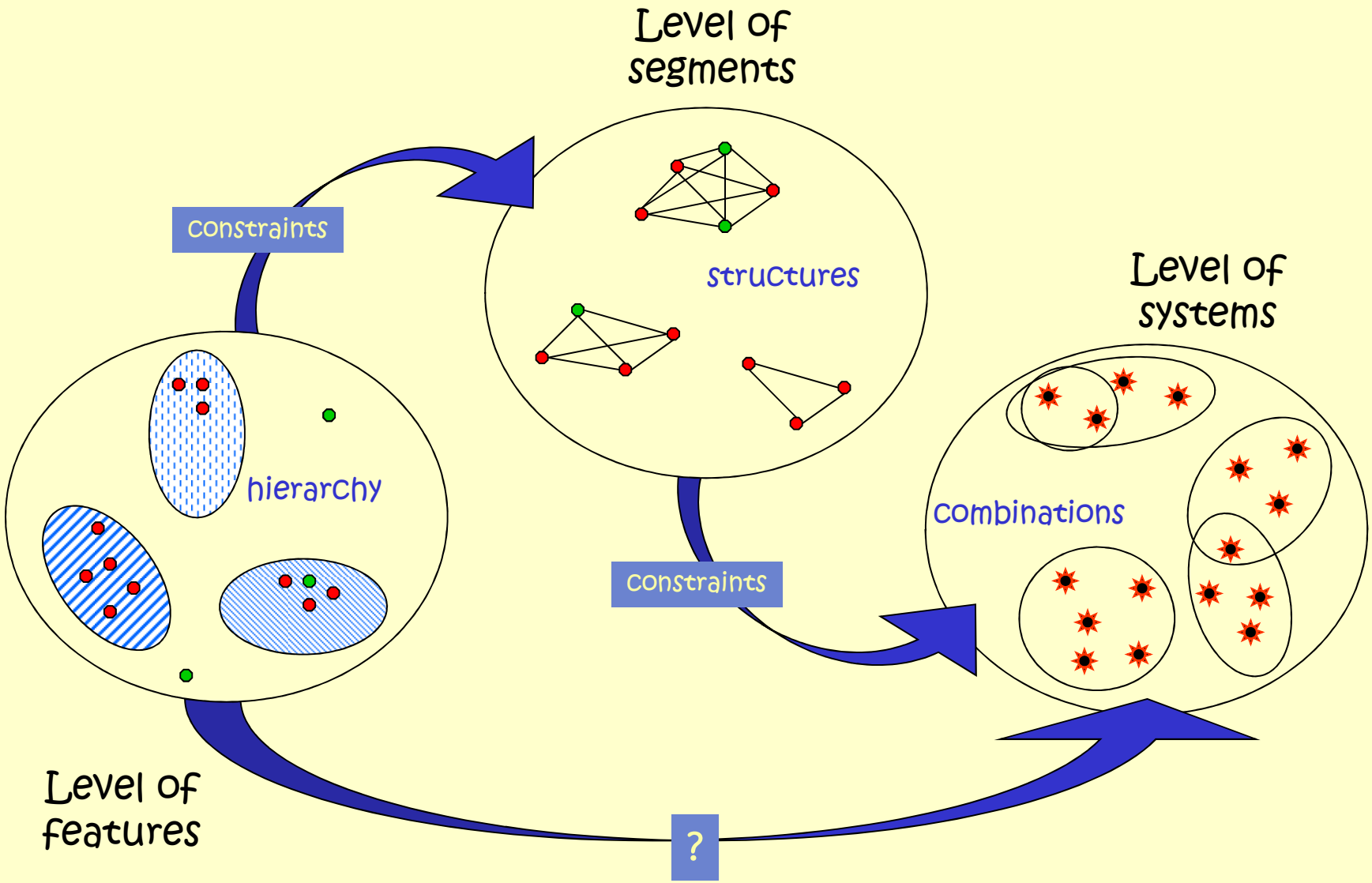
The frequency of distribution of a particular type is a function of:

1) its responsiveness to synchronic constraints

2) its positioning at the crossroads (or not) of evolutionary trajectories

3) its capacity of adaptation (number of possible extensions)

Possible phonological elements



Different indices

* Basicness:

how necessary is a feature in the definition of a segment

(Features, segments, systems)

The basicness of features is derived from the inventory of all segments (not an intrinsic property)
i {high front unrounded}
i: {long high front unrounded}

The basicness of a feature is a function of its ability to belong to the set of features that can minimally define a segment.
This value is normalized by the number of segments in which the feature appears.
(→ index from 0 to 1)

* Generativity:

how many different segments are based on the same one

(segments)

Given a segment, we calculate how many existing segments are derived from it by addition of features

Different indices

* Redundancy:

how economical is the set of features of a particular system

(systems)

The distance between each segment and its nearest neighbor averaged over the system

→ a system containing only minimal pairs will have a redundancy of 1

* Plasticity:

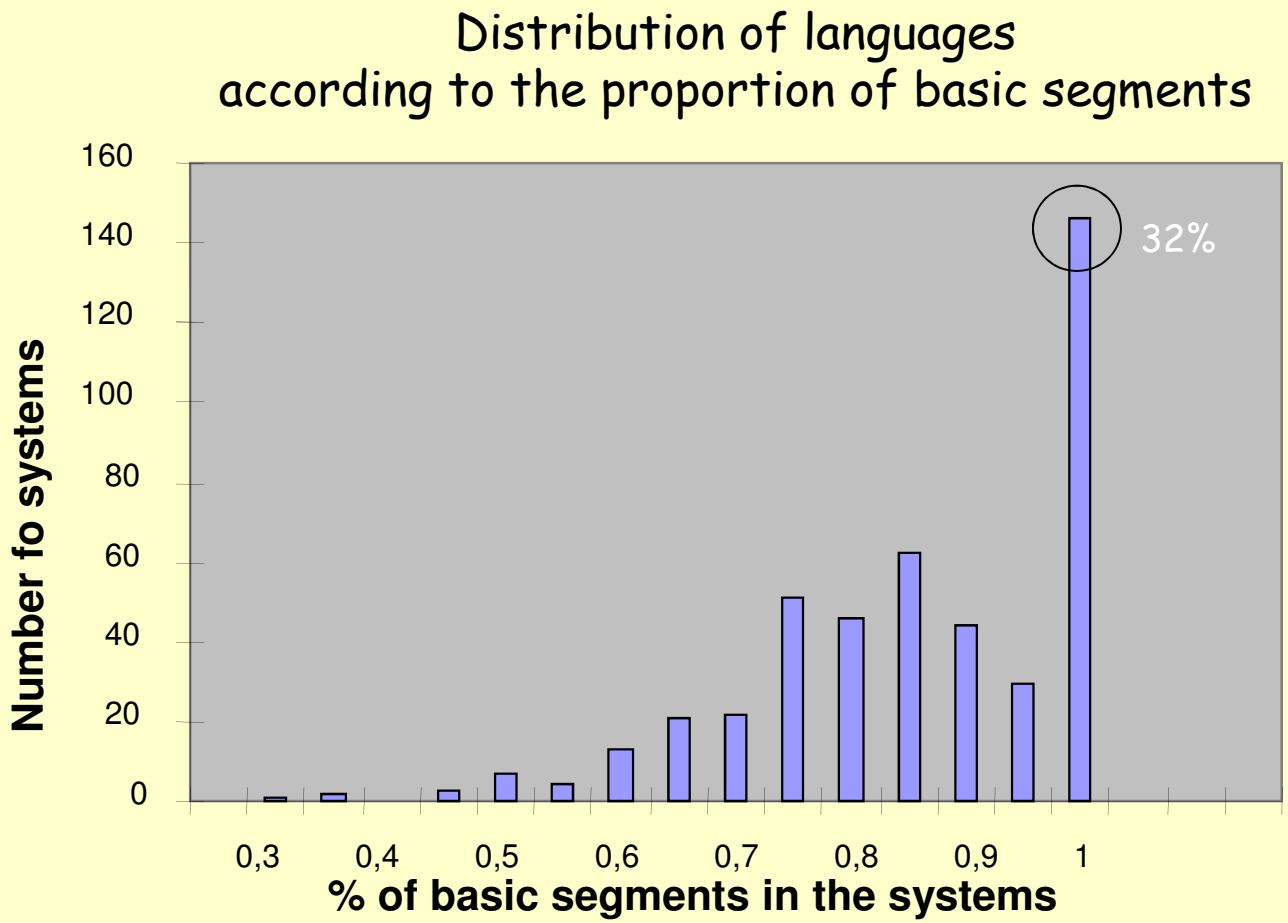
how many extensions can we have from one particular system

(systems)

The number of possible new segments a system can have, based on the index of generativity of its segments

- ✓ Introduction
- ✓ Presentation of our approach
- ✓ First results
- ✓ Complexity measurement
- ✓ Conclusion and perspectives

Basicness


Distribution of languages
according to the proportion of basic segments
160
140
120
100
80
60
40
20
0
Number fo systems
32%
0,3
0,4
0,5
0,6
0,7
0,8
0,9
1
% of basic segments in the systems

Generativity for vowels

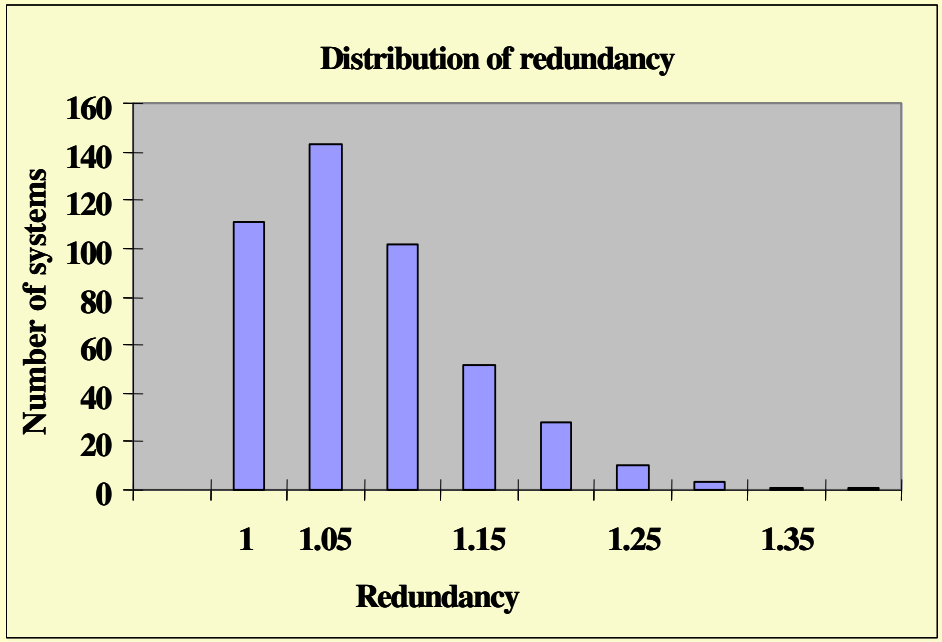
i
a
o
u
e

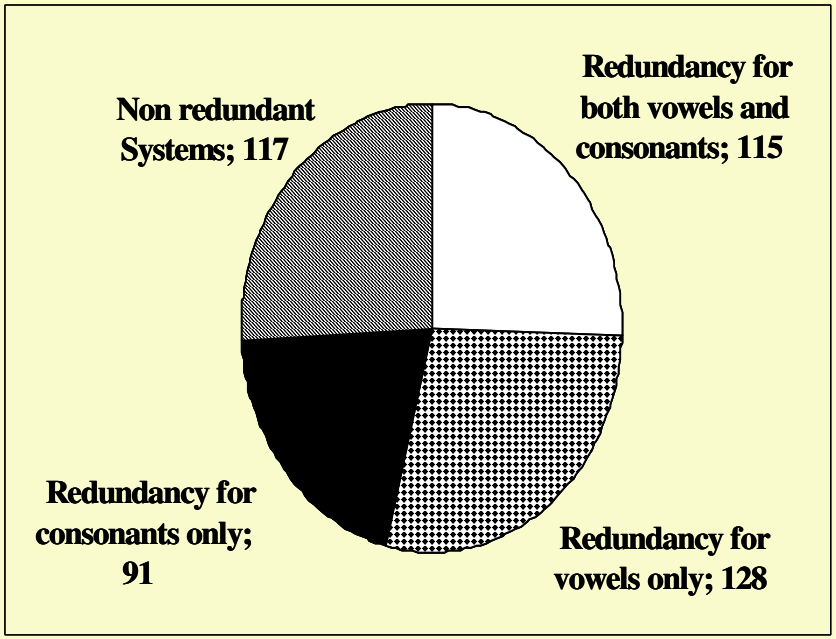
Segments	Generativity	Derivation degree	Frequency (in languages)
voiced high front unrounded	14	0	**0.87**
voiced low central unrounded	14	0	**0.87**
voiced higher-mid back rounded	12	0	**0.69**
voiced high back rounded	11	0	**0.82**
voiced higher-mid front unrounded	10	0	**0.65**
voiced lower-mid back 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

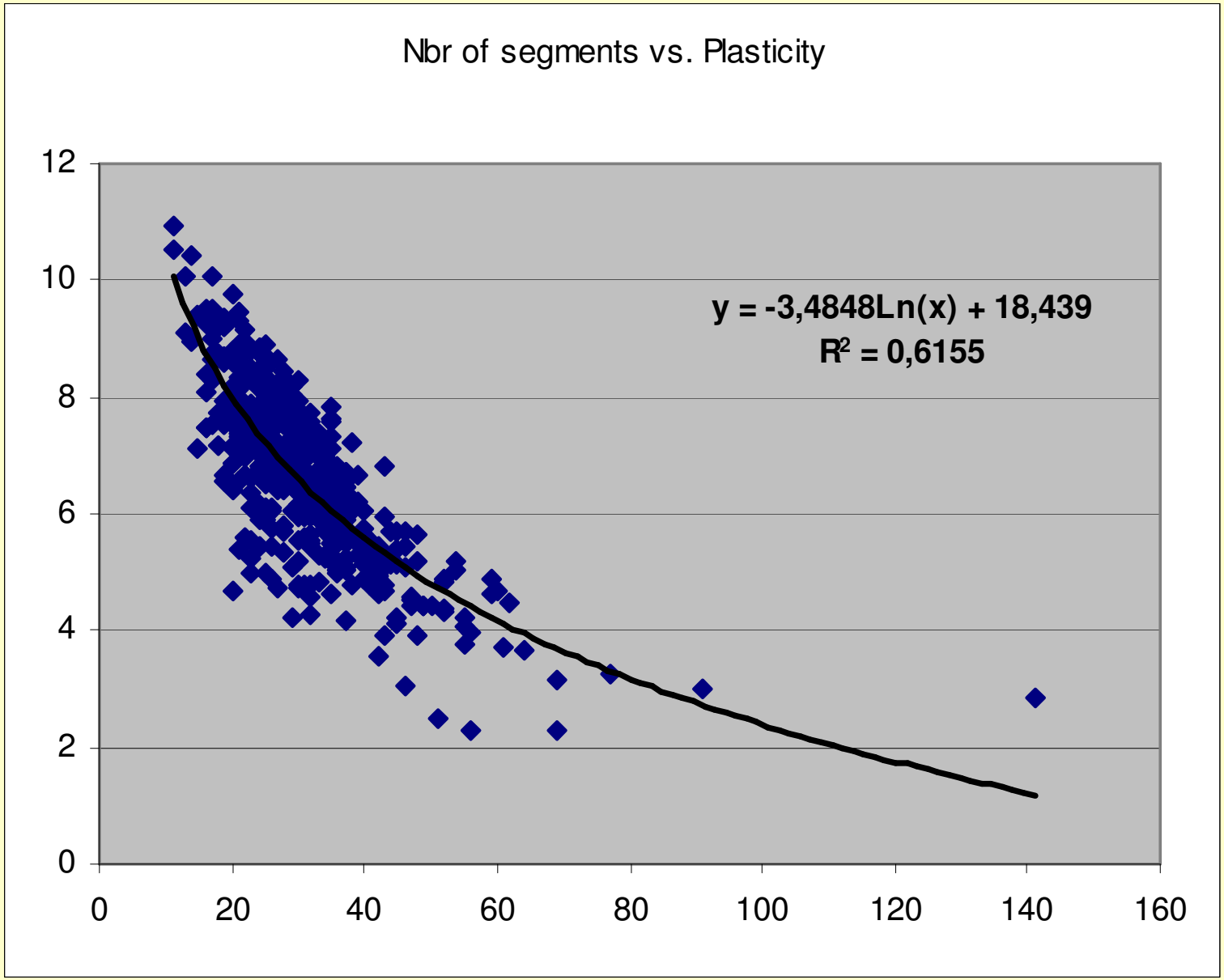
Segments	Generativity	Derivation degree	Frequency (in languages)
voiceless velar stop	18	0	**0.89**
voiceless alveolar stop	14	0	**0.74**
voiceless postalveolar sibilant-affricate	14	0	0.42
voiceless uvular stop	13	0	**0.12**
voiceless bilabial stop	12	0	**0.83**
voiced bilabial stop	11	0	**0.64**
voiced 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**

Redundancy


Distribution of redundancy
Number of systems
160
140
120
100
80
60
40
20
0
1
1.05
1.15
1.25
1.35
Redundancy


Non redundant Systems; 117
Redundancy for both vowels and consonants; 115
Redundancy for consonants only; 91
Redundancy for vowels only; 128

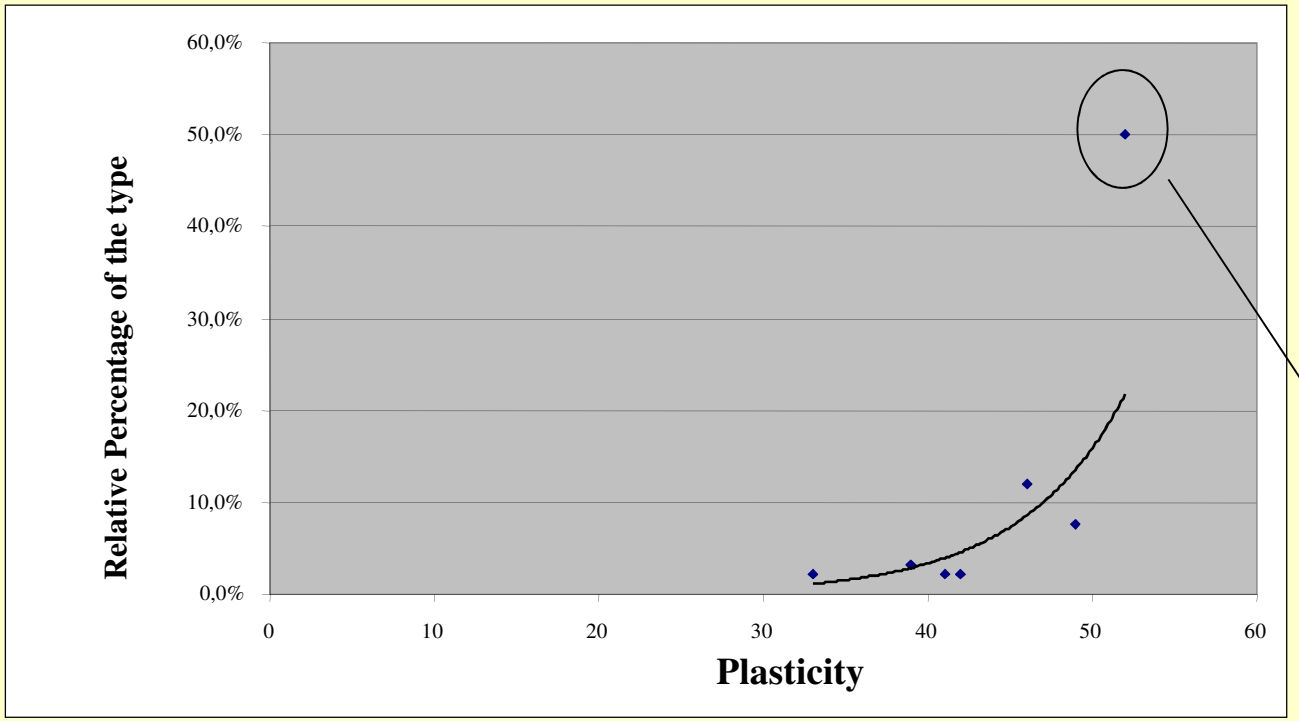
Plasticity


Nbr of segments vs. Plasticity
y = -3,4848Ln(x) + 18,439
R2 = 0,6155
12
10
8
6
4
2
0
0
20
40
60
80
100
120
140
160

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

- ✓ Introduction
- ✓ Presentation of our approach
- ✓ First results
- ✓ Complexity measurement
- ✓ Conclusion and perspectives

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

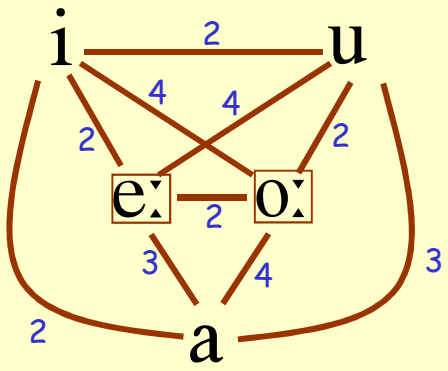
high – higher mid | long – ø | front – back
rounded - unrounded | front – back
rounded – unrounded
high – higher mid | long – ø
nasalized - ø

(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



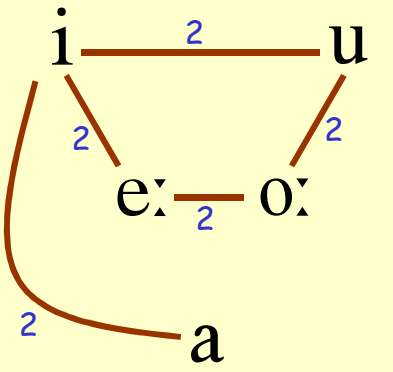
2. Compute shortest paths for all pairs of segments

	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

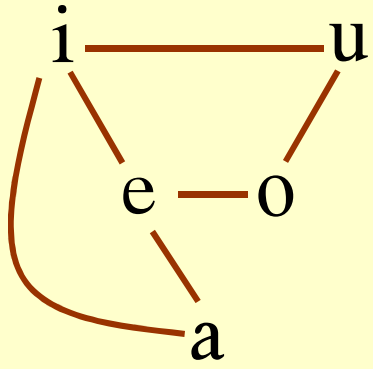
length of a path = maximum distance on this path

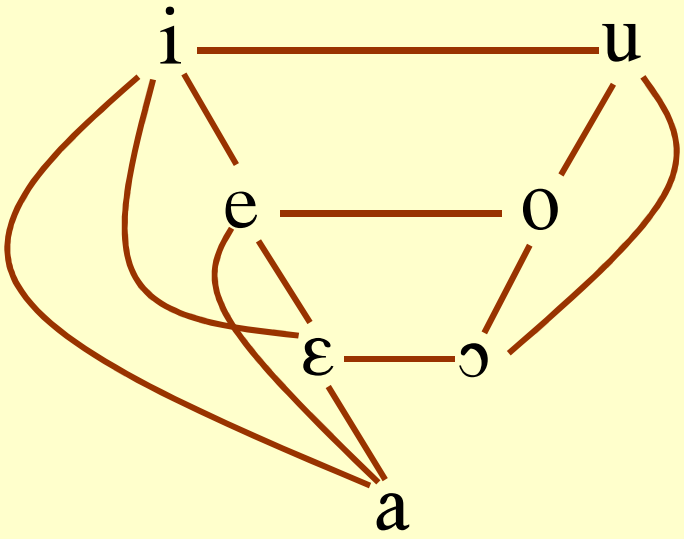
3. For all pairs of segments, remove direct link if it exceeds the length of the shortest path

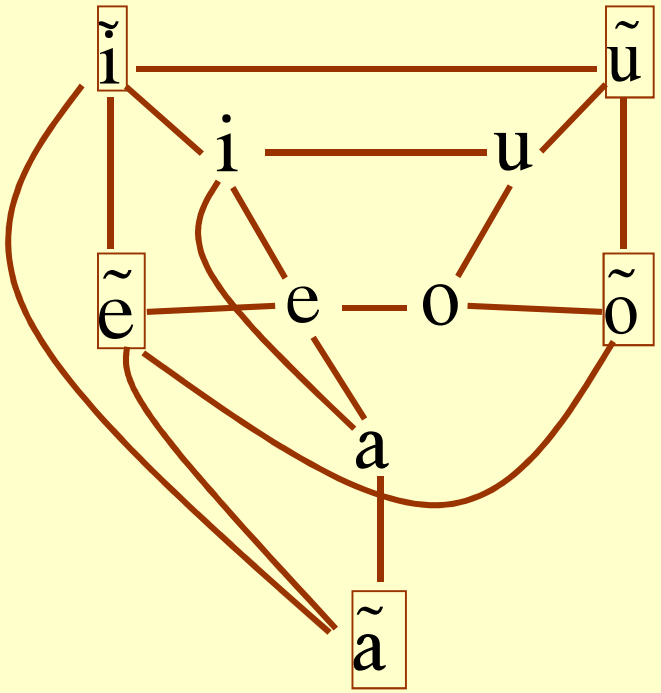
	i	u	e:	o:	a
i	0	2	2		2
u		0		2	
e:			0	2	
o:				0	
a					0

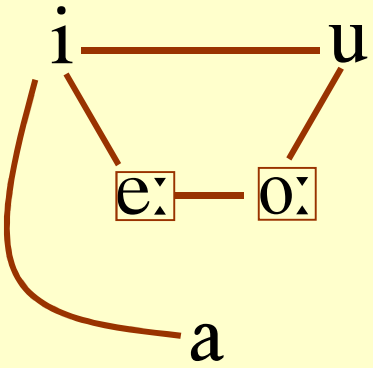


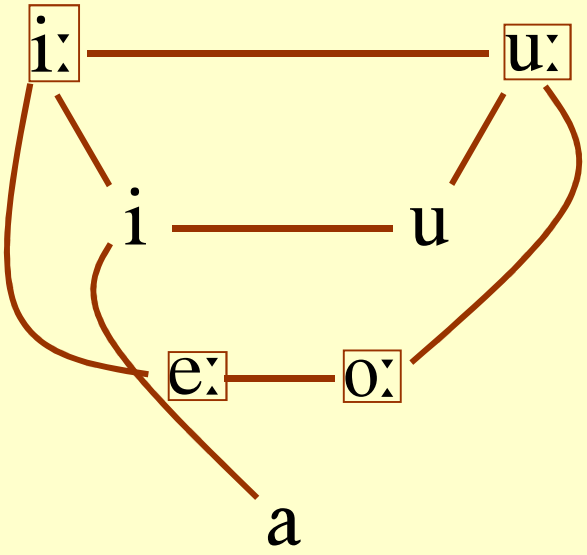
Examples of networks











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):

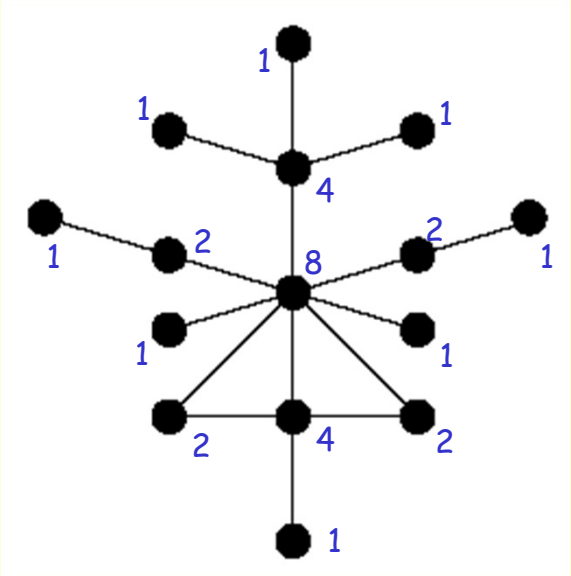
Properties:

- not related to graph size

- sensitive to hierarchical structures

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

Offdiagonal complexity: Examples

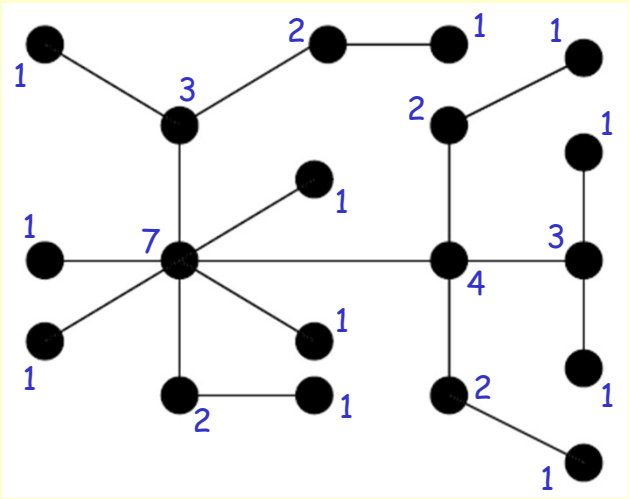


	1	2	4	8	k_2
1	0	2	4	2	2
2		0	2	4	8
4			0	2	6
8				0	0
k_1	C=0.974				

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

An attempt to measure *structural* complexity (only)

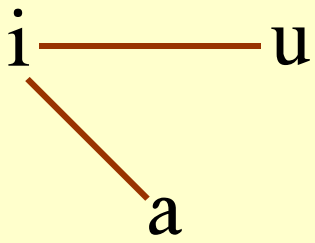
. Different from the overall complexity of a system

Offdiagonal complexity works with non-valued graphs…

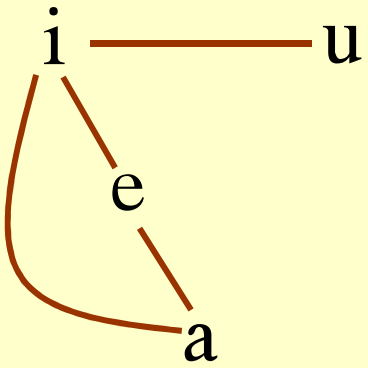
Discard the values of the links of the graph

(and think to something better later…)

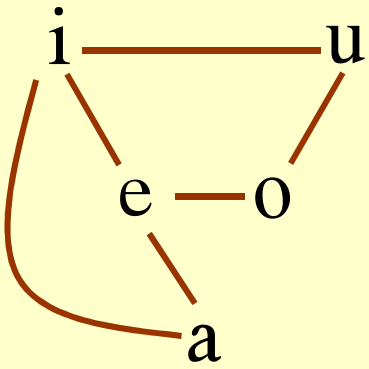
Examples of networks



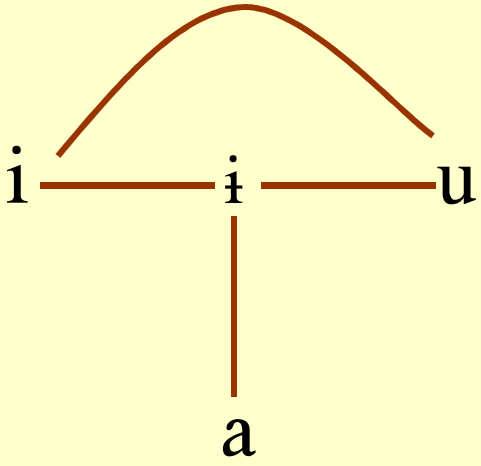
C = 0



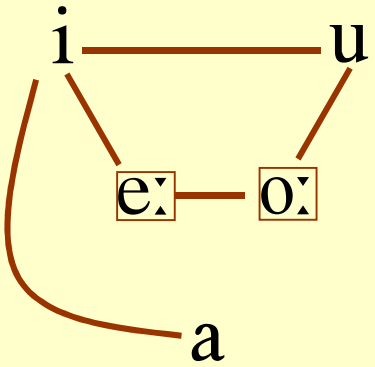
C = 1.04



C = 0.64

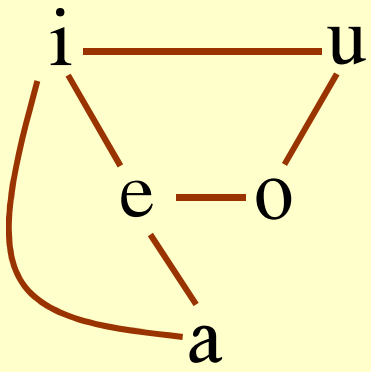


C = 1.04

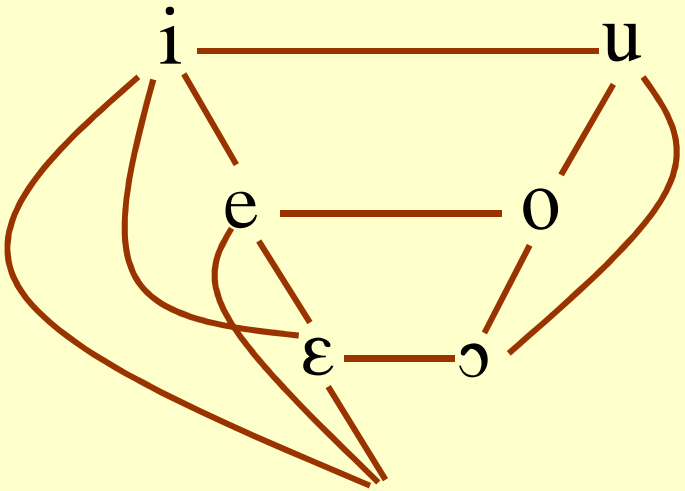


C = 0.69

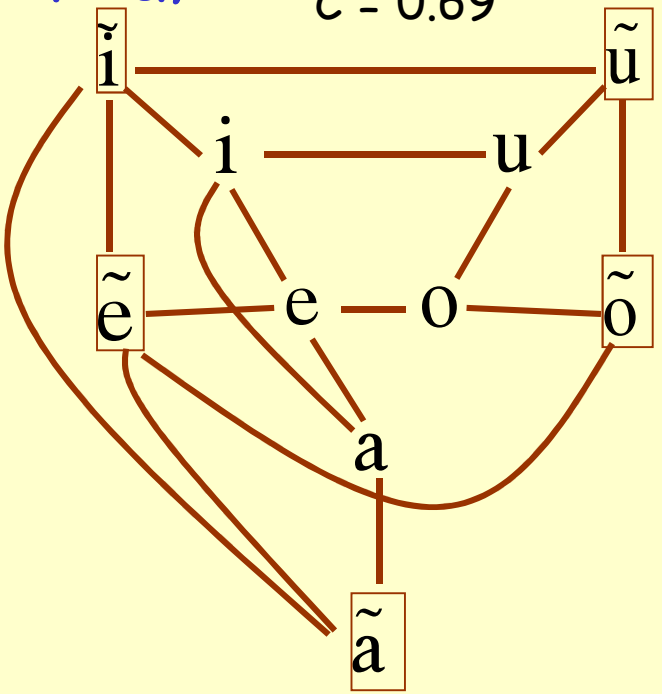
Examples of networks (cont'd)



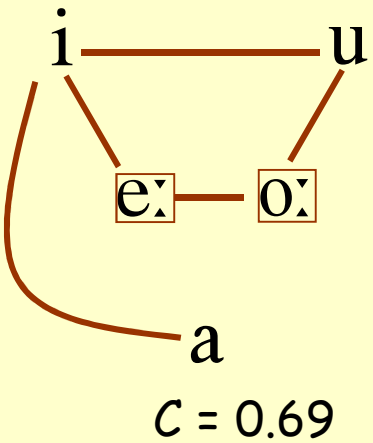
C = 0.64



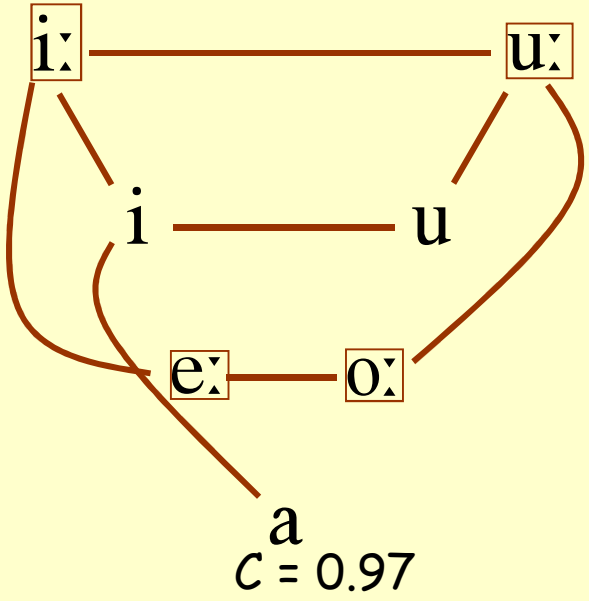
C = 0.69



C = 0.69



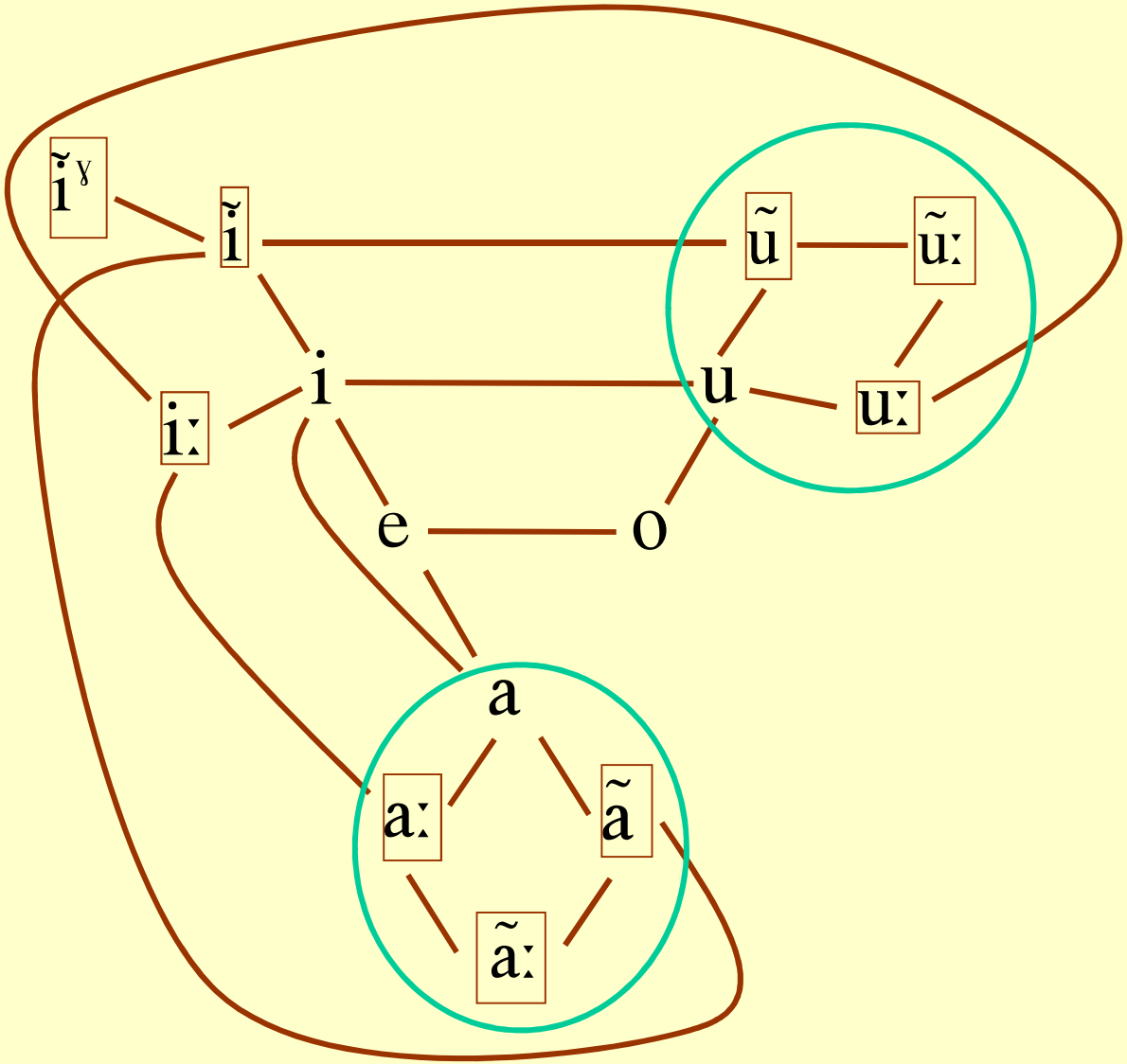
C = 0.69



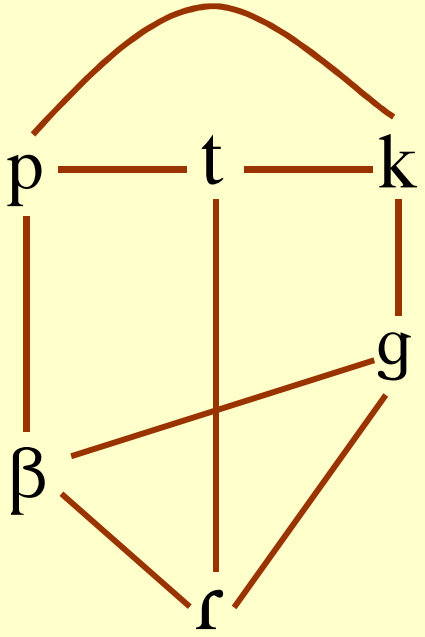
C = 0.97

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)

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

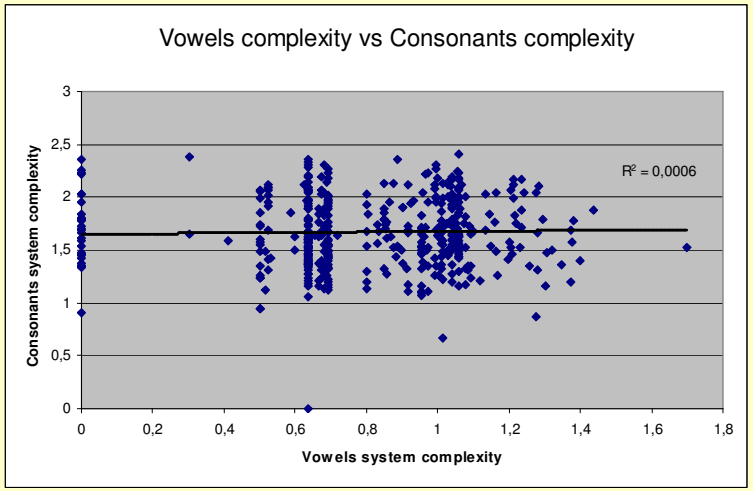
Significantly different: $t(450) = 9.85$, $p << 1$

Average complexity for consonantal systems

(clicks omitted for mental sanity reasons):

$C = 1.63$

No correlation between vocalic and consonantic structural complexity



- ✓ Introduction
- ✓ Presentation of our approach
- ✓ First results
- ✓ Complexity measurement
- ✓ Conclusion and perspectives

Conclusions

- Various indices can be computed to characterize phonological systems
  - ✓ Basicness and generativity
  - ✓ Redundancy and plasticity
  - ✓ Structural complexity

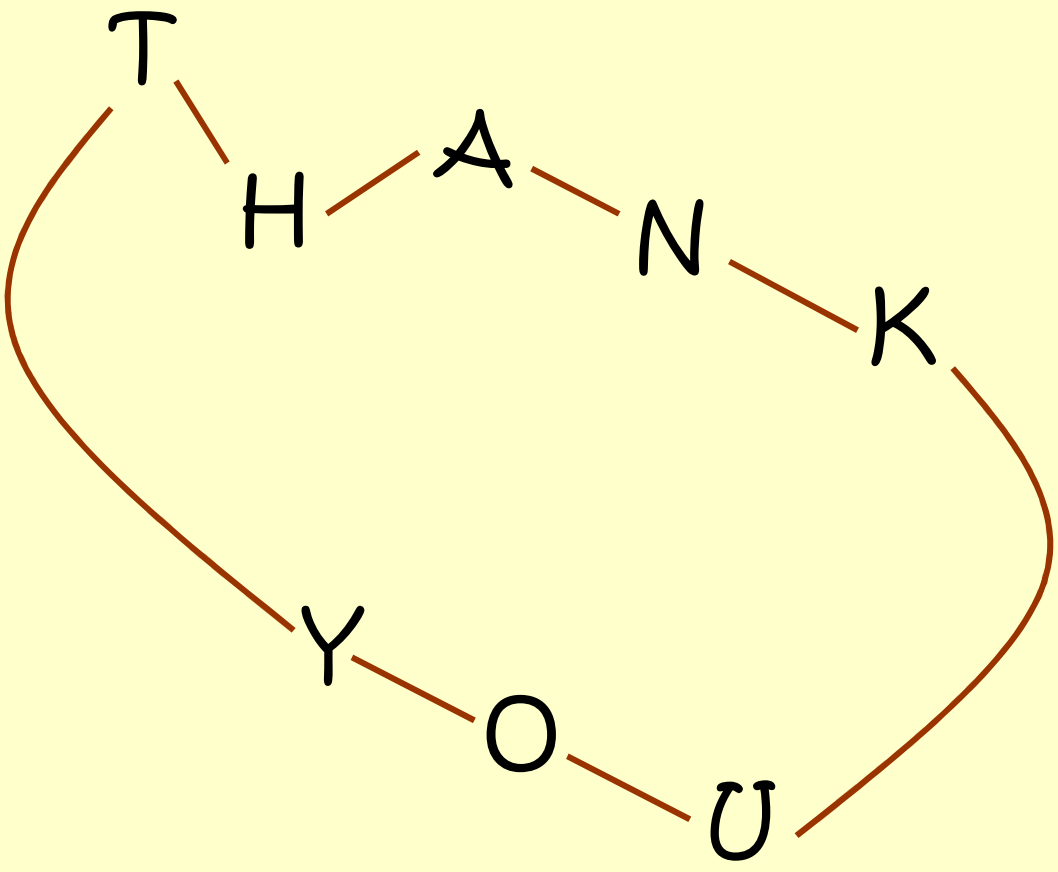
  - ✓ Enough to estimate "coherent" systems?
    - Enough dimensions? Relevant dimensions?
    - Threshold on possible values?

    → Combine structural complexity and previous indices

- Necessity to find the relevant level of description

Perspectives

- Structural complexity
  - ✓ Structural comparison of graphs (with ABSURDIST algorithm)
    → distances between systems
- 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