Form-meaning systematicity and language evolution

Mónica Tamariz
monica@ling.ed.ac.uk

Tübingen, 1 September 2009
Language as an adaptive system

Language structure is influenced by many pressures

- Express my identity!
- Convey varied meanings!
- Be easy to learn!
- Be easy to produce!
- Be easy to understand!
Language as an adaptive system

Language structure is influenced by many pressures

- Express my identity!
- Convey varied meanings!
- Be easy to learn!
- Be easy to produce!
- Be easy to understand!
Overview

1. Form-meaning systematicity

2. Exploring systematicity in Natural Language: A corpus study

3. Systematicity and compositionality in Artificial Languages: An experimental approach

4. Implications for language evolution
1. Systematicity between forms and meanings in language
Systematicity

• The relationship between two spaces is systematic if the structure of one space reflects the structure of the other

• Therefore knowing about one structure provides us with some knowledge about the other
**Systematicity in language**

Phonological forms systematically reflect morphosyntax

<table>
<thead>
<tr>
<th>present</th>
<th>past</th>
<th>3rd person</th>
</tr>
</thead>
<tbody>
<tr>
<td>talk</td>
<td>talked</td>
<td>talks</td>
</tr>
<tr>
<td>play</td>
<td>played</td>
<td>plays</td>
</tr>
<tr>
<td>study</td>
<td>studied</td>
<td>studies</td>
</tr>
</tbody>
</table>

Does it reflect semantics??
A systematic lexicon

Words that **sound similar**
tend to have **similar meanings**
Systematic lexicon – Why?

Systematicity reflects a preference for isomorphic / topographic / systematic representations in the nervous system

(From Tootell, Silverman, Switkes and De Valois, 1982)

(From Penfield and Rasmussen, 1950)
Systematic lexicon – What for?

Functions of systematicity:

• Help learn / understand new items
• Allow creativity and generalisation
2. Exploring systematicity in Natural Language: A corpus study

(Tamariz 2005, 2008)
**Looking for systematicity in the lexicon**

**Hypothesis:** The structure of the mental lexicon shows systematicity between forms and meanings

**Method:** Quantitative corpus study

**Data:** Three subsets of a 1M word Spanish transcribed speech corpus *(Marcos Marin, 1992)*
- 252 CVCV
- 146 CVCCV words of frequency >= 20
- 148 CVCVCV

**Two studies:**
- Syntax and semantics
- Semantics only
Measuring systematicity

Phonological similarity
(Tamariz 2005, 2008)

Distributional similarity
(Landauer & Dumais, 1997; MacDonald, 2000)
Measuring systematicity

Phonological similarity
(Tamariz 2005, 2008)

k-a-s-a v1, v2, v12
c1, c2, c12
k-i-t-a stress1, stress2
stress1, stress2 (syntax)

Distributional similarity
(Landauer & Dumais, 1997;
MacDonald, 2000)

Patterns of cooccurrence with:
394 words (Stx+Sem)
320 content words (Sem)
of freq >=200
Measuring systematicity

### Phonological similarity

<table>
<thead>
<tr>
<th></th>
<th>nene</th>
<th>kita</th>
<th>lima</th>
<th>kasa</th>
</tr>
</thead>
<tbody>
<tr>
<td>nene</td>
<td></td>
<td>0.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>kita</td>
<td>0.01</td>
<td></td>
<td>0.11</td>
<td>0.61</td>
</tr>
<tr>
<td>lima</td>
<td>0.11</td>
<td>0.61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>kasa</td>
<td>0.00</td>
<td>0.54</td>
<td>0.25</td>
<td></td>
</tr>
</tbody>
</table>

### Distributional similarity

<table>
<thead>
<tr>
<th></th>
<th>nene</th>
<th>kita</th>
<th>lima</th>
<th>kasa</th>
</tr>
</thead>
<tbody>
<tr>
<td>nene</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>kita</td>
<td>0.21</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lima</td>
<td>0.32</td>
<td>0.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>kasa</td>
<td>0.40</td>
<td>0.15</td>
<td>0.12</td>
<td></td>
</tr>
</tbody>
</table>

Fisher divergence
FD = 1.4

(significance established with Monte Carlo analysis)
## Result

<table>
<thead>
<tr>
<th></th>
<th>Syntax + Semantics</th>
<th>Semantics only</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FD</td>
<td>sig (p)</td>
</tr>
<tr>
<td>$cvcv$</td>
<td>5.03</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td>$cvccv$</td>
<td>2.18</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>$cvccv$</td>
<td>2.36</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>
Result

There is systematicity between form and meaning in the lexicon

Words that sound similar do tend to have similar meanings
Words that sound similar do tend to have similar meanings.

However, too much systematicity may pose a problem for comprehension!
Systematicity vs. disambiguation
The phonological correlates of systematicity

- The lexicon structure is under pressure
  1. To be systematic (which helps processing and learning)
  2. To avoid ambiguities derived from systematicity

Hypothesis:
The structure of the mental lexicon is shaped by these two opposed pressures
  - Some aspects of forms systematically reflect the structure of word meanings
  - Other aspects of forms should have “negative systematicity” with respect to word meanings
Method

Measure systematicity between

• Distributional similarity and

• Phonological similarity in terms of

  - Consonants
  - Vowels
  - Stress patterns

E.g. Do words that share the first consonant tend to have similar meanings?
Results

The impact of different phonological parameters on systematicity

CVCV words

All values $p<0.01$ except where stated
Results

The impact of different phonological parameters on systematicity

CVCV words

- Consonants have positive impact
- Vowels, negative impact
- Stressed vowel in the penultimate syllable, negative impact
- Other stress, positive impact

All values $p<0.01$ except where stated
Results

The impact of different phonological parameters on systematicity

CVCVVC words

- Consonants have positive impact
- Vowels, mostly negative impact
- Stressed vowel in the penultimate syllable, negative impact
- Other stress, positive impact

All values $p<0.01$ except where stated
Results

The impact of different phonological parameters on systematicity

CVCCV words

• Consonants have mixed impact
• Vowels, mostly negative impact
• Stressed vowel in the penultimate syllable, negative impact
• Other stress, positive impact

All values $p<0.01$ except where stated
Discussion

In the Spanish lexicon...

- The structure of consonants, stress pattern and the stressed final vowel has adapted to the pressure for systematicity
- The structure of vowels, especially the stressed vowel in the penultimate syllable has adapted to the pressure for easy disambiguation
- These results support the hypothesis that the lexicon is an adaptive system, evolving under multiple pressures

*(Other languages may have found different solutions to this conflict)*
3. Systematicity and compositionality in Artificial Languages: An experimental approach
Systematicity and compositionality

• Compositionality is a special type of systematicity between forms and meanings

“In a compositional system, the meaning of an expression is a function of the meaning of its component parts, plus the way the parts are combined together”

Compositional
lo- =
na- =
me- =
-pi =
-to = +

Holistic
lop =
mer =
sus =
pina =
e =
koti =
Compositionality and Expressivity

Compositionality facilitates *emergent categories* and *
*generalisation*

(If the meaning space is structured)

**Compositional**
- lo- = [square]
- na- = [circle]
- me- = [hexagon]
- -pi = [bullet]
- -to = [+]
- Extra meanings:
  1 = outside shape
  2 = insert
- Generalisation possible

**Holistic**
- lop = [bullet]
- mer = [+]
- sus = [bullet]
- pina = [+]
- e = [bullet]
- koti = [+]
- No extra meanings
- Generalisation not possible
Why is language compositional?

• The structure of language evolved over thousands of human generations who learn it and then use it.

• We assume that initially language was not compositional \((e.g. \ Wray, \ 2002)\).

• How did it come to have this structure?
  
  – The role of individual learning
  
  – The role of cultural transmission
The role of individual learning  
(Tamariz & Smith 2008)

- **Hypothesis**
  - There is a *learning bias* favouring compositional mappings between signals and meanings

- **Prediction**
  - More compositional languages are easier to learn

- **Test**
  - Artificial Language Learning experiment
  - Construct artificial languages with different levels of compositionality and see which are better learned
Artificial languages

MEANINGS

- ○
- □
- □
- □
- ○

FORMS

- 3-syllable pseudo words

27 different meanings

27 different forms
Procedure

Training

- hunelene
- letupi
- miko
- lekono

27 screens x 3 times

Testing

1. Signals
   Task: Type name

2. Meanings
   Task: Select picture
Variables

• Independent variable
  4 levels of input language compositionality

<table>
<thead>
<tr>
<th></th>
<th>LANG 1</th>
<th>LANG 2</th>
<th>LANG 3</th>
<th>LANG 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>RegMap</td>
<td>0.154</td>
<td>0.456</td>
<td>0.754</td>
<td>1.00</td>
</tr>
</tbody>
</table>

• Dependent variable
  Learnability
  – Measuring how well the compositional language structure was learned
RegMap measures the confidence that a signal and a meaning are unambiguously and consistently associated.

RegMap\( (X \mid Y) \) = \( 1 - \frac{H(X \mid Y)}{\log(n_x)} \)

### Co-occur

<table>
<thead>
<tr>
<th>Co-occur</th>
<th>Freq</th>
</tr>
</thead>
<tbody>
<tr>
<td>le -</td>
<td>3</td>
</tr>
<tr>
<td>miko</td>
<td>2</td>
</tr>
<tr>
<td>- ne</td>
<td>2</td>
</tr>
<tr>
<td>- ko</td>
<td>2</td>
</tr>
</tbody>
</table>
Results

Learnability of the compositional language structure

Similarity of compositional language structure of input and output languages
Results and discussion

More compositional language structures are learnt better.

Participants seem to have a learning bias for compositionality.

Individual learning may have played a role in language becoming compositional.
Results and discussion

Learnability of the compositional language structure

More compositional language structures are learnt better

Participants seem to have a learning bias for compositionality

Individual learning may have played a role in language becoming compositional

But where does this bias come from?
The role of enculturation

*(Brown & Tamariz, 2008)*

- **Hypothesis**
  - The learning preference for compositional structure is given by cultural knowledge such as literacy

- **Prediction**
  - Literates are better at learning compositional languages than illiterates

- **Test**
  - *Musical* Artificial Language Learning experiment
  - Construct artificial *musical* languages with different levels of compositionality and see which are better learned
Procedure

Training

Testing

Signals

Task: Sing tune

27 screens x 3 times

27 signals presented
Variables

• **Independent variables**
  – 3 levels of input language compositionality
    
    \[
    \begin{array}{ccc}
    \text{LANG 1} & \text{LANG 2} & \text{LANG 3} \\
    \text{RegMap} & 0.004 & 0.66 & 1.00
    \end{array}
    \]
  – 2 levels of musical literacy (literate/illiterate)

• **Dependent variable**
  – Learnability of the compositional structure of languages
Results

Learnability of the compositional language structure

It seems that literacy enhances the preference for learning compositional systems
Discussion:
The role of individual learning

• These results indicate that individual learning may increase language compositionality

• This effect may be enhanced by cultural knowledge such as literacy

• But language evolved as it was transmitted over many generations...
The role of cultural transmission
(Kirby, Cornish & Smith, 2008)

• Hypothesis
  – Cultural transmission contributes to the emergence of systematic structure in language

• Prediction
  – The effects of individual learning are enhanced in iterated learning

• Test
  – Iterated Artificial Language Learning experiment
  – Measure systematicity and learnability
Languages and procedure

MEANINGS

27 different meanings

Generation 0: random signals
- kimako
- kanige
- komako
- winige

Generation 1
- koni
- kuni
• One of the initial random languages

Results

- huhunigu kemuniwa kihupo
- wakiki pokihoku waguhiuki
- nihu niguki koni

muwapo powa hukinimu kekewa konihuke kopo
wako hukeko pohumu huwa hukike ponikiko
muko kokegu kimu kowagu kokihuko kiwanike
Results

- The same language some generations later

<table>
<thead>
<tr>
<th></th>
<th>wuneko</th>
<th>huneko</th>
<th>puneko</th>
</tr>
</thead>
<tbody>
<tr>
<td>wineko</td>
<td>huneko</td>
<td>poneko</td>
<td></td>
</tr>
<tr>
<td>wikeko</td>
<td>kuneko</td>
<td>poneko</td>
<td></td>
</tr>
<tr>
<td>winukuki</td>
<td>hunekuki</td>
<td>ponekuki</td>
<td></td>
</tr>
<tr>
<td>winekuki</td>
<td>kunukuki</td>
<td>ponekuki</td>
<td></td>
</tr>
<tr>
<td>wikekuki</td>
<td>kunekuki</td>
<td>ponekuki</td>
<td></td>
</tr>
<tr>
<td>winekiko</td>
<td>kunekiko</td>
<td>puniko</td>
<td></td>
</tr>
<tr>
<td>winekiko</td>
<td>kunkiko</td>
<td>pokiko</td>
<td></td>
</tr>
<tr>
<td>wikiko</td>
<td>kunekiko</td>
<td>pokiko</td>
<td></td>
</tr>
</tbody>
</table>
• Languages are increasingly systematic and learnable
Results

- Languages are increasingly systematic and learnable
- Compositionality is apparent, but it is not quantified
Quantifying compositionality
(Cornish, Tamariz & Kirby, in press)

• For each generation, calculate RegMap of all pairs of meaning and form dimensions

E.g. for Generation 2:

<table>
<thead>
<tr>
<th>Form</th>
<th>Meaning</th>
<th>Shape</th>
<th>Colour</th>
<th>Motion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Segment 1</td>
<td></td>
<td>0.154</td>
<td>0.884</td>
<td>0.095</td>
</tr>
<tr>
<td>Segment 2</td>
<td></td>
<td>0.365</td>
<td>0.250</td>
<td>0.216</td>
</tr>
<tr>
<td>Segment 3</td>
<td></td>
<td>0.313</td>
<td>0.366</td>
<td>0.452</td>
</tr>
</tbody>
</table>
Results

Segment 1

Emergence of compositionality over cultural transmission
Results

Emergence of compositionality over cultural transmission
Results

Emergence of compositionality over cultural transmission
Discussion

• We have observed a preference for individuals to learn systematic languages

• That preference may be enhanced by cultural institutions such as literacy and teaching

• Cultural transmission can mediate the accumulation of language systematicity over time
Conclusion

• Many pressures, sometimes conflicting, have left their mark on language structure over its long evolutionary history.

• The balance of systematicity in language is an adaptation to many factors
  
  – Neural constraints (preference for isomorphic, systematic mappings)

  – Learned biases (living in a structured, cultural world, we learn to expect structure)

  – Language function (the need to produce and understand flexibly)

  – Cultural transmission over generations
References


• Tamariz (in press). Could arbitrary imitation and pattern completion have bootstrapped human linguistic communication? *Interaction Studies*.


Form-meaning systematicity and language evolution

Mónica Tamariz
monica@ling.ed.ac.uk