Why sentences are more complex than words

Jeffrey Heinz\textsuperscript{1} \quad William Idsardi\textsuperscript{2}

\textsuperscript{1}heinz@udel.edu
University of Delaware

\textsuperscript{2}idsardi@umd.edu
University of Maryland

Parallel Domains
University of Southern California
May 6, 2011
Is phonology different from syntax?

Jean-Roger Vergnaud

No
Is phonology different from syntax?

Jean-Roger Vergnaud
No

Morris Halle
Yes (Bromberger and Halle 1989)
Is phonology different from syntax?

Jean-Roger Vergnaud
No

Morris Halle
Yes (Bromberger and Halle 1989)
Is phonology different from syntax?

Jean-Roger Vergnaud
No

Morris Halle
Yes (Bromberger and Halle 1989)

Elan Dresher, p.c.

*If two things are different, make them similar. If they are similar make them the same.*
This talk

There is an important **computational** difference between phonology and syntax that requires explanation.

**Hypothesis**
Humans make different kinds of generalizations over words than they do over sentences and this explains this difference.

**Linguistics and Cognitive Science**
We suggest this difference can play a key role in larger debates in cognitive science between domain-general and domain-specific learning.
Phonology \parallel Syntax

Formal Learning Theories

Conclusion
Strings

Strings are sequences of more basic units.

Sentences are sequences of morphemes.

John laughed while Mary talked.

Words are sequences of sounds.

b l i ñ
Language Patterns

Language patterns are sets of strings, or relations among strings.

No coda: *Coda

- \{a, ka, ta, pi.kou, ba.du.pi\} ⊂ *Coda
- \{bliŋ, męp.ka, karp\} ∩ *Coda = ∅
Language Patterns

Language patterns are sets of strings, or relations among strings.

Word final obstruent devoicing: $R = [-\text{son}] \rightarrow [-\text{voice}] / #$

- $\{\text{pad} \rightarrow \text{pat}, \, \text{pat} \rightarrow \text{pat}, \, \text{pabau} \rightarrow \text{pabax}\} \subset R$
- $\{\text{pad} \rightarrow \text{pad}, \, \text{pad} \rightarrow \text{dap}\} \cap R = \emptyset$
Language Patterns

Language patterns are sets of strings, or relations among strings.

Conjunction: $S \rightarrow S$ and $S$

- $\{\text{John swam and Mary laughed, They talked and they talked}\} \subset S$
- $\{\text{John swam and Mary, They talked and they}\} \cap S = \emptyset$
Language Patterns

Language patterns are sets of strings, or relations among strings.

Conjunction: $S \rightarrow S$ and $S$

- $\{\text{John swam and Mary laughed, They talked and they talked}\} \subset S$
- $\{\text{John swam and Mary, They talked and they talked}\} \cap S = \emptyset$

What kinds of sets and relations are natural language patterns?
The Chomsky Hierarchy
The Chomsky Hierarchy and natural language patterns

- Chumash sibilant harmony
  Applegate 1972
- English nested embedding
  Chomsky 1957
- Swiss German
  Shieber 1985
- Yoruba copying
  Kobele 2006
- English consonant clusters
  Clements and Keyser 1983
- Kwakiutl stress
  Bach 1975

Finite → Regular → Context-Free → Mildly Context-Sensitive → Context-Sensitive
The Chomsky Hierarchy and natural language patterns

- Chumash sibilant harmony
  Applegate 1972

- English nested embedding
  Chomsky 1957

- Swiss German
  Shieber 1985

- Yoruba copying
  Kobele 2006

- English consonant clusters
  Clements and Keyser 1983

- Kwakiutl stress
  Bach 1975
The Chomsky Hierarchy and natural language patterns

- Chumash sibilant harmony
  Applegate 1972
- English nested embedding
  Chomsky 1957
- Swiss German
  Shieber 1985
- Yoruba copying
  Kobele 2006

- English consonant clusters
  Clements and Keyser 1983
- Kwakiutl stress
  Bach 1975

Finite
Regular
Context-Free
Mildly Context-Sensitive
Context-Sensitive
Phonology is regular (Kaplan and Kay 1994)

\[ F_1 \times F_2 \times \cdots \times F_n = P \]

1. Optional, left-to-right, right-to-left, and simultaneous application of rules \( A \rightarrow B / C \quad D \) (where \( A, B, C, D \) are regular expressions) describe regular relations, provided the rule cannot reapply to the locus of its structural change.

2. Rule ordering is functional composition (finite-state transducer composition).

3. Regular relations are closed under composition.

4. SPE grammars (finitely many ordered rewrite rules of the above type) can describe virtually all phonological patterns.

5. Therefore, phonology is regular (both \( F_i \) and \( P \)).
Phonology is regular (Kaplan and Kay 1994)

\[ F_1 \times F_2 \times \cdots \times F_n = P \]

1. Optional, left-to-right, right-to-left, and simultaneous application of rules \( A \rightarrow B / C \quad D \) (where \( A,B,C,D \) are regular expressions) describe regular relations, provided the rule cannot reapply to the locus of its structural change.

2. Rule ordering is functional composition (finite-state transducer composition).

3. Regular relations are closed under composition.

4. SPE grammars (finitely many ordered rewrite rules of the above type) can describe virtually all phonological patterns.

5. Therefore, phonology is regular (both \( F_i \) and \( P \)).
Phonology is regular (Kaplan and Kay 1994)

\[
F_1 \times F_2 \times \cdots \times F_n = P
\]

1. Optional, left-to-right, right-to-left, and simultaneous application of rules \( A \rightarrow B / C \underbrace{\ldots \rightarrow} D \) (where \( A,B,C,D \) are regular expressions) *describe regular relations*, provided the rule cannot reapply to the locus of its structural change.

2. Rule ordering is functional composition (finite-state transducer composition).

3. Regular relations are closed under composition.

4. SPE grammars (finitely many ordered rewrite rules of the above type) can describe virtually all phonological patterns.

5. Therefore, phonology is regular (both \( F_i \) and \( P \)).
Phonology is regular (Kaplan and Kay 1994)

\[ F_1 \times F_2 \times \cdots \times F_n = P \]

1. Optional, left-to-right, right-to-left, and simultaneous application of rules \( A \rightarrow B / C \quad D \) (where \( A, B, C, D \) are regular expressions) describe regular relations, provided the rule cannot reapply to the locus of its structural change.

2. Rule ordering is functional composition (finite-state transducer composition).

3. Regular relations are closed under composition.

4. SPE grammars (finitely many ordered rewrite rules of the above type) can describe virtually all phonological patterns.

5. Therefore, phonology is regular (both \( F_i \) and \( P \)).
Phonology is regular (Kaplan and Kay 1994)

$$F_1 \times F_2 \times \cdots \times F_n = P$$

1. Optional, left-to-right, right-to-left, and simultaneous application of rules $A \rightarrow B / C \quad D$ (where $A,B,C,D$ are regular expressions) describe regular relations, provided the rule cannot reapply to the locus of its structural change.
2. Rule ordering is functional composition (finite-state transducer composition).
3. Regular relations are closed under composition.
4. SPE grammars (finitely many ordered rewrite rules of the above type) can describe virtually all phonological patterns.
5. Therefore, phonology is regular (both $F_i$ and $P$).
Phonology is regular (Kaplan and Kay 1994)

\[ F_1 \times F_2 \times \cdots \times F_n = P \]

1. Optional, left-to-right, right-to-left, and simultaneous application of rules \( A \rightarrow B / C \longrightarrow D \) (where \( A, B, C, D \) are regular expressions) describe regular relations, provided the rule cannot reapply to the locus of its structural change.

2. Rule ordering is functional composition (finite-state transducer composition).

3. Regular relations are closed under composition.

4. SPE grammars (finitely many ordered rewrite rules of the above type) can describe virtually all phonological patterns.

5. Therefore, phonology is regular (both \( F_i \) and \( P \)).
What about reduplication?

- It’s morpho-syntax (Inkelas and Zoll 2000, Roark and Sproat 2007).
Phonology is subregular

Proper inclusion relationships among subregular language classes (indicated from top to bottom).

TSL Tier-based Strictly Local
LTT Locally Threshold Testable
LT Locally Testable
SL Strictly Local
SP Strictly Piecewise

PT Piecewise Testable

Phonology is subregular

Proper inclusion relationships among subregular language classes (indicated from top to bottom).

- TSL Tier-based Strictly Local
- LTT Locally Threshold Testable
- LT Locally Testable
- SL Strictly Local
- SP Strictly Piecewise
- PT Piecewise Testable

The Chomsky Hierarchy and natural language patterns

Chumash sibilant harmony
Applegate 1972

English nested embedding
Chomsky 1957

Swiss German
Shieber 1985

Yoruba copying
Kobele 2006

So why the difference?
The problem of induction and generalization

Philosophy
(Plato, Aristotle, Hume, Mill, Russell, Carnap, Quine, Goodman, ...)

Linguistics

Computer Science

So how can language patterns be learned?
Define “Learning”

Figure: Learners are functions $\phi$ from experience to languages.
Results: Do feasible learners exist?

1. Identification in the limit from positive data (Gold 1967)
2. Identification in the limit from positive and negative data (Gold 1967)
3. Identification in the limit from positive data from r.e. texts (Gold 1967)
4. Learning context-free and r.e. distributions (Horning 1969, Angluin 1988, Chater and Vitanyi 2007)
### Results: Do feasible learners exist?

1. **Identification in the limit from positive data** (Gold 1967)
2. **Identification in the limit from positive and negative data** (Gold 1967)
3. **Identification in the limit from positive data from r.e. texts** (Gold 1967)
4. **Learning context-free and r.e. distributions** (Horning 1969, Angluin 1988, Chater and Vitanyi 2007)
Results: Do feasible learners exist?

1. Identification in the limit from positive data (Gold 1967)
2. Identification in the limit from positive and negative data (Gold 1967)
3. Identification in the limit from positive data from r.e. texts (Gold 1967)
4. Learning context-free and r.e. distributions (Horning 1969, Angluin 1988, Chater and Vitanyi 2007)
Results: Do feasible learners exist?

1. Identification in the limit from positive data (Gold 1967)
2. Identification in the limit from positive and negative data (Gold 1967)
3. Identification in the limit from positive data from r.e. texts (Gold 1967)
4. Learning context-free and r.e. distributions (Horning 1969, Angluin 1988, Chater and Vitanyi 2007)
Results: Do feasible learners exist?

1. Identification in the limit from positive data (Gold 1967)
2. Identification in the limit from positive and negative data (Gold 1967)
3. Identification in the limit from positive data from r.e. texts (Gold 1967)
4. Learning context-free and r.e. distributions (Horning 1969, Angluin 1988, Chater and Vitanyí 2007)
Positive Results

Many classes which cross-cut the Chomsky hierarchy and exclude some finite languages are feasibly learnable in the senses discussed (and others).

Lessons from formal learning theories

Learning requires a structured hypothesis space, which excludes at least some finite-list hypotheses.

Gleitman 1990, p. 12:

‘The trouble is that an observer who notices everything can learn nothing for there is no end of categories known and constructable to describe a situation [emphasis in original].’
Lessons from formal learning theories

Learning requires a structured hypothesis space, which excludes at least some finite-list hypotheses.

Gleitman 1990, p. 12:

‘The trouble is that an observer who notices everything can learn nothing for there is no end of categories known and constructable to describe a situation [emphasis in original].’
Hypothesis spaces for language learning

- Chumash sibilant harmony
  - Applegate 1972
- English nested embedding
  - Chomsky 1957
- Swiss German
  - Shieber 1985
- Yoruba copying
  - Kobele 2006

- English consonant clusters
  - Clements and Keyser 1983
- Kwakiutl stress
  - Bach 1975
Strategy #1: learn everything (e.g. Chater and Vitányi 2007)

Problems

1. Possible in principle, not feasible in practice
2. Predicts any pattern is possible with sufficient data
Strategy #2: Single hypothesis space for language

(e.g. Clark 2010)

Problems

1. Predicts syntactic patterns ought to be found within words.
Strategy #3: Distinct hypothesis spaces for phonology and syntax

1. The complexity differential between phonology and syntax can be explained if language-learning itself is *modular.*

2. People make *different* kinds of generalizations over words than they do over sentences.
Strategy #3 accords with recent research within linguistics

- Recent computational models for learning phonology are successful in part because the generalization strategies employed do not consider every finite pattern nor do they extend beyond the regular boundary (Hayes and Wilson 2008, Albright 2009, Heinz 2010, Goldsmith and Riggle to appear...).

- Likewise, the learners for syntax are successful in part because the learners’ generalizations are constrained to the right, non-superfinite classes of nonregular patterns (Yang 2000, et seq., Clark and Eryaud 2007, Yoshinaka and Clark 2010, Becerra-Bonache et al. 2010, ...).
Advocates of general purpose learners

Challenges

1. They must present a single learner capable of learning phonological and syntactic patterns from reasonably-sized sets of words and sentences, respectively (to our knowledge no such demonstration exists).

2. They must also either offer an explanation for the complexity differential or deny it.
One possibility: articulatory/perceptual grounding

Hypothesis
Sound sequences within words are constrained by psychophysical properties of the human nervous, motor, and auditory systems in ways that word sequences within sentences are not.
Long-distance patterns in phonology


Samala Chumash (Applegate 1972)

\[\text{stoyonowanowas}\] ‘3s stood upright’

*\text{stoyonowanowas}

*\text{stoyonowanowas}

Long distance disagreement (Suzuki 1998)

Grassman’s Law

\[t^h\text{rík-s}\] ‘hair’

\[t^h\text{rík}-es\] ‘hairs’

*\[t^h\text{rík}^h\text{-es}\]

Latin Liquid dissimilation (Jensen 1974, Odden 1994)

\[\text{nav-alis}\] ‘naval’

\[\text{lun-aris}\] ‘lunar’

\[\text{flor-alis}\] ‘floral’

*\[\text{flor-aris}\]
Is “long-distance” the right generalization?

Perhaps all long distance cases can be reduced to chained instances of strictly local generalizations.

1. Research exists which examines to what extent intermediary sounds in long-distance assimilation patterns are truly transparent and finds in many instances that the posture of the relevant articulator is maintained throughout pronunciation (Gafos 1996, Ní Chiosáin & Padgett 1997, Gordon 1999, Gafos and Benus 2003, Walker et al. 2009)

2. On the other hand, in Guaraní nasal harmony, research also exists which confirms the oral obstruent realization for voiceless stops that act transparent (Walker 1998).

3. What about the dissimilation cases?
If they deny the complexity differential... 

We expect to find syntactic patterns in phonology.

1. Nested embedding patterns in phonological words

```
C V C . C V C
```

```
C V . C V V . C V C
```

2. Multiple crossing dependencies in phonological words

```
C V . C V
```

```
```
Testable Hypothesis

Artificial Language Learning Experiments

Conclusion

There are substantial similarities between phonology and syntax.

1. Both are generative.

2. Both are richly structured domains which subsequently limit the cross-linguistic variation.

But there is a significant difference.

1. Phonological patterns can be described with regular grammars, but syntactic patterns cannot.

2. The hypothesis that language-learning itself is modularized currently offers the best explanation for this fact.
Conclusion

There are substantial similarities between phonology and syntax.

1. Both are generative.
2. Both are richly structured domains which subsequently limit the cross-linguistic variation.

But there is a significant difference.

1. Phonological patterns can be described with regular grammars, but syntactic patterns cannot.
2. The hypothesis that language-learning itself is modularized currently offers the best explanation for this fact.

Thank You.