SVO languages and pro-drop

How pro-drop affects learnability
Emergence of Language Structures Workshop

Ezra Van Everbroeck, Gary Cottrell, Maria Polinsky
UC San Diego
ezra@ucsd.edu, gary@cs.ucsd.edu, mpolinsky@ucsd.edu

Goal

• Determine the impact of pro-drop on languages with basic SVO word order.
  
  Why?
  Why pro-drop?
  Why SVO?

Method

• Compare how well computer models can learn each type of SVO language, with or without pro-drop.

• Assess simulation results against cross-linguistic data.

Assumptions

• Cross-linguistic distribution: A language type may be unattested because it is unlearnable.

• Learnability: The acquisition of who did what to whom structures is a crucial test of learnability.

Outline

• Introduction
• Linguistic parameters
• Connectionist model
• Network results
• Crosslinguistic comparison
• Conclusion

Who did what to whom?

• Word order
• Dependent-marking (= case)
• Head-marking (= verb marking)
• Tense/Aspect/Modality markers:

  E.g. payum narma ŋ
  man-PL woman-SG 3PL-3SG-see
  "The woman saw the men."

  E.g. payum narma ŋ
  man-PL woman-SG 3PL-3SG-see
  "The woman saw the men."

  E.g. payum narma ŋ
  man-PL woman-SG 3PL-3SG-see
  "The woman saw the men."

  E.g. payum narma ŋ
  man-PL woman-SG 3PL-3SG-see
  "The woman saw the men."
Pro-drop

- In many languages, subjects can be left unexpressed.

  E.g. Ø Estoy comiendo
      'I am eating.'
  Ø E.stando comiendo
      'You are eating.'
  Ø Estás comiendo
      'He/She is eating.'

  This is 'rich verb agreement' pro-drop.

Pro-drop

- In many languages, subjects can be left unexpressed.

  E.g. Ø Estoy comiendo
      'I am eating.'
  Ø Estando comiendo
      'He/She orders him/her to use a knife.'
  Ø (Li & Thompson 1981; Huang 1989)

  This is discourse (topic) pro-drop.

Not real pro-drop

- Common: expletives
  E.g. Ø está lloviendo
       'It is raining.'
  Ø seems like we have a problem

- Common: ‘diary-drop’
  E.g. Ø Gone fishing
       'Want to play?'

Not real pro-drop

- Common: expletives
  E.g. Ø está lloviendo
       'It is raining.'
  Ø seems like we have a problem

- Common: ‘diary-drop’
  E.g. Ø Gone fishing
       'Want to play?'

Parameter Space

<table>
<thead>
<tr>
<th>Head-marking</th>
<th>-pro</th>
<th>-case</th>
<th>+T/A/M</th>
<th>+Agr</th>
</tr>
</thead>
<tbody>
<tr>
<td>-case</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+case</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+pro</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-case</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+case</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Artificial languages

- A context-free grammar generates (in)transitive sentences

  s -> subj verb_i.
  subj -> noun_animate + affix_subject
  verb_i -> verb_intransitive + affix_verb-subject
  noun_animate -> one of many lexical forms
  verb_intransitive -> idem
  affix_subject -> a unique suffix
  affix_verb-subject -> idem

- Lexicon: 300 nouns, 100 verbs, 8 agreement markers, 1 T/A/M marker, 2 case markers

Experiment design

- Generate artificial languages for each language type.

- Train 20 networks on 3,000 sentences for each type; test on 3,000 new ones.

- A sentence is parsed correctly if all words are in right slot (S, V or O).
Pro-drop implementation

Based on averaged data for actual pro-drop languages (Ueno and Polinsky 2006):

<table>
<thead>
<tr>
<th></th>
<th>Noun</th>
<th>Pronoun</th>
<th>Pro</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject (Intrans)</td>
<td>25%</td>
<td>50%</td>
<td>25%</td>
</tr>
<tr>
<td>Subject (Trans)</td>
<td>25%</td>
<td>25%</td>
<td>50%</td>
</tr>
<tr>
<td>Object</td>
<td>75%</td>
<td>25%</td>
<td></td>
</tr>
</tbody>
</table>

Connectionism

- Neural networks appeal:
  - Have brain-like architecture
  - Can learn complex tasks
  - Show interesting acquisition profiles
  - Are robust to damage
  - Produce unpredictable results

- Many cognitive models of linguistic phenomena - e.g. past tense

Why connectionism?

- When other statistical methods ...
  - Learn faster
  - Are more predictable
  - Are more linguistics friendly
- Because ...
  - They do better math than children
  - Their cognitive interest is questionable
  - They don’t like novel words (P = 0)
  - They really can’t do better than 99%

Basic network

- Simple units (neurons)
- Connections (synapses)
- Activation (electricity)

Our network

- Elman network with extra recurrent layer at the output (≈ short term memory)

Sample sentence

*The man saw the cat.*
Let's see the results ...

<table>
<thead>
<tr>
<th>Head-marking</th>
<th>-pro</th>
<th>-</th>
<th>+T/A/M</th>
<th>+Agr</th>
</tr>
</thead>
<tbody>
<tr>
<td>-case</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+case</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-pro</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+case</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Experiment 1:
Familiar words (10 epochs)

<table>
<thead>
<tr>
<th>Head-marking</th>
<th>-pro</th>
<th>-</th>
<th>+T/A/M</th>
<th>+Agr</th>
</tr>
</thead>
<tbody>
<tr>
<td>-case</td>
<td>99.5%</td>
<td>99.4%</td>
<td>99.4%</td>
<td></td>
</tr>
<tr>
<td>+case</td>
<td>99.5%</td>
<td>99.5%</td>
<td>99.4%</td>
<td></td>
</tr>
<tr>
<td>-pro</td>
<td>73.4%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+case</td>
<td>98.8%</td>
<td>99.3%</td>
<td>99.0%</td>
<td></td>
</tr>
</tbody>
</table>

Experiment 2:
Familiar results (10 epochs)

<table>
<thead>
<tr>
<th>Head-marking</th>
<th>-pro</th>
<th>-</th>
<th>+T/A/M</th>
<th>+Agr</th>
</tr>
</thead>
<tbody>
<tr>
<td>-case</td>
<td>99.9%</td>
<td>99.9%</td>
<td>99.9%</td>
<td></td>
</tr>
<tr>
<td>+case</td>
<td>99.9%</td>
<td>99.9%</td>
<td>99.9%</td>
<td></td>
</tr>
<tr>
<td>-pro</td>
<td>97.8%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+case</td>
<td>98.8%</td>
<td>99.9%</td>
<td>99.9%</td>
<td></td>
</tr>
</tbody>
</table>

Experiment 3:
Novel Nouns/Verbs (30 epochs)

<table>
<thead>
<tr>
<th>Head-marking</th>
<th>-pro</th>
<th>-</th>
<th>+T/A/M</th>
<th>+Agr</th>
</tr>
</thead>
<tbody>
<tr>
<td>-case</td>
<td>98.7%</td>
<td>99.1%</td>
<td>99.0%</td>
<td></td>
</tr>
<tr>
<td>+case</td>
<td>99.2%</td>
<td>99.3%</td>
<td>99.0%</td>
<td></td>
</tr>
<tr>
<td>-pro</td>
<td>44.2%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+case</td>
<td>97.3%</td>
<td>99.0%</td>
<td>98.4%</td>
<td></td>
</tr>
</tbody>
</table>

The trouble with pro-drop

<table>
<thead>
<tr>
<th>pro</th>
<th>Order</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Intransitive</td>
<td>V</td>
</tr>
<tr>
<td>Transitive</td>
<td>V</td>
</tr>
<tr>
<td></td>
<td>3</td>
</tr>
</tbody>
</table>

• Linear word order predicts both word category and function.

The trouble with pro-drop

<table>
<thead>
<tr>
<th>pro</th>
<th>Order</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Intransitive</td>
<td>V</td>
</tr>
<tr>
<td>Transitive</td>
<td>V</td>
</tr>
<tr>
<td></td>
<td>3</td>
</tr>
</tbody>
</table>

• Linear word order is no longer predictive. Others cues (marking) are needed.
Network results summary

• No pro, no problem
• pro with marking is easy
• pro without marking requires rote learning of every word
  - Slow (vs fast mapping)
  - Useless for novel words
  - Insufficient for N-V homonyms

Network vs Language types

<table>
<thead>
<tr>
<th>Head-marking</th>
<th>Spanish, Au, Swahili</th>
<th>Polish, Albanian, Estonian</th>
</tr>
</thead>
<tbody>
<tr>
<td>-pro</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>+T/A/M</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>+Agr</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>-case</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>+case</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Mandarin</td>
<td>Vietnamese, English, Dutch, Swedish, Russian, Bulgarian, Lithuanian, Finnish, Estonian, German</td>
<td></td>
</tr>
<tr>
<td>Creoles</td>
<td>? S-E Asian languages</td>
<td></td>
</tr>
<tr>
<td>S-E Asian</td>
<td>?</td>
<td></td>
</tr>
</tbody>
</table>

Mandarin Chinese

• SVO word order (some SOV)
• No case or agreement
• Extensive pro-drop

ølai-le pro
'I/You/He/She/We/They' came.'

Zhangsan shuo [ø i,k bu renshi Lisi]
Zhangsan say pro not know Lisi
'Zhangsan said that [he/.../I/You/He/She/We/They] did not know Lisi.'

Constraints on pro-drop

• Only discourse topics are left unexpressed
ø yidian dou bu xihuan ø
'I didn’t like [it] a bit.'

• Structural exceptions (avoid N/V decision)
Wo gen *(ta) xue Yingwen
'I study English with him/her.'

Identifying nouns & verbs

• Cues for verbs: auxiliaries, co-verbs, aspect markers
  Tamen fa le wu shin ge qingtie
  'They sent out fifty invitations.'

• Cues for nouns: prepositions, classifiers, ba particle (in SVO)
  Zuotian you yi chang dianying
  'Yesterday there was a movie.'

Acquisition of Mandarin

• Word order: rigid SVO in production and comprehension

‘Morphology’: early acquisition of aspectual -le and classifier ge

"In child Chinese there's nothing worse, Than using nouns as verbs, or vice versa." (James Matisoff)
Mandarin summary

- Pro-drop is constrained
- Noun/Verb cues are available
- Acquisition is sensitive to available grammar
- Homonymy may be a non-issue

Mandarin is not a counter-example to the model results

Conclusions

- Pro-drop is independent of rich agreement
- Pro-drop requires distinct noun categories
- Morphological/Functional word marking can identify lexical categories
- Pro-drop creates surface ambiguities
- Ambiguities impede language learning

Questions?