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
  E.g. *Tom kissed Sue*
  Sue kissed Tom
- Dependent marking (= case)
  E.g. *Homo puellam planxit*
  man-NOM girl-ACC hit
  'The man hit the girl'
  E.g. *Hominem puella planxit*
  man-ACC girl-NOM hit
  'The girl hit the man'
- Head marking (= verb marking)
  - Rich agreement:
    E.g. payum narmaŋ na-mpu-tay
    man-PL woman-SG 3SG-3PL-see
    'The men saw the woman.'
    E.g. payum narmaŋ pu-n-tay
    man-PL woman-SG 3PL-3SG-see
    'The woman saw the men.'
  - Tense/Aspect/Modality markers:
    E.g. *Tom kissed Sue*

Pro-drop

- In many languages, subjects can be left unexpressed.

E.g. Ø Estoy comiendo
pro be-1sg eating
'[I] am eating.'

Ø Estás comiendo
pro be-2sg eating
'[You] are eating.'

Ø Está comiendo
pro be-3sg eating
'[He/She] is eating.'

3rd person + referential

This is 'rich verb agreement' pro-drop.

Pro-drop

- In many languages, subjects can be left unexpressed.

E.g. Ø mingling ta yong daozi
pro order 3sg use knife
'[He/She] orders him/her to use a knife.'

(Li & Thompson 1981; Huang 1989)

referential

This is discourse (topic) pro-drop.

Not real pro-drop

- Common: expletives

not referential

E.g. ø está lloviendo
'It is raining'

ø seems like we have a problem

- Common: 'diary-drop'

not 3rd person

E.g. Gone fishing
Want to play?

Parameter Space

Head-marking			
-pro	-	+T/A/M	+Agr
-case			
+case			
+pro	-	+T/A/M	+Agr
-case			
+case			

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).

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

Pro-drop implementation

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

	Noun	Pronoun	Pro
Subject (Intrans)	25%	50%	25%
Subject (Trans)	25%	25%	50%
Object	75%	25%	-

Connectionism

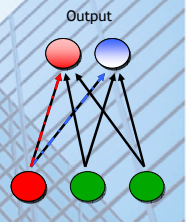
- Neural networks appeal:
  - Have brain-like architecture
  - Can learn complex tasks
  - Show interesting acquisition profiles
  - Are robust to damage
  - Produce unpredictable results
- Many cognitive models of linguistic phenomena – e.g. past tense

Why connectionism?

- When other statistical methods …
  - Learn faster *Vs acquisition profiles*
  - Are more predictable *Vs people are different*
  - Are more linguistics friendly *Vs kids are not*
- 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%

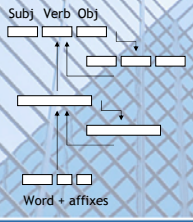
Basic network

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



Our network

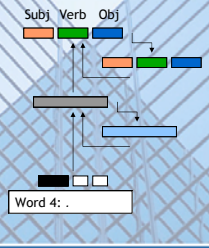
- Elman network with extra recurrent layer at the output (= short term memory)



Output: 16 / 16 / 16
Recurrent: 16 / 16 / 16
Hidden: 75
Recurrent: 75
Input: 16 / 8 / 8

Sample sentence

The man saw the cat .



Let's see the results …

Head-marking			
-pro	-	+T/A/M	+Agr
-case			
+case			
+pro	-	+T/A/M	+Agr
-case			
+case			

Experiment 1: Familiar words (10 epochs)

Head-marking			
-pro	-	+T/A/M	+Agr
-case	99.5%	99.4%	99.4%
+case	99.5%	99.5%	99.4%
+pro	-	+T/A/M	+Agr
-case	73.4%	99.2%	98.7%
+case	98.8%	99.3%	99.0%

no-pro? no-prob!

with pro? marking needed!

Note: Nice generalization

Note: Double marking not useful

Experiment 2: Familiar results (30 epochs)

Head-marking			
-pro	-	+T/A/M	+Agr
-case	99.9%	99.9%	99.9%
+case	99.9%	99.9%	99.9%
+pro	-	+T/A/M	+Agr
-case	97.8%	99.2%	98.7%
+case	98.8%	99.9%	99.0%

Learning every word separately takes time

Experiment 3: Novel Nouns/Verbs (30 epochs)

Head-marking			
-pro	-	+T/A/M	+Agr
-case	98.7%	99.1%	99.0%
+case	99.2%	99.3%	99.0%
+pro	-	+T/A/M	+Agr
-case	44.2%	98.8%	98.4%
+case	97.3%	99.0%	98.4%

Lexical identity does not, despite fast mapping

Word order and marking allows easy transfer of N/V detection

The trouble with pro-drop

-pro	Order		
	1	2	3
Intransitive	S	V	
Transitive	S	V	O
	✓	✓	

• Linear word order predicts both word category and function.

The trouble with pro-drop

+pro	Order		
	1	2	3
Intransitive	S	V	
with pro-drop	V		
Transitive	S	V	O
with pro-drop	V	O	
	✖	✖	

• 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

Head-marking			
-pro	-	+T/A/M	+Agr
-case	Vietnamese	Norwegian, English, Dutch, Swedish	French, Bulgarian
+case			Icelandic, Finnish, Russian, German
+pro	-	+T/A/M	+Agr
-case	? Creoles ? S-E Asian languages	Hebrew, Sundanese, Indonesian	Spanish, Au, Swahili
+case			Polish, Albanian, Estonian

Mandarin Chinese

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

ø lai-le
pro come-*le*
'[I/You/He/She/We/They] came.'

Zhangsan$_i$ shuo [ø$_{i,k}$ bu renshi Lisi]
Zhangsan say *pro* not know Lisi
'Zhangsan said that [he/...] did not know Lisi.'

Constraints on pro-drop

- Only discourse topics are left unexpressed

ø yidian dou bu xihuan ø
pro a.little all not like *pro*
'[I] didn't like [it] a bit.'

- Structural exceptions (avoid N/V decision)

Wo gen *(ta) xue Yingwen
I with 3sg study English
'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 issue PFV five then CL invitation
'They sent out fifty invitations.'

- Cues for nouns: prepositions, classifiers, *ba* particle (in SOV)

Zuotian you yi chang dianying
yesterday exist one CL movie
'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 worser,
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 ric agreement
- Pro-drop requires distinct nou categories
- Morphological/Functional word marking can identify lexical categories

Caution: Meaning required at all times

- Pro-drop creates surface ambiguities
- Ambiguities impede language learnin

Questions?