Issues and methods in the scientific study of spoken language

Štefan Beňuš

Seminar in research methods

Outline

- Study of language in general
- Issues in spoken language
  - Phonetics
  - Socio-linguistics
  - Discourse & pragmatics
- Data
  - Recording
  - corpora
- Methods
  - Acoustic analysis (Praat)
  - Transcription

What do linguists do?

- Common perception
  - They tell us what's correct
- Alternative approach
  - Language provides a window into our minds
  - By trying to understand how language work, we may learn about what goes on in our minds

English plural

- What do we know about plural formation in English?

- The alternation [s]/[z]/[əz] is **systematic** because the realization of plural morpheme can be predicted by the context: when a preceding sound is produced with vocal fold vibration (such as the [g] in *dog*) the plural ending is [z]; otherwise, it is [s].

- 4-year olds consistently pronounce the plural ending in *cats* and *dogs* as [s] and [z] respectively. Why?

Wug test
(Berko-Gleason 1957)
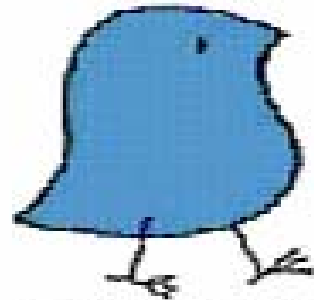
This is a wug.
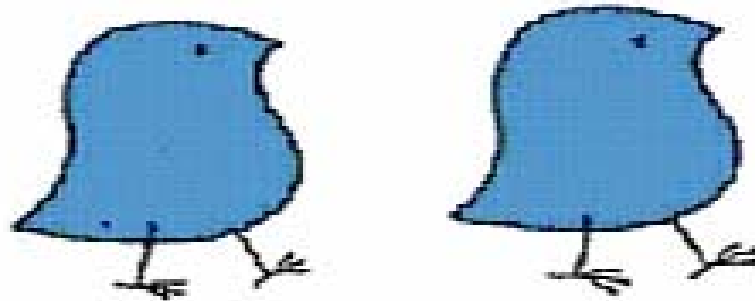
Now there is another one.
There are two of them.
There are two ___.

Other words tested

- Plural/possesive: lun, bik, tor, tass, gutch,…
- Past tense: blinged, ricked, mott,…

- Results: children productively reproduce the morphological alternations in words they have never heard.
- When tested, they call two toys named *glick*, as *glick*[s], and two instances of a *wug*, as *wug*[z].
- For other patterns, see the paper (Berko-Gleason, J. 1957. "The Child's Learning of English Morphology," *Word* 14:150-77)

Why is this interesting?

- The consistency in the wug-data can't be the result of memorizing the plural forms of the words themselves.

- Rather, the plural is formed in a rule-like fashion depending on phonetic properties of the final sounds in the stem.

- Native speakers know this, and small kids know this too without anybody telling them!

Language instruction

But surely, parents must TEACH language to their children…

Child:	My teacher holded the baby rabbits and we patted them
Adult:	Did you say your teacher held the baby rabbits?
Child:	Yes
Adult:	What did you say she did?
Child:	She holded the baby rabbits and we patted them.
Adult:	Did you say she held them tightly?
Child:	No, she holded them loosely.

Child:	Want other one spoon, Daddy.
Father:	You mean, you want "the other spoon"
Child:	Yes, I want other one spoon, please, Daddy.
Father:	Can you say "the other spoon"?
Child:	Other ... one ... spoon.
Father:	Say ... "other"
Child:	Other.
Father:	Spoon.
Child:	Spoon.
Father:	Other ... spoon.
Child:	Other ... spoon. Now give me other one spoon?

Poverty of the stimulus

- Children know things that they never heard, could not memorize, were not taught to know

- Imagine what kind of language input the kids hear

- Crucially, kids hear some *positive* input (i.e. how things should be said) but they don't hear *negative* input (i.e. things should not be said)

- Real linguists ask
  - How is this knowledge acquired?
  - How is it represented in our minds?
  - ...

Keep this in mind

- Language is a very complex system of our knowledge
- The available data are sparse & inaccurate

"In our endeavor to understand the reality we are somewhat like a man trying to understand the mechanism of a closed watch. He sees the face and the moving hands, even hears its ticking, but he has no way of opening the case. If he is ingenious, he may form some picture of a mechanism which could be responsible for all the things he observes, but he may never be quite sure his picture is the only one which could explain his observations. He may never be able to compare his picture with the real mechanism and he cannot even imagine the possibility of the meaning of such a comparison." (A. Einstein & L. Infeld: The evolution of physics)

Spoken vs. written language

- Spoken language is primary
  - Historically
  - Socially
    - Individual identity
    - emotions
  - Biologically
- Hence, spoken language is better suited for trying to understand our minds

Some examples of systematic patterns

- In general, we are interested in the functions/meaning of sound contrasts that we produce and perceive in speech

What do Americans know?

American	British
am[yu]se	am[yu]se
b[yu]ty (beauty)	b[yu]ty
c[yu]be	c[yu]be
d[u]pe	d[yu]pe
f[yu]me	f[yu]me
l[u]rid	l[yu]rid
n[u]ws (news)	n[yu]ws
p[yu]ny (puny)	p[yu]ny
pre[zu]me (presume)	pre[zyu]me
st[u]pid	st[yu]pid
s[u]t (suit)	s[yu]t

Features of non-native speech

- Quality of vowels and consonants
- Fluency
- Pitch range
- …

What does 'whatever' mean?

- 'you can have whatever you like'
- In a store:

  A: Hey, Richie, where do you want these?

  B: Anywhere.

  A: You want them over here?

  B: Yeah, whatever, just put them down
- At work, describing a client:

  A: So she ordered all this stuff and two days ago she changed her mind. I was like, whatever.

Prosody of 'whatever' affects its meaning in discourse

- A: Do you want get some Chinese   for dinner?

- B: Whatever.

- What phonetic features can cue the differences in meaning?

Other ambiguities

- Dogs must be carried
- You can *um* take the *uh* round shape an put it *um* …

How to get production data

- Record speech yourself
  - Interview
  - Reading
- Record speech of native speakers available on the internet
  - Radio, TV, movies, speeches, blogs, …
- Use available corpora
  - Buckeye
  - Columbia games
  - ICE

How to get perception data

- Questionnaires
- Perception experiments

What data

- Native speakers
- Native vs. non-native speakers
- Compare different dialects

How to analyze data

- Praat
  - http://www.fon.hum.uva.nl/praat/
  - Label and extract phonetic information
  - illustration

How to analyze data (cont'd)

- Transcriber
  - http://trans.sourceforge.net/en/presentation.php
  - Code for things of interest
  - illustration

Scientific approach

- Identify an interesting point/question/issue
  - Try to form a question or a hypothesis
- Do research, read available literature on the topic, se what's already known
- Adjust/focus your question to something that is still not known
- Identify the type of data and the way to collect them
- How would you analyze data, what features would you measure/count
- What would the outcome (both positive and negative) of the analysis mean for broader issues, for our understanding of language system?