

The interaction of syntactic and lexical information sources in language processing

Ev Fedorenko & Edward Gibson (Brain and Cognitive Sciences, MIT)
 evelina9@mit.edu, egibson@mit.edu

BACKGROUND AND MOTIVATION

Research Questions:

- How do lexical frequencies interact with syntactic expectations? (Experiments 1-2)
- What is the nature of syntactic expectations? (Experiments 3-4)

Tabor, Juliano & Tanenhaus (1997)

Context 1: Sentence initial

DETERMINER: *That cheap hotel is clean and comfortable to our surprise.*
CLAUSE-MARKER: *That cheap hotels are clean and comfortable surprised us.*
 Determiner preference: "hotels are clean" is read slower than "hotel his clean".

Context 2: Post-verbal (NB: verbs that can take clauses as complements)

DETERMINER: *The lawyer insisted that cheap hotel is clean and comfortable.*
CLAUSE-MARKER: *The lawyer insisted that cheap hotels are clean and comfortable.*
 Clause-marker preference: "hotel is clean" is read slower than "hotels are clean"

These results are not predicted by: (1) lexical category frequencies (e.g., MacDonald, 1993); (2) structure-based principles (e.g., Frazier, 1979, 1987).

THE CONTEXT-DEPENDENT LEXICAL FREQUENCY HYPOTHESIS

People are sensitive to lexical-category frequencies for each word (e.g., determiner vs. clause-marker reading for "that") within different syntactic environments.

Sent-initial: more determiners than clause-markers. **Post-verbal:** more clause-markers than determiners.

BUT: *The lawyer visited that cheap hotel / those cheap hotels to stay for the night.*

Preference: "that heap hotel" is read slower than "those cheap hotels".

People treat contexts following all verbs similarly (regardless of the lexical properties of the verb-class).

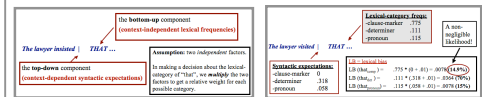
A two-component model of the interaction between lexical frequencies and syntactic expectations (Gibson, 2006):

CONTEXT-INDEPENDENT LEXICAL FREQUENCIES (c.f. Swinney et al., 1979)

People are sensitive to lexical-category frequencies for each word, independent of context.

CONTEXT-DEPENDENT SYNTACTIC EXPECTATIONS (c.f. MacDonald et al., 1994)

People are sensitive to syntactic expectations at each word.



Goal: to test the predictions of the two-component model using a common lexical-category ambiguity in English (see also Boland & Blodgett, 2001)

NORMING STUDIES

1. METALINGUISTIC JUDGMENTS STUDY.

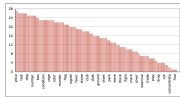
240 noun-verb ambiguous items were identified. Four raters judged whether each word was more likely to be a Noun, a Verb, or equally likely to be both.

Selection criteria: *Noun-biased* (3 or 4 N-responses, 0 V-responses; *Verb-biased* (3 or 4 V-responses, 0 N-responses; *Equi-biased* (no more than 2 N-responses or 2 V-responses).

60 words were selected in each group, for a total of 180 items.

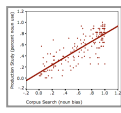
2. PRODUCTION STUDY.

56 participants were given a list of 90 items (30 from each category) and instructed to make short sentences with each word (90 unambiguous fillers were used, too).



3. CORPUS SEARCH.

CELEX database. 175/180 items (5 items not found). Lemma frequencies (normalized out of 1 mil) were used. Relative bias = Noun or Verb reading / (Noun+Verb readings).



RESULTS: Highly significant correlations across the three norming studies $R^2 > .56$, $p < .001$.

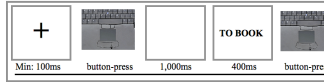
EXPERIMENTS 1 and 2

Design & Materials

63 participants, 180 items (+180 non-word fillers)

Design: 3 x 3 (*Bias* (between-items) - Noun-, Verb-, and Equi-biased x *Context* (N (the), V (to), Null)

Modified lexical decision paradigm:



24 participants, 24 items

Design: 2 x 2 (*Bias* (between-items) - Noun-, Verb- x *Context* (N (the), V (to))

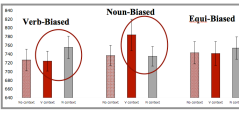
Self-paced word-by-word reading.

Materials:

Mary had a / to look at the documents ...

Larry had a bet with a friend ... / to bet on the outcome ...

Results

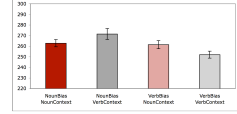


Expt 1 accuracies:

96.5%

Expt 2 accuracies:

95.4%



Interaction: $F(2,284)=5.45$; $p < .001$

Interaction: $F(1,123)=12.2$; $p < .01$

EXPERIMENT 3

Design & Materials

We can use the noun-verb ambiguity to investigate the nature of syntactic expectations.

The data in Experiments 1 and 2 are consistent with two different views of syntactic expectations:

- global syntactic expectations** that take into account the context of the entire sentence;
- n-gram frequencies** (e.g., Corley & Crocker, 2000; Tabor et al., 2004), whereby syntactic expectations are computed based on the context of the immediately preceding word(s).

Logic:

Examine cases where global and local expectations conflict (similar to Tabor et al.'s (2004) materials) but using the noun-verb ambiguity.

Design: 2 x 2 (Self-paced word-by word reading paradigm (48 participants, 16 items))

Factors: (1) **Adverb** (Absent / Present) [NB: Local expectation: Verb] (2) **Local expectation** (Verb / Noun)

Materials:

NoAdv/LocExp-Verb: *Some professors ...*

NoAdv/LocExp-Noun (!): *Some professors who are comfortable ...*

Adv/LocExp-Verb: *Some professors often ...*

Adv/RC:LocExp-Noun(Verb) *Some professors who are comfortable often ...*

... chair some sessions at conferences in their fields.

Control for local vs. non-local dependencies.

Results

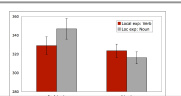
Reading times at the critical word + next word ("chair some"):

A ME of adverb presence: $F(1,46)=8.204$; $p < .01$

$F(2,115)=9.34$; $p < .01$

An interaction: $F(1,46)=3.46$; $p = .07$

$F(2,115)=2.601$; $p = .13$



EXPERIMENT 4

Design & Materials

Design: 2 x 2 (Self-paced word-by word reading paradigm (32 participants, 16 items))

Factors: (1) **Ambiguity** (Ambiguous / Unambiguous) (2) **Local expectation** (Verb / Noun)

Materials:

Amb/LocExp-V (!): *The doctor is prescribing Mrs. Jones ...drops to be taken two times a day...*

Amb/LocExp-N: *The doctor is prescribing Mrs. Jones ...some drops to be taken two times a day...*

Unamb/LocExp-V: *The doctor is prescribing Mrs. Jones ...pills to be taken two times a day...*

Unamb/LocExp-N: *The doctor is prescribing Mrs. Jones ...some pills to be taken two times a day...*

Results

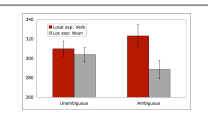
Reading times at the critical word + next word ("drops/pills to"):

A ME of LocExp: $F(1,31)=10.04$; $p < .005$

$F(2,115)=6.59$; $p < .05$

An interaction: $F(1,31)=3.89$; $p = .06$

$F(2,115)=1.82$; $p = .19$



CONCLUSIONS AND FUTURE WORK

Lexical-category ambiguities provide a useful way to investigate (1) the interaction between syntactic expectations and lexical frequencies, and (2) the nature of syntactic expectations.

Gibson's (2006) proposal provides a formal model of the interaction between context-dependent syntactic expectations and context-independent lexical frequencies.

We plan to investigate the syntactic expectations component on the model in more detail, as well as the lexical frequencies component.

Furthermore, we hope to characterize and quantify interactions among other information sources in on-line sentence comprehension.

SUMMARY

Experiments 1-2:

- Context-independent lexical frequencies and context-dependent syntactic expectations are two independent factors that affect the process of lexical-category ambiguity resolution.

Experiments 3-4:

- There are both global and local components to syntactic expectations.