

It's not the words, but how you say them: Effects of referential predictability on the production of names and pronouns

Elsi Kaiser, Edward Holsinger & David Li
{emkaiser, holsinge, lidc} @ usc.edu, University of Southern California

Introduction

- Language processing system is sensitive to **predictability**.
 - Words have shorter durations and more phonological reduction when predictable based on preceding words/collocational frequencies (Bell et al., 2003; Jurafsky et al., 2001, see also Bard & Aylett, 2001; Fowler et al., 1997, inter alia).
 - Also other domains, including syntax (Hale, 2001; Levy, 2008)
- **Our research:** Two studies on predictability in reference-tracking, especially the relation between (i) *how* an entity is referred to and (ii) *how likely* it is to be mentioned, i.e., how predictable it is.
- In domain of reference-tracking, researchers disagree regarding effects of predictability:
 - **View #1:** Referents' predictability is connected to the choice of referring expression: Reduced forms (e.g. pronouns) are used for highly predictable, expected referents (e.g. Arnold, 2008, Givón 1989).
 - **View #2:** Likelihood-of-mention is separate from choice of referring expression.
 - *Fukumura & van Gompel (2010)*: Likelihood-of-mention influenced by semantics; referring expression choice influenced by structural prominence
 - *Kaiser (2010)*: Discourse-level factors influence likelihood-of-mention, likelihood-of-prominentalization sensitive to subjecthood.
 - *Kehler et al (2008)* separate likelihood-of-mention, likelihood-of-prominentalization

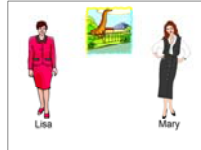
Exp 1 Pronoun interpretation: When given a pronoun, how do people interpret it?

Exp 2 Production of referring expressions:

- (i) When do people opt to use pronouns rather than names?
- (ii) Which referents do participants choose to talk about? => measure of predictability
- (iii) Acoustic duration of names => acoustic reduction?

Exp 1: Prompt pronoun ('comprehension')

- **Q =** When faced with a pronoun, how do people interpret it?
- Sentence continuation w/ **agent-patient verbs** (e.g. kicked, tickled, slapped).
- We tested active and passive sentences, to see how pronoun interpretation is influenced by syntactic roles vs. thematic roles.
- (a) **Mary**_{agent} slapped **Lisa**_{pat} at the zoo. **As a result** she...
- (b) **Lisa**_{pat} was slapped by **Mary**_{agent} at the zoo. **As a result** she...
- **Task:** See picture, hear auditory prompt, provide natural-sounding spoken continuation. Also eye-tracked (data not reported here). 24 targets, 36 fillers, 24 participants
- Coherence relations influence likelihood-of-mention -- control this by using causal relations => w/ agent-patient verbs, **causal relation** focuses **patient** (object in active voice) (e.g. Kehler 2002)
- Also tested ambiguous 'then'; preliminary analyses suggest it patterns like 'as a result' when used causally.
- Terminology: **Subject** => grammatical subject. **Object** => direct object or object in by-phrase
- **Coding:** Continuations double-coded by two independent coders: *Does pronoun refer to preceding subject, object or is antecedent unclear?* Coders instructed to be conservative, to choose 'unclear' if not enough info to determine antecedent. (Coding procedure similar to Rohde, 2008; Kehler et al., 2008; Kaiser, 2010). 25% 'unclear', excluded from analyses.

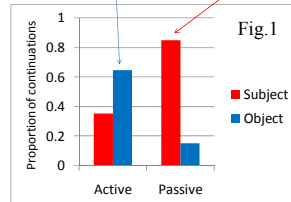


Results

Do people interpret the prompt pronoun as referring to the preceding subject or object?

Active voice: Preference for object (patient), p 's < .05

Passive voice: Preference for subject (patient), p 's < .001



Patient preference stronger with passives than actives

- Thematic role: **Patient** preferred in both active and passive.
- No main effect of grammatical role/subjecthood
- Why patient preference stronger with passives? Due to patient's promotion to a non-canonical position? (passives & topicality)

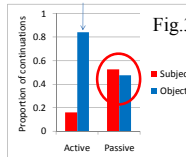
Exp 2: No prompt pronoun ('production')

- **Q1 Pronoun production** => When people choose to produce a pronoun, does it refer to the most predictable referent?
- **Q2 Predictability** => Which referent will people continue talking about? [Who is most likely to be mentioned next, most predictable?]
- **Q3 Acoustic duration** => When people produce names, are their acoustic durations connected to referent predictability?
- Same design, stimuli as Exp 1, but no pronoun (sound files truncated at end of connective). 24 new participants
- (a) **Mary** slapped **Lisa** at the zoo. **As a result...**
- (b) **Lisa** was slapped by **Mary** at the zoo. **As a result...**
- **Coding:** Continuations double-coded like Exp1.
- Coders noted (i) what kind of referring expression was used (e.g. pronoun, name, full noun phrase) and (ii) what it refers to (subject, object, unclear, another referent).
 - Pronouns produced on 11% of trials, 27% of these 'unclear'
- **Name durations:** measured by phonetically-trained annotator w/ Praat

Results

When people chose to produce a pronoun: Data looks like Exp1, patient preference =>

Q2 Likelihood-of-mention



Collapsing across referential forms, who do people start their continuations with?

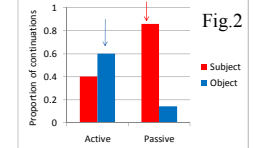
Actives = object (patient), p 's < .001
Passives = competition between object (agent) and subject (patient)

- Passives reveal **dissociation between likelihood-of-mention and likelihood-of-pronominalization** => Patients most likely to be pronominalized, but not most likely to be mentioned next.

Q3 Acoustic duration of names

- Duration of names produced in subject position (e.g. *Mary slapped Lisa at the zoo. As a result Lisa / Mary...*).
- Compared **active condition** (object has a high likelihood of subsequent mention) and **passive condition** (no clearly expected referent, since continuations split b/w subj and obj).
- **Result:** Names are shorter after actives (283ms) than after passives (320ms) (by subjects $p = .0858$, by items $p < .03$). **Name duration showed effects of predictability.**

Q1 Pronoun Production



Why more agent continuations in passives than actives?

- *Kaiser et al 2011*: Not a recency effect
- Due to **syntax-semantics mismatch** (semantically-prominent agent demoted to syntactically non-prominent by-phrase)

Conclusions

Thanks to Jackie Kim, Monica Do and Heidi Mettler for help with the experiments. This research was supported by NIH grant 1R01HD061457.

- Findings from various domains suggest that predictability is an important component of language processing.
- Our data suggest that it can also influence referential processing => **reduced acoustic durations**
- However, **we do not find evidence that predictability consistently influences choice of referring expressions**, contrary to some earlier claims.
 - Instead, we found that use and interpretation of pronouns is influenced by **thematic role**, independently of which referent is most predictable.
- We suggest that referents' likelihood-of-mention is influenced by an interplay of syntactic and semantic factors => mapping between syntactic and thematic roles (consequences of **syntax-semantics mismatches**)
- As a whole, our results highlight the benefits of exploring both lexical and acoustic aspects of referential production.