The role of context in young children’s comprehension of negation
Nordmeyer & Frank, 2014

Introduction

Problem: Negation is pervasive in human language and the word “no” is acquired early by children, yet studies suggest that it is difficult for even adults to process.

Suggestion: Pragmatic context makes processing negation easier.

Different types of negation

Early development (~12 months): nonexistence (“no more juice”) and refusal (“no go outside”)
Later development (~24 months): denial (“that not lollipop”)
As late as 4 years: difficulty with implicitly negative terms (like “less”)

Adults processing (comprehension) of negation

Adults are slower to process (true) negative sentences than non-negatives.
N400 (component sensitive to semantic expectancy of meaningful elements like words) is greater for negated elements in unexpected sentences (“A robin is not a truck”) compared to expected sentences, even when they are false (“A robin is/ is not a bird”).

Wason (1965):

Subjects faster to respond correctly to “Exactly one circle is not…” compared to “Exactly one circle is…”

Supportive contexts: N400 is reduced in negative sentences that are pragmatically licensed “In moderation, drinking red wine isn’t bad/good.” (N4 smaller to bad than good)
But N400 is the same to each type of ending (true/false) in sentences that are not: “A baby bunny’s fur isn’t very hard/soft.” (Similar N4 to both)

Experiment 1: Non-contrastive contexts; Experiment 2: Contrastive contexts
Participants: ~80 2-4-year-olds; ~16 adults in each experiment

Hear: “See these boys? Look at the boy who {has / has no} apples. Can you find him?”
**Results: Experiment 1**

For negative sentences:
- 2-y.o.’s don’t get it; 3-y.o.’s are better than 2-y.o.’s but still at chance; 4-y.o.’s look at correct target about 70% of time, but are slower than adults; adults look at target immediately after hearing “no”

**Results: Experiment 2**

For negative sentences:
- 2-y.o.’s don’t get it; 3-y.o.’s do better here, but not until ~1600 ms after “no”; 4-y.o.’s by 1200 ms; and adults are actually worse (slower) at this task than at Experiment 1
- But 3-, 4-y.o.’s, adults all switch back to competitor before settling on target.

Takeaway: Kids were better at processing negative sentences in supportive contexts.

“Children may have an easier time comprehending negation in contexts where there is a strong expectation that an action will occur or an object will be present, or when the negated object is highly salient to the child.”
A pragmatic account of the processing of negative sentences
Nordmeyer & Frank, 2014

Introduction

- Negation is informative when expectations are violated. (“We don’t have any chai today.”)
- As predicted by Grice: supportive context makes processing negative sentences easier.

Proposal: Contexts that set up strong expectations which are then violated make negative sentences more informative, and therefore will have a lower processing cost (and smaller RTs).

Definitions and assumptions:
- Informativeness defined as picking out the smallest subset of a context (Frank & Goodman, 2012)
- Reaction times are proportional to surprisal
- Surprisal: information-theoretic measure of the amount of information carried by an event based on probability of occurrence

Primary claim: context affects negative sentence processing by modulating listeners’ expectations.

Experiment 1: Context effects in processing negative sentences

83 participants recruited through Mechanical Turk; 28 trials.

Items: “Bob {has / has no} apples.” (Half positive, half negative.)
Half of participants saw no context. Half saw the following context:
Finding: Accepting negative statements such as “Bob has no apples” took longer when preceded by no context. The authors argue this is because an expectation has been set up that boys in the experiment will have apples; without this context, there is no reason to talk about apples at all.

Experiment 2: Parametric variations of context effects

Contexts:

**Exp’t 2a:** See pictures of three people holding objects (N=177)
Contexts: 0/3, 1/3, 2/3, 3/3 of characters holding an object (between subject)

**Exp’t 2b:** See pictures of four people holding objects (N=339)
Contexts: 0/4, 1/4, 2/4, 3/4, 4/4 of characters holding an object (between subject)

Target image was identical to that used in Experiment 1; 28 trials.

- no main effect of negation
- interaction between sentence type and truth value: difference between true positive and true negative was greater than difference between false positive and false negative
- as number of people with target item increased, RT decreased
- Claim: as proportion of people in the context with the target item increases, describing the picture as not having the target becomes more informative
Model 1

Surprisal of an utterance $w$, given the context and the speaker’s intended referent (Levy, 2008):

\begin{equation}
RT \sim -\log(P(w|r_s, C))
\end{equation}

Probability of an utterance is proportional to its utility (Frank & Goodman, 2012):

\begin{equation}
P(w|r_s, C) \propto e^{U(w;r_s, C)}
\end{equation}

Utility is the informativeness of the utterance $w$ minus its cost $D(w)$, here defined as number of words in the utterance multiplied by a cost-per-word parameter.

\begin{equation}
U(w; r_s, C) = I(w; r_s, C) - D(w)
\end{equation}

Informativeness in context is bits of information conveyed by the word $w$, which is assumed to have a uniform probability distribution over its contextual use (e.g., “boy with apples” applies equally to any boy with apples in the context).

\begin{equation}
I(w; r_s, C) = -( -\log(|w|^{-1}) )
\end{equation}

Combining equations (2)-(4) and normalizing over all words in vocabulary $V$:

\begin{equation}
P(w|r_s, C) = \frac{e^{\log(|w|^{-1})-D(w)}}{\sum_{w' \in V} e^{\log(|w'|^{-1})-D(w')}}
\end{equation}

Model predicts that as number of boys with apples increases, informativeness of “Bob has no apples” will increase because it selects an increasingly smaller subset of the context.

- Model 1 fails to capture U-shaped curve of reaction times, underestimating surprisal of 0/3 and 0/4 contexts for positive sentences and 3/3 and 4/4 contexts for negative sentences.
Model 2

This model includes the surprisal of the trial referent $r_S$ in addition to the surprisal of the utterance (similar to capturing a prior probability of the referent). This is estimated by the count of the target property in the context with smoothing parameter $\lambda$:

$$P(r_S|C) = \frac{\#MatchingPeople + \lambda}{\#TotalPeople + 2\lambda}$$

This results in the following, where the $\beta$ parameter can adjust relative contributions of the two surprisal effects (utterance, referent):

$$RT \sim - \log(P(w|r_S, C)) - \beta \log(P(r_S|C))$$

This model captures the surprisal of seeing, e.g., a boy with no apples after having seen a display of boys who all have apples.

- What (if anything) is to be made of the different scales (1-3, 2.0-2.4) of surprisal resulting from the two models?
- What would happen with different costs ($D(w)$) – both with a different cost-per-word parameter, and with different costs for different types of words? (E.g., maybe “boy” is easy since it has been produced/heard before, but “shirt,” or whatever other words were in their lexicon, might be harder.)
- Speaking of lexicon: what exactly is in their lexicon? How might this matter?
- Other parameters: $\lambda$ (smoothing) and $\beta$ (for weighting of utterance vs. referent)—how might varying these systematically change model’s predictions?