Recent findings suggest that language processing is sensitive to predictability (Bell et al., 2003; Hale, 2001; Hume & Mailhot to appear, Jurafsky et al. 2001, Levy, 2008). We investigated predictability in reference-tracking, focusing on how an entity is referred to and how likely it is to be mentioned. Currently, the role of predictability in reference-tracking is controversial: Some claim that predictability (likelihood-of-mention) is connected to referring-expression choice: Reduced forms (e.g. pronouns) are used for highly-predictable referents (Arnold, 2008; Givón, 1989). However, others (e.g. Fukumura & van Gompel, 2010) claim likelihood-of-mention is separate from choice of referring expression (see Kehler et al., 2008; Kaiser, 2010 for related work). To explore this, we investigated effects of likelihood-of-mention (i) on the choice of expressions and (ii) the acoustic duration of referring expressions.

In Exp1 (Prompt-pronoun), participants (n=24) heard sentences with agent-patient verbs while viewing scenes (see sample image). The study contained 4 male, 4 female characters. The task was to produce a continuation sentence. The prompt sentences were active or passive, and ended in a pronoun (ex.1). For consistency, we used the ‘as a result’ connective, which is known to focus the patient argument. Exp2 (No prompt-pronoun) was the same, but sentences did not end with a pronoun (24 new participants). In Exp1, coders noted what the prompt-pronoun referred to. In Exp2, coders noted the referring expression chosen and its referent. Durations of these expressions were recorded.

(1a) Mary\textsubscript{AGENT} slapped Lisa\textsubscript{PATIENT} at the zoo. As a result (she)...  
(1b) Lisa\textsubscript{PATIENT} was slapped by Mary\textsubscript{AGENT} at the zoo. As a result (she)...

Results/Referring expressions: In Exp1, participants mostly interpreted the pronoun as referring to the character that is the patient in the preceding sentence, i.e., the object in actives and the subject in passives (p’s<.05).

In Exp2, when participants produced pronouns (approx.11% of trials), we again see a patient-preference (>60% patient-continuations). However, when we look more broadly at which referent participants are most likely to continue with in Exp2, independent of the referring expression that is used, we see that the patient is not consistently the entity with the highest likelihood-of-mention. In active-voice conditions, overall we find that most continuations (collapsing pronoun-initial and name-initial continuations) do indeed start with the preceding object/patient (p’s<.01). However, in passive-voice conditions, these continuations were split

1 Exp1 and 2 also included ambiguous then connectives (temporal/causal). Preliminary analyses suggest that when used causally, then resembles as a result. However, due to then’s ambiguity and the resulting unbalanced data points, we focus here on as a result.
between the patient (the subject) and the agent in the by-phrase (no significant differences)—both have a comparable likelihood-of-mention. In other words, in passives, the patient and the agent are equally predictable.

Put together, the pronoun-initial and non-pronoun-initial continuations in Exp2 reveal a dissociation between likelihood-of-mention and likelihood-of-pronominalization. More specifically, the connection between pronominalization and the patient thematic role does not extend to likelihood-of-mention: Though actives show a bias for patient-initial continuations, in passives the presence of the agent in the by-phrase results in a boost in the overall likelihood of agent-initial continuations. (We have conducted other studies showing these results cannot be reduced to recency. We suggest they stem from the clash that arises when semantically-prominent arguments – agents – are placed in the low-prominence by-phrase, see Kaiser, Holsinger & Li, 2011).

Results/Acoustic duration: Given the dissociation between likelihood-of-mention and likelihood-of-pronominalization, does this mean predictability plays no role in reference-tracking? We also looked at the duration of names that participants produced in subject position (e.g. the second occurrence of ‘Lisa’ in Mary slapped Lisa at the zoo. As a result Lisa stormed off angrily). We compared names produced in the active condition, where the object of the prompt-sentence has a high likelihood of subsequent mention (i.e. it’s the expected referent), and names produced in the passive condition, where there is no clearly expected referent, since continuations were split between subject and object. Analyses of duration show that names in subject position are shorter (283ms) after active sentences than names in subject position (320ms) after passive sentences (p<.03 by items, p=.0858 by subjects). Thus, the duration of the names showed effects of how predictable the referent was.

Conclusions: Effects of likelihood-of-mention/predictability do exist in the domain of reference-tracking, in the form of acoustic reduction, although they do not appear to influence referring-expression choice. This work highlights the importance of exploring both lexical and acoustic aspects of referential production.

References:
Jurafsky et al. (2001). Probabilistic relations between words. In Frequency and the emergence of linguistic structure.