

Edge Cases of Discourse Salience in American English Casual Dialogs: A New Window into the Co-Constructed Nature of Social Conversation

Alex Luu

Brandeis University

alexluu@brandeis.edu

1 Discourse Salience in Social Dialog

¹ According to functional-cognitive literature, **salience** in dialog can be defined as a real-time property of mental entities which accounts for the fact that these mental entities come to be in the foreground of an interlocutor’s attention at the given time and therefore are referable from that interlocutor’s perspective (cf. Nolan, 2022). The term “**discourse salience**” is adopted to refer to salient content derivable from linguistic cues available in conversational discourse, such as morphosyntactic marking, noun phrase form and definiteness, syntactic role and construction, and discourse structure co-constructed by interlocutors.

I am conducting an empirical study of discourse salience in social conversation, using linguistic annotation and discourse analysis of **NEWT-SBCSAE**, a publicly accessible corpus of naturally occurring casual dialogs in American English (Du Bois et al., 2000; Riou, 2015; Luu and Malamud, 2020a). Performed by Linguistics majors and native speakers of English², the annotation focuses on the arguments of coherence relations in **NEWT-SBCSAE** and covers different linguistic aspects characterizing the “main point” of utterances constituting these arguments³. The discourse analysis includes statistical examination of annotation results and conversation analysis of target instances filtered based on annotated categories.

This paper focuses on analyzing edge cases of discourse salience identified based on two main aspects of information packaging at the utterance level, namely the given-new ordering of information and syntactic variations for realizing that ordering. The investigation scope is narrowed down to inter-speaker coherence relations, the first choice to explore the co-constructed nature of social dialog.

The examples examined in this paper are formatted as shown in Table 1. The center of each example is the bold utterance, which encapsulates an instance of edge cases, and its surrounding utterances which are connected to it via annotated coherence relations (in parentheses) (Luu and Malamud, 2020a). The index of each utterance reflects its chronological order (with the increment of 1).

| Utterance | Simplified transcript |
|--------------------|--|
| 38-AL (entity) | <i>Bill comes over with his leatherman toolman or or whatever it is.</i> |
| 39-AL | <i>Few minutes he had it undone.</i> <i>[laughter]</i> |
| 40-AN (concession) | <i>So she can’t use it now though.</i> |

Table 1: A contextualized utterance (in bold) from dialog *SBC043Spoonfuls* in **NEWT-SBCSAE** between Alice (AL) and her daughter Annette (AN).

2 Edge Cases of Discourse Salience

There is a scholarly consensus that the given-before-new ordering of information in discourse is preferred; and among syntactic variations for realizing that ordering, canonical word order (CWO), e.g. subject-verb-object (SVO) in English, is unmarked (Prince, 1992; Birner, 2012, inter alia). Generally, CWO is felicitous even in the context where it doesn’t adhere to the preferred ordering of information, while noncanonical word order (NWO)⁴ is felicitous only when it is used to realize the preferred information structure. In this work, new-before-given ordering and NWO characterize the edge cases of discourse salience as illustrated in Table 1. Knowing that the pronoun *it* in the utterance **39-AL** refers to an attaché case in prior discourse, we can conclude this utterance features the new-before-given ordering in the information

¹This paper’s live version is located at <https://osf.io/cedvx/>.

²From North-Eastern US. They were paid \$16/hour.

³This is grounded in the concept of at-issueness in formal semantics and pragmatics (e.g. Koev, 2018, inter alia).

⁴Including preposing (e.g. topicalization and focus-movement), postposing (e.g. existential and presentational *there*), argument reversal (e.g. inversion and passivization), their combinations, and cleft constructions (*wh-*, *it-*, *th-*).

exchange dimension: *few minutes* is newer than *he had it undone*. Moreover, the utterance is a NWO sentence as *few minutes* is preposed.

Among 1920 annotated arguments of interspeaker coherence relations available in NEWT-SBCSAE there are 14 new-before-given cases (0.73%) and 95 NWO cases (4.95%).

New-before-given Two clear new-before-given categories emerge from conversation analysis:

- dialogic resonance (Du Bois, 2014) (T. 2, 3)
- non-epistemic emphasis (Lu, 2022) (T. 4)

| Utterance | Simplified transcript |
|-------------------|---|
| 2886-K (temporal) | <i>I left my bag there.</i> |
| 2890-S | <i>Now the ghosts'll get it.</i> |
| 2892-K (entity) | <i>Ghosts'll get it.</i> |

Table 2: An example in *SBC034Times* between a couple.

The utterance **2890-S** in Table 2 shares a similar sequence of information slots with 2886-K – a triple of an agent, an action, and the bag (referred to by the noun phrase *my bag* and pronoun *it*). While this similarity results in the change of information ordering from given-before-new in 2886-K to new-before-given in **2890-S**, it preserves dialogic resonance at the syntactic level between two arguments of a coordinating coherence relation (temporal).

| Utterance | Simplified transcript |
|-----------------|---|
| 225-AL (entity) | <i>Oh and you know how I get when my heart just beats really fast?</i> |
| 229-AN | <i>Cathleen has to wear a heart monitor because of that mom.</i> |
| 230-AL (entity) | <i>When did she get that?</i> |
| 236-AL (entity) | <i>Would hers do that stop and then get real fast and?</i> |

Table 3: An example in *SBC043Spoonfuls*.

In Table 3 dialogic resonance happens at the social level: the utterance **229-AN** is an interlocutor-decentric move (Lu and Malamud, 2020b) from topics focusing on the hearer, Alice – the mother, to *Cathleen*. As the immediately preceding discourse of **220-AN** is solely dedicated to how overwhelmed Alice was at work, it is reasonable for Annette, the daughter and the speaker of **229-AN**, to lighten the conversation by switching the social focus to a third person at this moment. The resonance pattern here is someone had some trouble recently.

Different from above examples, the preposed new content in utterance **2582-D** (Table 4) is the speaker's strong self-positioning (*I do know*) and

| Utterance | Simplified transcript |
|---------------------|---|
| 2580-P (concession) | <i>You haven't read the book so you don't know.</i> |
| 2582-D | <i>Yeah but I do know it it's an awfully it's it's an awfully presumptuous thing to sit down and write a book about death when you haven't died.</i> |
| 2583-P | <i>But.</i> |
| 2584-P (concession) | <i>It has it has it has stories in there from from the Zen and.</i> |

Table 4: An example in *SBC005Book* between a couple.

expressive evaluation (*it's an awfully presumptuous thing*), demonstrating the dominance of the normative and affective dimensions in discourse salience. This non-epistemic emphasis allows the speaker to detach from the preferred information ordering.

Noncanonical word order All new-before-given cases examined above, except for the one in Table 1, preserve CWO and therefore confirm its felicity in the context of non-preferred information ordering. Examining the sole exception (**39-AL**), we can observe that it ends with laughter and therefore involves non-epistemic dimensions, similar to the case in Table 4. The difference is **39-AL** doesn't involve strong self-positioning, which is usually realized by *I* in the subject position. Thus, we can argue that the epistemically older information in **39-AL**, *he had it undone [laughter]*, is actually a new focus in the affective dimension. Consequently, NWO expressed in this utterance is still used to realize the preferred given-before-new structure, but in a non-epistemic dimension.

It's worth noting that the minor portion of NWO (4.95%) in annotated data supports CWO as the easiest and preferred way to produce salient content in spontaneous conversation, in which interlocutors constantly faces pressure of real-time interaction.

3 Conclusion

The new findings based on examining the edge cases of discourse salience are directly relevant to social dialog system modeling and evaluation. They confirm the importance of non-epistemic dimensions and relational work interlocutors rely on to co-construct their utterances' meaning. The findings also reveal concrete discourse configurations of these understudied aspects. As a result, this work demonstrates how theoretical work both underpins and arises from the empirical.

Acknowledgements

I am extremely grateful to my annotators, Eben Saveson and Tali Tukachinsky, for their curiosity, diligence and creativity. My deepest gratitude goes to [Sophia A. Malamud](#) for her active encouragement and thorough feedback on this paper.

References

- Betty J Birner. 2012. *Introduction to Pragmatics*. John Wiley & Sons.
- John W. Du Bois. 2014. [Towards a dialogic syntax](#). *Cognitive Linguistics*, 25(3):359–410. Publisher: De Gruyter Mouton.
- John W Du Bois, Wallace L Chafe, Charles Meyer, Sandra A Thompson, and Nii Martey. 2000. [Santa Barbara corpus of spoken American English](#). CD-ROM. Philadelphia: Linguistic Data Consortium.
- Todor Koev. 2018. [Notions of at-issueness](#). *Language and Linguistics Compass*, 12(12):e12306.
- Alex Lutu. 2022. [Sketching a linguistically-driven reasoning dialog model for social talk](#). In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, pages 153–170, Dublin, Ireland. Association for Computational Linguistics.
- Alex Lutu and Sophia A. Malamud. 2020a. [Annotating coherence relations for studying topic transitions in social talk](#). In *Proceedings of the 14th Linguistic Annotation Workshop*, pages 174–179, Barcelona, Spain. Association for Computational Linguistics.
- Alex Lutu and Sophia A. Malamud. 2020b. [Non-topical coherence in social talk: A call for dialogue model enrichment](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, pages 118–133, Online. Association for Computational Linguistics.
- Brian Nolan. 2022. *Language, Culture and Knowledge in Context: A Functional-Cognitive Approach*. Equinox Publishing Ltd.
- Ellen Prince. 1992. [The ZPG letter: Subjects, definiteness, and information status](#). In William C. Mann and Sandra A. Thompson, editors, *Discourse Description: Discourse Analyses of a Fundraising Text*, pages 295–325. Amsterdam: John Benjamins.
- Marine Riou. 2015. *The Grammar of Topic Transition in American English Conversation. Topic Transition Design and Management in Typical and Atypical Conversations (Schizophrenia)*. Ph.D. thesis, Université Sorbonne Paris Cité.