Sketching a Linguistically-Driven Reasoning Dialog Model for Social Talk

Alex Lưu
Brandeis University
alexluu@brandeis.edu

Abstract

The capability of holding social talk (or casual conversation) and making sense of conversational content requires context-sensitive natural language understanding and reasoning, which cannot be handled efficiently by the current popular open-domain dialog systems and chatbots. Heavily relying on corpus-based machine learning techniques to encode and decode context-sensitive meanings, these systems focus on fitting a particular training dataset, but not tracking what is actually happening in a conversation, and therefore easily derail in a new context. This work sketches out a more linguistically-informed architecture to handle social talk in English, in which corpus-based methods form the backbone of the relatively context-insensitive components (e.g. part-of-speech tagging, approximation of lexical meaning and constituent chunking), while symbolic modeling is used for reasoning out the context-sensitive components, which do not have any consistent mapping to linguistic forms. All components are fitted into a Bayesian game-theoretic model to address the interactive and rational aspects of conversation.[1]

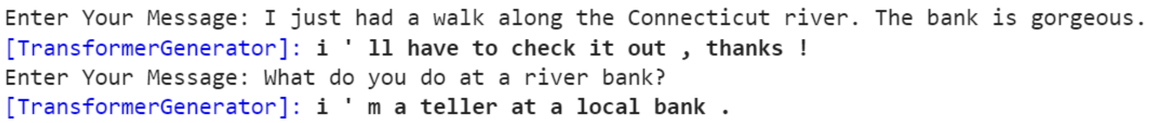
1 Introduction and Background

Developing dialog systems that can socially communicate with humans and make sense of conversational content would demonstrate that we are able to put together all linguistic knowledge and skills in action in a truly personalized manner, i.e. the dialog systems can use the same language competence to produce different coherent contents in different conversational situations. Such domain-independence would allow a dialog system to be robustly used across multiple content and task domains. While the benefit of using social interaction style in real-life dialog systems is a controversial topic (e.g. Chattaraman et al., 2019; Clark et al., 2019), the availability of social dialog agents can alleviate the critical shortage of human resources, e.g. in language education (Swanson and Mason, 2018), among other domains. Worldwide language learners often have little time to develop their communication skills with their teachers in the classroom setting; consequently teachers do not have enough clues to analyze their students' communication competence and their own teaching effectiveness. Having a dialog system that can socially converse with students at least about the learning materials outside the classroom and being able to access the content of these conversations for further analysis would be a plausible solution to the bottleneck of human resources in language education.

Research on social dialog systems has thrived in recent years thanks to optimism about the robustness of Natural Language Processing (NLP) techniques based on big data and deep learning. Within this period, emerging dialog systems either (1) seek to accommodate the open-domain information exchange by expanding the list of predefined interlocutors' intents and conversational topics for modular architectures, e.g. systems competing in Alexa Prize Socialbot Challenges (Khatri et al., 2018), or (2) focus on generating utterances to prolong the conversation based on neural network sequence-to-sequence approaches, such as Google's Meena (Adiwardana et al., 2020) and Meta's BlenderBot (Roller et al., 2021). Primarily motivated by the available data and statistical techniques but not the sociocultural characteristics of social talk, these dialog systems focus on fitting pre-existing training datasets, but not tracking what is actually happening in a conversation. As a consequence, they easily derail in a new context, as illustrated in Figure 1: the chatbot seems to interpret *bank* in the question *What do you do at a river bank?* as a financial institution, which is statistically preferred given the training data, instead of the land alongside a river, which is more felicitous in this particular context.

[1] The live version of this publication is located at https://osf.io/xz65k/.

*Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics
Student Research Workshop*, pages 153 - 170
May 22-27, 2022 ©2022 Association for Computational Linguistics



```
Enter Your Message: I just had a walk along the Connecticut river. The bank is gorgeous.
[TransformerGenerator]: i ' ll have to check it out , thanks !
Enter Your Message: What do you do at a river bank?
[TransformerGenerator]: i ' m a teller at a local bank .
```

Figure 1: A dialog between me and a state-of-the-art (SOTA) chatbot developed by Meta Research (Roller).

To address that shortcoming, this work takes a novel approach, relying on in-depth analysis of naturally occurring data to sketch out a more linguistically-informed[2] architecture to handle social talk in English, in which corpus-based methods form the backbone of the relatively context-insensitive components (e.g. part-of-speech (POS) tagging, approximation of lexical meaning, and constituent chunking), while symbolic modeling is used for reasoning out the context-sensitive components, which do not have any consistent mapping to linguistic forms. All components are fitted into a Bayesian game-theoretic model to address the interactive and rational aspects of conversation.

This paper is organized as follows. Section 2 provides a working definition of interlocutors' shared goal in social talk and its implications. Section 3 presents the key aspects of coherence with respect to the shared goal[3]. Section 4 outlines the conversational context that accounts for the coherence-based and goal-directed nature of social talk. Section 5 describes a simple worked example of the proposed model. Section 6 discusses the key differences between the proposed approach and current popular approaches to social dialog systems, analyzing its advantages and limitations, research priorities, and ethics and social impact considerations. Section 7 concludes and presents a plan for future work.

2 Interlocutors' Shared Social Goal

Lưu and Malamud (2020b) provides evidence of non-content based coherence in social talk that is not constrained by the purpose of information exchange. Specifically, the new-topic utterances in social talk, which begin a new topic not linguistically correlating with the content of prior discourse, signal certain sequential adjustment of the distances between the active conversational topic and each interlocutor, such as switching social focus from one interlocutor to another. This finding suggests that the definition of interlocutors' shared goal in social talk must be based on a social interaction formalism that goes beyond an information exchange framework (cf. Hovy and Yang, 2021 – a recent advocate for incorporating language's social factors into computational models of language use, given the SOTA NLP advancements).

Following the literature on intersubjectivity in communication (e.g. Rommetveit, 1976; Schiffrin, 1990; Wertsch, 2000; Tirassa and Bosco, 2008), I propose that the shared goal of interlocutors in social talk is **to create a coherent experience of together making sense of Self, the Other, and the relationship between them** (but not necessarily to share the same perspective on any aspect of the conversational content). This shared goal is not only primarily addressed by social talk but also forms a part of natural task-oriented conversation when the interlocutors attempt to build mutual rapport. Even the task-related conversational goals can be considered instantiations of this shared goal when Self and the Other are playing specific social roles in the task domains, e.g. seller – buyer or consultant – client. Within this shared goal, performing a conversational move implies taking a stance, i.e. a public social act of simultaneously **evaluating** objects (directly or indirectly) discussed in the conversational move, **positioning** subjects (Self or the Other or both), and **aligning** with the other subject, with respect to any salient dimension of the sociocultural field (Du Bois, 2007, p. 163). Regarding the sociocultural field, I adopt the proposal in Stevanovic and Peräkylä (2014) and Stevanovic and Koski (2018) for representing conversational interactions between Self and the Other as falling into one of three dimensions: **epistemic** (knowledge/information exchange - *how knowledgeable the interlocutors are*), **normative**[4] (power/social distance - *how powerful the interlocutors are*), and **affective** (affect/emotion - *how emotional the interlocutors are*). By explicitly paying attention to the normative and affective domains in the conversational context, we can expand current models of dialog that involve a representation of context but focus on the epistemic domain (such as those that build on Stalnaker, 1974, 1978; Roberts, 1996/2012, inter alia), and therefore can adequately

[2] As theory-neutral as possible.

[3] Detailed discussion on the concepts of "social talk" and "coherence" can be found in Lưu and Malamud (2020b).

[4] The original term, **deontic**, can be confusing since it expresses duty or obligation in the linguistics literature.

handle the multifaceted coherence in social talk and the corresponding social reasoning.

3 Multifaceted Coherence in Social Talk

Lưu and Malamud (2020b) shows that conversational coherence in social talk arises from at least two different sources, depending on whether the target utterance bears any content-based coherence relations to prior discourse.

Where there is at least one content-based coherence relation between the target utterance and prior discourse, this relation is shaped by certain **discourse hooks**[5] (located in the target utterance) that are pragmatically accessible to the hearer, and can be discourse-old, discourse-new bearing inferential relation to discourse-old, or discourse-new and related to discourse-old in a non-inferential manner (cf. Prince, 1992; Birner, 2012, inter alia). For example, in the social dialog in Table 1 the utterance 132-A is connected to the utterance 131-A by a coherence relation that is explicitly triggered by the conjunction *and* and shaped by several discourse hooks: the pronoun *it* is evoked as discourse-old information, referring to *a lovely red dress* which first appeared in 131-A, and *everything* is arguably inferrable from that *dress* via the entity/attribute inferential relation. Another coherence relation can be established between 133-A and 136-B since the clause *she totally ditched it* in the former utterance presupposes that there is a reason behind that action, which becomes the focus of the latter.

When there is no content-based coherence relation between the target utterance and prior discourse, conversational coherence is demonstrated by the **shift of social focus** created by certain explicit positioning or alignment signals in the target utterance. For example, the utterance 147-A of the excerpt shown in Table 1 switches the social focus from the speaker A to the hearer B by raising a question related to B, given that the preceding topic of discussion is A's difficulty in searching for a dress. By extending the conversation with new content relevant to the social subject who has received less focus in the preceding discourse, this utterance does contribute to the process of 'together making sense of Self, the Other, and the relationship between them' in a coherent way and seems to get its motivation from the non-epistemic domains: speaker A probably wants to show her attentiveness to speaker B (affective dimension), and her social closeness to B makes her think that this move is appropriate (normative dimension).

Utt.	Simplified transcript
131-A	*Well Rosemary and I went in for a look and uhm I found a lovely red dress*
132-A	*And I was like delighted with it and everything*
133-A	*And I brought mum up to see it and she totally ditched it*
134-B	*Yeah*
135-B	*Yeah*
136-B	*Why*
137-A	*She said it looked like she was she was saying it didn't do anything for my hips*
138-A	*It made my hips look big and like you know my bum and hips and everything*
139-A	*I was really excited cos I had the dress and then I just*
140-B	*But did you like it*
141-A	*Yeah*
142-B	*But she turned you off it*
143-A	*Yeah well I mean I'm hardly going to wear it now seeing everyone thinking I've big hip*
144-A	*Hip girl*
145-A	*I'll be called hippy*
146-A	*Hippo*
147-A	***Ah so how are you anyway***

Table 1: An excerpt, with indexed utterances, from telephone dialog *P1A-099* in the SPICE-Ireland corpus[6] (Kallen and Kirk, 2012) between two students A and B.

Coherence and Relevance It is worth noting that to be coherent, an utterance must not only be connected to the prior discourse via certain inferences, but also be **relevant** to the conversational goals, which coordinate the sequences of action performed by interlocutors' utterances (cf. Clift, 2016, pp. 89-94 for the discussion on coherence in interaction). Within the shared goal of interlocutors in social talk defined in Section 2, the expression of this relevance varies according to sociocultural dimensions. To be epistemically relevant, an utterance must, at least, introduce a new focus of discussion instead of simply repeating the old information. To be affectively relevant, an utterance can mimic the emotional intensity or support the sentiment of the immediately preceding discourse. For example, the interlocutors clearly show their matching emotional intensity in the excerpt of a face-to-face social dialog in Table 2 whose second

[5] I follow Birner to step away from the term 'topic' which "has not succeeded in becoming a unified concept within linguistic theory" (Birner, 2012, p. 214).

[6] This corpus can be obtained upon request to its directors.

half is full of laughter (see Ginzburg et al., 2020 for the discussion on emotive aspects of laughter). Finally, to be normatively relevant, the utterance should not, for example, provoke any controversial discussion that may hurt the social relationship between interlocutors, which is usually handled by profanity filters in current dialog systems (e.g. Khatri et al., 2018).

Coherence and Consistency Another aspect of conversational coherence is the consistency of the interlocutors' conversational contents and psychological behaviors, which are subsumed in the term '**speaker type**' in this paper. An utterance is incoherent if it commits an object evaluation, e.g. *I like cats*, that conflicts with another evaluation of the same object by the same speaker in prior discourse, e.g. *I hate cats*. One of the popular attempts to address this problem is the creation of the PERSONA-CHAT dataset for training and testing the aspects of persona consistency in chatbot models (Zhang et al., 2018). An utterance is also less coherent if it demonstrates some dramatic change in its speaker's behaviors, e.g. a rude statement from a speaker who is very polite in prior discourse. From the production perspective, a speaker would like to maintain their behavioral consistency; while from the interpretation perspective, a hearer would assume this consistency from the speaker to effectively decode the meaning of the speaker's utterance. Previous work such as Fang et al. (2018) shows that understanding the speaker's personality in the dialog helps the hearer in having better interaction strategies. It's worth noting that interlocutors' psychological behaviors vary according to different factors of the speech situation such as the cultural conditions, the interlocutors' personalities, and the relationship between interlocutors. For example, the social distance between interlocutors can affect the course and topics of discussion. Comparing the dialog in Table 2 between two friends and the dialog in Table 3 between a couple, we see that even though both of them are casual, the higher intimacy in the latter can be observed in all sociocultural dimensions:

Utt.	**Simplified transcript**
1442-M	*You know I wish I was uh the person whose voice they used in the telephone when it tells you the number has been changed*
...	...
1458-M	*They certainly use her a lot*
1459-M	*But I mean they only use what as uh five seconds total or something*
1460-M	*You know it's a*
1461-J	*Probably took her a long time to to say every possible combination*
1462-M	*Oh but they the computer does that*
1463-M	*All she has to do is say each digit*
1464-M	*And the computer*
1465-J	*Oh that's all it is*
1466-M	*Yeah*
1467-M	*It's like a series of samples*
1468-J	*And it automatically sorts em*
...	...
1474-M	*It would be much more pleasant if they had done all the combinations though*
1475-M	*You know call it up and there's something that actually says your number*
1476-M	*In toto*
1477-M	*You know [**laughter**]*
1478-J	*Yeah*
1479-J	*Or because it recognizes your phone number it automatically goes into the computer finds that*
1480-M	*Yeah that sample*
1481-J	*And and names the name*
1482-J	*Thank you Mister Smith for calling Pacific Bell*
1483-J	*[**laughter**]*
1484-M	*Yeah right*
1485-M	*You know [**laughter**]*
1486-J	*I am your personal computer representative*
1487-J	*[inhalation]*
1488-M	*That'd be great*
1489-J	*[**laughter**]*
1490-M	*[**laughter**]*

Table 2: An excerpt, with indexed utterances, from face-to-face dialog *SBC017Notions*[7] in the NEWT-SBCSAE corpus (Lưu and Malamud, 2020a; Riou, 2015; Du Bois et al., 2000) between two friends Michael and Jim.

- epistemic: the discussion topics are more personal (e.g. *two things I got out of my marriage, the marriage itself I mean as hellish*) and involve more creative association (e.g. *it pulled me under like a giant octopus or a giant giant shark, it's not the way with food*)
- affective: more instances of highly expressive language such as *really interesting*, *really got me grounded*, *as hellish as it was*, *like a giant octopus or a giant giant shark*, *the silent scream*, *so much better*, *very hellish*
- normative: the fact that the interlocutors are

[7] The original audio recording and transcript of the dialog can be conveniently browsed here.

comfortable with more personal topics and more expressive language; and the emphasis on positioning by explicitly involving Self in the story (e.g. *I used to have ..., two things I got out of my marriage, ... got me grounded, ... pulled me..., there I was, then I found that I was on my own two feet again, a way out of me*) but not on alignment as in the other dialog in which the interlocutors use the phrase *you know* as an alignment signal more frequently.

The difference in the normative dimension confirms that explicit positioning and alignment play an important role in the dynamics of social relationship between the interlocutors.

Utt.	**Simplified transcript**
2494-P	*I used to have this sort of standard line that there were two things I got out of my marriage*
2495-P	*One was a name that was easy to spell and one was a a child*
2496-P	*That really got me grounded*
2497-P	*But the fact of the matter is*
2498-P	*That the marriage itself I mean as hellish as it was it's like it pulled me under like a giant octopus*
2499-P	*Or a giant giant shark*
2500-P	*And it pulled me all the way under*
2501-P	*And then*
2502-P	*And there I was*
2503-P	*It was like the silent scream*
2504-P	*And then then I found that I was on my own two feet again*
2505-P	*And it really was what was hell in that that marriage became became a way out of me*
2506-P	*It was the flip side*
2507-P	*It's like sometimes you go through things and you come out the other side of them*
2508-P	*You come out so much better*
2509-P	*And if I hadn't had that if I hadn't had*
2510-P	*[inhalation]*
2511-D	*It's not the way with food*
2512-P	*What do you mean*
2513-D	*What goes in one way doesn't come out* ***[laughter]***
2514-P	***[laughter]***
2515-P	***[laughter]***
2516-P	*[inhalation]*
2517-P	*Comes out very hellish*

Table 3: An excerpt, with indexed utterances, from face-to-face dialog *SBC005Book*[8] in the NEWT-SBCSAE corpus (Lưu and Malamud, 2020a; Riou, 2015; Du Bois et al., 2000) between a couple, Pamela and Darryl.

4 Context Representation and Update

To be capable of reasoning about multifaceted coherence in social talk presented in Section 3, a linguistically-driven dialog model needs an adequate representation of the conversational context that consists of essential linguistic information obtained from either neural or symbolic knowledge. To optimally exploit both sources of knowledge, the relatively context-insensitive components of the conversational context are deduced by machine learning techniques; while the more context-sensitive components, which do not have any consistent mapping to linguistic forms, are reasoned out by symbolic methods. This division of labor takes advantage of the knowledge of pretrained statistical models as prior experiences to approximate linguistic meanings, at the same time separate them from the real-time meanings co-constructed by interlocutors in a specific conversational context via symbolic reasoning. To facilitate the reasoning, the conversational context has direct access to knowledge sources including linguistic dictionaries and thesauri, and world knowledge bases. An all-in-one option for knowledge sources is Wolfram Engine.

Specifically, using statistical models of off-the-shelf NLP libraries such as spaCy, we can automatically obtain basic linguistic annotations of an utterance including word tokens, their POS tags and contextual embeddings, syntactic relations between word tokens (as the result of dependency parsing), and linguistic constituents (including named entities). Based on these pieces of linguistic information, the discourse hooks in an utterance are identified by various heuristics such as:

Relying on linguistic definitions and relations:

- use dictionaries to obtain the senses of a word token and the corresponding definitions and examples of their usage in context
- select the most probable senses of that token in the target utterance based on the similarity scores between the contextual embeddings of the token and each of its senses
- use linguistic thesauri, such as WordNet (Fellbaum, 2010), to obtain the set of related lexical items of each selected sense, e.g. its synonyms, hypernyms and hyponyms
- identify and weigh potential discourse hook relations between each selected sense or related lexical item of the examined token and

[8] The original audio recording and transcript of the dialog can be conveniently browsed here.

other tokens in prior context based on the similarity scores between their embeddings

Relying on world knowledge bases:

- map a linguistic constituent to a concept in knowledge bases based on the similarity scores between their embeddings
- obtain a set of neighbor concepts of that linguistic constituent in knowledge bases
- identify and weigh potential discourse hook relations between each neighbor concept and other concepts in prior context based on the similarity scores between their embeddings

Relying on discourse knowledge:

- use the conversational context itself as a knowledge source to infer those potential discourse hook relations such as co-references between a pronoun in the target utterance and entities in immediately preceding discourse, and the temporal and spatial relations between an event or object in the target utterance and other events or objects in preceding discourse

These heuristics mainly address the first two types of discourse hook discussed in Section 3, discourse-old and discourse-new bearing inferential relation to discourse-old; the final type, discourse-new and related to discourse-old in a non-inferential manner, requires more sophisticated linguistic reasoning about presupposed content of an utterance. It is worth noting that by establishing potential discourse hook relations we not only connect two utterances but also lengthen various conversational threads which reflect different sequences of actions performed by the interlocutors, and therefore provide a deeper contextual structure in comparison with the contextual representation in which prior discourse is treated as a single conversational thread, usually called the dialog history.

Further, to represent non-epistemic dimensions, it is necessary to annotate at least the following:

Affective: intances of highly expressive language in the target utterance such as adjectives and idioms (which can be identified by analyzing their definition and properties recorded in the linguistic dictionaries) and their sentiments (which can retrieved from off-the-shelf sentiment analysis models)

Normative: default and emphasized positioning and alignment in the target utterance which can be identified based on the clause type of the utterance and the absence or presence of Self and the Other in its linguistic content; for example:

- If the target utterance is a declarative:
  - if Self is present in the utterance: emphasized Self positioning
  - else: default Self positioning
- If the target utterance is an interrogative:
  - if the Other is present in the utterance: emphasized alignment
  - else: default alignment
- If the target utterance is an imperative:
  - emphasized alignment
  - if Self is present in the utterance: emphasized Self positioning
  - else: default Self positioning

As discussed in Section 3, instances of highly expressive language help the dialog model estimate the emotional intensity or sentiment conveyed by its partner and flavor its own utterances with appropriate affective connotations; while emphasized positioning and alignment assist the model in recognizing a potential shift of social focus or of social distance expressed by its partner. Using clause types to reason out emphasized positioning is a basic pragmatic calculation of social acts encoded in an utterance in the proposed architecture. System designers can enrich the pragmatic calculation with additional normative rules for a more fine-grained representation of social acts[9]. Although clause type classification is not a current component of a typical automatic linguistic annotation pipeline, this task should not be as challenging as speech act/intent classification and should be robustly handled by statistical models because clause types are distinguished by specific form-based features, at least in English (Siemund, 2018).

A Minimally Viable Dialog Model A minimally viable linguistically-driven reasoning dialog model for social talk is an honest conversational companion in that it behaves as a conversing computer without wearing any superficial persona. It is equipped with all components listed in this section. From the interpretation perspective, whenever it receives the transcript of an utterance from the human interlocutor, it will obtain the automatic linguistic annotations of the utterance and apply predefined heuristics (1) to establish potential discourse hook relations between the words/constituents of

[9] Which can be informed by additional sociolinguistic knowledge, e.g. variant linguistic forms of the English suffix (ING) signal different levels of formality (an embodiment of social distance): the standard form *-ing* is more formal than the marked form *-in'* (Labov, 2012).

the utterance and other words/constituents in prior discourse, (2) to mark the instances of highly expressive language in the utterance with their sentiments, and (3) to capture the default/emphasized positioning and alignment in the utterance.

The model is aware that there may exist different alternatives in its interpretation; for example, each word in the utterance can have different discourse hook relations for different senses and therefore the number of alternatives for the whole utterance is the product of the numbers of senses. To select the best interpretation alternative, the model assigns a discourse salience score to each alternative.

This salience score is compositionally calculated based on how strongly an alternative is grounded in the context, including, for example, the weights of the discourse hook relations characterizing that alternative, and the recentness of discourse threads they participate in. Each alternative is also indexed with the sociocultural dimension that is most relevant to it: epistemic if it is full of discourse hooks, affective if it stands out with plenty of highly expressive language, or normative if it is highlighted by emphasized positioning/alignment. The interpretation alternative that has the highest discourse salience score will be added to the conversational context. Its salience factors and relevant sociocultural dimension provide human-readable evidence of what makes it a coherent move within the shared social goal, as discussed in Section 3.

From the production perspective, the model can heuristically generate a set of utterances as production alternatives which are salient with respect to the current conversational context and relevant to the conversational goal in at least one sociocultural dimension. For example, if the model knows that the human interlocutor just evaluated some aspect of an object and manages to find other information about the object in its knowledge bases, it can generate an utterance evaluating the object in the newly found aspect. In another scenario when the model has nothing else to comment on the object under discussion, it can switch social focus to the human interlocutor using the emphasized alignment technique, e.g. *What else are you interested in?* Similarly to the case of interpretation, the model can index each production alternative with the sociocultural dimension that is most relevant to it, and heuristically assign discourse salience scores to the alternatives in order to select the best one and update the context with its content.

Game-Theoretic Reasoning The selection of the best alternative from either interpretation or production perspectives can be formalized in a game-theoretic style, which pairs Lewis (1969/2002)'s signaling games (between two communicators) with the Bayesian approach to speaker/listener reasoning (see Tenenbaum et al., 2011 for an overview). Specifically, the probability $P(m|u, C)$ that the model assigns the hidden meaning m to the observable utterance u in the conversational context C depends on the prior probability $P(m)$ of the human interlocutor having m in mind and the utility value $U(u, m, C)$, corresponding to the salience of m with respect to u in C[10].

$P(m|u, C) \propto P(m) \times exp(\alpha \times U(u, m, C))$

(where α is a normalizing constant)

The prior probability $P(m)$ is used to account for the consistency of the speaker type discussed in Section 3. Specifically it captures the personal inclination of the human interlocutor towards a particular sociocultural dimension (cf. Yoon et al., 2020 for a different way to integrate these dimensions into a game-theoretic model and Asher and Lascarides, 2013 for a similar way to integrate a different aspect of speaker types into a game-theoretic model). There are three values of $P(m)$ for the three sociocutural dimension indexes:

- $P_{epi}(m) + P_{aff}(m) + P_{nor}(m) = 1$

These values are paired with the utility values of alternatives which share the same sociocultural dimension index. They can be learned offline based on a sample of human interlocutors or assigned by the interlocutor at the beginning of a conversation. These values can also be updated in a real time manner, e.g. if the human interlocutor produces a series of conversational moves that are highly relevant to the conversational goal in the affective dimension, $P_{aff}(m)$ will be increased accordingly.

5 A Worked Example

To demonstrate how the proposed dialog model works, a proof-of-concept text-based dialog system was developed based on the Free Wolfram

[10] This formalism simplifies the Bayesian inference in that it doesn't require the separation between speaker and listener behaviors as in recent popular game-theoretic frameworks for pragmatic reasoning, e.g. Iterated Best Response (Franke, 2009), Rational Speech Act (Frank and Goodman, 2012), and Social Meaning Game (Burnett, 2019). That simplification results from the fact that the model reasons based on a predefined shared goal and a rich representation of conversational context which accounts for all relevant aspects of real-time meanings co-constructed by interlocutors.

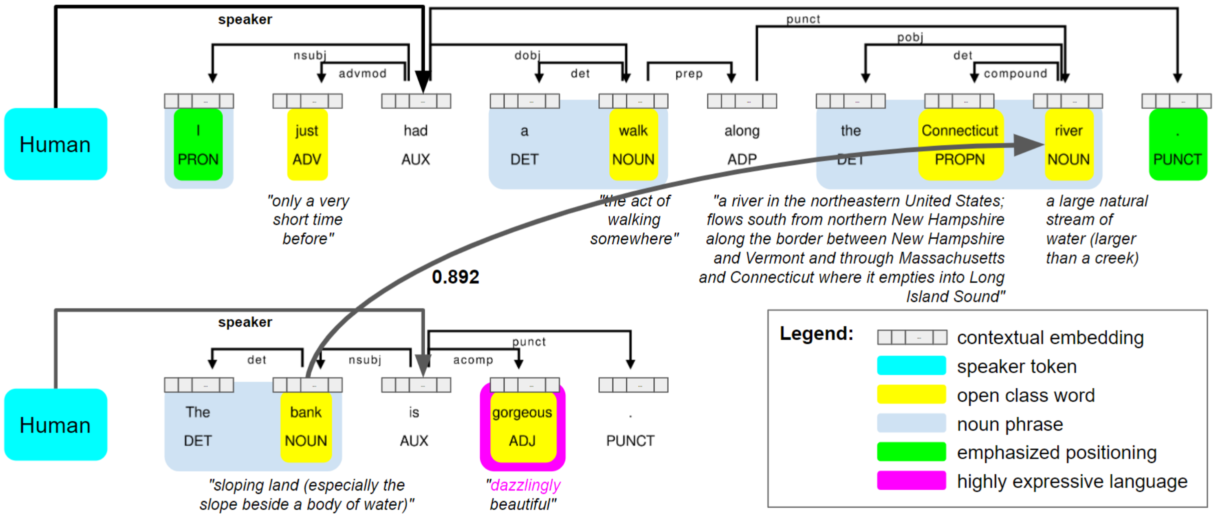

Human
speaker
nsubj
advmod
dobj
det
prep
punct
pobj
det
compound
I PRON
just ADV
had AUX
a DET
walk NOUN
along ADP
the DET
Connecticut PROPN
river NOUN
. PUNCT
"only a very short time before"
"the act of walking somewhere"
"a river in the northeastern United States; flows south from northern New Hampshire along the border between New Hampshire and Vermont and through Massachusetts and Connecticut where it empties into Long Island Sound"
a large natural stream of water (larger than a creek)
0.892
Human
speaker
det
nsubj
acomp
punct
The DET
bank NOUN
is AUX
gorgeous ADJ
. PUNCT
"sloping land (especially the slope beside a body of water)"
"dazzlingly beautiful"
Legend:
contextual embedding
speaker token
open class word
noun phrase
emphasized positioning
highly expressive language

Figure 2: Contextual representation of a dialog turn.

Engine for Developers and spaCy v2.3.5[11], including its small core model for English and the DistilBERT model (Sanh et al., 2019), accessed via spacy-transformers v0.6.x[11]. Figure 2 shows the conversational context created by the system after the human interlocutor enters the text string *I just had a walk along the Connecticut river. The bank is gorgeous.* (please refer to Appendix A for the snapshots of step-by-step context update).

The system uses spaCy's core model to tokenize the text string into word tokens (including punctuation marks), provide their POS tags, and then segment the sequence of tokens into sentences with their dependency structures. The system relies on DistilBERT to obtain the contextual embeddings of tokens and sequences of tokens so that it can calculate similarity scores between these embeddings to reason out the most appropriate senses of semantically ambiguous words as well as the potential discourse hook relations between linguistic constituents. For each open class word, i.e. an adjective, adverb, interjection, noun or verb, the system first retrieves all of WordNet senses from Wolfram knowledge base (via Wolfram Engine), and then identifies the real-time context-sensitive sense, as shown under each of these words in Figure 2. Each real-time sense is the one whose contextual embedding (calculated based on its textual definition) is the most similar to the contextual embedding of the corresponding word. They not only represent the human-readable meanings, but also participate in the creation of future discourse hook relations. Next, the system adds a meta-data token storing speaker information to each sentence before performing more context-sensitive reasoning.

Based on the pronoun *I*, the punctuation "." and the dependency links, the system recognizes that the first sentence is a declarative which has Self as the subject. Consequently, this sentence features the emphasized Self positioning. Moving to the second sentence, the system first examines alternative discourse hook relations between its sole noun phrase and other noun phrases in prior discourse, which results in the selection of the most salient relation between *the bank* and *the Connecticut river*, corresponding to the highest similarity score (0.892) between the embeddings of the real-time senses of the head nouns *bank* and *river*. The system then marks the adjective *gorgeous* as an instance of highly expressive language because its definition contains a degree adverb (*dazzlingly*).

To produce the most relevant response, the system puts more weight on the candidates addressing the second sentence as it is more recent. The highest salience score is achieved when the response includes both *the bank* and the emotional resonance of *gorgeous*, an instance of positive sentiment. Consequently, the system adds positive elements such as the predicate *like* to its planned response. A possible template for this planned response is "It seems that you like ... a lot, right?", which results in the ultimate response as *It seems that you like the bank a lot, right?*[12]

6 Discussion

Departing from current popular approaches to social dialog systems, which rely on available mod-

[11] Under the MIT License.

[12] Using templates is the simplest technique for the proof-of-concept system, but not a categorical implementation choice.

els for similar tasks[13] and conversational data created in artificial or asynchronous settings (Huang et al., 2020), this work starts with empirical analysis of naturally occurring data, i.e. human–human casual conversation in real life, to systematically define key linguistic characteristics of social dialog which can be modeled based on SOTA NLP techniques. This approach is in line with the preregistration practice promoted by van Miltenburg et al. (2021), entailing both advantages and limitations. By specifying what I want to capture in my model before the actual implementation, I can avoid the post-hoc problems faced by heavily data-driven architectures (e.g. Henderson et al., 2018). However, not relying on benchmark data and their corresponding techniques, I can not prove the practicality and reproducibility of my model in an actionable way before a full-blown dialog system is implemented. In addition, while this work starts with human–human conversation, its ultimate outcome is human–computer conversation which definitely diverges from the input guiding data and can potentially direct the research agenda into unplanned territories. It is also worth noting that while the modularity of the proposed architecture allows independent and simultaneous improvements of its components, its effectiveness can suffer from cumulative parsing errors caused by its pipeline design. Moreover, the statistical models of off-the-shelf NLP libraries used in the proposed architecture, mostly trained on planned text (e.g. Weischedel et al., 2013), may not work well on spontaneous conversation.

Research Priorities As the ultimate goal of the proposed dialog model is to truly facilitate mutual understanding in human–computer social communication, the model must aim at effectively coconstructing the real-time conversational context with its interlocutors and reasoning about their conversational moves (Kopp and Krämer, 2021). Thus, within the proposed framework I will focus on coherence-based context modeling and discourse salience calculation, taking into account the shared social goal. In other words, the research question that captivates me most is how to dynamically construct meaning in the context of social conversation (cf. Trott et al., 2020 for a broader research agenda). This priority implies the necessity of novel evaluation protocols to validly and reliably assess human–computer mutual understanding, which is ignored in current evaluation practices for in social dialog systems (Finch and Choi, 2020).

Another direction for exploration, which is more application-oriented, is how to optimally incorporate additional knowledge sources into the dialog model or spotlight a portion of the existing ones to seamlessly change salience calculation results, which conform to the system owner's desire. For example, imagine the scenario in which a language learner want to chat with the dialog system to enhance their vocabulary on a specific topic, they would definitely want the system to pay more attention to the area of knowledge sources which covers that topic. Ultimately, the dialog model could be systematically adapted for task-oriented dialog by integrating domain-specific knowledge bases.

Ethics and Social Impact Considerations The proposed dialog model is explainable in both its development approach and its interactions with different stakeholders (Kaur et al., 2022). First, its design is explicitly informed by empirical analysis of relevant data and its operational decisions are interpretable, using human-readable symbolic representation of conversational context. Second, the transparency of the proposed architecture with well-defined functional components can provide adequate and personalized explanations to the involved developers, domain experts, and end users.

Relying on publicly accessible NLP resources and featuring a widely integrable structure, the proposed dialog model can be freely implemented and used by independent end users, and continuously developed and enhanced by domain experts.

7 Conclusion and Future Work

This paper sketches out a novel dialog model for social conversation in English, motivated by a thorough investigation of the nature and linguistic characteristics of the phenomenon, including the shared goal between interlocutors and multifaceted coherence across different sociocultural dimensions. Next, I will implement a full-blown dialog system based on this model and develop adequate evaluation protocols, before iteratively evaluating and improving the system until it can consistently hold casual conversations with humans. Subsequently, I will use these conversations and their contextual representation as a new window into the social interaction between humans and reasoning machines.

[13] Either in the application aspect, e.g. task-oriented dialog models, or technical aspect, e.g. sequence-to-sequence machine translation models.

Acknowledgments

My deepest gratitude goes to Sophia A. Malamud, who exhaustively discussed every aspect of this paper with me. I am extremely grateful to Nianwen Xue, Anton Benz, Ralf Klabunde, and Malihe Alikhani for sharing their valuable perspectives on my work. Finally, I would like to thank the anonymous reviewers of EMNLP 2021, SCiL 2022, ARR (Dec 2021 deadline) and ACL-SRW 2022 for their detailed, constructive and actionable feedback.

References

Daniel Adiwardana, Minh-Thang Luong, David R So, Jamie Hall, Noah Fiedel, Romal Thoppilan, Zi Yang, Apoorv Kulshreshtha, Gaurav Nemade, Yifeng Lu, et al. 2020. Towards a human-like open-domain chatbot. *arXiv preprint arXiv:2001.09977*.

Nicholas Asher and Alex Lascarides. 2013. Strategic conversation. *Semantics and Pragmatics*, 6:1–62.

Betty J Birner. 2012. *Introduction to Pragmatics*. John Wiley & Sons.

Heather Burnett. 2019. Signalling games, sociolinguistic variation and the construction of style. *Linguistics and Philosophy*.

Veena Chattaraman, Wi-Suk Kwon, Juan E. Gilbert, and Kassandra Ross. 2019. Should AI-Based, conversational digital assistants employ social- or task-oriented interaction style? A task-competency and reciprocity perspective for older adults. *Computers in Human Behavior*, 90:315–330.

Leigh Clark, Nadia Pantidi, Orla Cooney, Philip Doyle, Diego Garaialde, Justin Edwards, Brendan Spillane, Emer Gilmartin, Christine Murad, Cosmin Munteanu, Vincent Wade, and Benjamin R. Cowan. 2019. What Makes a Good Conversation? Challenges in Designing Truly Conversational Agents. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 1–12, New York, NY, USA. Association for Computing Machinery.

Rebecca Clift. 2016. *Conversation Analysis*. Cambridge University Press.

John W Du Bois. 2007. The stance triangle. *Stancetaking in discourse: Subjectivity, evaluation, interaction*, 164(3):139–182.

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*.

Hao Fang, Hao Cheng, Maarten Sap, Elizabeth Clark, Ari Holtzman, Yejin Choi, Noah A. Smith, and Mari Ostendorf. 2018. Sounding board: A user-centric and content-driven social chatbot. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*, pages 96–100, New Orleans, Louisiana. Association for Computational Linguistics.

Christiane Fellbaum. 2010. WordNet. In Roberto Poli, Michael Healy, and Achilles Kameas, editors, *Theory and Applications of Ontology: Computer Applications*, pages 231–243. Springer Netherlands, Dordrecht.

Sarah E. Finch and Jinho D. Choi. 2020. Towards unified dialogue system evaluation: A comprehensive analysis of current evaluation protocols. In *Proceedings of the 21th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 236–245, 1st virtual meeting. Association for Computational Linguistics.

Michael C. Frank and Noah D. Goodman. 2012. Predicting pragmatic reasoning in language games. *Science*, 336(6084):998–998.

Michael Franke. 2009. *Signal to Act: Game Theory in Pragmatics*. Ph.D. thesis, Universiteit van Amsterdam.

Jonathan Ginzburg, Chiara Mazzocconi, and Ye Tian. 2020. Laughter as language. *Glossa: a journal of general linguistics*, 5(1):104. Number: 1 Publisher: Ubiquity Press.

Peter Henderson, Koustuv Sinha, Nicolas Angelard-Gontier, Nan Rosemary Ke, Genevieve Fried, Ryan Lowe, and Joelle Pineau. 2018. Ethical Challenges in Data-Driven Dialogue Systems. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '18, pages 123–129, New York, NY, USA. Association for Computing Machinery.

Dirk Hovy and Diyi Yang. 2021. The importance of modeling social factors of language: Theory and practice. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 588–602, Online. Association for Computational Linguistics.

Minlie Huang, Xiaoyan Zhu, and Jianfeng Gao. 2020. Challenges in Building Intelligent Open-domain Dialog Systems. *ACM Transactions on Information Systems*, 38(3):21:1–21:32.

Jeffrey L Kallen and John Monfries Kirk. 2012. *SPICE-Ireland: A User's Guide; Documentation to Accompany the SPICE-Ireland Corpus: Systems of Pragmatic Annotation in ICE-Ireland*. Cló Ollscoil na Banríona.

Davinder Kaur, Suleyman Uslu, Kaley J. Rittichier, and Arjan Durresi. 2022. Trustworthy Artificial Intelligence: A Review. *ACM Computing Surveys*, 55(2):39:1–39:38.

Chandra Khatri, Anu Venkatesh, Behnam Hedayatnia, Raefer Gabriel, Ashwin Ram, and Rohit Prasad. 2018. Alexa Prize — state of the art in conversational AI. *AI Magazine*, 39(3):40–55.

Stefan Kopp and Nicole Krämer. 2021. Revisiting Human-Agent Communication: The Importance of Joint Co-construction and Understanding Mental States. *Frontiers in Psychology*, 12.

William Labov. 2012. *Dialect Diversity in America: The Politics of Language Change*. University of Virginia Press, Charlottesville, VA.

David Lewis. 1969/2002. *Convention: A Philosophical Study*. John Wiley & Sons.

Alex Lưu 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 Lưu 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.

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.

Craige Roberts. 1996/2012. Information structure: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics*, 5:6–1.

Stephen Roller. ParlAI tutorial (accessed on 10/10/2021).

Stephen Roller, Emily Dinan, Naman Goyal, Da Ju, Mary Williamson, Yinhan Liu, Jing Xu, Myle Ott, Eric Michael Smith, Y-Lan Boureau, and Jason Weston. 2021. Recipes for building an open-domain chatbot. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 300–325, Online. Association for Computational Linguistics.

Ragnar Rommetveit. 1976. On the architecture of intersubjectivity. In Lloyd H. Strickland, Frances E. Aboud, and Kenneth J. Gergen, editors, *Social Psychology in Transition*, pages 201–214. Springer US, Boston, MA.

Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2019. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. In *The 5th Workshop on Energy Efficient Machine Learning and Cognitive Computing*, Vancouver, Canada.

Deborah Schiffrin. 1990. The principle of intersubjectivity in communication and conversation. *Semiotica*, 80(1-2):121–185.

Peter Siemund. 2018. *Speech Acts and Clause Types: English in a Cross-Linguistic Context*. Oxford Textbooks in Linguistics. Oxford University Press, Oxford, New York.

Robert Stalnaker. 1974. Pragmatic presuppositions. In Milton K. Munitz and Peter K. Unger, editors, *Semantics and Philosophy*, pages 197–213. New York University Press.

Robert Stalnaker. 1978. Assertion. In P. Cole, editor, *Syntax and Semantics 9: Pragmatics*, volume 9, pages 315–332. Academic Press, New York.

Melisa Stevanovic and Sonja E Koski. 2018. Intersubjectivity and the domains of social interaction: Proposal of a cross-sectional approach. *Psychology of Language and Communication*, 22(1):39–70.

Melisa Stevanovic and Anssi Peräkylä. 2014. Three orders in the organization of human action: On the interface between knowledge, power, and emotion in interaction and social relations. *Language in Society*, 43(2):185–207.

Pete Swanson and Shannon Mason. 2018. The world language teacher shortage: Taking a new direction. *Foreign Language Annals*, 51(1):251–262.

Joshua B. Tenenbaum, Charles Kemp, Thomas L. Griffiths, and Noah D. Goodman. 2011. How to Grow a Mind: Statistics, Structure, and Abstraction. *Science*, 331(6022):1279–1285. Publisher: American Association for the Advancement of Science.

Maurizio Tirassa and Francesca M Bosco. 2008. On the nature and role of intersubjectivity in communication. In *Enacting intersubjectivity: A cognitive and social perspective to the study of interactions*, pages 81–95. Amsterdam: IOS Press.

Sean Trott, Tiago Timponi Torrent, Nancy Chang, and Nathan Schneider. 2020. (Re)construing meaning in NLP. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5170–5184, Online. Association for Computational Linguistics.

Emiel van Miltenburg, Chris van der Lee, and Emiel Krahmer. 2021. Preregistering NLP research. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 613–623, Online. Association for Computational Linguistics.

Ralph Weischedel, Martha Palmer, Mitchell Marcus, Eduard Hovy, Pradhan, Lance Ramshaw, Nianwen Xue, Ann Taylor, Jeff Kaufman, Michelle Franchini, Mohammed El-Bachouti, Robert Belvin, and Ann Houston. 2013. OntoNotes Release 5.0. Artwork Size: 2806280 KB Pages: 2806280 KB Type: dataset.

James V Wertsch. 2000. Intersubjectivity and alterity in human communication. *Communication: An arena of development*, pages 17–31.

Erica J. Yoon, Michael Henry Tessler, Noah D. Goodman, and Michael C. Frank. 2020. Polite Speech Emerges From Competing Social Goals. *Open Mind*, 4:71–87. Publisher: MIT Press.

Saizheng Zhang, Emily Dinan, Jack Urbanek, Arthur Szlam, Douwe Kiela, and Jason Weston. 2018. Personalizing dialogue agents: I have a dog, do you have pets too? In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2204–2213, Melbourne, Australia. Association for Computational Linguistics.

A Step-by-Step Context Update

Figures 3–20 capture the sequence of context changes discussed in Section 5.

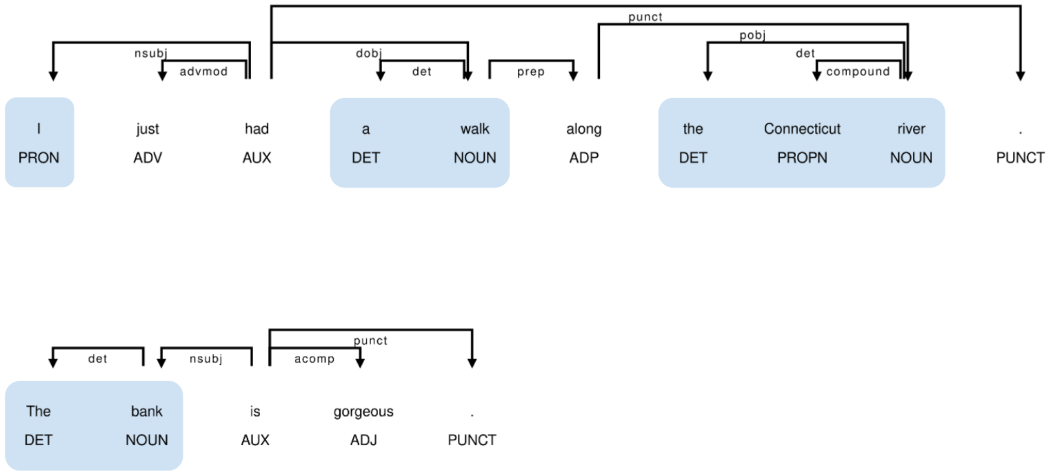


Figure 3: Add spaCy's linguistic annotations.

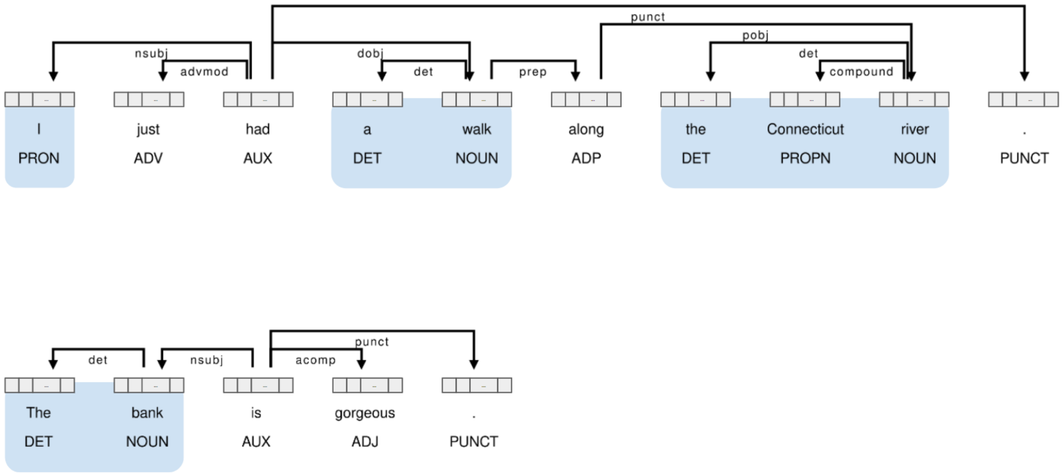


Figure 4: Add DistilBERT embeddings.

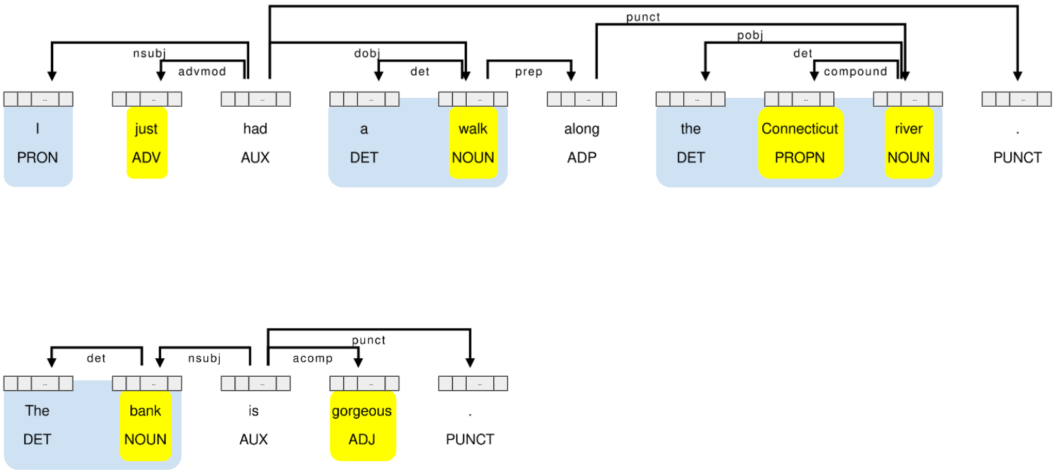


Figure 5: Navigate open class words, which are *just*, *walk*, *Connecticut*, *river*, *bank* and *gorgeous*.

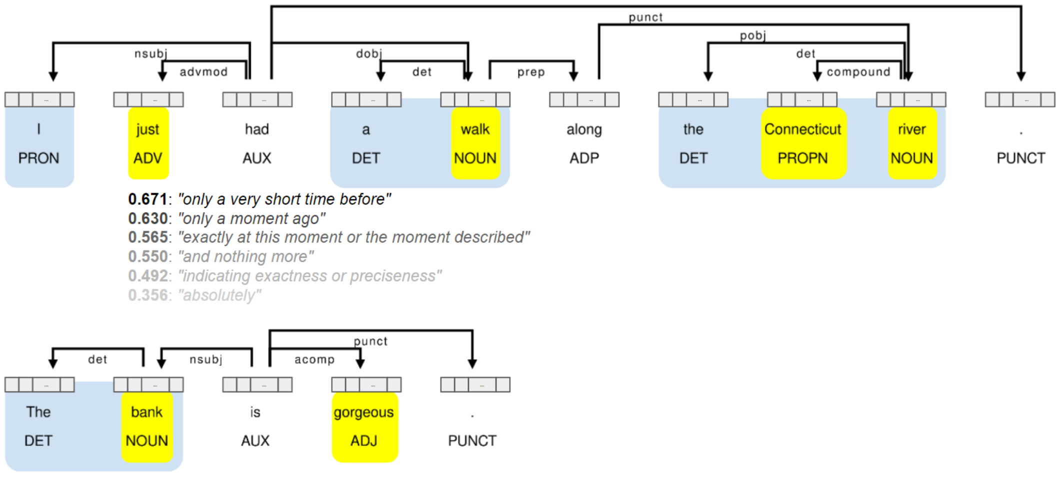


Figure 6: Calculate and rank similarity scores between *just* and each of its dictionary sense definitions.

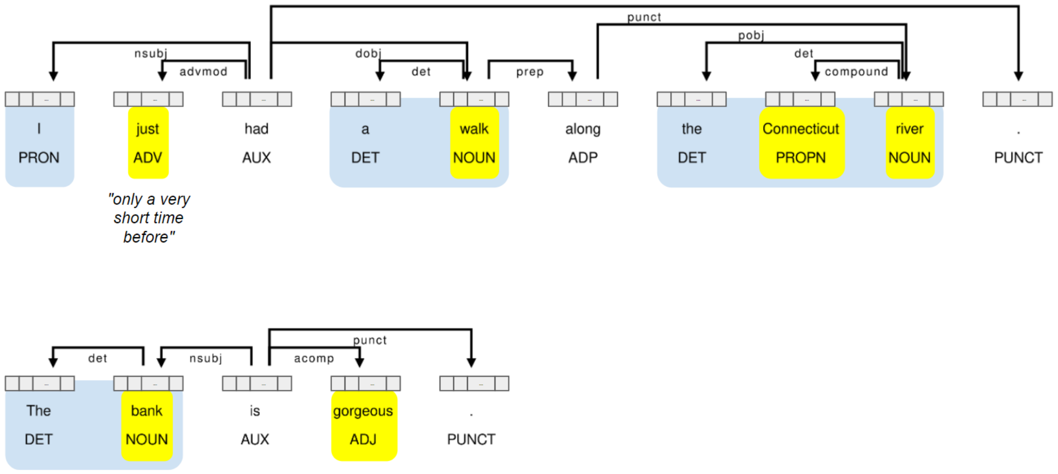


Figure 7: Add the contextually identified sense of *just*.

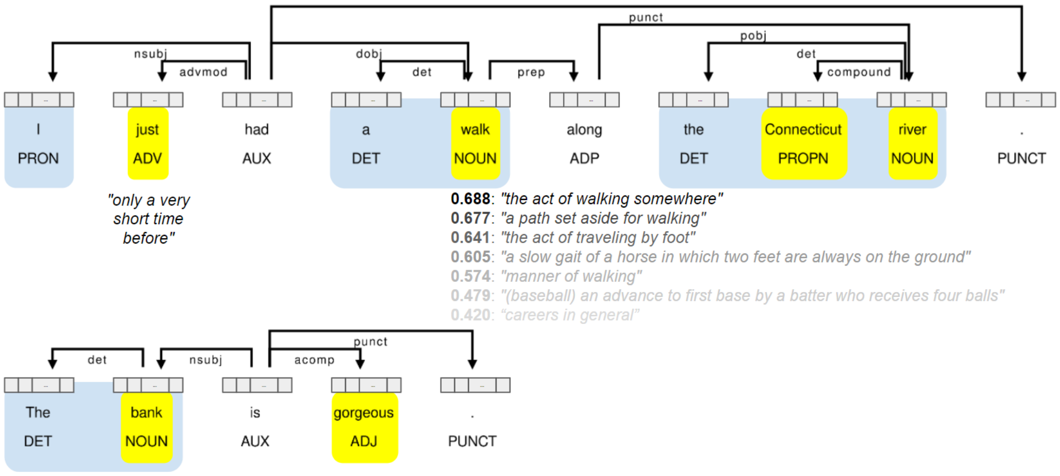

punct
nsubj
advmod
dobj
det
prep
pobj
det
compound
I
PRON
just
ADV
had
AUX
a
DET
walk
NOUN
along
ADP
the
DET
Connecticut
PROPN
river
NOUN
.
PUNCT
"only a very short time before"
0.688: "the act of walking somewhere"
0.677: "a path set aside for walking"
0.641: "the act of traveling by foot"
0.605: "a slow gait of a horse in which two feet are always on the ground"
0.574: "manner of walking"
0.479: "(baseball) an advance to first base by a batter who receives four balls"
0.420: "careers in general"
det
nsubj
punct
acomp
The
DET
bank
NOUN
is
AUX
gorgeous
ADJ
.
PUNCT

Figure 8: Calculate and rank similarity scores between *walk* and each of its dictionary sense definitions.

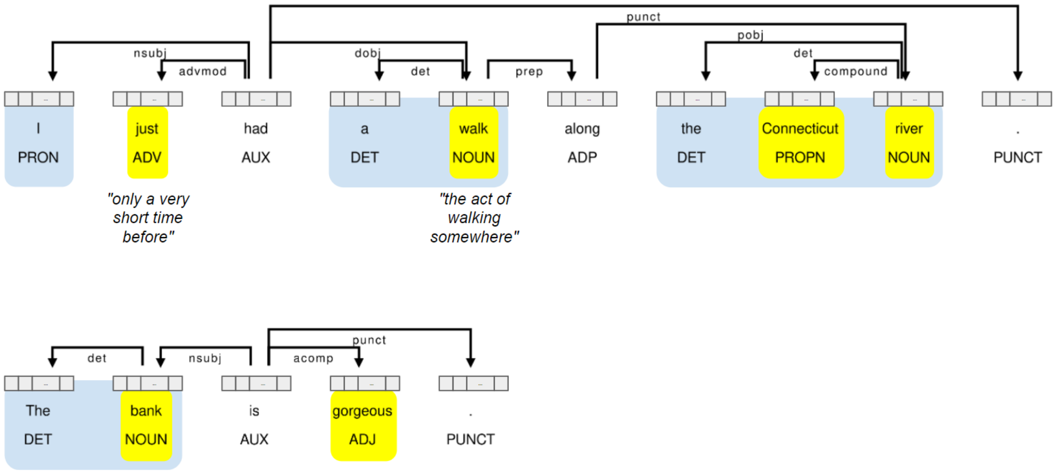


Figure 9: Add the contextually identified sense of *walk*.

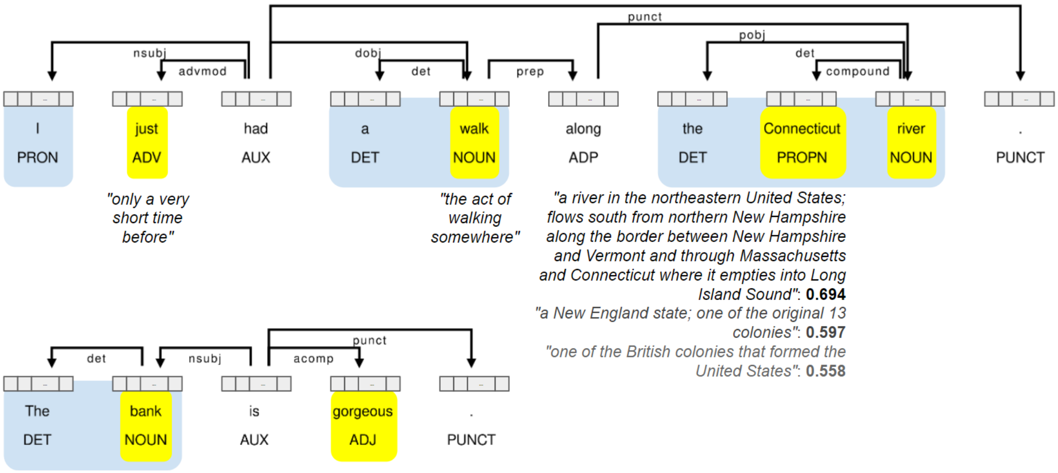

punct
pobj
nsubj
dobj
det
advmod
det
prep
compound
I
PRON
just
ADV
had
AUX
a
DET
walk
NOUN
along
ADP
the
DET
Connecticut
PROPN
river
NOUN
.
PUNCT
"only a very short time before"
"the act of walking somewhere"
"a river in the northeastern United States; flows south from northern New Hampshire along the border between New Hampshire and Vermont and through Massachusetts and Connecticut where it empties into Long Island Sound": 0.694
"a New England state; one of the original 13 colonies": 0.597
"one of the British colonies that formed the United States": 0.558
det
nsubj
punct
acomp
The
DET
bank
NOUN
is
AUX
gorgeous
ADJ
.
PUNCT

Figure 10: Calculate and rank similarity scores between *Connecticut* and each of its dictionary sense definitions.

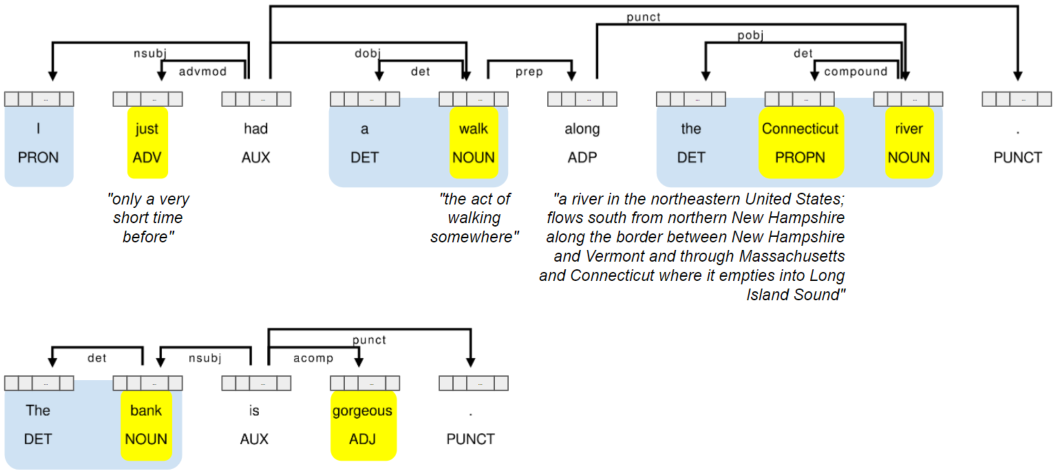

punct
pobj
nsubj
dobj
det
advmod
det
prep
compound
I
PRON
just
ADV
had
AUX
a
DET
walk
NOUN
along
ADP
the
DET
Connecticut
PROPN
river
NOUN
.
PUNCT
"only a very short time before"
"the act of walking somewhere"
"a river in the northeastern United States; flows south from northern New Hampshire along the border between New Hampshire and Vermont and through Massachusetts and Connecticut where it empties into Long Island Sound"
det
nsubj
punct
acomp
The
DET
bank
NOUN
is
AUX
gorgeous
ADJ
.
PUNCT

Figure 11: Add the contextually identified sense of *Connecticut*.

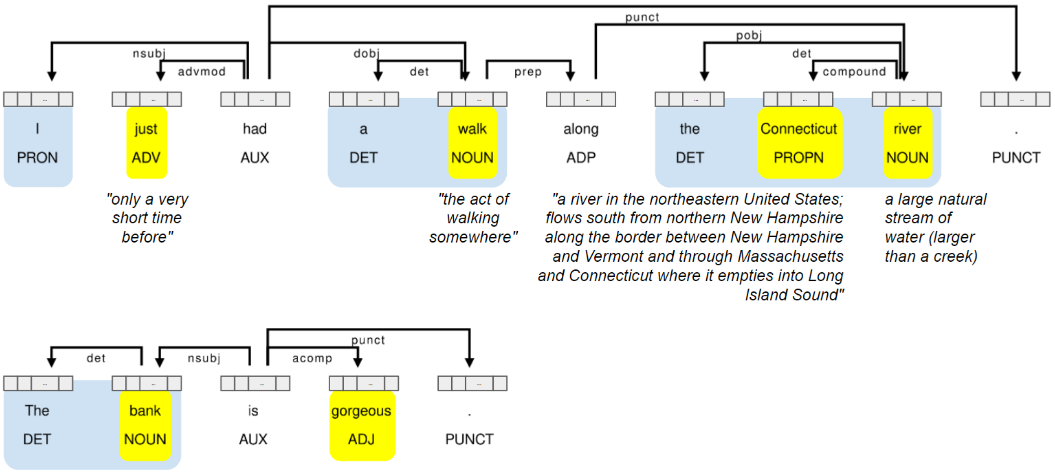

punct
nsubj
advmod
dobj
det
prep
pobj
det
compound
I
PRON
just
ADV
had
AUX
a
DET
walk
NOUN
along
ADP
the
DET
Connecticut
PROPN
river
NOUN
.
PUNCT
"only a very short time before"
"the act of walking somewhere"
"a river in the northeastern United States; flows south from northern New Hampshire along the border between New Hampshire and Vermont and through Massachusetts and Connecticut where it empties into Long Island Sound"
a large natural stream of water (larger than a creek)
det
nsubj
acomp
punct
The
DET
bank
NOUN
is
AUX
gorgeous
ADJ
.
PUNCT

Figure 12: Add the sole sense of *river*.

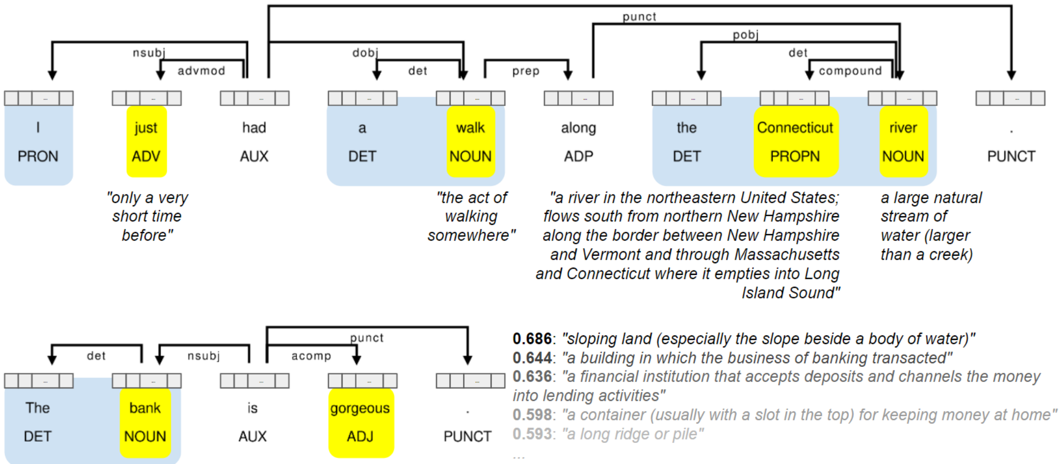

punct
nsubj
advmod
dobj
det
prep
pobj
det
compound
I
PRON
just
ADV
had
AUX
a
DET
walk
NOUN
along
ADP
the
DET
Connecticut
PROPN
river
NOUN
.
PUNCT
"only a very short time before"
"the act of walking somewhere"
"a river in the northeastern United States; flows south from northern New Hampshire along the border between New Hampshire and Vermont and through Massachusetts and Connecticut where it empties into Long Island Sound"
a large natural stream of water (larger than a creek)
det
nsubj
acomp
punct
The
DET
bank
NOUN
is
AUX
gorgeous
ADJ
.
PUNCT
0.686: "sloping land (especially the slope beside a body of water)"
0.644: "a building in which the business of banking transacted"
0.636: "a financial institution that accepts deposits and channels the money into lending activities"
0.598: "a container (usually with a slot in the top) for keeping money at home"
0.593: "a long ridge or pile"
...

Figure 13: Calculate and rank similarity scores between *bank* and each of its dictionary sense definitions. Before that, the contextual embedding of *bank* was recalculated based on a modified version of the second sentence, which is *The Connecticut river, the bank is gorgeous.* This enhancement of the real-time context-sensitive meaning of *bank* is informed by the fact that *the Connecticut river* is the noun phrase in the first sentence whose head noun, i.e. *river*, is the closest to *bank* in terms of similarity scores between their contextual embeddings.

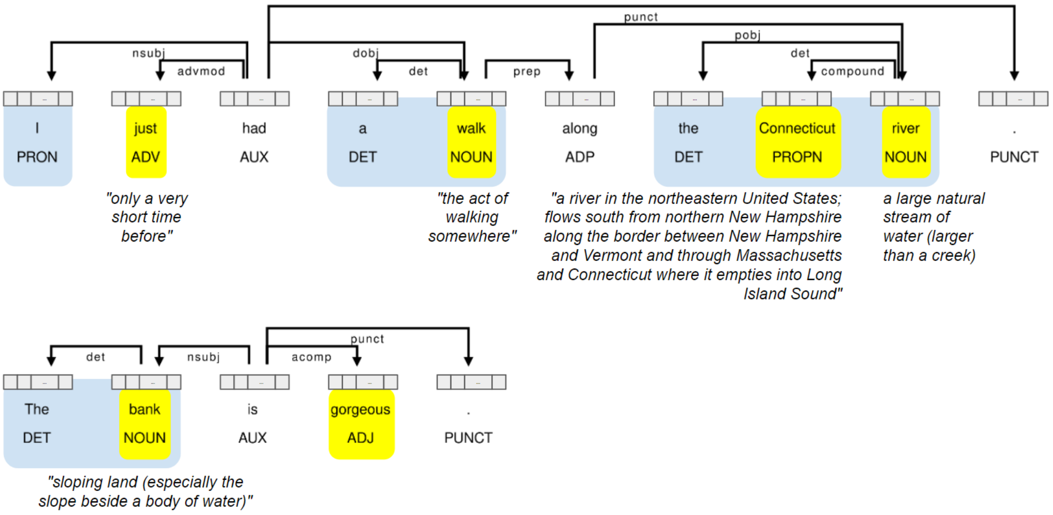

punct
nsubj
advmod
dobj
det
prep
pobj
det
compound
I
PRON
just
ADV
had
AUX
a
DET
walk
NOUN
along
ADP
the
DET
Connecticut
PROPN
river
NOUN
.
PUNCT
"only a very short time before"
"the act of walking somewhere"
"a river in the northeastern United States; flows south from northern New Hampshire along the border between New Hampshire and Vermont and through Massachusetts and Connecticut where it empties into Long Island Sound"
a large natural stream of water (larger than a creek)
det
nsubj
acomp
punct
The
DET
bank
NOUN
is
AUX
gorgeous
ADJ
.
PUNCT
"sloping land (especially the slope beside a body of water)"

Figure 14: Add the contextually identified sense of *bank*.

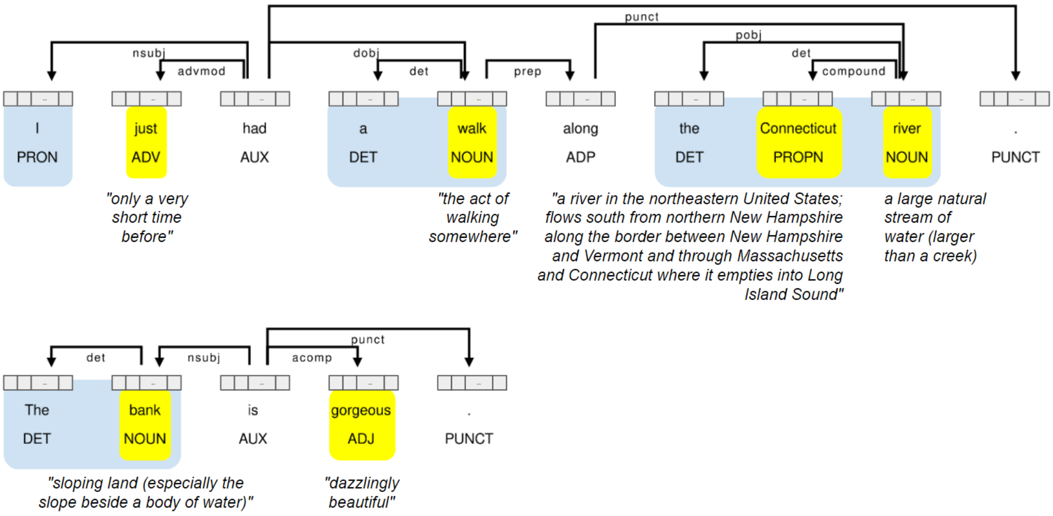

punct
nsubj
advmod
dobj
det
prep
pobj
det
compound
I
PRON
just
ADV
had
AUX
a
DET
walk
NOUN
along
ADP
the
DET
Connecticut
PROPN
river
NOUN
.
PUNCT
"only a very short time before"
"the act of walking somewhere"
"a river in the northeastern United States; flows south from northern New Hampshire along the border between New Hampshire and Vermont and through Massachusetts and Connecticut where it empties into Long Island Sound"
a large natural stream of water (larger than a creek)
det
nsubj
punct
acomp
The
DET
bank
NOUN
is
AUX
gorgeous
ADJ
.
PUNCT
"sloping land (especially the slope beside a body of water)"
"dazzlingly beautiful"

Figure 15: Add the sole sense of *gorgeous*.

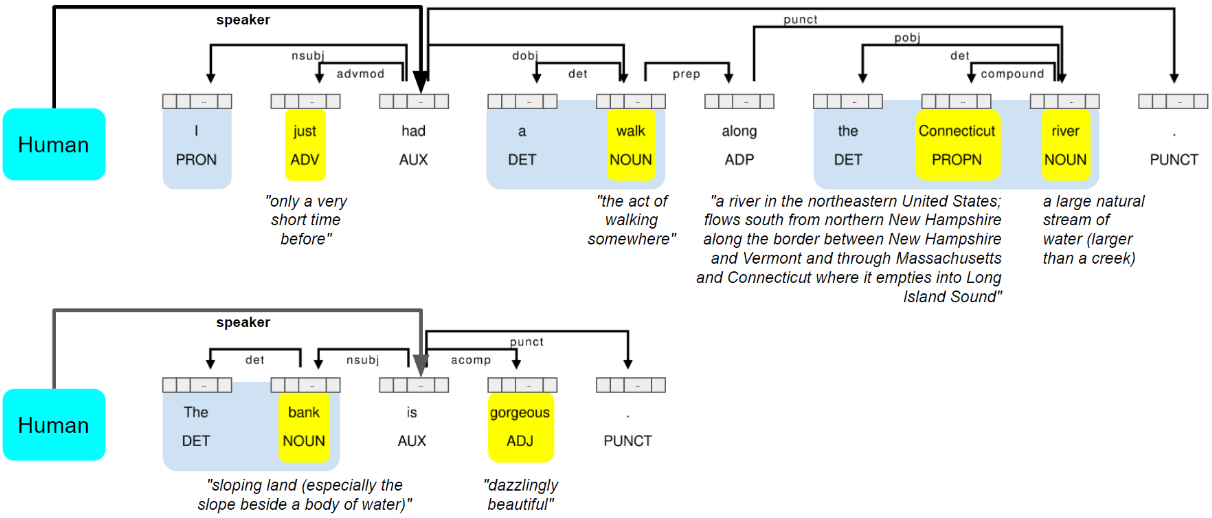

Human
speaker
punct
nsubj
advmod
dobj
det
prep
pobj
det
compound
I
PRON
just
ADV
had
AUX
a
DET
walk
NOUN
along
ADP
the
DET
Connecticut
PROPN
river
NOUN
.
PUNCT
"only a very short time before"
"the act of walking somewhere"
"a river in the northeastern United States; flows south from northern New Hampshire along the border between New Hampshire and Vermont and through Massachusetts and Connecticut where it empties into Long Island Sound"
a large natural stream of water (larger than a creek)
Human
speaker
det
nsubj
punct
acomp
The
DET
bank
NOUN
is
AUX
gorgeous
ADJ
.
PUNCT
"sloping land (especially the slope beside a body of water)"
"dazzlingly beautiful"

Figure 16: Add speaker tokens **Human** to each sentence.

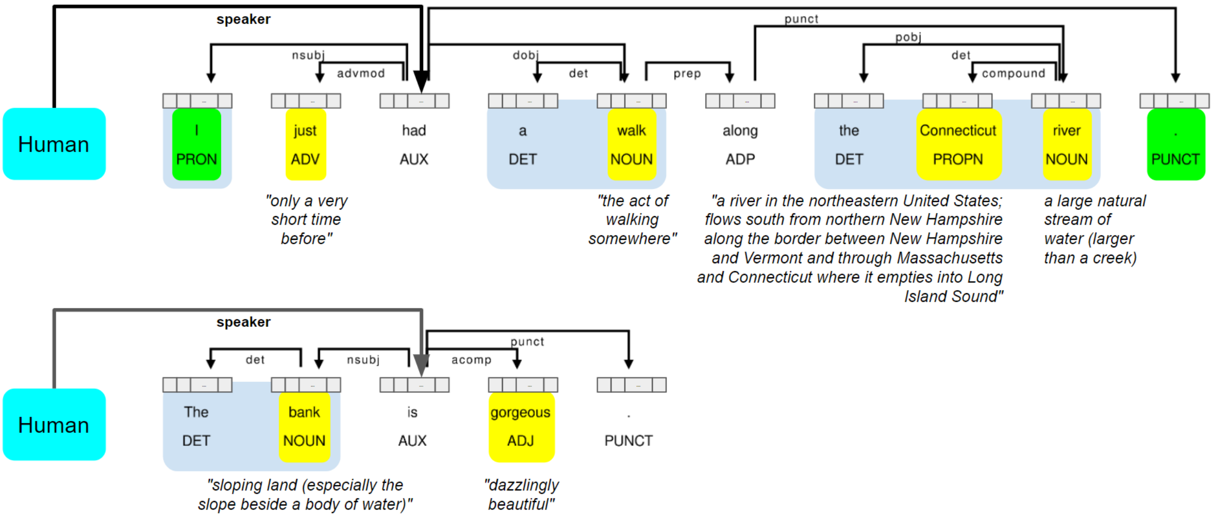

Human
speaker
punct
nsubj
advmod
dobj
det
prep
pobj
det
compound
I
PRON
just
ADV
had
AUX
a
DET
walk
NOUN
along
ADP
the
DET
Connecticut
PROPN
river
NOUN
.
PUNCT
"only a very short time before"
"the act of walking somewhere"
"a river in the northeastern United States; flows south from northern New Hampshire along the border between New Hampshire and Vermont and through Massachusetts and Connecticut where it empties into Long Island Sound"
a large natural stream of water (larger than a creek)
Human
speaker
det
nsubj
punct
acomp
The
DET
bank
NOUN
is
AUX
gorgeous
ADJ
.
PUNCT
"sloping land (especially the slope beside a body of water)"
"dazzlingly beautiful"

Figure 17: Identify emphasized positioning present in the first sentence. This is an instance of emphasized Self positioning, embodied by the first person pronoun *I* in a declarative sentence.

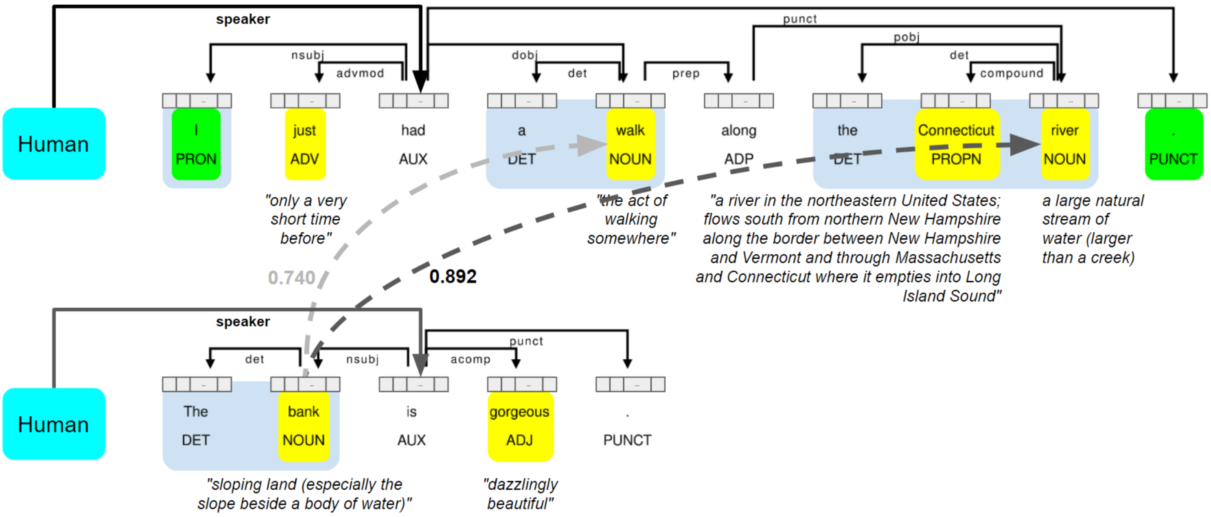

Human
speaker
nsubj
advmod
dobj
det
prep
punct
pobj
det
compound
I
PRON
just
ADV
had
AUX
a
DET
walk
NOUN
along
ADP
the
DET
Connecticut
PROPN
river
NOUN
.
PUNCT
"only a very short time before"
"the act of walking somewhere"
"a river in the northeastern United States; flows south from northern New Hampshire along the border between New Hampshire and Vermont and through Massachusetts and Connecticut where it empties into Long Island Sound"
a large natural stream of water (larger than a creek)
0.740
0.892
Human
speaker
det
nsubj
acomp
punct
The
DET
bank
NOUN
is
AUX
gorgeous
ADJ
.
PUNCT
"sloping land (especially the slope beside a body of water)"
"dazzlingly beautiful"

Figure 18: Identify potential discourse hook relations which connect the second sentence to the first sentence by calculating relevant similarity scores between the definitions of identified senses of head nouns of noun phrases.

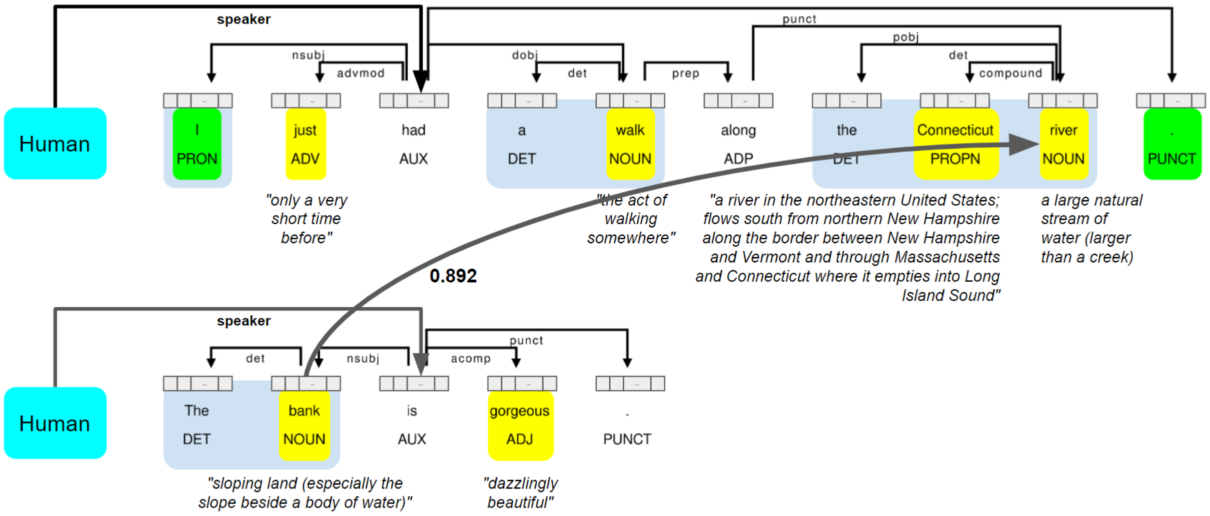

Human
speaker
nsubj
advmod
dobj
det
prep
punct
pobj
det
compound
I
PRON
just
ADV
had
AUX
a
DET
walk
NOUN
along
ADP
the
DET
Connecticut
PROPN
river
NOUN
.
PUNCT
"only a very short time before"
"the act of walking somewhere"
"a river in the northeastern United States; flows south from northern New Hampshire along the border between New Hampshire and Vermont and through Massachusetts and Connecticut where it empties into Long Island Sound"
a large natural stream of water (larger than a creek)
0.892
Human
speaker
det
nsubj
acomp
punct
The
DET
bank
NOUN
is
AUX
gorgeous
ADJ
.
PUNCT
"sloping land (especially the slope beside a body of water)"
"dazzlingly beautiful"

Figure 19: Select the most salient discourse hook relation, shaped by the similarity score between *bank* and *river*.

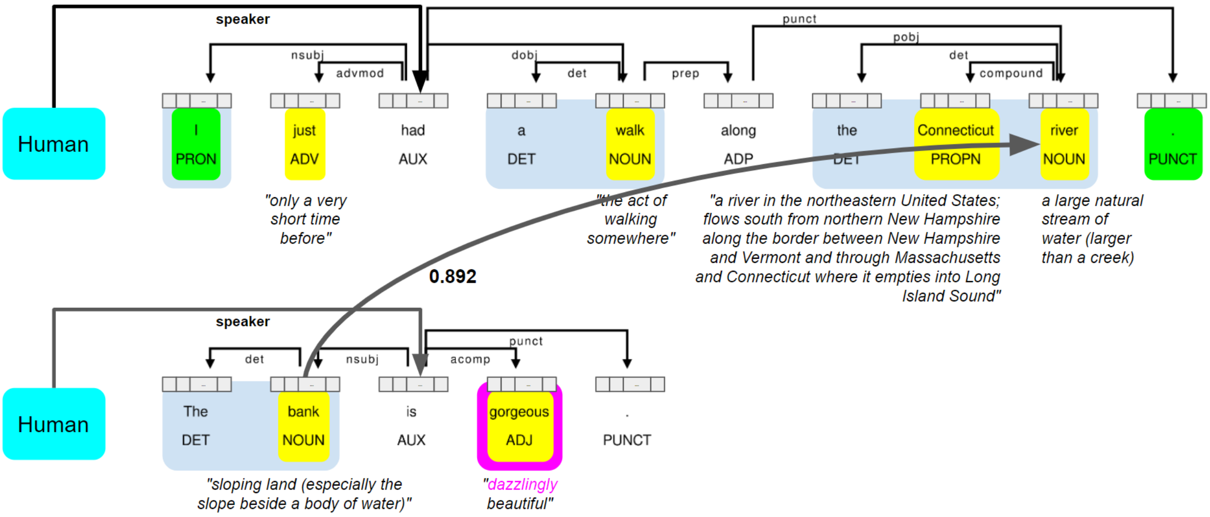

Human
speaker
nsubj
advmod
dobj
det
prep
punct
pobj
det
compound
I
PRON
just
ADV
had
AUX
a
DET
walk
NOUN
along
ADP
the
DET
Connecticut
PROPN
river
NOUN
.
PUNCT
"only a very short time before"
"the act of walking somewhere"
"a river in the northeastern United States; flows south from northern New Hampshire along the border between New Hampshire and Vermont and through Massachusetts and Connecticut where it empties into Long Island Sound"
a large natural stream of water (larger than a creek)
0.892
Human
speaker
det
nsubj
acomp
punct
The
DET
bank
NOUN
is
AUX
gorgeous
ADJ
.
PUNCT
"sloping land (especially the slope beside a body of water)"
"dazzlingly beautiful"

Figure 20: Identify highly expressive language present in the second sentence. This is an instance of positive sentiment expressed by the adjective *gorgeous*.