

## Pragmatic uses of *I don't know*, boosters, and hedges in text and talk

Allison Nguyen

Department of Psychology  
Illinois State University

ANGUYE9@ILSTU.EDU

Jean E. Fox Tree

Department of Psychology  
University of California Santa Cruz

FOXTREE@UCSC.EDU

Submitted 10/2024; Accepted 03/2025; Published online 05/2025

### Abstract

We examined how the phrases *I don't know*, *I dunno*, and *idk* are used in spontaneously produced speech and writing. We compared functions to the related phrases *totally*, *absolutely*, *sorta*, and *kinda*. We assessed usage across modalities (face to face, instant messaging, audiovisual), goals (tasks versus casual chat), and relationships (friends versus strangers). We also assessed where the phenomena occurred in a sentence, what words co-occurred with the phenomena, and what functions the phenomena served in the conversations. Communicators use phenomena differently depending on modality, goals, and relationships. We found that *I don't know* was used more often when people could access cues beyond the voice, and that both *I don't know* and *I dunno* can perform a variety of pragmatic functions. In instant messaging, *I don't know* has been lexicalized to *idk*, but *idk* does not have as many pragmatic functions as *I don't know* and *I dunno*.

**Keywords:** *I don't know*, discourse markers, hedges, boosters, communication modality

## 1 Introduction

In a commercial from 2007 a mother is talking to her daughter about their high phone bill, asking “Who could you be texting?” The daughter responds by saying out loud letters that are normally spoken, “*i d k my b f f Jill*” (Solarmax, 2007). In order for the mother to understand what the daughter means, the two have to engage in the process of mutually attempting to understand one another, particularly around what it means when the daughter says *idk*. While *bff* is an acronym for *best friend forever* (which the mother may or may not understand), *idk* can have multiple meanings, including marking uncertainty, prefacing disagreement, or highlighting commitment to the answer.

Conversations are processes that participants mutually engage in. A primary goal is to reach mutual understanding (Clark, 1996; Clark & Brennan, 1991; Stalnaker, 2002). In the process of reaching mutual understanding, or *common ground*, speakers and listeners must have access to previously grounded-upon information, as well as knowledge about their interlocutors, such as the type of relationship the two speakers share (Metzing and Brennan, 2003; Myrendal, 2019). Interlocutors use this knowledge to engage in small negotiations on things like referential expressions and conceptual pacts (Brennan & Clark, 1996; Soler et al., 2023). This is a delicate process, where negotiators must navigate politeness and social dynamics (Brown & Levinson, 1987; Beltrama & Schwarz, 2024; Scheibman, 2000; Tsui, 1991) while suggesting to their partner that they wish to negotiate on meaning. One way that speakers can navigate the complex social

©year Nguyen and Fox Tree

This is an open-access article distributed under the terms of a Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0/>).

aspects of conversations is by choosing to use words that cue their conversational partner about upcoming negotiations.

*Negotiation words* are a set of words that speakers can choose to deploy to highlight the negotiation aspects of establishing common ground. While negotiation is a key feature of reaching common ground, this process is not always explicitly expressed. When someone says “it’s fun,” the addressee understands that the speaker found something fun. However, saying “it’s kinda fun” requires negotiation about what is meant by “fun.” In spontaneously produced speech, words that highlight negotiation include those commonly identified as hedges (*sorta*, *kinda*) and words commonly identified as boosters (*absolutely*, *totally*), as well as items like *I don’t know*, which fall outside the classification of *hedge* or *booster*.

If conversational partners do not take contextual factors into account as part of the grounding process, we should expect to see no differences across text and talk. That is, speakers should use negotiation words in the same manner, with the same meaning across discourse modalities. Additionally, these expressions should not vary across discourse contexts, such as task or non-task conversation, or across friendship status. Said another way, an *I don’t know* should signal a lack of information to the same degree across modality type, task type, and friendship type.

In this paper, we assessed how conversation medium (in person face to face, audio-only, text-only), conversation type (on-task or off-task), and relationship type (friends or strangers) affected use of *I don’t know*, *absolutely*, *totally*, *kinda*, and *sorta*. Our assessment included examination of where *I don’t know*, boosters, and hedges were used and what they co-occurred with. We included four words of negotiation (two boosters, *absolutely* and *totally*, and two hedges, *sorta* and *kinda*) to understand how *I don’t know* contrasts with these similar expressions. The medium, conversation type, and relationship type varied across three corpora. The Roommates corpus was of face to face conversation in a laboratory, the Artwalk corpus was of on-task and off task audio conversation across friends and strangers in a partially-in-the-lab and partially-in-the-world setting, and the Instant Messaging corpus was of texted communication in a laboratory.

## 1.1 Boosters and Hedges

Negotiation words occur in both speech and writing, although the set of words found in different communicative settings vary widely. In writing, for example, the top boosters and hedges might be modal verbs like *could*, *would*, and *will* (Takimoto, 2015: 98) or *emotionally-charged intensifiers* such as *dishearteningly weak* and *particularly encouraging* (Salager-Meijer, 1994: 7). In speech, in contrast, the top boosters and hedges might be single words like *just* and *could* (Nuraniwati et al., 2021: 210) or two word expressions like *I think*, *you know*, and *sort of* (Holmes, 1990: 202). Defining boosters and hedges is a difficult challenge as the words used in different settings vary greatly; consider, for example, words used in academic writing (e.g., the words *thoroughly*, *undeniably*, *hypothetically*, and *reportedly*, Farrokhi et al., 2008: 97-98), versus prepared speaking like political speeches (e.g., the phrases *have to make sure*, *will never*, *can probably*, and *believe that*, Kashiha, 2020: 92-93), versus spontaneously produced dialogues (e.g., the words *seemingly* or *usually* or the phrase *in my view*, Islam et al., 2020: 3111). We now turn to definitions of boosters and hedges, along with information about the subset we explore in this report, which focuses on spontaneously-produced speech and writing.

Boosters are words that are canonically considered to mark conviction and knowledge on a stance (Hyland, 1998a, 2000b) – in other words, boosters are able to maximize commitment to an utterance. In certain scientific disciplines, for example, boosters are strategically deployed in writing in order to increase writer authority and to comment on the strength of arguments (Hyland, 1998a; Namaziandost, 2017). In planned and spontaneous speech, speakers use boosters to emphasize certainty (Holmes, 1990; Jalilifar & Alavi-Nia, 2012) or to make their words credible (Kashiha, 2020). Speakers and writers can also use boosters to exaggerate, which can sometimes imply sarcasm (D’Arcey et al., 2019).

Hedges, in contrast, are words or short phrases like *sort of* and *kind of* that communicate that the speaker may not be sure about the information being presented or how the addressee will receive it (Hyland, 1998a). In speech, many think of hedges as undesirable (e.g., Berger, 2024). However, there is evidence that hedges provide meaningful information, such as marking information as unsure (Jucker & Smith 1996) or indicating that speakers want to distance themselves from claims (Kashiha, 2020). Hedges also affect language processing, such as by increasing memory for items (Liu & Fox Tree, 2012). Liu and Fox Tree (2012) found that in storytelling contexts, hedged information was more likely to be omitted in a retelling, but more likely to be retained in recall tests.

There is an extensive list of both boosters and hedges. We selected four of the most common expressions for our study: the boosters *absolutely* and *totally* and the hedges *kinda* and *sorta*.

## 1.2 I don't know, I dunno, and idk

In addition to hedges and boosters, there are words that are neither hedges nor boosters, and yet serve some of the same functions. *I don't know* is a common expression in this category.

On its face, *I don't know* means the absence of knowing. The literal meaning is “I don't have the requested information or knowledge.” But people often use *I don't know* for a variety of pragmatic purposes. For example, they can say *I don't know* to convey that they both know the information and that they don't want to say what they know. Saying *I don't know* throws the ball into the addressees' court in a way that no other negotiation word does. *I don't know* is the opposite of a booster; on its face, it states a lack of knowledge as opposed to a certainty of knowledge. But at the same time, *I don't know* is not necessarily a hedge. *I don't know* can be said with certainty of lack of knowledge, for example. That is, speakers can say *I don't know* to indicate both they don't know and that they don't want to say (Pichler & Hesson, 2016; Grant, 2010; Brennan & Williams, 1995; Tsui, 1991). *I don't know* can also function as a marker of epistemic certainty and as a discourse marker (Grant, 2010; Kärkkäinen, 2010; Doehler, 2016). In addition, *I don't know* can be used to create distance between the speaker and a statement (Grant, 2010) or to steer the conversation away from a topic (Doehler, 2016).

The steering-away use of *I don't know* can be considered an *ostensible* use of *I don't know*. Ostensible language is language that has a particular use in its literal, on-record form (such as indicating lack of knowledge) but another use in an off-record form. Importantly, the off-record interpretation is meant to be recognized by the addressee as an insincere act by the speaker (which differentiates ostensible language from lying). For example, ostensible invitations occur when addressees are invited to an event with the implication that the invitee is expected to turn down the invitation, while maintaining social relationships (Isaacs & Clark 1990; Link & Kreuz 2005). By issuing the invitation, the inviter is off-record indicating that they would still like to spend time with the invitee, and by accepting the pretense, the invitee is willing to leave things off the record. Ostensible language, which often occurs with hedges and vague language, requires the two interlocutors to mutually recognize and collude together as to what the off-record reasons are. With an ostensible use of *I don't know*, the speaker wants to indicate that they know the information but they don't want to say it, with the reason remaining off-record. Such a reason could be that providing the information would hurt the feelings of the addressee, break someone else's confidence, or cause the spread of information the speaker wants kept private.

Unlike lying, ostensible uses of *I don't know* can convey subtler information than simple denials of knowledge. Indeed, *I don't know* in combination with pausing can be indicative of knowledge: When answering trivia questions, people who took time before saying *I don't know* were judged as more likely to know the answer than those who said *I don't know* more quickly (Brennan & Williams, 1995). Other uses of *I don't know* include managing turns, expressing attitudes, and reducing commitment to statements (Diani, 2004; Grant, 2010; Weatherall, 2011).

*I don't know* can also be used in turn organization. It can indicate that the speaker is planning to add additional information or mark a turn as containing non-standard information (Doehler 2016). *I don't know* also allows the speaker to exit a turn, even when the turn is not complete (Doehler 2016). Thus, *I don't know* has multiple uses in conversation, both literal and pragmatic. What interpretation communicators adopt might depend on many contextual factors, including what the communicators know about each other, what communicative modality is being used, and the purpose of the conversation.

There are two other forms of *I don't know* that can appear. *I dunno* is a well attested reduced form of *I don't know* that appears in spoken speech and should be considered a *bound phrase*, just like *I don't know* (Aijmer, 2009; Scheibman, 2000). Like *I don't know*, *I dunno* can serve a variety of pragmatic functions. It can be used in a literal sense (“I don't have the information requested”), but other attested uses include hedging, softening or avoiding disagreement, and turn exchange (Schiebman, 2000). Likewise, *idk* is a written form of *I don't know* that appears in instant messaging, text messaging, and other forms of technology-mediated communication, and likely serves similar functions, though usage has not been studied.

### 1.3 Differences Across Settings

Negotiation words are primarily produced in unprepared, unrehearsed settings. Many of them can be categorized as *discourse markers*, or words that aid the smooth flow of conversation (Fox Tree, 2010a, 2015b). Other spontaneous elements that contribute to conversation management include *backchannels* (Nguyen et al., 2024; Tolins & Fox Tree, 2014; Zellers, 2021) and *fillers* (Clark & Fox Tree, 2002; Goodwin, 1981; Walker et al., 2014). Spontaneous phenomena generally vary across conversational settings. For example, the backchannel *yeah*, which is used to indicate a speaker should continue telling a story, is more commonly used in face to face communication compared to telepresence communication (Fox Tree et al., 2021). As another example, discourse markers including *well*, *you know*, and *like* were found more frequently in hyperpartisan communication compared to nonhyperpartisan communication (Nguyen et al., 2022). Additionally, discourse markers are used more frequently in tasks compared to casual conversations over the phone, but used less frequently in tasks compared to casual conversations over instant messaging (Guydish et al., 2024). For these reasons, we believe that *I don't know*, *absolutely*, *totally*, *sorta* and *kinda* may also vary across settings.

One setting where a difference has been found is in different types of English for *I don't know*. New Zealand English speakers tended to use *I don't know* to avoid disagreement more often than British English speakers, and British English speakers used *I don't know* to avoid making a commitment more often than New Zealand speakers (Grant, 2010). In our study, we investigated English spoken in the same community (American English), but across different modalities (video chat, phone, text) and different types of conversations (task-related, chit-chat). We also examined potential differences in friendship status.

There is also reason to think that friends and strangers may use negotiation words differently. Within close social relationships (friends), it might be more acceptable to be vague, which would minimize the use of *I don't know* as a marker of lacking knowledge (Bristol & Rossano, 2020), and increase the number of hedges (*sorta*, *kinda*). In contrast, casual acquaintances or strangers may prefer *I don't know* over vague statements (Bristol & Rossano, 2020). Friends and strangers might use boosters (*absolutely*, *totally*) to emphasize commitment, but it is possible that close social bonds influence use: If emphasis on commitment is considered rude, then friends may be more likely to use them as friends are willing to tolerate more rudeness from friends than strangers (Gupta et al., 2007).

*Collocations*, or co-occurring words, may also vary across settings. *I don't know* often co-occurs with discourse markers such as *well*, *oh*, *I mean* and *you know* (Grant, 2010, see also Diani, 2004, and Aijmer, 2009). When used together with other markers, negotiation markers may add

nuance to conversational negotiation and management. For example, previous researchers suggested that *well* and *oh* commonly appear before disagreements or before expressing hesitation and that *I mean* and *you know* increase tentativeness and distance the speaker from the opinion (Grant, 2010). If a *well* is used to indicate an upcoming disagreement, a co-occurring *I don't know* or hedge might help distance the speaker from the upcoming disagreement in order to soften the impact of the disagreement.

#### 1.4 Current Study

While a lot of work has been done on *I don't know*, *absolutely*, *totally*, *kinda* and *sorta*, their uses have not been studied across conversation modalities, types of conversations, and friendship statuses. Word co-occurrences for boosters and *I don't know* have also not been explored. Co-occurrences have been explored for the hedges *sort of* and *kind of/kinda* in the British National Corpus (Gries & David, 2007). The British National Corpus is composed of 90% written language and 10% transcripts of spoken language produced between 1960 and 1994, with material from 1960 to 1975 being written only (Burnage & Baguley, 1996). Some of the conclusions were that *kind of* is more common in writing and *sort of* in speaking, that *kind of* co-occurs with verbs but *sort of* co-occurs with nouns, that *kind of* co-occurs with emotional states and *sort of* with colors, that *kind of* co-occurs with negative words like *depressing* more often than *sort of* but that both occur frequently with positive words like *fun*, and that *sort of* modifies quantities and people.

In the current report, we tested how *I don't know* varied across medium, relationship, and task type, including how it was used in comparison to common boosters and hedges. We also compared uses across our chosen hedges (*kinda* and *sorta*) and boosters (*totally* and *absolutely*).

There are many ways to use these words that go against conventional thought. Without examining what people say while communicating, it might seem that the subject and verb phrase *I don't know* would either stand alone or occur at the beginning of a sentence (such as before *that* or *whether*). But examples of *I don't know* being used in other ways can be found in spontaneously produced speech. In the following, *I don't know* occurs within a prepositional phrase:

- (1) “we got home like really really late like at at like I don't know like 2:30 or 3:00” (spoken example in Fox Tree, 2006, p. 739)

In the following it occurs within a modifying phrase:

- (2) “he pulls out an ant house um with all this ant furniture in it and stuff, a little ladder for the ant to com- climb up to and ring a bell, and little just- I don't know, kinda like what you'd see in a gerbil cage I guess just for ants” (spoken example in Fox Tree & Schrock, 1999, p. 285; the example is from a corpus collected by Herbert Clark of speakers retelling a Monty Python sketch)

In the following it occurs in a verb phrase:

- (3) “I've been trying to like I don't know understand the differences between Northern California and Southern California” (spoken example from Roommates)

In the following *I don't know* and *I dunno* occur within adjectival phrases:

- (4) “the giant sculpture that I'm looking at is not small or gumdrop like I mean it looks like, I don't know uh oh a pendulum that's inside a giant grandfather clock” (spoken example from Artwalk)
- (5) “it make a noise exactly- it sounds exactly like I dunno like a dog or a horse” (spoken example adapted from Overstreet, 2005, p. 1865)

We can also see *I dunno* open a turn without being followed by a subordinate clause (such as a clause beginning with *that* or *whether*):

- (6) “I dunno it's very abstract, like the bottom part” (spoken example from Artwalk)

On the flip side, a modifier like *kinda* might seem like it should only be found preceding what it modifies in the middle of a sentence. That is, a sentence that starts or ends with *kinda* would not be produced. But examples of such uses can be found in spontaneously produced communication. In the following two examples, *kinda* appears at the beginning and *sort of* at the end:

- (7) “sometimes I have to like, force it out of her. (laughs) so like, are you mad at me!? I know you’re mad at me! (laughs) kinda like that” (spoken example adapted from Thorne et al., 2009, p. 640)
- (8) “it’s kind of walking up stairs to the left sort of” (spoken example in Fox Tree & Mayer, 2008, p. 167)
- (9) “okay yeah I think I see, the cut out corner is like wobbly cut kinda?” (texted example from IM)

The following ungrammatical structure has both *I dunno* and *sort of*, where *I dunno* comes after an adjective in what appears to be a noun phrase, and *sort of* appears to begin a new phrase:

- (10) “on the top is kind of this like big I dunno sort of has a rectangle on its left side (spoken example in Fox Tree & Mayer, 2008, p. 174)

By not defining phenomena of interest by conventional syntactic form (e.g., *I don’t know* is a sentence with a noun and a verb and *kinda* is a modifier), we allow for productions that violate conventional expectations about grammaticality and acceptability. As the examples provided here show, in everyday language use people do violate syntactic expectations.

We predicted that uses will vary across task contexts. In the storytelling chit-chat context, we predicted that *I don’t know* would be more likely to mean “I don’t want to talk about it” than “I don’t have the information” because people are offering their own stories, which, presumably, they could simply not express instead of stating that they have no information about them. In the task contexts, in contrast, we predicted that *I don’t know* will be more likely to be used to mean “I don’t have information.” This would be observed in the on-task portions of the Artwalk and IM corpora. In object-identification tasks, people literally need to express that they don’t know what an object is. That is, the tasks afford the “I don’t have access to the information” meaning over the “I don’t want to talk about it” meaning. In both Artwalk chit-chat and IM chit-chat, we predict that *I don’t know* will mean “I am marking uncertainty,” where a participant avoids answering or alternatively, knows the answer and gives it but willingly distances themselves from it. In Roommates (audio-visual, face to face), we predict that there will be fewer expressions of *I don’t know* than in Artwalk chit-chat. We also predict that *I don’t know* will serve the same purpose as in the chit-chat — it will mean “I am marking uncertainty” more often than any other meaning, due to the ostensible use of *I don’t know*. Thus, in Roommates, which is a conversational task that began with storytelling, it would be conversationally odd for a speaker to mark their own uncertainty about a story that happened to them via *I don’t know*.

We also predict that uses will vary based on communicative modality. We predict fewer numbers of *I don’t know/idk* in texted communication compared to spoken communication. In the on-task communication, we predicted fewer texted expressions *I don’t know/idk* compared to spoken *I don’t know*. There is a start-up cost to sending a message, so people are more likely to send informative messages rather than messages like *I don’t know/idk* (unless they truly do not know!). So, while we predicted that *I don’t know/idk* would more often mean lack of knowledge in the texted task-based communication, we predicted fewer numbers of *I don’t know/idk* overall in text communication. We also predicted that there will be more *I don’t knows* in task-based Artwalk compared to task-based IM due to the start-up costs associated with typing compared to speaking. In both corpora, in the task-based portions, we predict that *I don’t know* will more often mean “I don’t have information.”

Of the non-texted modalities, we predicted that there will be more *I don’t know/I dunnos* in face to face communication than phone communication because addressees will be better able to indicate their lack of knowledge or willingness to answer without seeming rude when they have

additional facial cues. That is, in a conversation where the speakers have access to non-verbal visual cues, using an expression with multiple meanings like *I don't know* can be unambiguous because of the additional cues. Alternatively, with only voice cues, *I don't know/I dunno* may be interpreted as a rude unwillingness to answer rather than an ostensible distancing from an answer.

For Artwalk chit-chat and task, where we are able to look at conversations between friends and strangers, we predict that friends will use *I don't know/I dunno* to mean “I am marking uncertainty” and strangers will use it to mean “I don't have information.” See Table 1 for a list of predictions.

Artwalk - Task-based	More <i>I don't knows</i> than in Artwalk chit-chat More <i>I don't knows</i> than in IM task <i>I don't know</i> means “I don't have information”
Artwalk - Chit-chat	<i>I don't know</i> means “I am marking uncertainty”
IM - Task based	More <i>I don't knows</i> than in IM chit-chat <i>I don't know</i> means “I don't have information”
IM - Chit - chat	<i>I don't know</i> means “I am marking uncertainty”
Roommates	More <i>I don't knows</i> than in Artwalk chit-chat <i>I don't know</i> means “I am marking uncertainty”
Friends	<i>I don't know</i> means “I don't want to talk about it”
Strangers	<i>I don't know</i> means “I don't know”

**Table 1.** Predictions.

## 2 Method

*I don't know* and its variants (*I dunno*, *idk*) were analyzed across three corpora for their quantity and function. Comparison expressions were analyzed (*absolutely*, *totally*, *sorta*, *kinda*) for their occurrence across corpora.

### 2.1 Corpora Assessed

Use of *I don't know* and the four comparison markers were assessed in three corpora: (1) the Roommates corpus of in person face to face communication (Bryant, 2010), (2) the Artwalk corpus of audio-only communication (Liu et al., 2016), and (3) the Instant Messaging corpus (henceforth the IM corpus) of texted communication (Guydish & Fox Tree, 2022). Please see Table 2 for a breakdown of the corpora.

Corpus	Word Count	Dyad Type	Conversation Type	Modality
Artwalk	236,629	Friends / Strangers	Task / Chit-chat	Audio-only
IM	82,691	Strangers	Task / Chit-chat	Text-only

Roommates	58,441	Strangers	Storytelling Chit- chat	Face to face chat
-----------	--------	-----------	----------------------------	----------------------

**Table 2.** Word Count, Dyad Type, Conversation Type, and Modality Type for the Three Corpora Assessed.

The Roommates corpus consisted of in-person spoken communication where participants were asked to come into the lab and have a conversation (see Bryant, 2010, for a full description). The corpus was collected at a West Coast research university. There were 22 transcripts of different dyads in the Roommates corpus. Participants were initially asked to tell a story about a roommate, and the conversation was allowed to continue from there, a task we will refer to as *storytelling chit-chat*. Because this was an open-ended task, there is no defined measurement of success, as there is in referential tasks. However, the dialogue was prompted by a task in a laboratory setting. This type of spontaneous conversation resembles both task and non-task related communication with respect to other conversational phenomena, such as backchannels (see Nguyen et al., 2024 for more). We predict that storytelling chit-chat will more closely resemble naturalistic conversation (the *chat* portions of the Artwalk and IM corpora) than task conversation (the *task* portions of the Artwalk and IM corpora) for *I don't know*.

Artwalk is a corpus of conversations between dyads via telephone (Liu et al., 2016). There are 69 transcripts total; 48 are part of the original corpus and are publicly available (Liu et al., 2016), and 21 are part of an expansion (Guydish et al., 2021). One participant was in the lab (the *director*) and the other was in a West Coast downtown location with public art (the *follower*). The director was responsible for giving the follower directions to different artworks installed in the location. Conversational contributions related to finding the artworks comprises the task-oriented portion of the corpus. Artwalk also, importantly, has spontaneously generated off-task conversation as well, the chat or chit-chat portion of the corpus (see Nguyen et al., 2024, for an argument as to why these sections of conversation should be considered naturalistic). Out of the 69 dyads, Artwalk has 47 dyads with information recorded about the relationship between the people in the dyad; of these 47, 25 dyads were composed of friends and 22 dyads were composed of strangers.

Like Artwalk, the IM corpus has both on-task and off-task sections (see Guydish & Fox Tree, 2022). There are 65 transcripts in the IM corpus. This corpus was collected at a West Coast research institution. Pairs of dyads engaged in a referential task card-matching activity where they had to identify abstract shapes together. Participants also had sections where they could have conversations unrelated to the task. These conversations are analogous to the chat portions of Artwalk. All conversations in the IM corpus took place between strangers.

## 2.2 Coding

*I don't know*, *I dunno*, *idk*, *absolutely*, *totally*, *sorta*, and *kinda* and their surrounding contexts were extracted using an automated method. Extracted information was double checked by a research assistant. The assistant ensured that the correct items of interest were pulled and also ensured that the surrounding context was obtained.

Occurrence (how often the phrase occurs) and co-occurrence (what it occurs with) rates were counted. Functions were coded using an adapted schema based on Grant (2010). All coding was done by two research assistants, with a third coder brought in to resolve disagreements. Location and co-location information was also coded for these words. Utterances were looked at individually, but the phrases of interest were coded in the context of the utterance they were spoken in. Research assistants were trained on definitions developed and refined over several iterations of coding (see Table 3 for the definitions the RAs were provided), and then coded all expressions of



interest individually for occurrence, co-occurrence, location, and co-location. IRRs were calculated for all coding and interpreted using the ranges proposed by Landis and Koch (1977).

Grant's (2010) coding scheme was adapted to code the expressions of interest. In addition to Grant's original categories, four more were added: *prefacing agreements*, *expressing agreement*, *highlighting commitment to the answer*, and *maximizing compliments*. Because one suggested use of *absolutely* and *totally* is that they increase commitment, we added these four categories to capture these uses. There is no "seeking assessment" (the opposite of *avoiding assessment*) because assessment is focused on evaluating the interlocutor's contributions, not one's own. This means that "seeking assessment" would look like agreement or disagreement, both behaviors captured under other codes. There is no "seeking commitment" (the opposite of *avoiding commitment*) because "seeking commitment" can be interpreted as simply making a statement (e.g., "he's kinda tall" is avoiding commitment, but "he's tall" is making a statement). That is, speakers are assumed to commit as much as they can to every utterance, and it is only by marking it in some sense that commitment is lowered. In contrast, "highlighting commitment" can be interpreted as adding extra commitment, which isn't the counterpart to avoiding commitment (cf., "he's absolutely tall"). Please see Table 3 for a synopsis of the coding scheme. Coding categories from Grant (2010) are listed first, and the newly developed categories are listed after.

	Category	Definition	Example
Grant (2010)	Inability to provide information / Insufficient knowledge	The expression is used when the speaker has a lack of knowledge	"yeah it wo- it won't let me take the picture right now <b>I don't know</b> what to do" (Artwalk corpus)
	Prefacing Disagreement	The expression is used to manage the social relationship while disagreeing with the speaker	"[laugh] it's <b>kinda</b> not i uh i mean it think it's just supposed to be just suggestive and amorphous i don't i don't know i'm it could be just me" (Artwalk corpus)
	Avoiding Disagreement	The expression is used when the speaker wants to avoid giving a negative response	"oh I <b>totally</b> listen . didn't I remember that you live off campus[?]" (Roommates corpus)
	Avoiding Assessment	The expression is used to avoid judging the truth of their interlocutor's statements	D: yeah they can represent people, <b>I don't know</b> but uh they are very simplified. you only see the oval shape
	Avoiding Commitment	The expression is stressing the speaker's lack of confidence of the truth of the utterance	"D: uhh it looks <b>kinda</b> like uhm little brown and yellowish? i think" (Artwalk corpus)
	Minimizing Compliment	The expression is downplaying the speaker's confidence in a	[no examples in corpora]

		compliment (from the interlocutor)	
	Marking Uncertainty	The expression is stressing the uncertainty of the utterance	“Yeah but um the bottom of it its [sic] <b>kinda</b> like the stone texture and everything” (Artwalk corpus)
	Unclear / Missing data	The data can't be coded due to an inability to make a judgment (lack of context, etc)	“I <b>kinda</b> wan” (Artwalk corpus)
New addition to coding scheme	Prefacing Agreement	The expression is used to manage the social relationship while agreeing with the speaker	“yeah <b>kinda</b> yeah” (Artwalk corpus)
	Expressing Agreement	The expression is used when the speaker wants to give a response in agreement	“Yeah, <b>*totally!*</b> ” (Roommates corpus)
	Highlighting commitment to the answer	The expression is emphasizing the speaker's confidence in an expressed compliment	“there was <b>absolutely</b> no drama at all and then to *go from that to*” (Roommates corpus)
	Maximizing Compliment	The expression is emphasizing the speaker's confidence in an expressed compliment (from the interlocutor)	[no examples in corpora]

**Table 3.** Coding scheme for analysis with the marker of interest bolded. The first 8 categories are from Grant (2010). The bottom 4 categories are created for this analysis.

Because location can be a cue for pragmatic function (Aijmer, 2009), location was also coded. *Start of utterance* indicated the word/phrase started the utterance or was a standalone phrase. *End of utterance* indicated the word/phrase ended the utterance, or appeared within the last sentence of the utterance. *Middle of utterance* was selected for all other locations. Co-locators were coded by counting the bigrams immediately to the left and right of the word/phrase.

### 3 Results

Results for each corpus are presented below. Location information is presented across corpora in Table 4. Following that, cross-corpora comparisons are reported.

#### 3.1 Roommates

All 21 transcripts in the Roommates corpus were coded following the procedure laid out above. Due to the difficult nature of the task, interrater reliability across the entire corpus was fair (Fleiss's kappa = .32).

##### 3.1.1 Hedges and Boosters

Of the four words besides *I don't know*, only *totally* and *kinda* appeared more than a handful of times. About 40% of the uses of *kinda* was to mark uncertainty, and about 30% were to avoid commitment (see Table 5 in section 3.4 for uses). The most common words to precede *kinda* were *it* and *I*; the most common word to follow was *like* (see Table # for collocations). About 60% of the uses of *totally* were *highlighting commitment to the answer*. *Like* was the most common word before *totally* ("that's like totally the way [she] is"). About half the words after *totally* were verbs. *Absolutely* appeared once and *sorta* appeared twice.

##### 3.1.2 I don't know

*I don't know* occurred 219 times in the Roommates corpus. *I dunno* and *idk* appeared zero times. *I don't know* was most commonly coded as *marking uncertainty* (56 occurrences) followed by *indicating insufficient knowledge* (33 occurrences). *I don't know* was difficult to code because it was often used as a stand-alone particle, resulting in a code of ( *missing information* (23 occurrences). The rest of the data was split between the rest of the coding scheme. In Roommates, people told each other stories and chatted. They may have been unsure of how much their interlocutor might agree, distancing themselves from their statements with *I don't know*.

All 219 appearances of *I don't know* were coded for co-located words. *Well*, *oh*, and *I mean* – words that had previously co-occurred with *I don't know* in Grant (2010) – did appear with *I don't know* in the Roommates corpus. *Well* appeared twice, once following *I don't know* and once before. *Oh* appeared 4 times, twice before *I don't know* and twice after *I don't know*. *I mean* appeared 4 times, 3 times before *I don't know* and once after. *You mean* was not present. Other common co-appearing words included *like*, which appeared 32 times before *I don't know* ("like I don't know") and 22 times after ("I don't know like this girl [name]"), as well as *wh-* words, which appeared three times before, and 19 times after ("I don't know why"). Phrases like *I think*, *I feel*, and *I guess* also appeared before and after *I don't know*. These are phrases that express a thought or opinion, so it makes sense that a speaker using them might qualify them with *I don't know*, and also highlight the function *I don't know* plays in softening the commitment to an upcoming statement.

#### 3.2 Artwalk

All 59 transcripts in the Artwalk corpus were coded following the procedure laid out above. Inter-rater reliability was moderate (Fleiss' kappa = .55). Below, we discuss the hedges and boosters (*absolutely*, *totally*, *sorta*, *kinda*) first, followed by *I don't know*. We then discuss marker use by friendship status.

##### 3.2.1 Hedges and boosters

Like in Roommates, only *kinda* and *totally* appeared in large numbers. *Kinda* was the most frequent marker used in Artwalk (306 occurrences). The other three markers were used five times less frequently (61 occurrences). *Kinda* was much more frequent in task-related communication

(89% of use cases) compared to chit-chat (11% of use cases). *Totally* appeared 41 times in the Artwalk corpus, with a roughly even split between task-related (54% of uses) and chit-chat conversation (46% of uses). *Sorta* was rare in the corpus, appearing 13 times. It appeared 11 times in task-related conversation, and twice in chit-chat. *Absolutely* was also rare in the corpus, appearing 7 times in the corpus, with 71% of those uses in task-related conversation and 29% in chit-chat. Please see Table 5 in section 3.4 for a detailed breakdown.

### 3.2.2 I don't know

*I don't know* appeared 168 times in the Artwalk corpus, 138 times in task-related conversation, and 30 times in chit-chat conversation. All instances of *I don't know* were coded for co-located words.

In task-related conversation, 40 uses of *I don't know* were coded as *indicating insufficient knowledge*, 28 were coded as *marking uncertainty*, 7 were coded as *avoiding commitment* or *avoiding disagreement*, and 8 were coded as *missing or insufficient data*. *I don't know* appeared as a reduplication multiple times, with 7 sequences of an *I don't know* followed by another *I don't know*. *I don't know* was preceded by *um* or *uh* a total of 11 times (“um, I don't know”). A common co-locator with *I don't know* was *if*, following *I don't know* 28 times (“I don't know if”). *Wh-* words were also common after *I don't know*, appearing 17 times. *How* also appeared 11 times after *I don't know*. *I think*, *I feel*, and other expressions of thought, feeling, or belief did occur after *I don't know*, which is in line with *I don't know* serving as a marker of uncertainty (“I don't know, I think we're doing a good job”).

In chit-chat conversations, 10 uses were coded as *indicating insufficient knowledge*, 5 as *marking uncertainty*, and the rest were split between *highlighting commitment*, *prefacing disagreement*, and *missing or insufficient data*. *I don't know* was preceded by *but* 4 times (“but I don't know”), *yeah* 3 times, and *um* and *like* 1 time each. *I don't know* was followed by *if* 4 times, *wh-* words 3 times, *like* 2 times, and *how* 1 time.

*I dunno* is a spoken, reduced form of *I don't know*. Because Artwalk is a spoken corpus, *I dunno* was examined for usage. *I dunno* occurred 136 times in the Artwalk corpus, 100 times in task-related conversation and 36 times in chit-chat conversation. All instances of *I dunno* were coded for co-located words.

In the task-related conversation, 27 uses were coded as *marking uncertainty*, 17 uses were coded as *indicating insufficient data*, and the rest of the uses were split between *avoiding disagreements* and *avoiding commitment*. Before *I dunno*, *like* was the most common word, with 11 uses (“like I dunno”). Other common words include *uh/um* (9 occurrences), *it* (3 uses) and *well* (2 uses; for example, “well I dunno”). *You know* occurred once. After *I dunno*, *wh-* words were very common (15 uses), followed by *how* (8 uses), *if* (7 uses), and *like* (6 uses). There were 7 occurrences of expressions like *I guess* occurring after *I dunno*, highlighting *I dunno*'s use as a marker of knowledge state (“I dunno, I guess”).

In the chit-chat conversation, *I dunno* was coded as *marking uncertainty* 14 times and as *indicating insufficient data* 13 times. The rest of the data was split between *avoiding commitment*, *avoiding disagreement*, and *prefacing disagreement*. Looking at words that co-locate to the left of *I dunno*, *uh* preceded *I dunno* 4 times (“uh I dunno”), *but*, *so*, and *I/I'm* appeared 3 times each, and *yeah* appeared twice. Common words that followed *I dunno* were *I/I'm* (8 times as in “I dunno, I have a bunch of loans”), *if* (4 times), *wh-* words (3 times), and *but* (2 times).

### 3.2.3 Friends and strangers

For the friends and strangers analysis, 12 transcripts were excluded because they did not have information on whether the participants were friends or strangers, resulting in 47 transcripts used in the analysis. Of those 47 remaining transcripts, 25 were friend dyads and 22 were stranger dyads. We examined the use of *I don't know* across task-related and chit-chat conversation. Two

different things were analyzed – the number of occurrences of *I don't know* / *I dunno* and the use of *I don't know/I dunno*.

Starting with numbers used, there was no difference in the number of combined *I don't know* and *I dunnos* between friends and strangers across both chit-chat and task related conversation,  $t(44.96) = .89, p = .37$ . There was no difference in the number of combined *I don't know* and *I dunnos* in chit-chat related conversations ( $t(34.40) = -.83, p = .41$ ) or in task related conversation ( $t(44.57) = 1.47, p = .15$ ). Breaking apart *I don't know* and *I dunno*, there were no differences in the number of *I don't knows* used by friends and strangers in either chit-chat ( $t(21.98) = -1.40, p = .17$ ). or task-related conversation ( $t(41.547) = -.31, p = .76$ ). There was also no difference in the number of *I dunnos* used in chit-chat by strangers and friends ( $t(42.62) = .53, p = .60$ ). However, there was a difference in the number of *I dunnos* used in task related conversation, with more *I dunnos* in friend dyads working on tasks compared to dyads of strangers ( $t(25.45) = 2.37, p = .03$ ).

To look at uses, we identified the two most common uses of *I dunno* and *I don't know*, which were *marking uncertainty* and *inability to provide requested information*. We were particularly interested in how usage frequency (how many times a marker of uncertainty or a literal marker of non-information was used) was affected by conversation type and relationship.

The first analysis collapsed across *I don't know* and *I dunno* for the *I don't have the information* use. Within this use, there was no significant difference in how friends and strangers used these words across conversation types,  $X^2(1, N = 69) = .357, p = .550$ . Breaking apart *I don't know* and *I dunno*, for the *I don't have the information* uses, there was no difference in how friends and strangers used *I don't know* across conversation types ( $X^2(1, N = 45) = 3.794, p = .05$ ). However, for the *I don't have the information* uses, *I dunno* did differ in how it was used across relationship and conversation types, with *I dunno* being more likely to be used among friends in task-settings ( $X^2(1, N = 24) = 5.71, p = .017$ ).

The second analysis collapsed across *I don't know* and *I dunno* for the *marking uncertainty* use. There was no difference in usage across relationship or conversation type, ( $X^2(1, N = 67) = .45, p = .50$ ). Breaking apart *I don't know* and *I dunno*, there was also no difference in frequency of *marking uncertainty* use across relationship or conversation type, ( $X^2(1, N = 36) = .789, p = .387$ ). There was also no difference in usage frequency for *I dunno* across relationship or conversation type, ( $X^2(1, N = 31) = .003, p = .959$ ).

Looking within chit-chat conversation, there was no difference in how *I don't know/I dunno* was used across relationships, ( $X^2(1, N = 58) = .395, p = .530$ ). Looking at the individual forms, there was again no difference in usage frequencies across relationships for *I don't know*, ( $X^2(1, N = 35) = .412, p = .521$ ), and for *I dunno* ( $X^2(1, N = 23) = .059, p = .809$ ).

Looking within task-based conversation, there was no difference in how *I don't know/I dunno* was used across relationships, ( $X^2(1, N = 78) = .423, p = .515$ ). For *I don't know*, there was no difference in how it was used across relationships, ( $X^2(1, N = 46) = 3.05, p = .08$ ). For *I dunno*, there was a significant difference in how it was used across relationships, with both the *I don't have the information* and *marking uncertainty* uses more likely among friends than strangers.

### 3.3 IM

All 65 transcripts in the IM corpus were coded following the procedure laid out above. Due to the difficult nature of the task, interrater reliability across the entire corpus was fair (Fleiss's kappa = .28).

#### 3.3.1 Hedges and boosters

*Kinda* was, again, the most frequent of the discourse markers used in the corpus (213 occurrences). Taken together, the other markers – *absolutely*, *totally*, and *sorta* – were used six times less frequently (35 occurrences). *Kinda* appeared 213 times in the IM corpus, with about

70% of the uses in task-related conversation and approximately 30% of the uses in chit-chat. *Totally* appeared 17 times in the IM corpus, with most of its uses in chit-chat (about 70% of the uses). *Sorta* appeared 14 times in the IM corpus, with 70% of the uses in task-related conversation.. *Absolutely* was rare, appearing 4 times in the IM corpus, with all uses in the chit-chat portion. See Table 5 in section 3.4 for a detailed breakdown.

### 3.3.2 I don't know

*I don't know* appeared 6 times, and all six occurrences were in chit-chat. Three were used to indicate insufficient knowledge and three were coded as “I don't want to say.” *I don't know* was followed by *how much* twice, *why* twice, *too much* once, and *how many* once. There were two occurrences of *I dunno*, both in chit-chat conversations and both at the start of an utterance (“I dunno”). Both of them were coded as *Missing or insufficient data*. *I dunno* appeared with *lol* once.

Because this was a text-based conversation, a written, abbreviated form of *I don't know* (*idk*) was also analyzed. *Idk* appeared 71 times, 7 times in task-related conversation and 64 times in chit-chat related conversation. In the task-related conversation, 3 uses of *idk* were coded as *indicating insufficient knowledge*, 3 were coded as *marking uncertainty*, and 1 was coded as *avoiding commitment*. In the chit-chat based conversation, there were 64 appearances of *idk*. The most common use was *indicated insufficient knowledge*, with 35 occurrences. The next most common was *avoiding commitment*, with 13 instances. The rest of the data was split between *missing or unable to code* (8), *marking uncertainty* uses (7), and avoiding assessment (1). The most common word to precede *idk* was *but*, with 6 instances. Other words that preceded *idk* include *although*, *because*, *oh* and *yeah*. The most common word to follow *idk* was *if*, with 13 occurrences (“idk if we can go back”). *How* also followed *idk* at a high rate (10) and so did *wh-* phrases (7).

<i>Kinda</i>	<i>Totally</i>	<i>Sorta</i>	<i>Absolutely</i>	<i>I don't know</i>	<i>I dunno</i>	<i>idk</i>
Roommates						
Start: 21	Start: 18	Start: 2	Start: 0	Start: 87	None	None
Middle: 101	Middle: 25	Middle: 0	Middle: 1	Middle: 104		
End: 34	End: 9	End: 0	End: 0	End: 29		
Artwalk Task						
Start: 41	Start: 4	Start: 0	Start: 1	Start: 49	Start: 45	None
Middle: 173	Middle: 12	Middle: 10	Middle: 1	Middle: 68	Middle: 43	
End: 57	End: 6	End: 1	End: 3	End: 10	End: 12	
Artwalk Chat						
Start: 9	Start: 2	Start: 2	Start: 1	Start: 16	Start: 14	None
Middle: 16	Middle: 12	Middle: 0	Middle: 1	Middle: 11	Middle: 15	
End: 10	End: 5	End: 0	End: 0	End: 3	End: 7	
IM Task						
Start: 47	Start: 2	Start: 3	None	None	None	Start: 5
Middle: 72	Middle: 1	Middle: 4				Middle: 2
End: 27	End: 2	End: 3				End: 0
IM Chat						

Start: 23	Start: 6	Start: 2	Start: 1	Start: 5	Start: 2	Start: 47
Middle: 33	Middle: 5	Middle: 2	Middle: 3	Middle: 1	Middle: 0	Middle: 24
End: 12	End: 1	End: 0	End: 0	End: 0	End: 0	End: 2

**Table 4.** Locations of marker

### 3.4 Cross-Corpora Comparisons

In Table 5, the uses of hedges and boosters across corpora are presented. Results are discussed for each corpus in their respective sections, above.

Corpus	Word	Use type	# of uses	Top use	Most common preceding word	Most common following word/part of speech
Roommates	Absolutely	N/A	1	<i>Intensify commitment</i>	was	no
	Sorta	N/A	2	<i>Avoid commitment, avoid assessment</i>	I'm, was	like, a
	Totally	N/A	52	<i>Highlighting commitment to the answer</i>	like	verbs
	Kinda	N/A	156	<i>Marking uncertainty</i>	it's	like
Artwalk	Absolutely	Task	5	<i>Express agreement</i>	yeah	no
	Absolutely	Chit-chat	2	<i>Highlight commitment, perform agreement</i>	I have, I know	No, its
	Sorta	Task	11	<i>Marking uncertainty</i>	it's	verb
	Sorta	Chit-chat	2	<i>Marking uncertainty</i>	I, it	Want, weird
	Totally	Task	22	<i>Highlighting commitment</i>	I	verbs
	Totally	Chit-chat	19	<i>Highlighting commitment</i>	I	verbs

	Kinda	Task	271	<i>Marking uncertainty</i>	its	like
	Kinda	Chit-chat	35	<i>Highlighting commitment</i>	its	Verbs, adjectives
IM	Absolutely	Task	0	-	-	-
	Absolutely	Chit-chat	4	<i>Highlighting commitment, expressing agreement</i>	yeah, we	have, but, breathtaking
	Sorta	Task	10	<i>Marking uncertainty</i>	looks	like
	Sorta	Chit-chat	4	<i>Marking uncertainty</i>	it, well, can, detective	verbs
	Totally	Task	5	<i>Highlighting commitment</i>	its	adjectives
	Totally	Chit-chat	12	<i>Highlighting commitment</i>	I	verb
	Kinda	Task	146	<i>Marking uncertainty</i>	its	look/looks/looked
	Kinda	Chit-chat	68	<i>Highlighting commitment</i>	its	adjectives

**Table 5.** Hedges and Boosters across corpora

To do statistical comparisons across corpora, raw counts were converted into percentages, where the number of instances was divided by the number of words in the transcript. We predicted that there would be more *I don't knows* in task-based portions of a conversation than in chit-chat portions. In Artwalk, there were more *I don't knows* in task-based conversation,  $t(58) = 2.55$ ,  $p = .01$ . *I dunno* was also used significantly more in Artwalk task-based conversation than chit-chat,  $t(58) = 2.70$ ,  $p < .001$ . In IM, in contrast, we found no evidence of a difference in the rate of *I don't knows* in task-based portions and chit-chat portions. However, *idk* was used more in chit-chat,  $t(64) = 2.69$ ,  $p = .009$ .

Across Artwalk and IM task-related conversation, we found no evidence of a difference in the rate of *I don't knows*,  $t(71.1) = 1.77$ ,  $p = .07$ . As *I don't know* appeared only 138 times across all the task-related conversations, or 0.0006 %, in Artwalk, and *I don't know* appeared no times in task-related conversations, this result is unsurprising. Combining the forms of *I don't know* used in Artwalk task-based conversations (*I don't know* and *I dunno*) and the forms used in IM task-based conversations (*I don't know* and *idk*), there are more overall in IM,  $t(118.77) = 2.89$ ,  $p = .004$ .



Across Artwalk chit-chat and Roommates conversations, there was a significant difference in the number of *I don't knows*, with more in Roommates,  $t(99.7) = 2.34, p = .02$ .

Artwalk - Task-based	More <i>I don't knows</i> than in Artwalk chit-chat. No difference in <i>I don't knows</i> compared to IM task. <i>I don't know</i> primarily means "I don't have information." <i>I dunno</i> primarily means "I don't have information."
Artwalk - Chit-chat	<i>I don't know</i> primarily means "I don't have information." <i>I dunno</i> means "I am marking uncertainty" and "I don't have information."
IM - Task based	No appearances of <i>I don't know</i> . <i>Idk</i> primarily means "I don't have information."
IM - Chit - chat	<i>I don't know</i> means "I am marking uncertainty" and "I don't have information." <i>Idk</i> primarily means "I don't have information." More <i>idk</i> in chat.
Roommates	More <i>I don't knows</i> than in Artwalk chit-chat. <i>I don't know</i> is used primarily to mark uncertainty. No uses of <i>idk</i> / <i>I dunno</i> .
Friends	No difference in use of <i>I don't knows</i> . More <i>I dunnos</i> in chit-chat and task-related conversation compared to strangers.
Strangers	No difference in use of <i>I don't know</i> . No difference in use of <i>I dunno</i> .

**Table 6.** Summary of Results

## 4 Discussion

When negotiating common ground, speakers are deliberate about what they want to signal to their interlocutors. Saying "it's totally fun" is different from saying "it's sorta fun" which is different from saying "it's fun, I don't know." The negotiation words qualify the level of fun and the speaker's commitment to what they are saying. We found that communicators use negotiation words differently depending on modality, goals, and relationships.

### 4.1 I don't know

*I don't know* (and the spoken and written forms *I dunno* and *idk*) were primarily used to indicate a lack of knowledge or inability to provide the requested information. Across task-related conversations, we found no evidence of differences in the frequency of *I don't know* across audio-only and text-only conversations. However, when looking at all the forms people could use, there were more *I don't knows* and associated forms (in this case *idk*) in the IM corpus. There were more *idks* than *I don't knows*, so this is likely due to the low start up cost of typing *idk* versus the start up cost of saying *I don't know* or *I dunno*. Additionally, *I dunno* and *idk* have different meanings, with *I dunno* being used to mark uncertainty more often than *idk* (see Table 6).

In spoken chit-chat conversations (Roommates, Artwalk chit-chat), there were more *I don't knows* used in Roommates, in line with our predictions. One explanation for this is that *I don't know* might have been less vague in a situation where people are able to access cues beyond the voice. This is supported by the usage of *I don't know* – most of the uses in Roommates were *marking uncertainty* uses, and most of the uses in Artwalk chit-chat were *I don't have the information* uses. Another reason there might be more *I don't knows* in Roommates compared to Artwalk chit-chat is that Roommates may function more like a task rather than true spontaneous communication. While it was focused on telling stories and chatting, participants were still brought into the lab and given instructions, whereas with Artwalk, participants freely chose to engage in chit-chat without prompting (see Nguyen et al., 2024] for similar behavior in backchannels). A third explanation is that the different uses of *I don't know* reflect differences between storytelling chit-chat and problem solving in communication, rather than differences between modalities (in person versus audio only) or differences in motivations (asked-to-chat versus chose-to-chat).

One interesting finding is that speakers in the IM corpus did not use *I don't know* or *idk* very often in the task-related portion of the conversation. When communicating in a medium that is lacking other cues a speaker might use to indicate uncertainty, it seems that speakers prefer to avoid expressing a lack of knowledge. While we did not look at what speakers actually did, it seems likely that rather than saying *I don't know*, interlocutors asked questions or expressed that they didn't understand their partner directly (e.g., “I can't find what you're talking about”).

Relationships between speakers did affect use of *I dunno*. Friends overall used *I dunno* in tasks more often, and to mean both *I don't have the information* and *I am marking uncertainty*. Friends might use *I dunno* in tasks more often than strangers because of the shared common ground between them; a friend who says *I dunno* during a difficult task is probably perceived as less rude than a stranger who says *I dunno*. Because speakers must negotiate on meaning while retaining politeness and limiting misunderstandings, it is safer to use *I dunno* with someone who already understands what you might mean by that.

*I don't know*, *I dunno*, and *idk* show different patterns for where people use them in utterances. *I don't know* appeared more in the middle of an utterance, but could also appear at the start of an utterance across the three corpora. In contrast, *I dunno* appeared at almost equal rates at the start and middle of an utterance across the three corpora. While all three are less likely to be used at the end of an utterance, suggesting they are less likely to be used as a floor-yielding marker (Aijmer, 2009), *idk* is also less likely to be used in the middle of an utterance compared to the start of an utterance across the three corpora. We found that *idk* rarely performed anything beyond the literal meaning, and tended to appear at the beginning of utterances. *Kinda*, one of the two common hedges, was more likely to appear in the middle of an utterance than the start or end, though it did appear in all three positions across corpora. *Absolutely* (a common booster), and *sorta* (the other common hedge) were rare, but when they did appear, it was across all three positions. *Totally* (the other common booster) tended to appear in the middle of an utterance rather than at the start or end.

It appears that *idk* is the form of *I don't know* that has become lexicalized to mean “I don't have the information” in instant messaging communication. This process might be similar to what has happened to *laughing out loud* (*lol*), which, like *idk*, started out as a full phrase and shortened to an acronym, and has evolved, like *idk*, past the original meaning (Herring, 2012; Tagliamonte & Denis, 2008; Varnhagen et al., 2010). This may be the result of pressures of technology: While typing on phones has gotten easier with the introduction of the full keyboard, it is still difficult to do. Likewise, there are time costs to typing on a keyboard on a computer (people can say more than they can type in a given amount of time), and so people who are using mediated communication might favor shorter messages as a result. This contrasts with *I dunno*, which acts pragmatically like *I don't know*.

As we move into a world where communication occurs online more than offline, we should expect to see more phrases lexicalize. Providing a preliminary study of the location and functions

of *idk* provides insight into how lexicalized phrases might be used, and when and how they can replace their spoken counterparts in online communication. Our work provides a snapshot in time against which future uses can be compared.

## 4.2 Hedges and boosters

Overall, there were very few boosters in the three corpora. Across all three corpora, *absolutely* appeared 12 times and *totally* appeared 110 times. When *absolutely* did appear, it was used to either highlight commitment or to express agreement. Thus, *absolutely* appears to function in the way that boosters are expected to – that is, it maximizes commitment to the utterance.

The picture around *totally* is less clear. Most uses of *totally* were to *highlight commitment*, and *express agreement* as expected from a booster. However, *totally* is able to perform a broader range of functions compared to *absolutely*. Some uses of *totally* were to *avoid disagreement* or to *preface agreement*. This suggests that *totally* may function in a broader sense compared to canonical boosters — it has an expanded use that can cover more functions.

Of the hedges, *sorta* (29 times total) was rare in comparison to *kinda* (676 times total). Looking at the uses of *sorta*, it was primarily used to *mark uncertainty*. Other uses of *sorta* were to avoid commitment or assessment. In other words, *sorta* is a true hedge – it is used in cases where people are interested in distancing themselves from a statement, or when they are unsure of the information they are reporting.

On the other hand, *kinda* has a broader set of uses. The most common uses of *kinda* were to *mark uncertainty* and to *highlight commitment*. *Kinda* could also be used to avoid commitment, or express agreement, but these uses were rare in comparison. *Kinda* and *sorta* are typically grouped together as a set of hedges that functionally mean the same thing, but these results suggest that *kinda* and *sorta* have different pragmatic functions and are not equivalent in use.

## 4.3 Collocations

Regarding collocations, the common hedges (*kinda*, *sorta*) frequently appear before and after *like* and *looks*. The hedges can be interpreted as modifying *like* and *looks* to suggest “roughly” or “approximately.” In addition, *weird* and other adjectives (which are easily qualified by approximation) tended to follow common hedges. *Totally* tended to appear after *I* and before verbs. This supports the idea that one of the main functions of *totally* is to highlight commitment. *Absolutely* most often appears with agreement, supporting its function of indicating agreement. As for *I don’t know*, *I dunno*, and *idk*, all three appeared before *wh-* phrases (*which*, *why*), echoing previous findings on *I don’t know* and *I dunno* (Aijmer, 2009). *I dunno* and *I don’t know*, both forms that are spoken aloud, often preceded expressions like *I feel*, suggesting that *I don’t know* can be used to mark uncertainty about the listener’s reception, not just the speaker’s beliefs. *idk*, which is typed, did not appear with expressions like *I feel*. It was, however, preceded by *but* and followed by *if*, which suggests a use in distancing the writer from the intentions conveyed.

## 5 Conclusions

Negotiation words are often assumed to be equivalent and interchangeable when they are of the same type (the booster *totally* can substitute for *absolutely*; the hedge *kinda* can substitute for *sorta*). The evidence provided here shows instead that the words have different pragmatic functions and are not as interchangeable as previously thought. One notable finding is that *I don’t know* appears infrequently in texted conversations, with the shortened form *idk* more likely to appear, providing evidence of the lexicalization of *idk* into a carrier of meaning beyond the original *I don’t know*. Another is that within the hedges and boosters, there are significant differences in how frequently *kinda* and *totally* are used compared to *sorta* and *absolutely*. While previous researchers have shown that these words cluster together on amounts of negotiation and correction

they imply (Nguyen & Fox Tree, under review), the current evidence highlights key differences: *totally* and *kinda* are more general than *absolutely* and *sorta*. That is, hedges and boosters are precisely selected for their meanings and cannot be interchanged without altering the meaning the speaker is trying to convey.

The findings reported here have broad-ranging applications. For example, there are significant implications for those learning English, whether they be new language learners or artificial agents. While one of the boosters and one of the hedges we looked at (*absolutely*, *sorta*) hewed closely to expected patterns, others (*totally*, *kinda*) had broader semantic ranges. In addition to needing additional information about these boosters and hedges, communicators would need to be taught or trained how to interpret the intended meaning of *I don't know*, or to produce *I don't know* in the appropriate way. There is a difference between wanting to convey a lack of knowledge versus wanting to convey a desire to keep information off record, among other uses. Contextual information such as the conversational modality, the type of conversation, or the acquaintanceship status of communicators could also be explicitly taught or programmed. Another implication applies to the broader understanding of negotiation words. *I don't know* is not like boosters or hedges. It wasn't used to intensify commitment to an answer or to express agreement, like boosters were. *I don't know* is also more flexible than a hedge because it can convey a literal lack of information along with marking uncertainty.

Future work might include assessing the nonverbal cues associated with boosters, hedges, and *I don't know*. For example, people might indicate uncertainty by pulling lips back with a grimace, certainty by furrowing brows, or lack of knowledge by raising shoulders and eyebrows. With respect to *I don't know*, the bodily movements for indicating a desire to keep information off record may be different (perhaps a mouth movement instead of a shoulder shrug). Examining these words in a larger context that not only includes turn by turn analyses but also body gestures would provide insight into how people metalinguistically mark the different uses of *I don't know*. Researchers might also consider assessing directly how boosters, hedges, and *I don't know*s (and its variants) are interpreted across different settings. For example, the uses of *I don't know*, boosters, and hedges may vary in legal contexts or among people of different ages.

## Acknowledgements

Funding for this project was received from the University of California Santa Cruz Social Sciences Division Quarter Dissertation Fellowship. We thank the research assistants who assisted on this project.

## References

- Karin Aijmer. "So er i just sort i dunno i think it's just because. . .": A corpus study of *I don't know* and *dunno* in learners' spoken english. In *Corpora: Pragmatics and Discourse*, pages 151–168. Brill, 2009.
- Andrea Beltrama and Florian Schwarz. (Im) precise personae: The effect of socio-indexical information on semantic interpretation. *Language in Society*, pages 1–28, 2024.
- Susan E. Brennan and Maurice Williams. The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Memory and Language*, 34(3):383–398, 1995.
- Rachel Bristol and Federico Rossano. Epistemic trespassing and disagreement. *Journal of Memory and Language*, 110:104067, 2020.
- Penelope Brown and Stephen C. Levinson. *Politeness: Some universals in language usage*.

- Number 4 in Cambridge Papers in Linguistics. Cambridge University Press, 1987.
- Gregory A. Bryant. Prosodic contrasts in ironic speech. *Discourse Processes*, 47(7):545–566, 2010.
- A. Gavin Burnage and Glynis Baguley. The British national corpus. *Library and Information Briefings*, (65), February 1996.
- Herbert H. Clark and Jean E. Fox Tree. Using *uh* and *um* in spontaneous speaking. *Cognition*, 84(1):73–111, 2002.
- J. Trevor D’Arcey, Shereen Oraby, and Jean E. Fox Tree. Wait signals predict sarcasm in online debates. *Dialogue & Discourse*, 10(2):56–78, 2019.
- Giuliana Diani. The discourse functions of I don’t know in English conversation. In *Pragmatics and Beyond New Series*, pages 157–172. 2004.
- Simona Pekarek Doehler. More than an epistemic hedge: French *je sais pas* ‘I don’t know’ as a resource for the sequential organization of turns and actions. *Journal of Pragmatics*, 106:148–162, 2016.
- Farahman Farrokhi and Safoora Emami. Hedges and boosters in academic writing: native vs. non-native research articles in applied linguistics and engineering. *Journal of English Language Pedagogy and Practice*, 1(2):62–98, 2008.
- Jean E. Fox Tree. Placing like in telling stories. *Discourse Studies*, 8(6):723–743, 2006.
- Jean E. Fox Tree. Discourse markers across speakers and settings. *Language and Linguistics Compass*, 4(5):269–281, 2010.
- Jean E. Fox Tree. Discourse markers in writing. *Discourse Studies*, 17(1):64–82, 2015.
- Jean E. Fox Tree and Sarah A. Mayer. Overhearing single and multiple perspectives. *Discourse Processes*, 45(2):160–179, 2008.
- Jean E. Fox Tree and Josef C. Schrock. Discourse markers in spontaneous speech: Oh what a difference an oh makes. *Journal of Memory and Language*, 40(2):280–295, 1999.
- Jean E. Fox Tree, Steve Whittaker, Susan C. Herring, Yasmin Chowdhury, Allison Nguyen, and Leila Takayama. Psychological distance in mobile telepresence. *International Journal of Human-Computer Studies*, 151:102629, 2021.
- Charles Goodwin. *Conversational Organization: Interaction Between Speakers and Hearers*. 01 1981.
- Lynn E. Grant. A corpus comparison of the use of *I don’t know* by British and New Zealand speakers. *Journal of Pragmatics*, 42(8):2282–2296, 2010.
- Stefan Th. Gries and Caroline V. David. This is kind of/sort of interesting: variation in hedging in English. *Studies in Variation, Contacts and Change in English*, 2, 2007.
- Swati Gupta, Marilyn A. Walker, and Daniela M. Romano. How rude are you? evaluating politeness and affect in interaction. In *Proceedings of the 2nd International Conference on Affective Computing and Intelligent Interaction*, 2007.
- Andrew J Guydish and Jean E Fox Tree. Reciprocity in instant messaging conversations. *Language and Speech*, 65(2):404–417, 2022.
- Andrew J Guydish, J Trevor D’Arcey, and Jean E Fox Tree. Reciprocity in conversation. *Language and Speech*, 64(4):859–872, 2021.
- Andrew J Guydish, Allison Nguyen, and Jean E Fox Tree. Discourse markers in small talk and tasks. *Discourse Studies*, 26(5):606–620, 2024.
- Janet Holmes. Hedges and boosters in women’s and men’s speech. *Language & Communication*, 10(3):185–205, 1990.
- Ken Hyland. Persuasion and context: The pragmatics of academic metadiscourse. *Journal of pragmatics*, 30(4):437–455, 1998.

- Ken Hyland. Hedges, boosters and lexical invisibility: Noticing modifiers in academic texts. *Language awareness*, 9(4):179–197, 2000.
- Ellen A. Isaacs and Herbert H. Clark. Ostensible invitations. *Language in society*, 19(4):493–509, 1990.
- Jumayel Islam, Lu Xiao, and Robert E. Mercer. A lexicon-based approach for detecting hedges in informal text. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 3109–3113, 2020.
- Alireza Jalilifar and Maryam Alavi-Nia. *We are surprised; wasn't Iran disgraced there?* A functional analysis of hedges and boosters in televised Iranian and American presidential debates. *Discourse & Communication*, 6(2):135–161, 2012.
- Andreas H. Jucker and Sara W. Smith. Explicit and implicit ways of enhancing common ground in conversations. *Pragmatics. Quarterly Publication of the International Pragmatics Association (IPrA)*, 6(1):1–18, 1996.
- Elise Kärkkäinen. Position and scope of epistemic phrases in planned and unplanned American English. In *New approaches to hedging*, pages 203–236. Brill, 2010.
- Hadi Kashiha. On persuasive strategies: Metadiscourse practices in political speeches. *Discourse and Interaction*, 15(1):77–100, 2022.
- J. Richard Landis and Gary G Koch. The measurement of observer agreement for categorical data. *Biometrics*, 33(1): 159–174, 1977.
- Kristen E. Link and Roger J. Kreuz. The comprehension of ostensible speech acts. *Journal of Language and Social Psychology*, 24(3):227–251, 2005.
- Kris Liu and Jean E. Fox Tree. Hedges enhance memory but inhibit retelling. *Psychonomic bulletin & review*, 19:892–898, 2012.
- Charles Metzing and Susan E. Brennan. When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory and Language*, 49(2):201–213, 2003.
- Jenny Myrendal. Negotiating meanings online: Disagreements about word meaning in discussion forum communication. *Discourse studies*, 21(3):317–339, 2019.
- Islam NamazianDost. A comparative study of boosters in the discussion section of medical and applied linguistics articles. *International Journal of Applied Linguistics and English Literature*, 6(7):1–5, 2017.
- Allison Nguyen and Jean E. Fox Tree. *Is kinda less sure than partially?*: Putting negotiation words on a scale. Under review.
- Allison Nguyen, Tom Roberts, Pranav Anand, and Jean E. Fox Tree. Look, dude: How hyperpartisan and non-hyperpartisan speech differ in online commentary. *Discourse & Society*, 33(3): 371–390, 2022.
- Allison Nguyen, Andrew J Guydish, and Jean E. Fox Tree. Backchannels in the lab and in the wild. *Interaction Studies*, 25(1):70–99, 2024.
- Tri Nuraniwati and Alfelia Nugky Permatasari. Hedging in TED talks: A corpus-based pragmatic study. *JEELS (Journal of English Education and Linguistics Studies)*, 8(2):203–226, 2021.
- Maryann Overstreet. And stuff *und so*: Investigating pragmatic expressions in English and German. *Journal of Pragmatics*, 37(11):1845–1864, 2005.
- Heike Pichler and Ashley Hesson. Discourse-pragmatic variation across situations, varieties, ages: *I don't know* in sociolinguistic and medical interviews. *Language & Communication*, 49:1–18, 2016.
- Françoise Salager-Meyer. Hedges and textual communicative function in medical English written discourse. *English for specific purposes*, 13(2):149–170, 1994.

- Joanne Scheibman. *I dunno*: A usage-based account of the phonological reduction of don't in American English conversation. *Journal of pragmatics*, 32(1):105–124, 2000.
- solarmax. bffjill. URL <https://www.youtube.com/watch?v=4nIUcRJX9-o>.
- Aina Garí Soler, Matthieu Labeau, and Chloé Clavel. Measuring lexico-semantic alignment in debates with contextualized word representations. In *Proceedings of the First Workshop on Social Influence in Conversations (SICon 2023)*, pages 50–63. Association for Computational Linguistics, 2023.
- Sali A. Tagliamonte and Derek Denis. Linguistic ruin? lol! instant messaging and teen language. *American speech*, 83(1):3–34, 2008.
- Masahiro Takimoto. A corpus-based analysis of hedges and boosters in English academic articles. *Indonesian Journal of Applied Linguistics*, 5(1):95–105, 2015.
- Avril Thorne, Lauren Shapiro, Kim Cardilla, Neill Korobov, and Paul A Nelson. Caught in the act: How extraverted and introverted friends communally cope with being recorded. *Journal of Research in Personality*, 43(4):634–642, 2009.
- Jackson Tolins and Jean E. Fox Tree. Addressee backchannels steer narrative development. *Journal of Pragmatics*, 70:152–164, 2014.
- Amy BM Tsui. The pragmatic functions of *I don't know*. *Text-Interdisciplinary Journal for the Study of Discourse*, 11(4):607–622, 1991.
- Connie K. Varnhagen, G. Peggy McFall, Nicole Pugh, Lisa Routledge, Heather Sumida-MacDonald, and Trudy E. Kwong. Lol: New language and spelling in instant messaging. *Reading and writing*, 23:719–733, 2010.
- Esther J. Walker, Evan F. Risko, and Alan Kingstone. Fillers as signals: Evidence from a question–answering paradigm. *Discourse Processes*, 51(3):264–286, 2014.
- Ann Weatherall. *I don't know* as a prepositioned epistemic hedge. *Research on Language & Social Interaction*, 44(4):317–337, 2011.
- Margaret Zellers. An overview of forms, functions, and configurations of backchannels in Ruruuli/Lunyala. *Journal of Pragmatics*, 175:38–52, 2021.