A Computational Analysis of Tentativeness and Causation in Design Talk

Chandrasegaran, R.S.K.; Akdag Salah, A.A.; Lloyd, P.A.

A computational analysis of tentativeness and causation in design talk

Senthil Chandrasegaran[a], **Almila Akdag**[b] and **Peter Lloyd**[a]

[a] Faculty of Industrial Design Engineering, TU Delft, Delft, The Netherlands, {r.s.k.chandrasegaran, p.a.lloyd}@tudelft.nl

[b] Department of Information and Computational Sciences, Utrecht University, Utrecht, The Netherlands, a.a.akdag@uu.nl

Abstract. Analysing records of design activity such as transcripts or documents have typically involved close reading of transcripts and manual identification of concepts and behaviours. We explore the applicability of a machine-learning based computational tool—called *Empath*—in identifying high-level patterns in design talk. Specifically, we use it to examine the datasets from the Design Thinking Research Symposium (DTRS) workshops for two contrasting aspects of design talk—the expression of tentativeness that characterises designers' exploration of the problem-solution space, and the expression of causal reasoning that characterises designers' analytical thinking. We find that such a tool can be effectively used as a means of "distant reading". However, the lack of design relevance in the tool's training data results in ambiguities and mis-categorisations that still need resolution through close reading.

Keywords: Computational Linguistics, Conversation Analysis, Design Talk, Design Thinking, Machine Learning

1 Introduction

Is design activity characterised by an equal balance between tentativeness and speculation on the one hand and by evaluation and justification on the other? Models of designing that describe a 'basic' cycle of activity generally describe projective activity, characterised by tentativeness or epistemic uncertainty, followed by explanatory (or rationalising) activity, characterised by evaluation and justification (Lloyd, 2019). For example Schön's (1992) description of designing as a series of reflective 'moving experiments' is premised on the idea that something has to materially be put into the world before the understanding of its implications can take place and therefore be justified. Similarly, Roozenburg (1993) describes a process initiated by the logic of abduction prior to a deduction of consequences. An equivalence of projective and explanatory processes in design activity is suggested by these models but is this borne out in actual practice? To attempt to answer these questions, we examine an aggregation of four datasets that resulted from the Design Thinking Research Symposium's shared-data workshops, and computationally categorise sections of the data as indicative of tentativeness and of causation. We then observe patterns in which these two categories occur across sessions and examine the contexts in which they occur together or separately. In doing so, we also explore the advantages and disadvantages of modern computational analysis tools that can supplement a close reading of text.

Over a period of nearly 30 years the Design Thinking Research Symposium series (Cross, 2018) has conducted four shared-data workshops, generating data from design activity of largely professional designers in a number of different study conditions. These datasets include think aloud protocols (Cross et al., 1996), naturally-occurring designer-client discussion (McDonnell & Lloyd, 2009), design education (Adams et al., 2016), and naturally-occurring co-creation (Christensen et al., 2017).

The disciplines of design from which this data has been generated have been industrial design engineering (DTRS2), architecture and engineering design (DTRS7), design education (DTRS10) and product design (DTRS11). Sec. 2 provides a short summary of each workshop.

In previous work (Lloyd et al., 2021) we have explored the idea of epistemic uncertainty using Linguistic Inquiry and Word Count (LIWC) (Pennebaker et al., 2015), a tool with pre-defined and curated lexical categories used to tag, categorise, and classify text based on the association of the text content with these categories. Specifically, we focused on the 'tentative' subcategory of LIWC where we found words associated with hypothetical, tentative thinking such as "if", "maybe", "might", "perhaps", "possibly", and "probably". The use of a term from this set is typically an indicator that the designer is considering or suggesting the exploration of a possibility or future conditional. For example, in the discussion between an architect and their client concerning the design of a crematorium (DTRS7) the architect asks:

*"...did you see this as a space that **might** have its own small lectern in it or some altar-like feature inside it?"*

In this and a sister paper for DTRS13 (Akdag-Salah et al., 2021) we draw on a corpus resulting from the combination of the four shared-data DTRS workshops to computationally explore the concepts of "tentativeness" and "causation" related to designing found in prior studies. In doing so, we also explore the pros and cons of such a computational approach when used to analyse concepts relating to designing. This is something that has only recently become relatively easy to do, with analyses of designers prior to this mainly focusing on smaller design 'protocols', and the manual identification of textual excerpts to identify new concepts. This traditional way of analysing design activity is akin to "close reading", a term in literary research where the goal is to focus on specific arguments, individuals, or ideas and trace their evolution across the document(s) (Jänicke et al., 2015). In contrast, "distant reading", a term coined by Moretti (2005) is an approach to take a global view of a text or texts, visualizing its (or their) global features. Distant reading thus relies on computational analyses of large text, the results of which are presented graphically in the form of simple charts or complex data visualizations.

Our examination of tentativeness and causation outlined above follows this distant reading approach at first, looking at global patterns of occurrence of these categories across datasets and sessions. Specifically, we use a tool called Empath (Fast et al., 2016) and its pre-trained machine-learning model to first generate lexical categories containing words corresponding to the notions of tentativeness and causation, and then find matches between these categories and speech turns in the dataset transcripts. We then drill down to patterns of interest and examine them through close reading to critically interpret not only the context of the exchange underlying the patterns, but also the accuracy of the computational approach used to identify the patterns. We posit that such an approach can help expand the contexts in which one might expect to find instances of design thinking and help train the next generation of artificial intelligence (AI)-based conversational systems to recognise designerly talk.

2   The DTRS Dataset

The DTRS series (Cross, 2018) has held a series of "common data workshops" that have resulted in four shared datasets, created with a view that different perspectives, methods, and theories can be proposed and tested using these common data. The four datasets cover the disciplines of industrial design engineering (DTRS2), architecture and engineering design (DTRS7), design education (DTRS10), and product design (DTRS11). Some details about the datasets are provided below.

DTRS02 consists of one 2-hour 'think-aloud' design session with a single designer and another 2-

hour session featuring a team of three designers. Both sessions work on the same design problem, a cycle pannier, verbalising their thoughts.

DTRS07 consists of four 2-hour meetings of 'naturally-occurring' design activity. Two of the meetings feature an architect communicating his designs to his client. The other two meetings feature a multidisciplinary design team discussing initial ideas for a 'digital pen'.

DTRS10 consists of 38 videos of varying length showing design reviews in five disciplines (industrial design, mechanical design, service-learning design, entrepreneurial design, and choreography). The videos are diverse and feature a range of interactions, but are primarily based around teacher-student discussion, both individually and in teams.

DTRS11 features 20 video recordings, again of varying length (up to 45 mins). In the first sessions the design of two co-creation session for a large car manufacturer are discussed. The co-creation sessions are filmed, and these are followed by videos discussing the co-creation sessions and the possible design products that might result.

Table 6 shows the session numbers and lengths for each of the four DTRS corpora. At a combined 373,983 words, these datasets—specifically the types of discussions—provide a composite picture of design activity. Together, the corpora provide opportunities for examining different kinds of discussions at scale. However, most contributions that have arisen from the transcripts and videos have involved close reading and manual analysis of specific datasets and sessions. These contributions, such as analyses of framing (Dong et al., 2017), storytelling (Lloyd & Oak, 2018), and spiderwebbing (Mabogunje et al., 2009), provide conclusions in the form of hypotheses that need testing with larger datasets.

Table 6. Dataset Statistics

Dataset	Sessions	Dataset Size (words)	Session Size (words)	
			Mean	S.D.
DTRS2	2	37,969	18,984	4,085
DTRS7	4	68,861	17,215	4,944
DTRS10	38	92,751	2,441	3,424
DTRS11	20	174,402	8,720	4,590
Total	64	373,983	5,843	6,162

3 Methodology

To illustrate our approach of using computational tools to examine designerly ways of thinking, we look at two kinds of behaviours that have been examined in prior research through close reading and qualitative text analysis. The two behaviours we focus on are expressions related to causal reasoning—where designers express or enquire about the reasoning behind decisions or observations (Cardoso et al., 2014)—and tentativeness—where designers use downtoners and other expressions of vagueness to soften their language to strengthen social ties (Glock, 2009) or to express epistemic uncertainty (Ball & Christensen, 2009). Using a tool such as LIWC provides a scalable alternative— or at the very least, a complementary approach—to close reading, which has been the traditional way to analyse such transcripts. With LIWC, one can identify the parts of the transcript that contain terms that are indicative of "movement", "positive emotion", "cognitive processes" or any of the 91 categories and sub-categories pre-defined in LIWC. A simple way that LIWC achieves this is by looking for words in the transcript that match words defined in a category of interest.

However, predefined lexical categories may not always match categories that researchers are seeking to identify for their work. For instance, in our prior work, we found that the LIWC subcategory of

"tentative" did not exactly match the ways in which tentativeness was explored in the DTRS datasets (Lloyd et al., 2021). Empath (Fast et al., 2016) is a tool that seeks to address LIWC's limitations of a predefined set of categories that may not address a researcher's specific requirement. In contrast to LIWC's human- generated and curated categories and associated dictionaries, Empath provides a set of nearly 200 machine-generated categories that are then validated by humans. Empath also provides users with the ability to generate a new category by supplying seed terms that they think are associated with the intended category.

Empath's pre-defined categories are generated from dependency relationships in ConceptNet (Liu & Singh, 2004). Terms in each category are generated from seed terms that are in turn used to query a vector space model, which is a representation of documents as vectors in a high-dimensional space such that proximal vectors represent similar documents. The vector space is created by mapping the neural embedding of a neural network based on a skip-gram architecture (see Mikolov et al., 2013) trained to predict words that co-occur in a large corpus of text. Empath uses a dataset of modern fiction[1] as the training corpus, as it (a) provides a better breadth of topical and emotional categories, and (b) correlates better with the categories defined in LIWC than corpora containing news articles or online discussions.

To generate a new category, the user provides a category name and a set of seed terms. Empath first generates vector representations of seed words based on the vector space model and sums the vectors to create an "analogy" vector that represents a combined association with all seed terms. It also queries the space for a vector corresponding to the category name and adds it to the above analogy vector. The vector space in the neighbourhood of this analogy vector is queried to find additional terms related to the category. For instance, providing a list of seed terms such as *"twitter"*, *"instagram"*, and *"facebook"* and a category name of *"social media"* returns such additional terms as *"timeline"*, *"hashtag"*, *"notifications"*, *"direct message"*, and *"Tumblr"*, all of which are terms associated with social media.

In this section, we describe our use of Empath to create lexical categories seeded with terms associated with tentative thinking and causal reasoning. We then use these categories as lenses with which to examine the DTRS datasets.

Table 7. Data format of the combined DTRS datasets used for the analysis in this paper

Dataset	Session	Speaker	Content
DTRS-02	Think-aloud	Interviewer	*"Before I give you the design brief, are there any questions about the procedure?"*
DTRS-02	Think-aloud	Dan	*"Um No, other than the... I have these two references here, the file of information that you have Um.. I will just ask you for..what?., identify..I say what kind of technical information I want?"*
…	…	…	...
…	…	…	...
DTRS-11	Follow-up interview	Interviewer	*"I understand that, alright, that was the end of it, eh anything you want to add? We've been through a lot, so I am gonna let you go"*
DTRS-11	Follow-up interview	Ewan	*"(laughs) no, I think I'm okay, I think that makes sense eh not anything related to this at least"*

---

[1] https://wattpad.com

3.1 Processing the Text

The transcripts for all 64 sessions across the four DTRS datasets were cleaned to remove time stamps, location descriptions, and descriptions of any seating arrangements since our focus was on the content and context of what was being said. For this same reason, in-line descriptions of subjects' actions such as pointing, or gesturing were retained. The transcripts were combined in such a form as to enable analysis of the combined dataset, while still being able to filter for individual datasets, sessions, or speakers. The format of the data set up for analysis is illustrated in Table 7. As shown in the table, the order of rows follows the order of speech turns from each session's transcript.

3.2 Creating Lexical Categories

To choose appropriate seed terms to create the lexical category of "causation", we refer to Cardoso et al. (2014) and their study of the DTRS10 dataset involving design feedback sessions between students and instructors. The study revealed that questions that help students think about their design rationale, the potential effects of their design choices, their interpretation of related phenomena, and other forms of "deep reasoning questions" aid students' reflection on their state of design. While their work involved qualitative coding of the transcript through close reading and reviewing of the corresponding video recordings, we attempt to explore the occurrence of such questioning or expressions of reasoning through a lexical category that we create through Empath.

Focusing on the "causation" (why) and "procedural" (how) components of rationale, we input seed words such as *"because"*, *"effect"*, *"explain"*, *"how"*, *"why"*, etc. to generate a category called *"Causation"* using Empath, resulting in a lexical category with the following 84 terms:

because, given, moreover, regardless, though, yet, affect, affected, affecting, affects, appeal, attachment, basis, causes, circumstance, complication, concerning, conclude, conditions, consequence, consider, context, conversion, crisis, critical, crucial, depends, determine, disastrous, downfall, effect, effected, end result, essentially, experience, explain, extent, function, illness, implies, imply, influence, justify, killing, kind, knowing, magnitude, main problem, mean, meaning, means, meant, mental state, method, might, mindset, motive, necessity, occur, outcome, part, possibly, potential, predict, proves, purpose, real problem, reason, regardless, relation, relevant, result, side effects, significance, significant, situation, specifics, suppose, surely, telling, terms, therefore, though, understand

Though some words such as "illness" do not quite fit into the intended lexical category—and while the category can benefit from removal of such terms—for the sake of replicability and to highlight the shortcomings of such approaches, we make no modification to the generated set of terms.

The second lexical category of "tentativeness" is motivated by an emerging sub-field in design research known as "epistemic uncertainty". This is the idea that creative behaviour is triggered in situations of uncertainty to lessen that uncertainty and progress the design process (Ball & Christensen, 2009; Ball et al., 2010; Cash & Kreye, 2018; Christensen & Ball, 2018; Christensen & Schunn, 2009; Paletz et al., 2017). Prior work examining the DTRS7 dataset (Ball & Christensen, 2009; Glock, 2009), the DTRS11 dataset (Christensen & Ball, 2018), and other studies (Cash & Kreye, 2018) have shown that designers typically use downtoners, hedges, modal adverbs and other expressions of tentativeness to express such uncertainty. This is linked to Schön's model of "naming, framing, moving, and evaluating" (Schön, 1984) where a designer makes educated guesses and suggestions as a way of moving forward in the design process. In our prior work (Lloyd et al., 2021) we examine tentative aspects of design conversations and find expressions of possibility such as modal adverbs (e.g., could, might, probably) to occur prevalently across all DTRS datasets. Using seed terms such as *"if"*, *"maybe"*, *"might"*, *"perhaps"*, *"possibly"*, *"probably"*, etc. in Empath, we

use the category name of *"Tentativeness"* to generate the lexical category corresponding to epistemic uncertainty that includes the following 59 terms:

able, actually, afraid, also, although, any, anyone, anything, anyway, assume, assuming, because, besides, case, definitely, doubt, either, else, exactly, expect, figured, guess, hoping, however, if, knew, knowing, maybe, mean, meant, might, must, now, obviously, only, or, otherwise, perhaps, personally, plus, possibly, probably, should, so, suppose, supposed, sure, surely, though, thought, unless, wanted, well, whether, willing, wish, wonder, wondering, yet

We refer to this category as "tentativeness" in this paper rather than "epistemic uncertainty" for two reasons: to connect to LIWC's "tentative" subcategory under "cognitive process" explored in prior work (Lloyd et al., 2021), and because the category name provided in Empath also influences the words generated. Since Empath is trained on domain non-specific text like fiction, "tentativeness" is a more interpretable category than "epistemic uncertainty".

Using discussion turns as the unit of our analysis, we counted the number of matches between each lexical category and the words in the turn, and normalized the count for every turn, dividing the number by the total words in the turn. We only looked for whole word matches and chose not to lemmatize the words, neither in the turns nor in the lexical categories, as the sense of such words (e.g., modal adverbs) is often linked to the specific form of the word. In the following section, we use the measures of matches between speech turns and the two lexical categories created to examine the question of whether design activity is tentative or rational. We attempt to answer this question at different levels of aggregation: at the level of the dataset, sessions, and then individual speakers.

4 Results

Simply counting the number of turns that contain at least one match for each lexical category and normalizing this count with the total turns for each dataset, we immediately see a higher incidence of speech turns that express tentativeness exclusively. In other words, of the turns matching at least one lexical category, most of the turns feature a match with *tentativeness* while having no matches with *causation* (Figure 4). The next highest set of matches belong to turns that match *both* categories (tentativeness and causation), while only a small number of turns match exclusively with *causation*. At the dataset level at least, design discourse appears to be characterised more by tentative thinking than it is by thinking about cause, effect, or rationale.
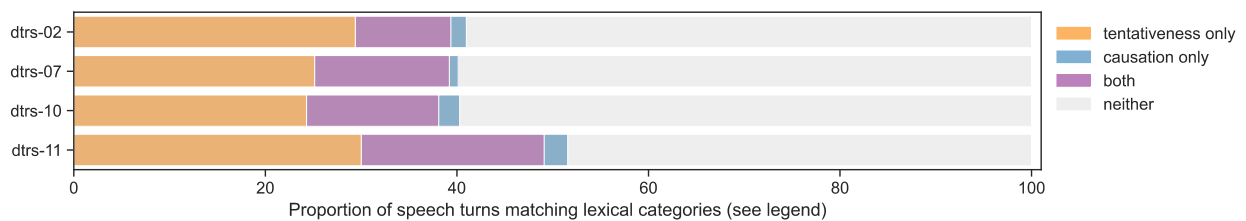

Figure 4. The proportion of speech turns in each dataset containing terms from the lexical categories of tentativeness, causation, and both. The highest proportion of the matched speech turns fall exclusively under tentativeness, followed by turns matching both categories. Only a small proportion of turns fall exclusively under causation

Figure 5 shows a general overview of the degree of correlation between the two lexical categories for the average turn in each dataset. Each dataset is represented by a scatterplot, with each dot in the scatterplot representing a turn of speech, and the dot's x- and y-positions corresponding to the normalised count of matches between the words in the speech turn and the lexical categories of

tentativeness and causation respectively. The slope of the regression line confirms what we see in Figure 4: on average a turn contains more matches with the *tentativeness* category than with the *causation* category. The counts for tentativeness and causation are not normally distributed, so correlations were computed using Kendall's τ_b.
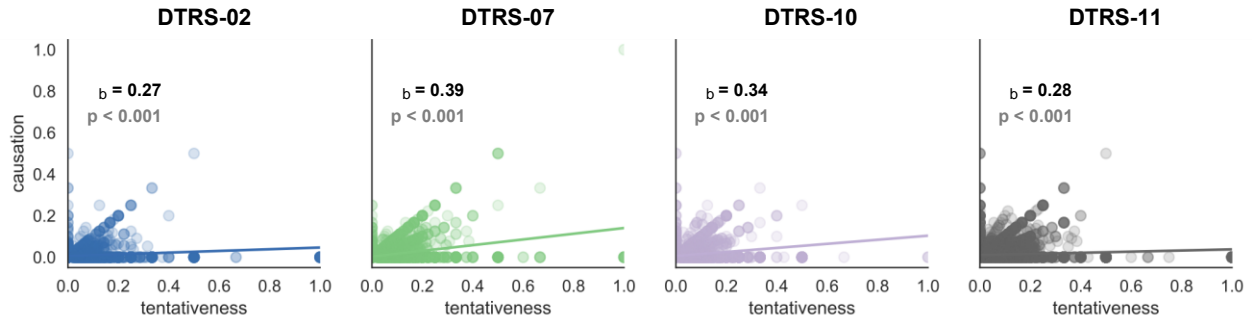

Figure 5. Scatterplots for each dataset showing the normalised number of matches in the "causation" and "tentative" lexical categories for each speech turn. Each dot in the scatterplot corresponds to one speech turn. The x-position of the dot shows the total matches between the words in that turn and those in the "tentativeness" category, divided by the total number of words in that turn. The y-position of the dot shows the corresponding measure for that turn with the "causation" category. A regression line is shown for each plot, with the slope indicating a greater number of matches with tentativeness rather than causation on average for the speech turns. Kendall's τ_b coefficients do not show a strong correlation between the two lexical categories for the datasets in general

Since the chart in Figure 4 is normalized for the number of speech turns, we can compare across datasets and see if it aligns with our understanding of the datasets at this level. Of the four datasets, DTRS10 is the one that involves design reviews as teacher-student discussions. We thus might expect a greater proportion of turns in this dataset centred around causation with the instructors asking questions about design decisions and the students attempting to explain them. However, we see that it is DTRS11 that has a higher proportion of turns related to (a) only tentativeness (close to DTRS2), (b) both tentativeness and causation, and (c) only causation (similar to DTRS10). Perhaps this high proportion of tentativeness and causation is because of the sessions that compose DTRS-11, which include co-creation workshops followed by sessions that involve reflective activities such as debriefing, sharing insights, clustering insights, linking insights to projects, to name a few. We next examine the lexical category matches at a session level to verify some of these conjectures.
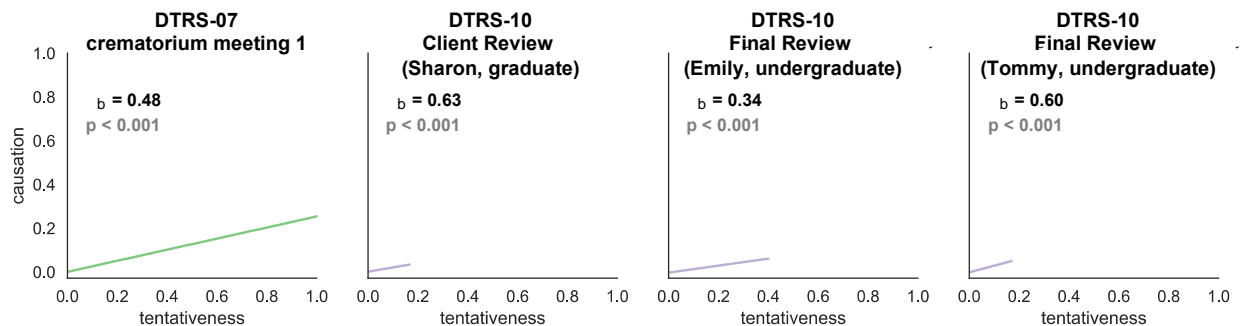

Figure 6. Scatterplots for four sessions in the entire corpus that show a moderate to strong association ($\tau_b > 0.34$, $p < 0.05$) between the normalised number of matches of the two lexical categories of "tentativeness" and "causation", for each turn. The plot is similar to that shown in Figure 5, including the regression lines and Kendall's τ_b coefficients. The plots are coloured by dataset

4.1 Examining higher incidences of tentativeness and causation

Figure 6 shows four sessions whose speech turns show a relatively stronger association between their matches to both lexical categories. The correlations are moderate-to-strong (Kendall's τ_b ranging from *0.34* to *0.63*). However, we refrain from making claims based on statistics alone, choosing instead to examine some of the exchanges in the sessions.
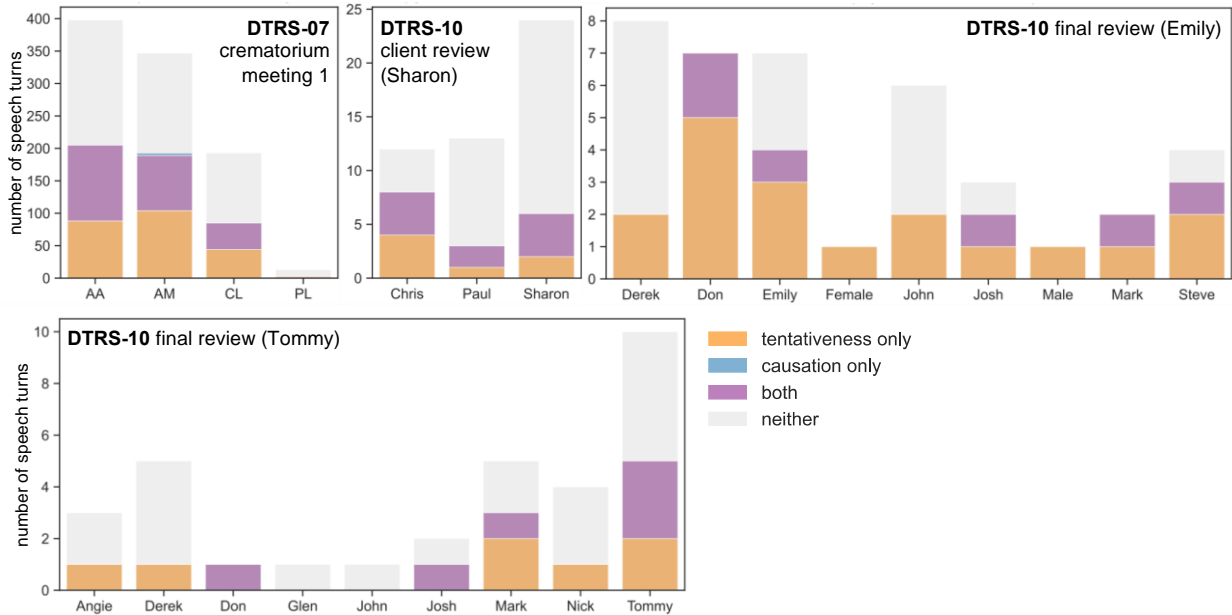

Figure 7. Speaker-wise turn counts for the four sessions identified in Figure 6 showing relatively higher association between causation and tentativeness in spoken turns. The proportion of turns that include a match that is exclusively with the tentativeness category (orange), exclusively with causation (blue), and with both categories (purple) are shown. Some anomalies are immediately revealed; for instance, all of Don's speech turns in Emily's final review session (DTRS10) have at least one match with the tentativeness lexical category, while Derek and John have no matches at all with the causation lexical category for the same session

Figure 7 shows the speaker-wise turn counts for these four sessions. Right away, the reason for the correlation is evident: there are almost no speech turns that match exclusively with the *causation* category, which means that if a turn matches with causation, it almost certainly also matches with tentativeness. DTRS7's first crematorium meeting has the highest correlation between the two categories and looking at Figure 7, it is possible to guess why: the overlap between the categories is the highest among the four sessions shown in the figure. DTRS10's client review session for Sharon comes close, but the turn count for the DTRS7 session is an order of magnitude higher, resulting in a stronger correlation between the two.

Examining the turn distribution among the speakers, as well as the proportion of tentativeness and causation (and both together) across speaker turns gives us an interesting perspective of the speakers themselves. For instance, in DTRS7's crematorium meeting 1, the two most dominant speakers are AA (the client) and AM (the architect). We can see from Figure 7 that a greater proportion of the architect's turns are associated with tentativeness, which makes sense: the architect is proposing a number of ideas based on requirements posed by the client, and is thus using expressions of tentativeness as part of the "naming, framing, moving, and evaluating" process (Schön, 1984). Consider for instance the exchange below. Words that match with the tentativeness category are highlighted in orange, words that match with causation are highlighted in blue, and words that match with both are highlighted in purple.

AM (ARCHITECT): *So you could have your cremulator as a bit at the back of the cremator*

AA (CLIENT): *Yes, what happens, what happens is when you rake down you rake it in to a cremulator on each machine so you don't have the removal of cremated remains until they've been cremulated and put in to a powder, it's all done on the machine itself, so you rake it in to an area and obviously the problem with that is that there's some issues with removal of large metal joints and hip joints and leg joints and other things, and also the concern of servicing each cremulator on each machine, quite a lot involved with it, they haven't got those quite up and running at the moment so it could be, by the time we get to this it might be, yep.*

AM (ARCHITECT): *Right, we've allowed a cremulator room anyway if you want a separate cremulator that would be like close raking but that would probably used for storing*

AA (CLIENT): *Where would the operators, would they have a separate office to sit in to, there's a switch room, control room would it be any, an area, or would the operator sit within that one area.*

AM (ARCHITECT): *We show a control desk here but certainly if you didn't need a cremulator room that could easily be converted into a control room, you could even have a glass wall on it if you wanted*

AA (CLIENT): *Are we talking about the room still being chilled there by the, because that was the original chill room wasn't it, but that's not now going to happen*

AM (ARCHITECT): *No we don't have a chill room, we have coffins stores as such here because, forgot to mention those, the idea was that, erm, that if you did have Sikh funeral you wouldn't really want to see other coffins hanging around, the idea is that you would store coffins in racks in here*

In the above exchange, the discussion around a cremulator—a device that is used for further processing the remains after cremation—follows a pattern of the client declaring requirements and asking questions, with the architect responding to the questions with design proposals or explanations. The responses by the architect show instances of tentativeness (e.g., *"that would probably (be) used for storing"*) and some causation (e.g., *"because... if you did have Sikh funerals you wouldn't want to see..."*). The questions and declarations of the client show similar patterns, with tentativeness indicating their interest in alternatives and causation when they are explaining the reasoning behind a requirement. What is interesting here is the text highlighted in purple, indicating that the word ("because" in the above exchange) appears in both categories. This overlap of words is not always the reason why the turns are shared with both lexical categories. Consider an exchange that happens earlier in the discussion between the architect and the council officer (CL).

CL (COUNCIL OFFICER): *We've got an issue [inaudible] considering following a training session that the staff went on this week at Stevenage crematorium to look at use of the audio, the new audio, digital audio system that had been installed, and a particular issue is the accommodation, and having accommodation where they can work almost like a theatre technician so that they can be part of the service yet be divorced from the service so that they can operate PCs and things but also be in sight of what's going on, so being more involved. So we're currently looking at bringing stuff out the back of the existing chapel so that they can sit at a desk away, but part of –*

AM (ARCHITECT): *Well that's the desk that I thought they could work at, I saw this as being combined organ and AV control, so someone might sit here and play the organ and there might be*

AV controls here, but essentially all the spaghetti and all the amplifiers and CD players, cassette players, you name it, computers could be in here –

The exchange concerns the creation of a space for the crematorium employees for managing the audio/video aspects of a funeral service. As in the earlier excerpt, there is an aspect of causation in the council officer's request: provide facilities *so that* the employees might be able to work at the equipment while monitoring the proceedings. The architect's response is a classic example of speculation and exploration of possibilities with terms such as "I thought", "might", and "well". There is an expression of causal reasoning as well, with "so" and "essentially"—used in the sense of "in effect" here—to explain the implications of the proposed choices. While both turns in the above exchange are categorised as belonging to both tentativeness and causation, it is not because the matched words themselves belong to both categories. When examining the matched words, the sense of causation is evidenced more in the word *"yet"* rather than the other matches (*"part"* and *"also"*).

The main word that conveys causation (*"so"*) forms part of the *tentativeness* category but not the *causation* category. The words "yet" and "might" do overlap, i.e., they are present in the lexical categories of both *tentativeness* and *causation*. We discuss possible causes and the implications of such terms in Sec. 5.

Visualizing the turns aggregated at the level of the speakers for each session also helps us identify any patterns or anomalies. In Figure 7 we saw some patterns, such as the greater incidence of tentativeness- related turns and the lower incidence of turns related only to causation in these datasets. The figure also helps identify anomalies. For instance, Glen and John from DTRS10's final review session for Tommy speak very less and neither of their speech turns are associated with tentativeness nor causation. This is a less interesting observation, but another anomaly can be seen in Don's speech turns in another DTRS10 session—the final review for Emily. All of Don's speech turns in this session—and he is one of the dominant speakers—are associated with tentativeness, either exclusively (orange), or along with causation (purple). In the following excerpts from the session, we focus only on Don's speech turns, with the rest of the speakers greyed out. We highlight only Don's speech turns in the way that we did earlier: words that fall under the categories of tentativeness, causation, or both.

EMILY: *So you could make it personalized or whatever color material*

DON: *Maybe different tops with the color–*

EMILY: *Mm-hmm.*

…

…

STEVE: *– start point, could you have a start at one end there –*

DON: *And then you go back and you put the radius on the top back edge after the fact, possibly.*

JOHN: *Instead of putting a joint there?*

[Crosstalk—more than one "Yeah"]

DON: *That's the only way I could see it. Even then, I'm not sure –*

…

…

DON: *I guess the question would be is could you do it a little differently again instead of laying entire thickness of sheets of plywood, could you have external solid wood frame cut out, and then your draping the plywood skin to solid wood.*

JOHN: *So each veneer bowed on you can get tied [unintelligible]*

[Crosstalk]

JOSH: *Well, we'd have to – well, I mean [unintelligible] –*

DON: *Just the outside? I mean that's close –well, that bottom one's probably [unintelligible]*

Even though these speech turns are largely stripped of context in the above exchange, it is clear that Don is providing some feedback to Emily about colours, finishes, and material for her design. The feedback takes the form of suggestions and not questions: he is not challenging Emily to articulate her rationale but offering ways to perhaps improve her design and share his knowledge, and in this way his role appears to be more of a design consultant rather than a design teacher. The high matches with *tentativeness* are indicative of this role, and the words that are categorised as *causation* are also categorised under tentativeness. These overlapping words— *"possibly"* and *"mean"*—in this context are indicative of tentativeness rather than causation.

4.2 Examining lower incidences of causation

In this section, we will examine sessions where spoken turns showed a lower correlation between the two lexical categories. Figure 8 shows four such sessions. In other words, there is *no consistent* co-occurrence or separation between the two lexical categories for a given speech turn.
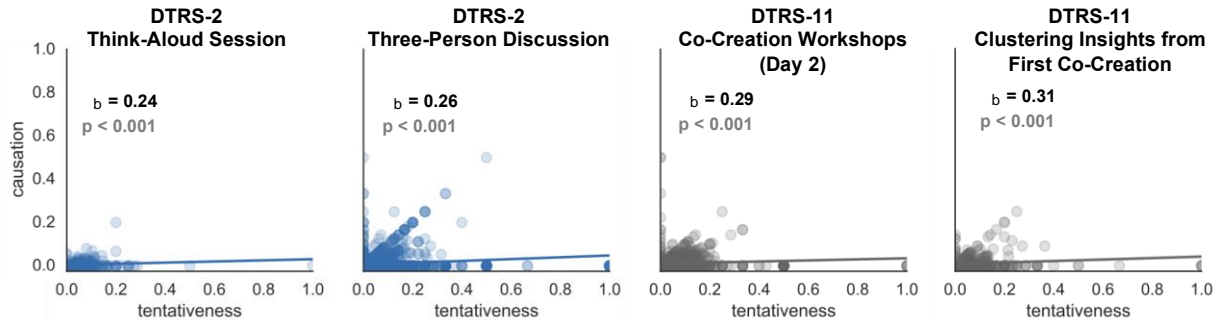

Figure 8. Scatterplots for four sessions with low-to-moderate correlation between the normalised number of matches between each turn and the two lexical categories of tentativeness and causation. Kendall's τ_b and regression line are shown for each session. The plots are coloured by dataset

Sessions with the lower correlations tend to belong to the DTRS2 and DTRS11 datasets, while higher correlations are seen in the case of sessions from the DTRS7 and DTRS10 datasets. This is perhaps because there was a greater opportunity to explore questions of causation and tentativeness together in a real-world designer-client discussion (DTRS7 crematorium meetings) or a teacher-student feedback session (DTRS10 datasets), as there would be more opportunities to challenge, question, and explain decisions. In contrast, perhaps co-creation sessions (DTRS11) or individual/collaborative design sessions (DTRS2) offer fewer opportunities to challenge and explain when exploring design processes? Answering such a question requires analysis that is beyond the scope of this work. However, we can examine these four sessions in closer detail to see if they share any patterns.
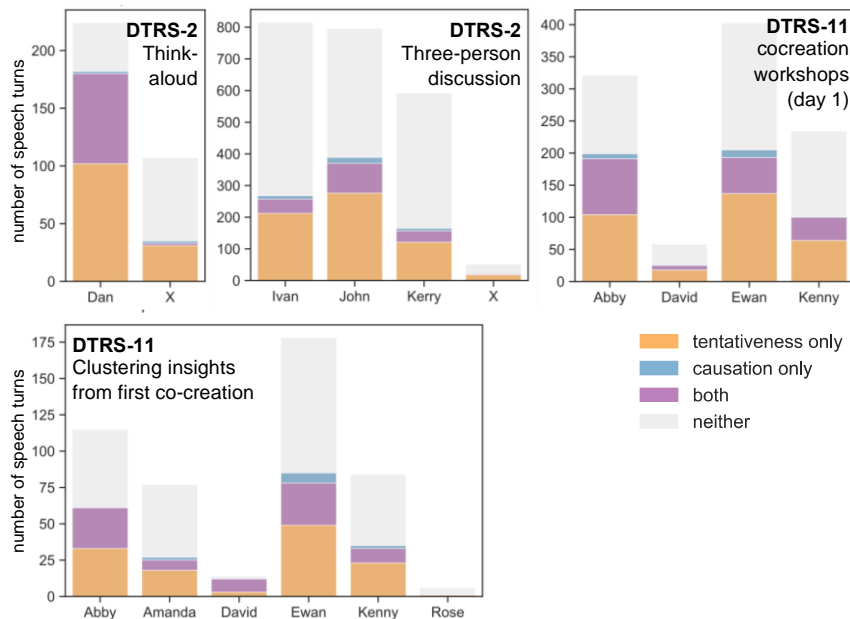

Figure 9. Speaker-wise turn counts for the four sessions identified in Figure 8 with lower correlations between causation and tentativeness matches. The proportion of turns that include a match that is exclusively with the tentativeness category (orange), exclusively with causation (blue), and with both categories (purple) are shown. A comparison with the sessions in Figure 7 shows a higher proportion of turn counts that match exclusively with the causation category, though this proportion remains the lowest of the three

Figure 9 shows the turn counts aggregated by speaker for each session. The first attribute of this figure that contrasts with Figure 7 is the greater proportion of "causation only" turns, especially for the dominant speakers. At first glance, this implies that at least some turns in these sessions relate to causation without relating to tentativeness at the same time. Examining these sessions at a turn-level, we see some evidence to support our conjecture about challenging vs. supplementing. Consider the following exchange from DTRS2. The discussion concerns the design of a bicycle pannier, with this part of the discussion concerning braze-ons—parts of a bicycle such as water bottle cage mounts, gear shifter bosses, or bicycle pump attachment pegs that are permanently attached to the frame. No matches for tentativeness are evident in this exchange, but the matches for causation and for both categories are highlighted below.

JOHN: *Ok they're a given I, I think we need a solution for people who don't have braze-ons, what if you don't have them*

KERRY: *there are solutions for that they they already make little bracket things that kinda*

IVAN: *Little junk*

JOHN: *They're kind like "wree wree"* IVAN: *Yeah just little junk things that y'can* JOHN: *I don't understand*
KERRY: *This goes around your tube that comes here*

JOHN: *Don't they seem cheesy though to you?*

First, the exchange itself is one where one speaker (John) is proposing the exploration of a certain solution space, and then challenging or questioning the suggestions from his teammate (Kerry). Kerry in response explains his solutions to John, who continues to pass judgment. On the face of it, this exchange is understandably one where John's questions and challenges might be categorised under *causation*, but let us take a closer look at the words that are tagged. The contexts for the utterances *"kinda"*, and *"kind"* suggest the tentativeness and vagueness associated with designerly thinking. Yet "kinda", which is a contraction for "kind of" is not part of the Empath category of tentativeness, and "kind" is part of the causation category. The final two turns by John asking for explanation and then judgment are closer to a reasonable interpretation of causation. Yet the last turn is categorised as both *tentativeness* and *causation* as the word *"though"* exists in both categories. As we noticed in the previous section, this miscategorisation can be attributed to the overlap of terms in the two lexical categories, though we do find evidence of "causation" in the snippet as a whole.

Compare the above snippet to the following exchange that occurs some minutes later:

JOHN: *We can assume Kerry has expert knowledge (laugh)*

KERRY: *But um we're gonna need to assume that but since we know the Batavus guys and the buddies with the backpack guy that, um, that we could design for that bike frame end*

IVAN: *Well, look the frame comes in different sizes, so –*

KERRY: *Yeah it comes in different sizes but that distance stays pretty constant, maybe we can still make it–*

IVAN: *Yeah, let's assume it's just this horizontal*

KERRY: *Maybe we can make it independent of this dimension somehow but I'm not sure if that's–*

This exchange is more collaborative: the designers are affirming each other's suggestions and building on each other's ideas and assumptions. As is evident from the highlighting, this exchange is entirely categorised under *tentativeness*. The categorisation here is fairly accurate with the exception of the word "so", which in this case could be seen as causation—the implication of the frame being offered in different sizes. Thus, a cursory inspection of the categorised turns appears to support our

conjecture: a stronger association between tentativeness and causation appears to occur when the discussion involves opportunities to challenge, question, and explain. These opportunities appear to arise more when people are explicitly or implicitly involved in roles to challenge or question—such as clients and teachers.

To further support this conjecture, let us examine one final exchange from DTRS-11, involving the clustering insights from a co-creation session.

EWAN: *I almost had this golden thing up here, like advertisement wise. 'Cause like everyone can buy a diamond, no one can buy commitment. (laughter) It is– and then it was like the next line "buy the company-product". (laughter) That was why it failed! "So buy our product!"*

ABBY: *"It's not a diamond, but eh"*

AMANDA: *"It's the commitment!"*

EWAN: *Yeah, it's embodied, it is yeah tangible commitment (laughter)*

ABBY: *It's a beautiful commitment!*

EWAN: *Yeah yeah, yeah. And the thing is- was that we elevate the eh, the kind of the exclusivity and the premiumness to the- it's not tangible anymore, it is the, it's a single value,*

ABBY: *Mhm.*

EWAN: *Eh, which is far beyond the golden diamond, which is like in the excessive part.*

ABBY: *(..) yeah.*

EWAN: *So I write here on the- yeah?*

ABBY: *But I kind of agree with will or whoever it was, asking if commitment and responsibility are- or how to make responsibility and commitment sexy.*

Putting aside the categorisation, we see that the exchange very much follows the affirm-and-build-upon pattern that we saw in the previous exchange. Abby, Amanda, and Ewan take turns to build on a take-away message (*"it's not a diamond, it's a (beautiful) commitment"*) from Ewan's story and link it to the problem they are working on, which is to make responsible decisions (similar to "commitment") surrounding the choice of a car to purchase more desirable. We would expect to see more tentativeness than causation here, and indeed we do, except they are not tagged as such. The words "kind of" and "almost" can be association with tentativeness, but "kind" is categorised as *causation* while "almost" is not categorised at all. On the other hand, "so" and "if"—words associated with causal reasoning—are both tagged as *tentativeness*.

It is evident that the lexical categories offered by Empath hold promise when applied to design discussion but require refinement when examined at the level of detail of close reading. In the next section, we discuss the implications of our approach and examine ways to address the nuance needed in applying lexical categories that are more relevant to design discussions.

5   Discussion

The goal of this paper was to illustrate the opportunities provided by new computational approaches for analysing design discourse through distant reading. To do this, we examine the balance between speculation on one hand and justification on the other, defining a lexical category of *tentativeness* as a marker for the former, and *causation* as a marker for the latter. We use Empath (Fast et al., 2016), a tool that provides a set of predefined machine-generated and human-refined categories and uses machine learning to generate new categories using seed terms. We use prior work on studying causation (Cardoso et al., 2014) and tentativeness (Ball & Christensen, 2009; Christensen & Ball, 2018; Glock, 2009) to provide seed words for Empath and generate the corresponding categories. We

use text-matching at the level of speech turns and aggregate categorised turns, visualising these categorisations at the level of datasets and sessions of interest.

While there is not a strong association between turns being categorised under tentativeness or causation, we find some patterns: higher incidences of both categories are found in sessions where the participants have roles and opportunities to challenge, question, and justify choices in addition to the user designerly approach of exploring the problem and solution space. We thus find a higher proportion of turns containing words from both categories in sessions under DTRS7—especially the exchanges between the client and the architect in the crematorium meetings—and DTRS10—with its teacher-student feedback sessions.

At the other end of the scale, the sessions that show a weak correlation between the two categorisations are characterised by more collaborative exchanges between participants where they affirm and build on each other's suggestions rather than challenge them. Such sessions—typically those under DTRS2 with its individual and collaborative design session, and DTRS11 with its co-creation sessions—exhibit exchanges more characterised by tentativeness than causation.

When drilling down into these exchanges, we also find cases where speech turns are miscategorised, due to the co-occurrence of words in both categories of *tentativeness* and *causation*. A close examination of both lexical categories reveals the words *because, knowing, mean, meant, might, possibly, suppose, surely, though*, and *yet* that appear in both categories. Polysemy—the existence of multiple meanings and/or senses for the same word—can be used to explain some of these co-occurrences. For instance, the word *"mean"* when used in the phrase "I mean" can be indicative of tentative thinking, but when used in a question, say, "What does this mean for the project?" can refer to the effect of a decision or choice. However, it does not explain the association of such words as "because" with tentativeness or "suppose" with causation. Perhaps the limitation is that of the training data. As mentioned earlier, Empath samples the vector space around the combined vector representations of a category name and seed terms. This vector space is created using a corpus of fiction as the training dataset. Are associations of tentativeness and causation in fiction the same as they are in design? Perhaps not. Fast et al. (2016) explain this limitation in their paper with the example of gloveboxes: it is possible that the word *glovebox* often co- occurs with the word *gun* in fiction with characters often taking guns out of gloveboxes, creating a strong association between the two that may not exist in other domains outside fiction. Thus, association-based lexical category population would work better if the associations in the training data are similar to the associations in the data to which the model is applied.

With the advent of new vector representations that provide different associations for the same words in different contexts of use (see Devlin et al., 2018), it may be possible to create design-specific word associations by using a vector space trained on design-related corpora such as designers' interviews, lectures/talks, as well as recordings of design discussions similar to the DTRS datasets. The pattern of examining the data at an aggregate level using visualizations, and then drilling down into detail for a closer inspection and verification of data is one that is well-established in the visual analytics community. Applications for interactive visualizations that combine distant and close reading of text data have gained traction in the digital humanities for examining documents (e.g., Jänicke et al., 2017; Jänicke et al., 2018; Koch et al., 2014) and conversations (e.g., Chandrasegaran et al., 2019; El-Assady et al., 2016). To a lesser extent, these approaches have been applied to examining design sessions (Chandrasegaran et al., 2017a; Chandrasegaran et al., 2017b). In the future, we plan to integrate computational approaches such as Empath with visual analytic approaches such as these to provide a fluid and interactive way for researchers to analyse design discourse at scale.

Our approach of combining distant reading approaches enabled by computational analyses of designers' speech with traditional close reading approaches has implications for pushing the boundary of our understanding of where design thinking may or may not occur. Qualitative analyses of design

thinking concepts have typically focused on explicitly "designerly" talk such as those captured in the DTRS datasets. There is now emerging work on studying design thinking concepts in ostensibly non-design scenarios such as, say, parliamentary debates (Umney & Lloyd, 2018). Our approach can provide researchers with ways to scale up such studies, to "search" existing records of conversations for occurrences of concepts relating to design thinking. Our work has also the potential for labelling larger conversation datasets, which can be used to train artificially-intelligent conversational agents to interact and work with designers productively, especially if the agents can recognise and respond to certain kinds of designerly talk.

6 Conclusion

In this paper, we present an examination of speculation contrasted with justification in design activity at a larger scale than typically conducted for such analyses. To do this, we use a machine-learning-based tool called Empath that we used to create lexical categories of words commonly associated with "tentativeness"—linked to speculation—and words associated with "causation"—linked to justification. By examining the matches between speech turns and these two lexical categories aggregated and visualized at the dataset level, the session level, and finally a close reading at the turn level, we identify patterns of dialogue between designers and clients or design students and their teachers that have a stronger association between speculation and justification. We also find patterns of dialogue between designers in collaborative sessions and co-creation sessions that are mostly speculative and only weakly associated with justification. Both these findings illustrate the value of using computational analysis for identifying patterns across design discourse and analysing associated text via a combination of distant and close reading techniques. Using examples of miscategorisation of speech turns in the same datasets, we also highlight the pitfalls of using lexical analysis incorporating machine learning data (fiction) that is different from the application data (design discussions). We propose a regular updating of such machine-learning models with contextual data, combined with a human-in-the-loop approach to verify the patterns highlighted by computational models. We close by discussing the implications of such work in expanding the boundaries of where one can find instances of design thinking, and in helping train AI agents to recognise and respond to designerly talk.

References

Adams, R. S., Cardella, M., & Purzer, S. (2016). Analyzing design review conversations: Connecting design knowing, being and coaching. *Design Studies*, *45*, 1-8.

Akdag-Salah, A., Chandrasegaran, S., & Lloyd, P. (2021). Storytelling in design activity: A Linguistic Inquiry and Word Count (LIWC) exploration of the DTRS datasets. In *Proceedings of the Design Thinking Research Symposium 13 (accepted).*

Ball, L. J., & Christensen, B. T. (2009). Analogical reasoning and mental simulation in design: two strategies linked to uncertainty resolution. *Design Studies*, *30*(2), 169–186.

Ball, L. J., Onarheim, B., & Christensen, B. T. (2010). Design requirements, epistemic uncertainty and solution development strategies in software design. *Design Studies*, *31*(6), 567-589.

Cardoso, C. C., Eris, O., Badke-Schaub, P., & Aurisicchio, M. (2014). Question asking in design reviews: how does inquiry facilitate the learning interaction? In *The 10th Design Thinking Research Symposium* (pp. 1–18). Purdue University, Indiana, USA.

Cash, P., & Kreye, M. (2018). Exploring uncertainty perception as a driver of design activity. *Design Studies*, *54*, 50–79.

Chandrasegaran, S., Badam, S. K., Kisselburgh, L., Peppler, K., Elmqvist, N., & Ramani, K. (2017a). VizScribe: a visual analytics approach to understand designer behavior. *International Journal of Human-Computer Studies*, *100*, 66–80.