Constructing design activity in words

Exploring linguistic methods to analyse the design process

Chandrasegaran, Senthil; Salah, Almila Akdag; Lloyd, Peter

DOI
10.1016/j.destud.2023.101182

Publication date
2023

Document Version
Final published version

Published in
Design Studies

Citation (APA)
Chandrasegaran, S., Salah, A. A., & Lloyd, P. (2023). Constructing design activity in words: Exploring linguistic methods to analyse the design process. *Design Studies*, *86*, Article 101182. https://doi.org/10.1016/j.destud.2023.101182

Constructing design activity in words: Exploring linguistic methods to analyse the design process

Senthil Chandrasegaran, Faculty of Industrial Design Engineering, TU Delft, The Netherlands

Almila Akdag Salah, Department of Information and Computational Sciences, Utrecht University, The Netherlands

Peter Lloyd, Faculty of Industrial Design Engineering, TU Delft, The Netherlands

Analysing transcripts of design activity typically involve either close reading or manual coding of data, which limits the amount of data that can be analysed. In contrast, we explore a machine-learning based linguistic analysis tool called Empath to identify patterns of reasoning in design talk. The data we use derives from the Design Thinking Research Symposium (DTRS) shared-data workshops which we analyse to look at two contrasting aspects of design talk: the expression of tentativeness, characterising designers' generative thinking; and the articulation of explanations, characterising their deductive or analytical thinking. We show, at the level of speech turns, how tentativeness and explanation relate to, and overlap, each other. Finally, we discuss the limitations of this 'linguistic analysis at scale' approach.

Keywords: design thinking, collaborative design, design activity, research methods, text analysis

Corresponding author:
Senthil Chandrasegaran
r.s.k.chandrasegaran@tudelft.nl

Is design activity characterised by an equal balance between speculation and rationalisation? Models of designing that describe a fundamental cycle of activity generally describe projective activity, characterised by tentativeness or epistemic uncertainty, followed by explanatory activity, characterised by evaluation and justification (Lloyd, 2019). For example, Schön's description of designing (1992) as a series of reflective 'moving experiments' is premised on the idea that something material must be put into the world before the understanding of its implications can take place and therefore be justified. Similarly, Roozenburg (1993), in his 'basic model of designing' describes a process initiated by the logic of abduction prior to the deduction

www.elsevier.com/locate/destud
0142-694X *Design Studies* **86** (2023) 101182
https://doi.org/10.1016/j.destud.2023.101182

of consequences. An equivalence of projective and explanatory processes in design activity is implied by these models but is this borne out in practice? In this paper we explore four datasets deriving from the Design Thinking Research Symposium's shared-data workshops to computationally categorise instances of tentativeness and of explanation. 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 build on the work of Menning et al. (2018) to further explore how contemporary computational analysis tools can complement 'close' reading of transcripts of design activity.

Over a period of nearly 30 years, the Design Thinking Research Symposium (DTRS) series, described by Cross (2018), has conducted four shared-data workshops, generating data from the 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). Section 2 provides a short summary of each workshop.

Analyses of these shared datasets have revealed insights into both the speculative and rationalising elements of design thinking by looking specifically at word usage in the transcripts. Aspects of speculation have been explored in, for example, how analogies and mental simulations are used to resolve uncertainty (Ball & Christensen, 2009), how vagueness allows space for negotiation (Glock, 2009), how professional roles are constructed (Oak, 2009), and how framing opportunities are taken up (McDonnell, 2017). Aspects of rationalisation have been explored through, for example, how judgements are given (Oak & Lloyd, 2016), how learning opportunities are created (Adams et al., 2015), and how problems and solutions are connected (McDonnell, 2009). However, the distinction between speculation and rationalisation is not always as clear cut as the 'basic' models of designing mentioned in the first paragraph might indicate, and as we will go on to show.

In this paper we draw on a corpus resulting from the combination of the four shared-data DTRS workshops to computationally explore the concepts of 'tentativeness' and 'explanation' relating, in turn, to elements of speculation and rationalisation in the above-mentioned studies. The machine-learning approach we apply is something that has only recently become viable, with analyses of designers prior to this mainly focusing on smaller design 'protocols', and the manual identification of textual excerpts to explore new theoretical

concepts. This 'traditional' way of analysing design activity is akin to 'close reading', a term from 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 that takes a global view of a text, analysing and visualizing its more general features. Distant reading thus relies on computational analyses of large amounts of text, the results of which are presented graphically in the form of charts or data visualizations. We will present our use of both approaches—i.e., close and distant reading—for the exploration outlined in this paper.

In previous work (Lloyd et al., 2021) we have explored the idea of tentativeness using a dictionary-based approach embodied in a software program called Linguistic Inquiry and Word Count or LIWC (Pennebaker et al., 2015). LIWC is a tool comprising of 104 human-curated socio-psychological and grammatical lexical categories[1] that is used to classify individual words in a text. LIWC has subcategories of both 'tentative' and 'causation' under the category of 'cognitive processes.'

The 'tentative' subcategory of LIWC includes words associated with speculative, projective thinking such as 'if', 'maybe', 'might', 'perhaps', 'possibly', and 'probably'. This is the idea that creative behaviour is triggered in situations of uncertainty to lessen that uncertainty and thus 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; Paletz et al., 2017), and other studies (Cash & Kreye, 2018) has shown that designers typically use downtoners, hedges, modal adverbs, and other expressions of tentativeness when proposing new ideas or interpretations. We refer to this category as 'tentativeness' in this paper rather than 'epistemic uncertainty' for two reasons: (a) to capture both the notion of epistemic uncertainty and the use of downtoners when proposing new ideas, and (b) to connect to LIWC's 'tentative' subcategory under 'cognitive process' explored in prior work (Lloyd et al., 2021). The use of a term from this lexical category is typically an indicator that the designer is considering or suggesting the exploration of a possibility or future conditional.

The 'causation' subcategory of LIWC contains words like 'because', 'how', 'allow', 'make', 'force' and so on. In LIWC, words in this subcategory are considered indicative of engagement with past experiences and processing them to move past the effects of these experiences. In the context of design, we look at 'causation' as the process of designers explaining, analysing, or justifying their design proposals or choices, or those of their collaborators, apprentices, or students. Cardoso et al. (2014) show that various forms of 'deep reasoning questions' aid students' reflection on the state of their design. Forms

of such questions include those that help students think about their design rationale, the potential effects of their design choices, or their interpretation of related phenomena. Work by Christensen and Ball (2016) on instructors' evaluations of student designs show the use of mental stimulations—"*reasoning about new possible states of a design object in terms of its qualities, functions, features or attributes*" (p. 124)—that accompany evaluations of functional aspects of the designs. Work by Adams et al. (2016) shows that asking students to articulate their reasoning and helping them to consider the limitations of their designs—seen as some of the patterns in how coaches engage in pedagogy—require coaches to ask for or provide explanations and reasons.

Given the imperfect alignment of the cognitive processes of 'tentativeness' and 'causation' as defined in LIWC to aspects of *'tentativeness'* and *'explanation'* as described in the context of designing, we explore the latter two categories using a machine learning approach called *Empath* (Fast et al., 2016). Empath supports the automated creation of custom categories through the use of seed words, which are then used to generate a lexical category of additional terms related to the seed words. We examine the generated terms, propose ways of refining them to suit our context, and use the new lexical categories to explore the DTRS datasets.

Our exploration of tentativeness and explanation uses a combination of close and distant reading approaches to (a) highlight established instances of *tentativeness* and *explanation* in prior work using computational approaches and (b) examine, using the same computational approaches, the global patterns of occurrence of these categories across datasets and sessions. We suggest 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-based conversational systems to recognise designerly talk.

1 The DTRS dataset

The DTRS series (Cross, 2018) includes a series of 'common data workshops' that have resulted in four shared datasets, created so that different perspectives, methods, and theories about designing can be proposed and tested. The four datasets, each consisting of audio and video material with corresponding transcripts, 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.

- **DTRS2** consists of a 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.

- **DTRS7** 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 min). In the first sessions the design of two co-creation sessions 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 1 shows the session numbers and lengths for each of the four DTRS workshop datasets. At a combined 373, 983 words, these datasets, forming a corpus, provide a composite picture of design activity. They cover such different forms of design activity as proposing, reflecting, and evaluating, different contexts of designing as client discussions, design reviews, think-aloud sessions, and cocreation sessions, and reflecting different situations of designing like educational and professional settings. The corpus thus provides opportunities for examining different kinds of design thinking at scale.

2 *The use of lexical categories in computational text analysis*

The traditional, qualitative approaches to analyse text have been complemented by various computational approaches in the past decades. These computational approaches include parts-of-speech identification (e.g., Toutanova et al., 2003) used for natural language understanding, topic modelling for identifying thematic content (see Vayansky & Kumar, 2020), and sentiment analysis for gleaning affect in large-scale text (Feldman, 2013), to name a few. Of relevance to this work is the use of lexical categories to identify markers of certain kinds of thinking and/or behaviour. Specifically, this

Table 1 The DTRS dataset statistics

Dataset	*Filmed sessions*	*Dataset size (words)*	*Session size (words)*	
			Mean	*S.D.*
DTRS2	2	37, 969	18, 984	4085
DTRS7	4	68, 861	17, 215	4944
DTRS10	38	92, 751	2441	3424
DTRS11	20	174, 402	8720	4590
Total	64	373, 983	5843	6162

involves the creation of psycholinguistic dictionaries of lexical categories and counting words from spoken or written text that match these categories, using increased word count as indicative of certain behaviour (Tausczik & Pennebaker, 2010). For instance, the use of a set of prepositions in speech or writing are found to be indicative of complex and concrete information being provided, while the use of 'exclusion words' like 'without', 'except', or 'apart (from)' are indicative of cognitive complexity (Tausczik & Pennebaker, 2010). One of the most widely used and validated applications for lexical analysis of text is Linguistic Inquiry and Word Count, or LIWC (Pennebaker et al., 2015).

LIWC comes with a predefined set of lexical categories that are curated and periodically revised by a team of experts in psychology and linguistics. In addition, it allows the user to create their own dictionaries for domain-specific analyses or to allow researchers to formulate and validate new lexical categories. However, the creation of new categories needs careful curation and iteration, which in turn demands time and effort (Donohue et al., 2014). In response to this challenge, Fast et al. (2016) proposed *Empath*, a machine learning approach that mines natural language text corpora to extract topical and emotional signals from text, embodied as a tool that allows the creation and population of new lexical categories when provided a few words as seed text. The topical, cognitive, and emotional signals in Empath are encoded in a multi-dimensional vector representation such that semantically related words are represented by vectors that are close to each other in Empath's vector space. For instance, words such as 'dog', 'cat', 'gerbil', and 'pet' would have vector representations that are close to each other because these words are related in meaning and/or context. Given a set of related seed words, Empath can query the space around the vectors corresponding to these words and generate additional words that are semantically related to the seed words. These additional words are used to populate the required category.

Note that to create this vector space representation, Empath uses as its training data a large corpus of modern fiction,[2] as it has been shown to provide a better breadth of topical and emotional categories. Thus, the semantic associations between words in the vector space would reflect how the words are often associated in fiction, and may be different from how the words are associated in specific domains.

Both LIWC and more recently Empath have seen widespread use in different text analysis applications. LIWC has been used to identify appropriate decision-making methods to address a problem based on the language used to describe the problem (McHaney et al., 2018), tracking emotional changes in a design team across different stages of a creative process (Ewald et al., 2019), study the effect of conflict and team diversity on team creativity using LIWC's 'insight' category (Paletz et al., 2018), and explore how the mode

and content of information given to novice designers could affect their creative process (Mertens & Toh, 2019). Empath has been used in studies identifying the kinds of information on Wikipedia entries for which readers engage with in-line references (Piccardi et al., 2020), changes in perceptions and emotions associated with artificial intelligence in journalism (Fast & Horvitz, 2017), and identify vocabulary entries distinct to hate speech on social media (Ribeiro et al., 2018).

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. As mentioned above, the two behaviours we focus on are expressions related to *explanation* and *tentativeness*.

Using a tool such as LIWC provides a consistent and scalable method of analysis. However, as discussed in the Introduction, predefined lexical categories may not always match categories that researchers are seeking to identify for their work. In this section, we describe the computational generation of alternative lexical categories by seeding terms associated with tentativeness and explanation. We then use these categories as lenses with which to examine the DTRS datasets.

3.1 Pre-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 other 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. Transcripts were combined to enable analysis of the entire corpus, while still being able to filter for individual datasets, sessions, or speakers.

3.2 Creating lexical categories

To choose appropriate seed terms to create the lexical category of explanation, we refer to studies of the DTRS10 dataset involving design review sessions between students and instructors (discussed above). Focusing on the 'rationale' (why) aspect of explanation, specifically causal reasoning, we input seed words such as 'because', 'effect', 'explain', 'how', and 'why' to generate a category called *'Explanation'* using Empath. This resulted in a lexical category with the following 84 terms[3]:

> *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

In this list, some words such as 'crisis', 'critical', 'disastrous', 'illness', etc. are terms that do not necessarily relate to 'explanation' or justification in typical design contexts. Recall that Empath is trained on text corpora composed of works of fiction. In the context of fictional writing, it is reasonable to associate these terms with causation: the cause or effect of events and decisions can be 'critical' or 'disastrous' depending on the storyline and genre of fiction. For the purpose of this analysis, we could manually remove these terms from the list. While manual examination of individual terms would incorporate a better understanding of nuance and context, it could also result in errors or inconsistencies.

We propose a combination of such a manual approach with a computational approach to identify semantic groups within this lexical category through the use of word embeddings. Word embeddings (Mikolov et al., 2013) are multidimensional vector spaces that represent a given language such that vector operations can be used to express semantic relationships between words. For instance, words that are closely related to each other in meaning or context when represented as vectors in a word embedding would also show up as physically close to each other in the vector space. There are several word embeddings available for use such as Word2vec (Mikolov et al., 2013), FastText (Bojanowski et al., 2017), and ConceptNet Numberbatch (Speer & Chin, 2016), to name a few, generated using different encoding techniques on different training datasets. Each word embedding has a set vocabulary or list of words encoded in the embedding. We use FastText (Bojanowski et al., 2017), as its vocabulary or word list has a reasonable overlap with the words in our lexical categories of *tentativeness* and *explanation*, thus allowing us to look up vector representations for most of the terms in the categories. We compute cosine distances between the vector corresponding to each word in the dictionary for the *explanation* category above and each of the vectors corresponding to the remaining word in that same dictionary. We then group the words hierarchically using agglomerative hierarchical clustering (Day & Edelsbrunner, 1984) such that groups of semantically-distinct words can be visually recognised and separated. Figure 1(a) shows the result of this grouping.

We can see semantically distinct groups emerge in Figure 1(a), such as cluster number 3 in gray, containing the words 'illness', 'mental', 'killing',

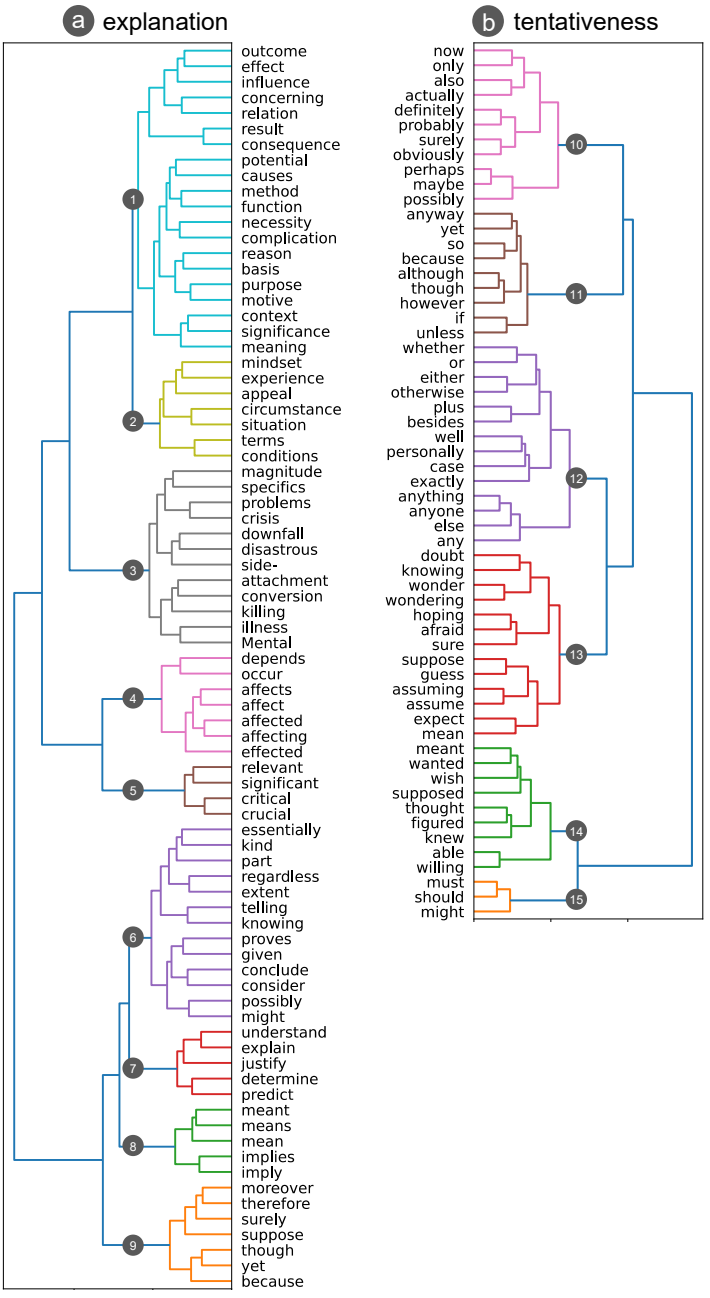

Figure 1 Words generated to form the categories of (a) 'explanation' and (b) 'tentativeness' clustered based on cosine distances of vectors corresponding to each word. A dendrogram representation is used to show hierarchies in clusters for manual inspection and coloured based on a threshold established through visual inspection. Words connected by similarly coloured lines indicate a cluster within which words are more semantically related to each other than they are to the remaining words outside the cluster. These clusters are also labelled by numbers (1–15)

Constructing design activity in words

'conversion', attachment', 'disastrous', 'downfall', 'crisis', 'problems', 'specifics', and 'magnitude'. Most of these words may be associated with causal reasoning in works of fiction, where cause and effect is often dramatic and may include illness, disaster, or even killing. However, these are not terms associated with explanation related to causal reasoning in design. We can thus remove this set of terms. Similarly, we also see some words in cluster 6 (purple) like 'might', 'possibly', 'conclude', 'consider', and 'essentially', which are more about tentativeness and uncertainty rather than explanation. The words corresponding to *tentativeness* can thus be examined and removed manually. This approach still involves some subjectivity (for instance, 'conclude' in cluster 6 is retained because it is closer to *explanation* than to *tentativeness* or *epistemic uncertainty*. However, the clustering allows for a structured way in which to examine the terms. Note that out-of-vocabulary terms (words that are not included in the list of terms in the word embedding) from the generated lexical categories such as 'mental state' and 'end result' are replaced with the closest term found in the vector space, such as 'mental' and 'consequence' respectively. These substitutions are also reflected in the dendrogram shown in Figure 1(a).

The second lexical category of 'tentativeness' is motivated by the emerging sub-field in design research known as 'epistemic uncertainty' mentioned earlier. Since Empath is trained on text from works of fiction, it again lacks some domain-specificity. Thus, 'tentativeness' in Empath is a more interpretable category than 'epistemic uncertainty', which is a term more specific to designing (e.g., Ball & Christensen, 2009; Ball et al., 2010). Using seed terms such as 'if', 'maybe', 'might', 'perhaps', 'possibly', 'probably', etc. in Empath, we thus 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*

Using a similar clustering approach as for the *explanation* category, we group terms in this category as shown in Figure 1(b). While most of the terms in this category appear related to tentativeness, we can again see terms in the brown cluster ('unless', 'if', 'because', 'so', 'yet') that seem to be closer to the *explanation* category. As before, we remove those terms and add them to the explanation category.

Finally, we examine the words common to both categories. These are 'because', 'knowing', 'mean', 'meant', 'might', 'possibly', 'suppose', 'surely', 'though', and 'yet'. To determine which of the remaining terms should be removed from which category, we perform the same clustering approach (see Figure 2).

Immediately, we see two clear clusters: the cluster at the bottom (in orange) with the terms 'because', 'knowing', 'suppose', 'surely', 'though', and 'yet', and the cluster at the top (in green) with the words 'mean', 'meant', 'might', and 'possibly'. While the top cluster (green) clearly belongs in the tentativeness category, the bottom (orange) cluster's most suitable category is not immediately apparent. Based on similarity to terms like 'because', 'though', and 'yet' which clearly belong in *explanation*, we also choose 'knowing', 'suppose' and 'surely' to also include in this category.

The final list of terms in the *explanation* category is thus:

affect, affected, affecting, affects, appeal, basis, because, causes, circumstance, complication, concerning, conclude, conditions, consequence, consider, context, critical, crucial, depends, determine, effect, effected, end result, essentially, experience, explain, extent, function, given, if, implies, imply, influence, justify, kind, knowing, magnitude, main problem, meaning, means, method, mindset, moreover, motive, necessity, occur, outcome, part, potential, predict, proves, purpose, real problem, reason, regardless, regardless, relation,

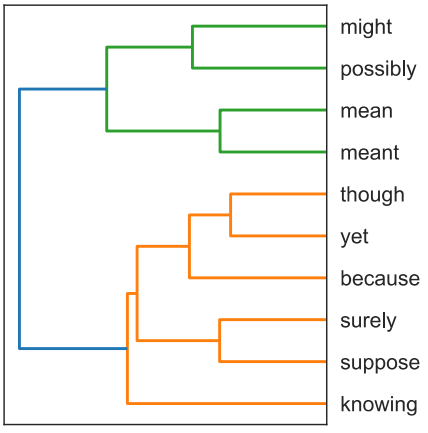

Figure 2 Clustering of terms shared between the two categories of 'tentativeness' and 'explanation' show two groupings with the green cluster (top) aligned more to the tentativeness category, while the orange cluster (bottom) tends more to the explanation category (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Constructing design activity in words

relevant, result, significance, significant, situation, so, specifics, suppose, surely, telling, terms, therefore, though, understand, unless, yet

Similarly, the final list of terms in the *tentativeness* category is:

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

With the categories now fixed, and using speech turns as the unit of our analysis, we count the number of matches between each lexical category and the words in the turn. We only looked for whole word matches, choosing not to lemmatize[4] the words, neither in the turns nor in the lexical categories, as the sense of such words is often linked to the specific form of the word. For instance, a verb-noun form such as "is not possible" is more assertive while a modal verb-modal adverb form such as "cannot possibly" is more tentative. In the following section, we use the measures of matches between speech turns and the two lexical categories created to examine the question of how far design activity is tentative, explanatory, or both, and in what contexts. We first illustrate how matches between a given transcript and either lexical category can be interpreted. We then attempt to answer the question at different levels of aggregation: at the level of the dataset, sessions, and then individual speakers.

4 Results

4.1 Explanation of word matches with an example

We first examine an exchange from an architect-client meeting from DTRS7. While there are other interlocutors in the conversation, the two dominant speakers are the architect (labelled 'AM' in Excerpt 1 below), and the client (labelled 'AA'). Words that match the *tentativeness* category are underlined and highlighted in orange, while words matching the *explanation* category are in **bold** and highlighted in blue. Note the sparsity of words in each turn that match each lexical category. In the context of this paper, we treat a speech turn with *one* match to a lexical category the same as a speech turn with *more than one* match to the category. All data on matches to lexical categories pertain to the number of turns with *at least one match* to either or both categories. While other measures such as the number of matches within turns or the percentage of words in a turn that match with a lexical category can also be used, we find that treating 'match/no match' as a binary measure for a turn is sufficient to illustrate our approach in this work.

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

392 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 into 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.

393 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

394 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.

395 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

396 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

397 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

— Excerpt 1: From the first Client-Architect meeting, DTRS7 Dataset

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 (marked by words indicative of tentativeness such as 'might' or 'probably') or explanations (marked by words indicative of explanation such as 'because' or 'if'). The questions and declarations of the client show similar patterns, with tentativeness indicating their interest in alternatives and explanation when they provide the reasoning behind a requirement.

The above example shows how lexical categories—once created and refined—can be used as lenses with which to closely examine verbal data from design sessions. The matching of terms between dictionaries and a given body of text or a transcript is computational, which means that it can scale to the level of turns, sessions, or even datasets, thus supporting a combination of overview and detail analyses.

Note that most of the turns across the dataset do not have word matches with either of the two lexical categories (see Figure 3 for distribution), as they are either quick exchanges of one to three words either multiple people agreeing

with what someone said, stating facts, or narrating real or imagined scenarios like the one below:

```
848 AM:  in his opinion, the options include and extra-large cremator for,
849 AA:  for bigger —
850 AM:  folks —
851 AA:  people —
852 AM:  my size —
853 AA:  and ladies my size too, I'm building up to getting stuck in the cremator
         doors, that's what I'm building up to [laughter]
854 AA:  They won't forget me when they cremate me that's what I'm looking for, yes
855 AM:  and that'll cost you another seven and a quarter thousand pounds.
```

— Excerpt 2: From the first Client-Architect meeting, DTRS7 Dataset

4.2 Overview analyses using lexical categories

One of the advantages of using dictionary approaches such as LIWC and Empath is that of scale. Once the lexical category words are finalised, performing a word match at the level of turns, speakers, sessions, and even datasets provide increasing levels of overview. For instance, Figure 3 shows the result from counting the number of turns across all the sessions in each DTRS dataset containing at least one match for each lexical category of *tentativeness* and *explanation*. Since each dataset and session are of varying sizes/lengths (see Table 1), we normalise this count with the total turns for each dataset to show turns with matches to one or both lexical categories as a proportion of the total turns.

Figure 3 shows an almost-equal distribution of turns matching *tentativeness* (orange with hatching from upper left to lower right), *explanation* (blue with hatching from lower left to upper right), and an overlap of both categories (purple with cross-hatching) together across the DTRS2, DTRS7, and DTRS10 datasets, with a slightly higher proportion of overlap for the DTRS11 dataset. Within each dataset the distribution of turns across *tentativeness*, *explanation*, and both together is similar. These patterns indicate a similarity in the way in which design conversations occur across different contexts of designing. This also provides—at least at an aggregate level—the answer to our initial question of "*is design activity characterised by an equal balance between speculation and rationalisation?*" The answer seems to be 'almost', with the balance tipping slightly towards *speculation*. On average across the datasets, 15% of speech turns have at least one word matching only the tentativeness lexical category, 12% of turns have at least one match with only the explanation category, 16% of turns have at least one match with both tentativeness and explanation categories, and 57% of speech turns have no matches with either category.

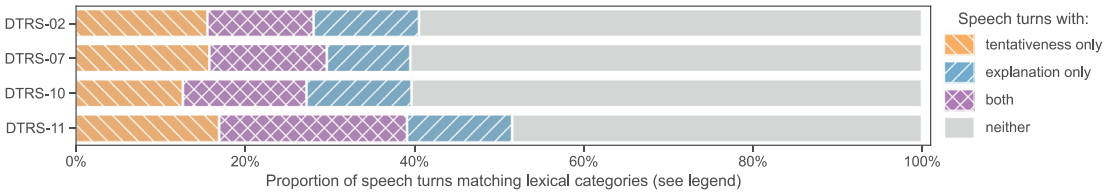

Figure 3 Stacked bar chart representing, for each dataset, the proportion of speech turns containing terms from the lexical categories of tentativeness, explanation, and both. The chart shows a uniformity in the distribution across the DTRS datasets. An exception is the DTRS11 dataset, which shows a higher proportion of speech turns matching both categories

Given the way in which we have structured the data (see Section 3.1), it is possible to attempt to answer this question at different levels of aggregation. One can create and examine such distribution charts for each session within a dataset, each speaker within a session or across multiple sessions, or even each speech turn in a session. We chose to look at two sets of sessions from each dataset representing two extremes in one of the patterns we see at this level of overview. Of these two sets, the first set of sessions represents a session from each dataset with the *highest* number of turns that match with both lexical categories, i.e., turns that feature words from both categories of *tentativeness* and *explanation*. The second set of sessions represents the other extreme—a session from each dataset with the *least* number of turns that match with the same two categories.

Figure 4 shows the sessions from each dataset showing the *most* number of turns featuring matches to *both* lexical categories. Note that each session has a different number of speech turns, so the charts show the percentage of turns within a session that match each category. The proportion of speech turns for each speaker is indicated by a stacked bar chart, with the total height of each stacked bar representing the proportion of speech turns taken by the corresponding speaker. Each stacked bar is also split into four parts, the height of each part representing the proportion of turns by that speaker that feature a match with the *tentativeness* category (orange with hatching from lower left to upper right), *explanation* (blue with hatching from upper left to lower right), both categories (purple with cross-hatching) and neither category (light grey). Within a session, the total height of each bar also indicates the proportion of speech turns taken by the corresponding speaker. The four sessions shown in the figure include a think-aloud protocol (DTRS2), an architect-client meeting (DTRS7), a design review session with an undergraduate student (DTRS10), and a recap of co-creation workshops with external consultants (DTRS11). These will be discussed in detail in Section 4.3.

Figure 5 shows the dataset from the set of sessions that show the *least* number of turns matching both lexical categories, with percentages matches between turns and categories represented in the same way as for Figure 4. The sessions

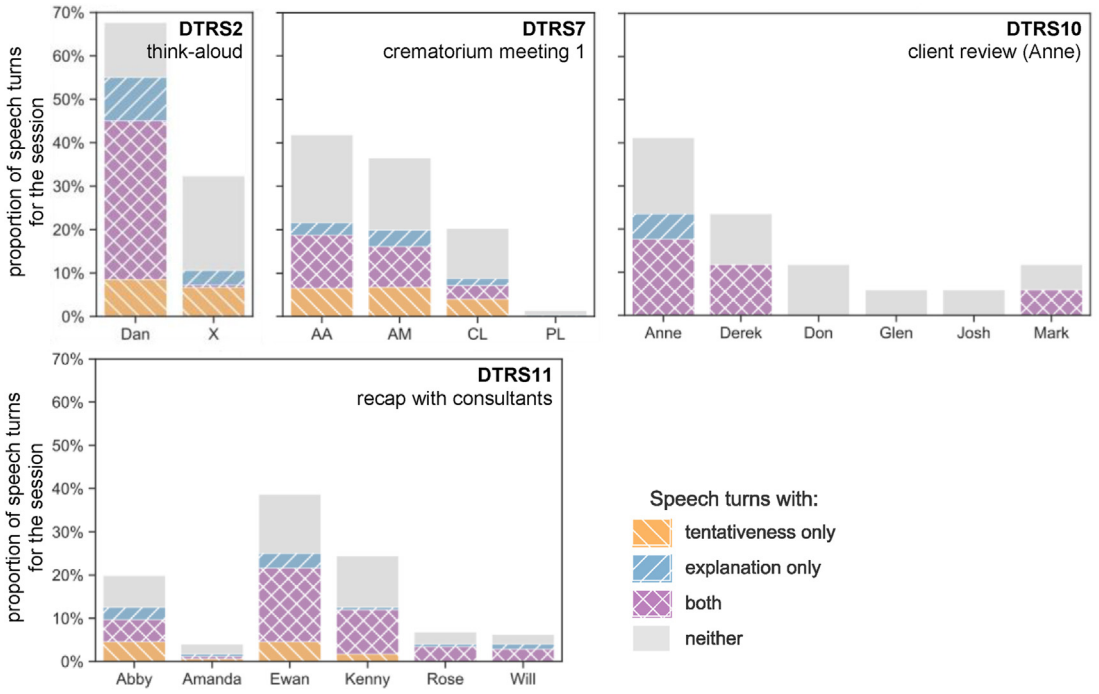

Figure 4 Speech turns separated by speaker and coloured according to whether a turn has words that match the 'tentativeness' category (see legend), the 'explanation' category, or each of both categories. Each session shown above—one from each dataset—features the most number of turns matching both. Turn counts are shown as a proportion of total speech turns for that session to enable comparison across the sessions. The turns matching both categories indicate instances of both generative and deductive thinking (Lloyd & Scott, 1994). See Section 4.3 for details and Table 2 for detailed turn distributions for these sessions

in this set include a collaborative design session (DTRS2), the final design review of an undergraduate student (DTRS10), and a meeting between designers, consultants, and an automotive accessories department to share insights from a prior co-creation session (DTRS11).

4.3 *Speech turns featuring both* tentativeness *and* explanation *together*

Comparing Figures 4 and 5, it is not immediately clear if the inherent nature of the session is the reason for the difference in the distribution of speech turns that match with tentativeness only, or with explanation only, or with both. At first glance, the sessions do not look very different in nature: both figures feature a review meeting between students and instructors (DTRS10), and a meeting between designers and consultants/other collaborators (DTRS11). On the other hand, the think-aloud session from DTRS2 features a high number of speech turns featuring words from both *tentativeness* and *explanation* categories (Figure 4), while the collaborative design session from the same

Table 2 Detailed breakdown of speaker turns and lexical category matches for sessions shown in Figure 4

Dataset	*Session*	*Speaker*	*Proportion of total turns with at least one match for:*				*Total % turns per speaker*
			Only Tentativeness	*Only Explanation*	*Both Categories*	*Neither Category*	
DTRS2	Think aloud	Dan	8%	10%	**37%**	13%	68%
		X	7%	3%	1%	22%	32%
DTRS7	Crematorium meeting	AA	7%	3%	**12%**	20%	42%
		AM	7%	4%	**9%**	17%	36%
		CL	4%	2%	3%	12%	20%
		PL	–	–	–	1%	1%
DTRS10	Client review Anne	Anne	–	6%	**18%**	18%	41%
		Derek	–	–	**12%**	12%	24%
		Don	–	–	–	12%	12%
		Mark	–	–	6%	6%	12%
		Glen	–	–	–	6%	6%
		Josh	–	–	–	6%	6%
DTRS11	Recap with consultants	Ewan	5%	3%	**17%**	14%	39%
		Kenny	2%	1%	**10%**	12%	24%
		Abby	5%	3%	5%	7%	20%
		Rose	–	1%	**3%**	3%	7%
		Will	–	1%	**3%**	2%	6%
		Amanda	1%	1%	1%	2%	4%

Note: Bold text indicates instances where there are more turns featuring a match with both tentativeness & explanation than turns featuring a match with either tentativeness or explanation.

dataset features a low number of speech turns with matches to both categories (Figure 5).

To explain similarities and differences in the use of the words in each lexical category in speech turns across these different sessions, we first examine representative speech turns from some of the sessions shown in Figure 4. The following excerpts are coded the same way as Excerpts 1 & 2, i.e., words matching the *tentativeness* category are underlined and highlighted in orange, while words matching the explanation category are in **bold** and highlighted in blue. The numbers next to the speakers indicate the turn number in that session.

238 DAN: we have er this thing pulling up er this is three inches em we have em a moment of, er, twenty times fifteen is equal to, em, three hundred inch-pounds, er, is gonna be that, em, on three **so** this here could have a hundred pounds of force up one hundred pounds force, ok, **so**- **so** that's just **telling** us what kinda numbers we're talking about here, em, I'm not scared by that **because** obviously a bicycle has much more than that that's lateral load and I'm assuming that that's the worst case of one g, one g side loads, lateral loads, em, assume em, thirty pound pack, ok now the **reason** I'm also gonna do that is, is **because** I'm sure that, em, people are gonna try and misuse it, in fact kids are gonna sit on this thing **so** it's probably gonna have to be able to handle a little bit more than that **so** maybe this optional pin here will work...

— Excerpt 3: From a think-aloud protocol session from the DTRS2 Dataset

Constructing design activity in words

48 AA (CLIENT): ... I mean the other suggestion that perhaps I could make at this stage would be perhaps for a small amount of outside seating, **because** people like to smoke at funerals, they like to have a, and the seat that we've got out by the car park at the moment, the half seat, even **if** it's cold and not very nice is, actually people feel more happier out there then they do sometimes in the waiting room...

49 AM (ARCHITECT): yes, well we can certainly add some outdoor seating out here **if** you wish. we have got some outdoor seating here, we've got a number of benches there, but we can add...

(Overlapping talk - turns 50 & 51 excised for brevity)

52 AA (CLIENT): ... especially **if** we have a no-smoking policy in the waiting room, which we have now, people like to mess up our grounds with their cigarette ends. Just a thought that they don't know where they're going most of the time even **if** you were to point them in the right direction, and they feel frightened, most people attend the funeral don't want to actually be right next to the family **because** they don't want to be in their face, and they don't want to get involved with it, but they want to attend the funeral, and **so** it's awkward they want to stand back a bit, they want to be seen that they're there, but they actually don't want to actually even perhaps sometimes speak to the other people, you know.

— Excerpt 4: From the first client-architect meeting, DTRS7 Dataset

Consider the two exchanges above, one from a think-aloud session from DTRS2, and one from a client-architect meeting in DTRS7. The turn from DTRS2—from a think-aloud design session—features a bicycle pannier design prompt and has the designer, Dan, making propositions: either about assumptions and requirements ("*...I'm assuming...one g side loads*"), or about his ideas for features ("*...maybe this optional pin here will work*"). Dan also explains the reasoning behind his assumptions ("*...the **reason** I'm also gonna do that is... **because** I'm sure that, em, people are gonna try and misuse it...*"). This is consistent with Lloyd and Scott's framework (1994) based on a think-aloud protocol of an individual's design processes, specifically generative utterances—what the authors describe as utterances involving "*creating something to reason about and advancing the solution*" (p. 127)—and deductive utterances—clarifying statements that involve "*perceiving and representing the problem*" (Lloyd & Scott, 1994).

The turn from DTRS7, displays similar instances of tentativeness in propositions (e.g. "*...the other suggestion that perhaps I could make ... for a small amount of outside seating*") and of explanation when justifying the propositions ("***because** people like to smoke at funerals...(they) feel happier out there... than they do sometimes in the waiting room*"). In their study of design team ideation sessions Cramer-Petersen et al. (2019) categorise such propositions of ideas suggested to address constraints as abductive reasoning, and the explanations justifying the proposed idea as deductive reasoning. They also show that in design sessions, abductive reasoning is typically followed by deductive reasoning, which aligns with what we see in the two examples discussed. A closer examination of similar speech turns from the two remaining sessions shown in Figure 5 also shows similar patterns despite the varied contexts.

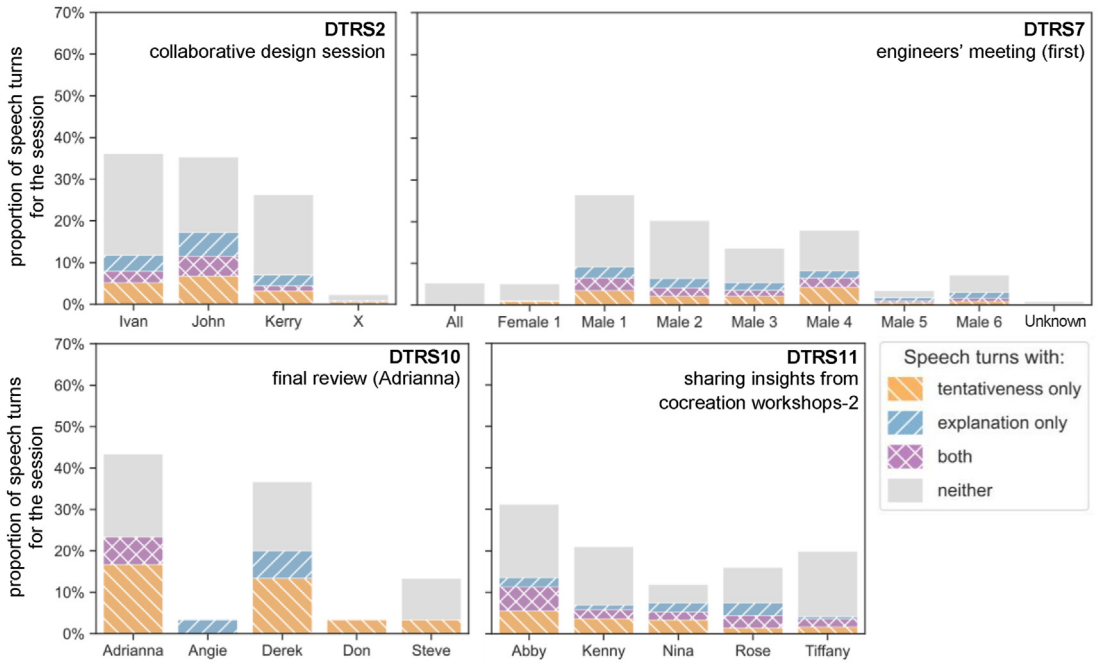

Figure 5 Speech turns separated by speaker and coloured according to whether a turn has at least one match with the 'tentativeness' category (see legend), at least one match with the 'explanation' category, or at least one match with each of both categories. Each session shown—one from each dataset—features the least number of turns matching both categories for that dataset. Compare this set of charts with those shown in Figure 4 above. The turns matching with only tentativeness or explanation show instances of 'building' up of creative or analytical ideas respectively across multiple speakers. See Section 4.4 for details and Table 3 for detailed turn distributions for these sessions

Though the speaker in Excerpt 4 is a client, not a designer, the inputs they are giving are contextual design inputs based on their experience. Oak (2009), in an analysis of an exchange immediately preceding Excerpt 4, explains how the client is performing the role of a 'client' by (a) responding to the architect's questions on specifics, such as room dimensions, with rich and contextual descriptions along with justifications of the context, and (b) implying that decisions on specifics such as room sizes should be made by the architect. While the client appears to emphasize their role of 'non-designer', from the nature of the exchange in Excerpt 4, it would seem that the client is indeed making designerly contributions to the process.

4.4 *Speech turns with matches to the* tentativeness *category only*

The sessions from Figure 5 are those with the least number of turns matching both *tentativeness* and *explanation* categories. Examining turns with words predominantly matching the *tentativeness* category, we see exchanges marked by shorter turns and interruptions or overlapping talk.

288 KERRY: It might be tough to get your junk, to, you'd have to make the—
289 IVAN: You'd have to pack it
290 KERRY: —backpack have two, two halves
291 IVAN: But that's— but people do that now, I mean you have the ones that have the big compartment on the bottom
292 KERRY: mm, mm, separate compartments
293 JOHN: Maybe, maybe a separate compartment's in there, just a zipper
294 IVAN: Oh, **so** like a middle (inaudible)
295 JOHN: Yeah, yeah, and you can — prrp — just pull it in half when you unzip
296 KERRY: mm, mm, that'd be good
297 IVAN: ok

— Excerpt 5: From the collaborative design session, DTRS2 Dataset

166 KENNY: Well, I think what I head from— I think one was that, what he wanted to have is a— someway where he could showcase that he was successful
167 ABBY: Yeah! Yeah! Exactly!
168 KENNY: His goals—
169 ABBY: —but not maybe in the way- not maybe to actually get a prize, it's like a physical item 'okay not just' but to show off that 'oh I totally'
170 KENNY: (overlapping) I think eh that actually value (overlapping speech from others saying 'yeah, yeah') this is a token of my, this and—
171 ABBY: Exactly! Exactly, and then that's easy to is as s— **kind** of a status symbol
172 KENNY: (overlapping) It's definitely a status symbol

— Excerpt 6: From a discussion with external consultants, DTRS11 Dataset

In both these sessions, we see the same kind of generative 'turns' but these span multiple turns, with more than one speaker adding on to the proposed ideas. For instance, John and Ivan add on to each other's idea in the engineers'

Table 3 Detailed breakdown of speaker turns and lexical category matches for sessions shown in Figure 5

Dataset	*Session*	*Speaker*	*Proportion of total turns with at least one match to:*				*Total % turns per speaker*
			Only Tentativeness	*Only Explanation*	*Both Categories*	*Neither Category*	
DTRS2	Collab. design session	Ivan	5%	4%	3%	24%	36%
		John	7%	6%	5%	18%	35%
		Kerry	3%	3%	1%	19%	26%
		X	1%	0%	0%	1%	2%
DTRS7	Engineers' Meeting (first)	Male 1	3%	3%	3%	17%	26%
		Male 2	2%	2%	2%	14%	20%
		Male 4	4%	2%	2%	10%	18%
		Male 3	2%	2%	1%	8%	14%
		All	—	—	—	5%	5%
		Female 1	1%	—	—	4%	5%
		Male 5	1%	1%	—	2%	3%
		Male 6	1%	1%	1%	4%	7%
		Unknown	—	—	—	1%	1%
DTRS10	Final Review (Adrianna)	Adrianna	17%	—	7%	20%	43%
		Derek	13%	7%	—	17%	37%
		Steve	3%	—	—	10%	13%
		Angie	—	3%	—	—	3%
		Don	3%	—	—	—	3%
DTRS11	Sharing insights from cocreation workshops 2	Abby	6%	2%	6%	18%	31%
		Kenny	4%	1%	2%	14%	21%
		Tiffany	2%	1%	2%	16%	20%
		Rose	1%	3%	3%	9%	16%
		Nina	3%	2%	2%	4%	12%

meeting from the DTRS dataset (John: "*Maybe, maybe a separate compartment is in there, just a zipper*", Ivan: "Oh, ***so*** like a middle —", John: "*Yeah, yeah, and you can — just pull it in half when you unzip*"). The 'so' uttered by Ivan in this exchange is not as much an explanation of John's idea, but an echoing of his thought, meant to convey an understanding. Similarly, the exchange between Kenny and Abby in DTRS11's insight-sharing session also shows a similar building up on each other's ideas. Here the generative utterances are not of new ideas, but new interpretations of what they observed from the co-creation workshops.

At the level of individual turns, these exchanges indicate only generative utterances or only deductive utterances, marked by terms matching only the *tentativeness* category or only the *explanation* category respectively. However, grouping successive turns together, the form becomes similar to the speech turns discussed in Section 4.3. This is indicative of a need for a flexibility in the granularity with which to categorise utterances, perhaps considering speech turns in some cases and word count in other cases.

4.5 Speech turns with matches to the explanation category only

In a similar vein, we see some turns in Figures 4 and 5 that show matches to only words in the *explanation* category. A closer look at the exchanges shows similar patterns to what was observed in Section 4.4, except that the conversation is predominantly analytical in these cases.

```
686 IVAN:    It, er, there's this big — on the child's seat there's this big piece
             of plastic that just sort of goes like that
687 JOHN:    So —
688 IVAN:    So then you just — you bring it up
689 KERRY:   It hooks
690 IVAN:    Yeah, and rotate it down and then there's this one clamp back there, J
             can I interject here
691 JOHN:    On our weight spec if you just wanna sorta look at these, they have
             weights — yeah they're between four hundred and thirty grams to, er,
             six hundred and thirty grams so
692 IVAN:    Four hundred and six hundred, say
693 JOHN:    Yeah so em I think that's — that's reasonable, five hundred grams is
             about one point one pounds ...
694 IVAN:    So we're talking er between about, say, one pound and er —
695 KERRY:   One and a half pounds
696 IVAN:    One and a half
697 JOHN:    Yeah that'd probably be the max that people would ...you know
             lighter's better so long as it's gonna work
```

— Excerpt 7: From the collaborative design session, DTRS2 Dataset

```
540 MALE 6:  — because if you want to teach a child something about pens, you'd
             give them a pen wouldn't you,
541 MALE 1:  Yeah
542 MALE 2:  No, well, it depends — it could be fun, it could teach you something —
543 MALE 1:  I suppose —
544 MALE 4:  — teach you something else. It's going to be quite a fat lump,
             probably, so it's not going to be that kind of pen that you use in
             future —
545 MALE 1:  mmm
546 MALE 4:  — so it doesn't feel much like a pen
```

— Excerpt 8: From the first engineers' meeting, DTRS7 Dataset

In both the above exchanges, there is the same sense of 'building upon each other's ideas' as we saw in Section 4.3, but these turns involve words matching the *explanation* category, which makes sense given the speakers are explaining the reasoning behind an idea or an interpretation. In the exchange from the DTRS2 collaborative design session, the exchange involves an analytical estimation of the kinds of loads that might be applied on a bicycle pannier, while in the exchange from the DTRS7 engineers' meeting, the discussion is about an analysis of whether giving a child a pen will impact their learning to use a pen in the future. In both cases, the dominant part of the conversation is indicative of analytical or deductive thinking, with a few words from the *tentativeness* category indicating expressions of approximation or uncertainty that would accompany analytical estimates.

While these observations are not conclusive, they are indicative of the value of lexical categories, generated through a computational approach and refined by a combination of manual judgement and additional computation, when applied in the analysis of design discussions. 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 overview charts shown in Figures 3 through 5 and the sample excerpts illustrate some of the advantages of a computational approach to analysing aspects of design thinking using lexical categories. For instance, Figure 3—showing at the dataset level a similar distribution of turns featuring at least one match to only the *tentativeness* category, turns featuring at least one match to only the *explanation* category, and turns featuring one match to each of *both* categories—suggests that certain aspects of design dialogue remain similar regardless of the context of designing. The excerpts showing similar lexical terms used in different domains and contexts of designing serve to strengthen this suggestion, while Figures 4 and 5 allow us to examine the nuances of these lexical category matches across speaker turns.

For the purpose of this work, we treated single and multiple matches between a given turn and words in a lexical category as the same, i.e., we counted only whether at least one of the words in a turn matched with one of the words in a lexical category. However, counting the number of words in a turn that match with words in a lexical category, or even the number of unique words that match, would provide us with additional nuances to explore. The turn matches to lexical categories are shown at the level of the dataset (Figure 3) and sessions (Figures 4 and 5), but they can also be shown aggregated at the speaker level. This may provide us with insights into whether we can infer an interlocutor's role across multiple sessions based on what aspects of

designerly thinking is evident in their speech. While we use *tentativeness* and *explanation* as categories and behaviours to examine, this approach could be extended to other aspects of designing such as reflection, collaboration, fixation, and so on, assuming that dictionary categories indicative of each aspect can be created.

At the same time, the process outlined in Section 3 identifies some limitations of this approach that need to be overcome. Due to the nature of the training data used in Empath, i.e., a dataset of modern fiction, we identify a few challenges in using this approach: (a) the words generated by Empath under each category is sensitive to the seed terms provided by the user, (b) Empath generates terms under both categories that find limited use in the context of designing, and (c) Empath also generates terms that appear in both categories.

The challenge posed by (a) and (b) may be related to the fact that Empath is trained on a corpus of fictional works, which creates word associations to lexical categories that may not be relevant in a design context. Issue (a) may be caused by a situation where seed word variations that may not be very different in a design context are very different in a fictional context, while issue (b) may reflect a similar contextual difference in the generated terms. In our approach, issue (b) is partly fixed by clustering the terms in a category based on semantic distance: we first looked up vectors corresponding to the terms from each lexical category in a multidimensional vector space, and clustered the words based on pairwise cosine distances between the vectors. This helps us identify clusters of words that are semantically more distant from the remaining words, and we then used our own judgement to remove words from these clusters to refine the dataset. Such an approach can be further refined by first qualitatively coding samples of the dataset and then computing matches between the same samples and the Empath lexical categories. This would potentially identify both false positives (speech turns not indicative of, say, tentativeness when Empath has tagged it as such) and false negatives (speech turns indicative of tentativeness that are not tagged by Empath). Donohue et al. (2014) adopted this approach in their analysis of political rhetoric preceding the Oslo peace accords, the results of which showed that while refined, dictionary-based approaches can be efficient, but human coding is more sensitive to context.

A more productive line of inquiry would be to investigate the creation of vector representations of words generated from different forms of design discourse such as designers' interviews, lectures/talks, as well as recordings of design discussions similar to the DTRS datasets. This can result in a more 'designerly word embedding' that capture semantic and contextual associations between words in the context of designing. Such approaches have been explored for specific domains (see Devlin et al., 2019). Domain- or context-relevant word embedding may also address issue (a), i.e., make the set of terms

generated by Empath less sensitive to the seed terms, especially if provided by experienced researchers.

The challenge posed by issue (c) is more complex: in our case, Empath generated terms that appeared in both categories of tentativeness and explanation. This is not necessarily a mistake. 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. For this paper, we combined overview visualizations with detailed examinations of the transcript to understand the patterns in the lexical matches. Our recommendation is to create overview visualizations, and then filter by criteria of interest and look at patterns at the level of sessions, turns, or by close reading.

Examining the data at an aggregate level using visualizations, and then filtering down into detailed views for a closer inspection and verification of data is well-established in the information visualization and visual analytics community (Shneiderman, 1996). 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; Menning et al., 2018) 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, 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. This could also combine the strength of dictionary-based approaches and human coding identified by Donohue et al. (2014).

Combining distant reading approaches enabled by computational analyses of designers' speech with traditional close reading approaches has implications for our understanding of where design thinking may or may not occur. Qualitative analyses of design thinking concepts have typically focused on talk in scenarios that explicitly involve designing, 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 methods 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, orient to, and respond to certain kinds of designerly talk.

6 Conclusion

In this paper, we theorised that speculative and rationalising aspects of designing can be explored through the linguistic concepts of tentativeness and explanation and we show how this can be achieved 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* and words associated with *explanation*. Looking at the matches between speech turns and these two lexical categories, we found that the balance between tentativeness and explanation in design conversation seems to tip slightly toward tentativeness. However, examining the distribution of the categories at the session level, and finally a close reading at the turn level, we identify patterns in a think-aloud design session and in dialogue between designers and clients or design students and their teachers that show a stronger association between tentativeness and explanation, often in the same speech turn. We also find patterns of dialogue between designers in collaborative sessions and co-creation sessions that are more speculative, while other analytical discussions in similar sessions indicate more rationalisation. 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. Finally, to overcome contextual biases, we propose updating general machine-learning models with appropriate contextual data, combining this with human validation to verify the patterns highlighted by computational models.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Notes

1. Reported number of categories is for LIWC 2015. Later versions may have more categories.
2. www.wattpad.com.
3. Note that the number, as well as types, of words generated by Empath for a lexical category can be sensitive to the seed words used.
4. Lemmatization in natural language processing is the process of treating all inflected forms of a word as the same. For instance, 'went', 'gone', 'going' etc. are inflected forms that can be lemmatized under 'go'.

References

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

Adams, R. S., Forin, T., Chua, M., & Radcliffe, D. (2015). Making visible the "how" and "what" of design teaching. *Analyzing Design Review Conversations* 431.

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.

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics, 5*, 135–146.

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) Indiana, USA: Purdue University.

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.

Chandrasegaran, S., Badam, S. K., Kisselburgh, L., Ramani, K., & Elmqvist, N. (2017b). Integrating visual analytics support for grounded theory practice in qualitative text analysis. *Computer Graphics Forum, 36*(3), 201–212.

Chandrasegaran, S., Bryan, C., Shidara, H., Chuang, T. Y., & Ma, K. L. (2019). TalkTraces: Real-time capture and visualization of verbal content in meetings. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems* (pp. 1–14).

Christensen, B. T., & Ball, L. J. (2016). Dimensions of creative evaluation: Distinct design and reasoning strategies for aesthetic, functional and originality judgments. *Design Studies, 45*, 116–136.

Christensen, B. T., & Ball, L. J. (2018). Fluctuating epistemic uncertainty in a design team as a metacognitive driver for creative cognitive processes. *CoDesign, 14*(2), 133–152.

Christensen, B. T., Ball, L. J., & Halskov, K. (Eds.). (2017). *Analysing design thinking: Studies of cross-cultural co-creation*. CRC Press.

Christensen, B. T., & Schunn, C. D. (2009). The role and impact of mental simulation in design. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 23*(3), 327–344.

Cramer-Petersen, C. L., Christensen, B. T., & Ahmed-Kristensen, S. (2019). Empirically analysing design reasoning patterns: Abductive-deductive reasoning patterns dominate design idea generation. *Design Studies, 60*, 39–70.

Cross, N. (2018). A brief history of the design thinking research symposium series. *Design Studies, 57*, 160–164.

Cross, N., Christiaans, H., & Dorst, K. (1996). *Analysing design activity*. Wiley.

Day, W. H., & Edelsbrunner, H. (1984). Efficient algorithms for agglomerative hierarchical clustering methods. *Journal of Classification, 1*(1), 7–24.

Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the ACL Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* (pp. 4171–4186).

Donohue, W. A., Liang, Y., & Druckman, D. (2014). Validating LIWC dictionaries: the Oslo I accords. *Journal of Language and Social Psychology, 33*(3), 282–301. https://doi.org/10.1177/0261927X13512485.

El-Assady, M., Gold, V., Acevedo, C., Collins, C., & Keim, D. (2016). Contovi: Multi-party conversation exploration using topic-space views. *Computer Graphics Forum, 35*, 431–440.

Ewald, B., Menning, A., Nicolai, C., & Weinberg, U. (2019). Emotions along the design thinking process. In *Design thinking research: Looking further: Design thinking beyond solution-fixation* (pp. 41–60).

Fast, E., Chen, B., & Bernstein, M. S. (2016). Empath: Understanding topic signals in large-scale text. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems* (pp. 4647–4657).

Fast, E., & Horvitz, E. (2017). Long-term trends in the public perception of artificial intelligence. In *Proceedings of the AAAI conference on artificial intelligence, Vol. 31(1)*.

Feldman, R. (2013). Techniques and applications for sentiment analysis. *Communications of the ACM, 56*(4), 82–89.

Glock, F. (2009). Aspects of language use in design conversation. *CoDesign, 5*(1), 5–19.

Jänicke, S., Blumenstein, J., Rücker, M., Zeckzer, D., & Scheuermann, G. (2018). TagPies: Comparative visualization of textual data. In *Visigrapp (3: Ivapp)* (pp. 40–51).

Jänicke, S., Franzini, G., Cheema, M. F., & Scheuermann, G. (2015). On close and distant reading in digital humanities: A survey and future challenges. In *Eurographics conference on visualization (EuroVis) - STARs* (pp. 83–103).

Jänicke, S., Franzini, G., Cheema, M. F., & Scheuermann, G. (2017). Visual text analysis in digital humanities. *Computer Graphics Forum, 36*, 226–250.

Koch, S., John, M., Wörner, M., Müller, A., & Ertl, T. (2014). VarifocalReader: In-depth visual analysis of large text documents. *IEEE Transactions on Visualization and Computer Graphics, 20*(12), 1723–1732.

Lloyd, P. (2019). You make it and you try it out: Seeds of design discipline futures. *Design Studies, 65*, 167–181.

Lloyd, P., Akdag Salah, A., & Chandrasegaran, S. (2021). How designers talk: Constructing and analysing a design thinking data corpus. In *Proceedings of the ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. Volume 6: 33rd International Conference on Design Theory and Methodology (DTM). Virtual, Online. August 17–19, 2021. V006T06A034*.

Lloyd, P., & Scott, P. (1994). Discovering the design problem. *Design Studies, 15*(2), 125–140.

McDonnell, J. (2009). Collaborative negotiation in design: A study of design conversations between architect and building users. *CoDesign, 5*(1), 35–50.

McDonnell, J. (2017). Design Roulette: A close examination of collaborative decision making in design from the perspective of framing. *Analysing Design Thinking: Studies of Cross-Cultural Co-Creation* 373.

McDonnell, J., & Lloyd, P. (Eds.). (2009). *About: Designing—analysing design meetings*. CRC Press: Taylor and Francis Group.

McHaney, R., Tako, A., & Robinson, S. (2018). Using LIWC to choose simulation approaches: A feasibility study. *Decision Support Systems, 111*, 1–12.
Menning, A., Grasnick, B. M., Ewald, B., Dobrigkeit, F., & Nicolai, C. (2018). Verbal focus shifts: Forms of low coherent statements in design conversations. *Design Studies, 57*, 135–155.
Mertens, A. T., & Toh, C. A. (2019). It rings a bell! Memory's impact on information utilization by novice designers in the early design process. In *Proceedings of the ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. Volume 7: 31st International Conference on Design Theory and Methodology (V007T06A03)*.
Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems* (pp. 3111–3119).
Moretti, F. (2005). *Graphs, maps, trees: Abstract models for a literary history*. Verso.
Oak, A. (2009). Performing architecture: Talking 'Architect' and 'Client' into being. *CoDesign, 5*(1), 51–63.
Oak, A., & Lloyd, P. (2016). 'Throw one out that's problematic': Performing authority and affiliation in design education. *CoDesign, 12*(1–2), 55–72.
Paletz, S. B., Chan, J., & Schunn, C. D. (2017). The dynamics of micro-conflicts and uncertainty in successful and unsuccessful design teams. *Design Studies, 50*, 39–69.
Paletz, S. B., Sumer, A., & Miron-Spektor, E. (2018). Psychological factors surrounding disagreement in multicultural design team meetings. *CoDesign, 14*(2), 98–114.
Pennebaker, J. W., Booth, R. J., Boyd, R. L., & Francis, M. E. (2015). *Linguistic inquiry and word count: LIWC 2015*. Pennebaker Conglomerates.
Piccardi, T., Redi, M., Colavizza, G., & West, R. (2020). Quantifying engagement with citations on Wikipedia. In *Proceedings of The Web Conference 2020* (pp. 2365–2376).
Ribeiro, M., Calais, P., Santos, Y., Almeida, V., & Meira, W., Jr. (2018). Characterizing and detecting hateful users on Twitter. In *Proceedings of the International AAAI Conference on Web and Social Media, Vol. 12(1)*.
Roozenburg, N. F. (1993). On the pattern of reasoning in innovative design. *Design Studies, 14*(1), 4–18.
Schön, D. A. (1992). Designing as reflective conversation with the materials of a design situation. *Knowledge-Based Systems, 5*(1), 3–14.
Shneiderman, B. (1996). The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of the IEEE Symposium on Visual Languages* (pp. 336–343).
Speer, R., & Chin, J. (2016). An ensemble method to produce high-quality word embeddings. *CoRR, abs/1604.01692*.
Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology, 29*(1), 24–54.
Toutanova, K., Klein, D., Manning, C. D., & Singer, Y. (2003). Feature-rich part-of-speech tagging with a cyclic dependency network. In *Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics* (pp. 252–259).
Umney, D., & Lloyd, P. (2018). Designing frames: The use of precedents in parliamentary debate. *Design Studies, 54*, 201–218.
Vayansky, I., & Kumar, S. A. (2020). A review of topic modeling methods. *Information Systems, 94*, 101582.