Proceedings of the ASME 2021 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference
IDETC/CIE 2021
August 17–20, 2021, Virtual, Online

DETC2021-71200

HOW DESIGNERS TALK: CONSTRUCTING AND ANALYSING A DESIGN THINKING DATA CORPUS

Peter Lloyd[1,*] **Almila Akdag Salah**[1,2] **Senthil Chandrasegaran**[1]
[1]Designing Intelligence Lab, Faculty of Industrial Design Engineering,
Delft University of Technology, Netherlands
[2]Faculty of Computer Science, Utrecht University, Netherlands
Email: {p.a.lloyd, a.a.akdagsalah, r.s.k.chandrasegaran}@tudelft.nl

ABSTRACT

A necessary condition of understanding how designers work is understanding how designers talk. In this paper we show how new methods of linguistic data analysis are beginning to reveal insights into the general nature of design conversations. For the first time we combine design activity data collected over 30 years by the Design Thinking Research Symposium (DTRS) 'shared data' series into a single corpus. We apply emerging techniques of analysis on this corpus and explore word forms, expressions, topics, and themes related to the particularities of how designers talk. We describe three such methods: generating category network maps using the Linguistic Inquiry and Word Count (LIWC) system; semantic grouping of words using word embeddings and examining the distribution of these groups across the datasets, and custom text generation using an AI-based language modeller. In applying these methods, we show that exploring design activity data at the corpus level can reveal more general patterns of design talk and raise key questions and hypotheses for further study. We see these methods as a first step in developing an understanding of how people not considered to be designers (e.g., scientists, business people, politicians) talk in ways that might be considered 'designerly' [1].

*Address all correspondence to this author.

1 INTRODUCTION

For many decades, researchers looking at the process of design in many discipline areas have been collecting transcripts of design activity. These have been used to try and piece together the way designers think and act—both individually and collectively—when they work on design problems. Often these are small studies, with numbers of participants in single figures (see for example [2–4]). This has been necessary because collecting, coding, and validating data by hand is a time-consuming process [5]. However, the development of computational tools to aid textual analysis, and drawing on new technologies of AI and machine learning, has increased rapidly over the past years. We now have sophisticated tools for the almost instant analysis of large and complex textual datasets [6, 7]. Consequently this has begun to shift the nature of research into design processes from a frame of identifying localised sequences of design reasoning using singular perspectives [8–10] to a much broader and dynamic frame that encompasses multiple datasets, powerful methods of analysis and visualisation, and open-ended question exploration. Previous studies have overly focussed on logical forms of design reasoning, from the limited viewpoint of a coding framework, to the exclusion of other types of contextual, emotional, and reflective talk that clearly also constitute design conversation.

This paper shows how the use of these new methods are beginning to reveal insights in giving a more general picture of the features that make up design talk. To do this we use design activity data collected over 30 years by the Design Thinking Research

Symposium (DTRS) shared data series. We apply emerging methods of analysis to these data to illustrate how word forms, expressions, topics, and themes related to the particularities of how designers talk can be explored. We conclude with key questions and hypotheses for future work, particularly in the area of how people not considered to be designers (e.g. scientists, engineers, business people, politicians) talk in ways that might be considered 'designerly' [1], and how we can model such attributes into machine-learning models for such applications as conversational agents and automated text analyses.

2 THE DTRS SERIES

2.1 Research Themes of the DTRS

The Design Thinking Research Symposium Series has held eleven workshops and conferences [11] since the original 'Research in Design Thinking' seminar was held at the TU Delft in 1991, featuring a cross-disciplinary collection of leading academics including Nigel Cross and Donald Schön [12]. The conferences and seminars have been influential but it is the four shared-data workshops that have provided the most important research contribution so far producing four books [13–16], seven journal special issues [16–22], and featuring nearly 50 published papers. All workshops have provided ground-breaking insights and new research methodologies with which to understand the uniqueness of design activity. DTRS2, for example, yielded the research method of linkography [23], now a widely used method in studies of creativity and designing [24].

Using a common dataset as a starting point, different research groups have shown and tested different perspectives, methods, and theories about designing during each workshop. Data-driven and theory-driven approaches have been used, along with quantitative and qualitative analyses looking at (to give a few examples): idea generation, use of analogies, imagination, social order, conversational turn-taking, gesture, embodied cognition, teamwork, paralinguistics, and many other topics.

Though there have been computational [6], non-verbal [25], and even choreographic [26] analyses of design activity, the most insightful findings have come from the close reading and interpretation of specific examples of talk from the transcripts and video material. Such analyses have highlighted such design concepts as framing [27], storytelling [28], vagueness [29], and spiderwebbing [30], though many conclusions remain as hypotheses waiting to be evidenced at scale in much larger datasets.

Computational data analysis tools and techniques have now created an opportunity to operationalise some of these hypotheses and build a more comprehensive evidence base to explore the linguistic features of design talk at this larger scale. The key is the creation of a data corpus of design conversation that is transferable between different software systems and amenable to computational analysis. In the following section we describe how we have done this with the DTRS shared workshop data.

2.2 The DTRS Dataset

In four separate shared-data workshops, the Design Thinking Research Symposium series [11] has generated data of design activity in the disciplines of industrial design engineering (DTRS2), architecture and engineering design (DTRS7), design education (DTRS10) and product design (DTRS11). The nature of these data has differed across workshops, from think aloud protocols [13], designer-client discussion [14], design education [15], and co-creation [16]. Although DTRS10 was set in the educational context of a university and included design students along with educators, all other workshops have featured professional experts. A short summary of each workshop is 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. [17]

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'. [31]

DTRS10 consists of a large number 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. [32]

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. [33]

We argue that these various forms of data and types of discussion (think aloud, team-based, designer-client, educational, etc.) provide a composite picture of design activity. No one dataset captures all elements of design activity, but together they present a more rounded—if not exhaustive—view of the types of activity that typically occur in a design process: thinking, discussing, reflecting and evaluating, and talking with clients. Table 1 shows the session numbers and lengths for each of the four DTRS datasets.

Though relatively small for a corpus, these data represent a unique and comprehensive record of professional design activity. They also form a valuable resource with which to generate and test hypotheses about how designers talk and the features of language and discourse that result from design dialogue.

TABLE 1. 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

2.3 Building a Design Thinking Data Corpus

To compile the four sets of workshop data into a single corpus we first obtained copies of all existing transcript files. In the case of DTRS2 this meant digitising hardcopy, but all other workshops had digital copies of their data. Appropriate permissions were obtained from the data owners where an existing data use agreement was not already in place. All data were anonymised.

All metadata—such as a description of the location, the speaker information, seating arrangement etc.—were removed from the transcripts as the focus of the analysis was on what was being said and the meaning surrounding it. For the same reason, descriptions of actions, such as *"[points to drawing]"* were retained to provide context. Also removed were transcription symbols, line numbers, and other formatting markers. This was done in order to maintain sentence continuity and render words recognizable by language-processing tools. Finally, each transcript was segmented using speech turns as the basic unit: each line of the final transcript consisted of the dataset name, session name, speaker name, and the utterance by that speaker for their entire turn. Stemming or lemmatisation, where inflected forms of a word are combined (e.g. design, designed, designs, designing are all considered to be a single lemma) was *not* performed so that idioms or phrases could be matched using dictionaries wherever applicable. The decision to perform an analysis at the level of turns, sessions, or datasets varied depending on the analysis performed and are detailed in the following sections.

As a proof-of-concept the four datasets were individually imported into the corpus analysis software Sketch Engine[1] as sub-corpora, and then combined to form one corpus. Sketch Engine parses the data into its linguistic and collocational components, (including lemmatisation) forming a database that can be used to build a sense of how the use and meaning of words or phrases are constructed. This helped us carry out an initial exploration of the data to understand what kind of results might be returned. Sketch Engine combines statistics with collocation to help explore word usage in what is termed a 'word sketch'. For example, exploring the use and meanings of the words *'problem'* (174 occurrences) and *'solution'* (58 occurrences) in the four DTRS datasets. Importantly Sketch Engine also means that the DTRS corpus can be compared with other corpora to further explore what might be distinctive features of design activity data. For example, by comparing the DTRS datasets with the British National Corpus (BNC) of Spoken Words [34] we find that the occurrences of the word *'problem'* are 295 per million words spoken (BNC) against 366 per million words spoken (DTRS). Designing as an activity, we might initially conclude, involves more talk about problems. Using Sketch Engine allowed us to explore some initial ideas about design talk including tentativeness, storytelling, past experience, and future imagination, some of which we pick up on in the following sections. Overall, using Sketch Engine showed us how powerful and productive it is to conduct an analysis at the level of the corpus. In the following three sections we explore three further methods with which to do this.

3 LINGUISTIC INQUIRY AND WORD COUNT (LIWC)

What are the global features of the DTRS datasets when compared to other linguistic datasets as well as with each other? We used Linguistic Inquiry and Word Count (LIWC) software to map out the word usage in the DTRS datasets. LIWC is a general and increasingly popular tool with which to analyse linguistic datasets from primarily a psychological viewpoint. Built up over several decades, it provides an extensive dictionary for coding words into a number of main and sub-categories and counting how often they appear in a text. Currently there are around 6,400 dictionary words and word stems in 91 categories and sub-categories, covering dimensions of linguistics (e.g. pronouns, prepositions), other grammar (e.g. quantifiers, common adjectives), and psychological processes (e.g. positive emotion, insight). Dictionary words can be counted in multiple categories. Table 2 shows a selection of categories and word examples; please see Pennebaker et al. [35] for the full list.

The advantage of using LIWC as a coding tool, rather than, say, coding frameworks originating in the design studies literature, is that a much wider range of linguistic phenomena can be explored. While design process coding frameworks largely focus on specific reasoning patterns, stages in the design process, or the design object itself, LIWC can reveal the emotional colouring of word usage alongside (for example) cognitive, perceptual and social processes. Furthermore, the fact that LIWC has been used so extensively in many other disciplines means that external comparisons can be easily made, something that has not been possible previously, with studies of design activity.

The power of LIWC lies in its dictionary, which assigns words to at least one (and usually more than one) category. The relationship between these assignments can be used to generate a network map, where each category is a node, with the relations between categories becoming links. Once such a network structure is built, the LIWC category map can be used as a visualization tool to show how word categories and relationships play out in a given corpus. Using LIWC to describe a dataset in this manner is a new approach; hence we generated a sample of network maps from a variety of linguistically different corpora.

[1] https://app.sketchengine.eu/

TABLE 2. LIWC Dictionary Categories relevant to design discourse.

Main category	Sub-category	Word examples
Linguistic Dimensions (function words)		
Pronouns	1st per. singular	I, me, mine
	3rd per. plural	they, their, they'd
Impersonal pronouns		it, it's, those
Articles		a, an, the
Prepositions		to, with, above
Negations		no, not, never
Other Grammar		
Common verbs		eat, come, carry
Common adjectives		free, happy, long
Interrogatives		how, when, went
Quantifiers		few, many, much
Psychological Processes		
Affective proc.	positive emotion	love, nice, sweet
	negative emotion	hurt, ugly, sad
Social proc.		mate, talk
	friends	buddy, neighbour
Cognitive proc.	Insight	know, think, realise
	Tentative	maybe, perhaps, could
	Differentiation	hasn't, but, else
Perceptual proc.	See	view, saw, seen
	Hear	listen, hearing
	Feel	feels, touch
Drives	Achievement	win, success, better
	Power	superior, bully
	Risk	danger, doubt, chance
Time orient.	Past focus	ago, did, talked
	Present focus	today, is, now
	Future focus	may, will, soon
Relativity	Motion	arrive, bike, go
	Space	down, in, thin
	Time	end, until, season

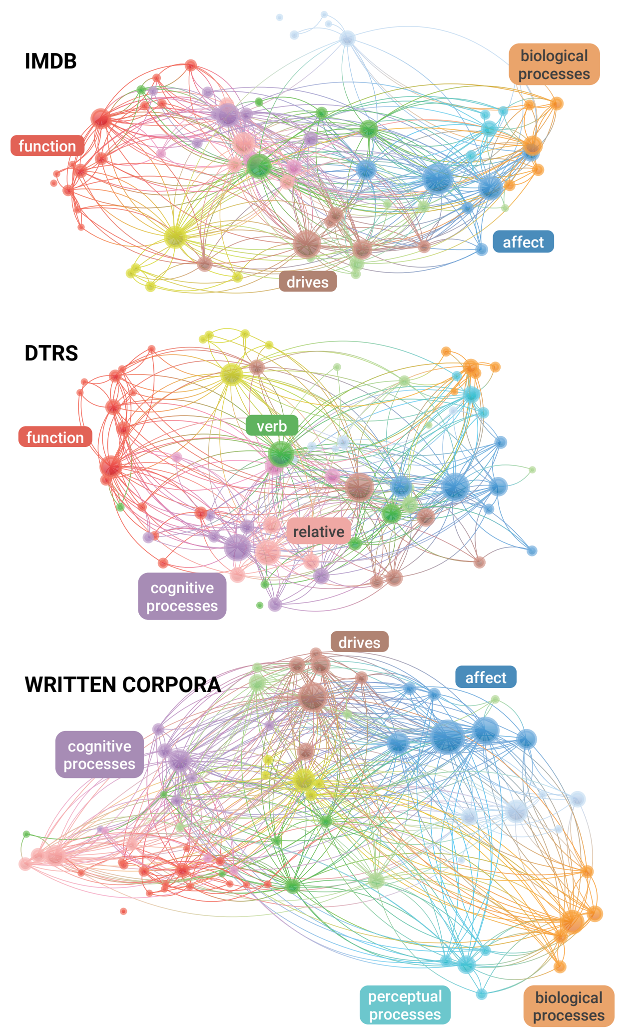

FIGURE 1. An overview of LIWC Category Network Maps for IMDB data (top), DTRS data (middle) and Written Corpora (bottom). Each network has 73 LIWC subcategories as nodes, and coloured according to the top 12 categories to which they belong. The dominant categories in each corpus are labelled. Relative node sizes represent the number of unique words from the dataset that fall into the corresponding LIWC category. Similarly, link thickness between two nodes represents the number of unique words from the dataset that appear in the two corresponding LIWC categories.

Our first comparison corpus was of conversation: a dataset of 1068 movie scripts collected from the Internet Movie Database (IMDB) [36]. The second comparison corpus was of written text combining a blog post dataset collected from 19,320 bloggers over a period of 6 months and covering about 681,288 blogposts [37] together with scientific publications (around 500 randomly selected papers from the computer science domain, and 100 randomly selected books from the humanities domain).

In Figure 1, we see the LIWC network maps for the three corpora together (IMDB, DTRS, and Written Corpora) each of which has 73 nodes, i.e. LIWC subcategories that are colored according to the top 12 categories they belong to in the LIWC dictionary. The datasets are quite different in size, and a simple statistical count of word occurrences for network conversion would have resulted in node/link sizes that would be difficult to compare. To normalize, we focused on the relational structure of categories within these datasets, where each word belonging to more than one category is counted only once (regardless of how frequently it appears in the corpus) as a link between these categories. Hence, the links between the categories do get stronger as long as a rich categorical representation is present in the corpus. The resulting networks show the distribution of LIWC category words and how

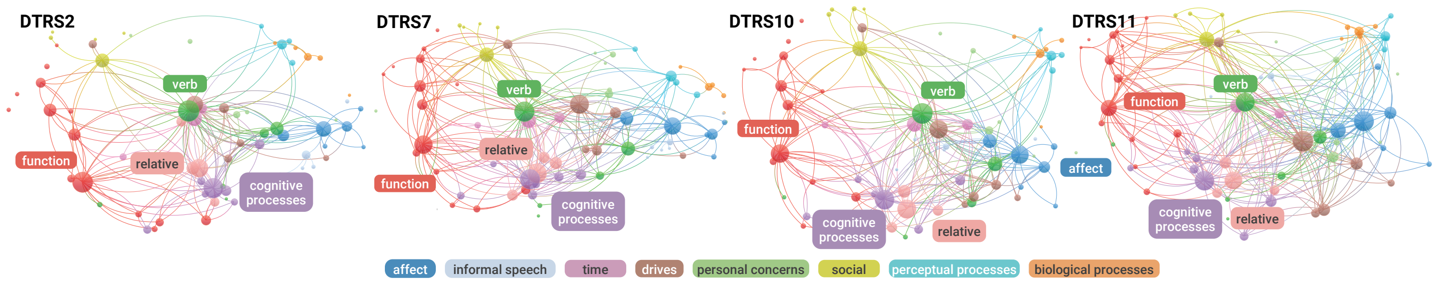

FIGURE 2. An overview of LIWC Category Network Maps of DTRS datasets. The network is visualized with VosViewer.

they are connected within each corpus. Since we retained phrases, idioms, and word forms such as conjugation of verbs, the resulting networks still reflect the links between the grammar and semantic categories of LIWC accurately.

A comparison between IMDB, DTRS and Written Corpora reveals striking differences between the written and spoken word. The Written Corpus has a network structure where the top LIWC categories turn into clusters of their own sub-categories, and each such cluster is connected to all the other clusters, but they are also closely placed to semantically closer categories. For example, the categories *biological processes* (orange) and *perceptual processes* (light blue) are next to each other at the right side of the network. Similarly *function* (red) and *relative* (dark pink) are neighbours at the left bottom side of the network. *Relative* category contains words from *time, space and motion* subcategories; words from these categories frequently also take on functional roles.

This clean network structure with top categories mapping into individual clusters does not hold for the conversational datasets. Both for the IMDB and DTRS networks, we observe a mingling between subcategories, especially between *cognitive processes* (purple), *verb* (dark green), and *relative* categories. Conversations, probably due to their informal nature where strict sentence structure or grammar rules are not followed, show a network structure where subcategories are much more interlinked with each other. Furthermore, if we look at the node sizes in written versus conversational datasets, we see that for the Written Corpora the *affect* (dark blue), *drives* (brown) and *cognitive processes* are the most important nodes in the network. In contrast, the conversational datasets feature a prominence of *function, relative, cognitive processes and verb* categories. *Social and affect* categories are also visible but hold a secondary place in the network.

A closer look between the conversational corpora (IMDB and DTRS) networks reveals the following differences: *affect and biological processes* categories are quite small for DTRS, whereas the IMDB and Written corpora have similar sized nodes for these categories. This might be expected as movie scripts (IMDB), reflecting many aspects of real life, generally have emotionally loaded words along with words relating to biological experiences. The same goes for blog posts where people recount everyday experiences. A second visible difference between the IMDB and DTRS networks is that the categories *cognitive processes, relative, function and verb* for DTRS are all bigger then the same categories in the IMDB network. The node size of the category *function* is three times bigger for IMDB than the Written Corpus; for DTRS, the node size difference is four times.

When we compare individual DTRS datasets to each other (Figure 2), we see that their network structure is quite similar. In all of them, the category *verb* is in the middle of the network, usually directly connected to *motion* and *focuspresent*, probably a direct outcome of the conversational nature of our corpus. We also observe that each network has the categories *function* and *affect* in separate clusters, indicating a similarity to the spoken word corpora. The remaining categories are generally interconnected. What is striking is that the category of *relative*, with subcategories of *time, space and motion* is always interlinked with the category of *cognitive process*. This is also similar to the spoken word corpora, and suggest ideas that combine conceptual terms with concrete ones, such as a combination of hypothesizing and conjecturing with experiences and the environment.

LIWC category networks enable us to compare large datasets by focusing on functional and semantic categories and their links to each other through language usage patterns. However, when the datasets get smaller, and their distinguishing characteristics show a lower overlap, the network maps can fail to reveal subtle differences in linguistic patterns. For the comparison of each of the 64 sessions of the four DTRS datasets, examined categories that were differently instantiated and networked between written and spoken corpora. These are the categories of *cognitive process*, *function*, *verbs*, and *relative* (all prominent categories in the DTRS dataset) in relation to categories that are dominant in the Written and IMDB corpora, i.e. the categories *drives*, *affect* and *social*. In doing this, we provide an overview of the differences and similarities in DTRS meetings for these categories.

Figure 3 shows a scatterplot of the above categories across all 64 sessions with individual DTRS datasets color-coded according to each dataset. Note that the size of the datasets are uneven, with two sessions for DTRS2 (blue), four sessions for DTRS7 (orange), 38 sessions for DTRS10 (red) and 20 sessions for DTRS11 (light

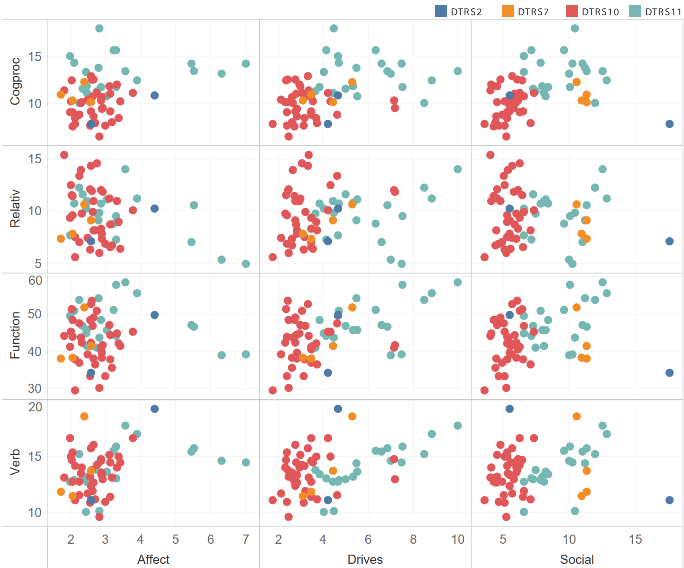

FIGURE 3. An overview of LIWC categories in all DTRS datasets, covering 64 individual sessions.

blue). Thus DTRS10 and DTRS11 dominate the plots. Each plot offers an aggregated view of each session's chosen categories. For example the plot at the left bottom corner (affect/verbs) shows the total number of words from these categories in a given session. The relational nature of the graph does not necessarily hold at the level of utterances, sentences or paragraphs. For example, the fact that a session has a high percentage of words from the *drive* and *cognitive process* category does not necessarily mean that all participants make use of words from these categories every time they speak. This chart needs to be read as a way to see if there are sessions that are outliers in more than one category.

The interesting patterns in Figure 3 are the differences between DTRS10 and DTRS11 across the categories of *social*, *affect* and *drives*. For example, four DTRS11 sessions have a high usage of *affect* words. Almost all DTRS11 sessions have a higher usage of socially related words as does DTRS7. As both datasets are obtained from naturally-occurring commercial design processes with multiple stakeholders this might be expected. Most of the DTRS11 sessions have a higher distribution of words assigned to the *drives* category, which might also be explained by the commercial context, and consequent financial investment, of the two datasets. We can also see that for DTRS2 there is a significant difference in the *social* category revealing the nature of the two sessions. One was a think-aloud text of just one designer, while the other was an ideation session between 3 designers familiar with each other, thus exhibiting a higher usage of *social* words. Similarly for DTRS2, we might expect a think-aloud protocol to be higher in the *cognitive process* category than a team ideation, and that is confirmed in the top middle plot of Fig. 3.

When we focus on the categories that are most interesting for our study, i.e. *verbs, function words*, time and space indicating words from the *relative* category, or the words related to thinking, discussing and imagining in the *cognitive process* category, we see a scattered distribution of these categories among all sessions and across all DTRS datasets. There are of course some outliers, but a pattern indicating a similar type of word usage from the two other corpora (IMDB and Written) is not evident.

To summarise, using the LIWC dictionary and creating network maps of different corpora and datasets has allowed us to both compare and contrast design thinking word usage with other types of spoken and written text, as well comparing different types of design thinking. What LIWC does is to reveal a much broader range of activities and processes present within the DTRS data than previous studies have been able to show. Word usage relating to the specific categories of *cognitive process* and *time orientation* are particularly evident and we go on to explore a subcategory from the former in more detail in the next section.

4 TENTATIVENESS: CHARACTERISING DESIGNERLY TALK

One aspect of designing that has been noted both in our initial proof-of-concept using Sketch Engine, but also in the literature, is the idea of tentative suggestion as a way of moving forward in the design process [29]. This forms the basis of what some have termed a 'basic design cycle' [38] which includes, for example, Schön's model of 'naming, framing, moving, and evaluating' [39, 40] and ideas about 'primary generation' [41]. Synthesis—the proposing of new solutions to an observed problem or need—is seen as one of the most essential aspects of design [42]. This synthesis is described as "an abductive sense-making process" [43, p. 17], characterized by hypothesising—making educated guesses based on the problem constraints and the designer's own experiences.

Schön [44] refers to the idea of 'moving experiments' where hypotheses are formed and tested as the design processes progresses: *"Faced with a particular site and a design task, the designer selects one or more prototypes from his/her repertoire, seeing the site in terms of the prototype carried over to it, and seeing the prototype in the light of the constraints and possibilities discovered in the site"* (p. 11). A hypothesis not being confirmed triggers a process of reflection and reframing.

Glock [29], in his close examination of the DTRS7 dataset, identifies markers such as modal adverbs (could, might, probably, etc.), downtoners or expressions of vagueness (kind of, sort of, a bit), and hedges (I think, I would, etc.). We find an overlap between such expressions and the *tentative* category in LIWC, under the larger umbrella of the *cognitive processes* category.

We posit that the proportion of speech turns that include at least one word from the LIWC category of *tentative* is indicative of hypothetical or tentative thinking. Figure 4 shows some unifor-

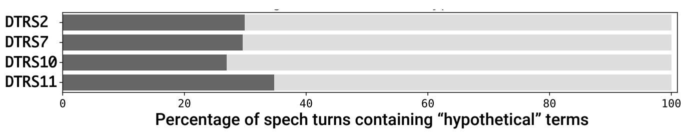

FIGURE 4. Normalised proportion of speech turns containing words associated with hypothetical thinking (dark grey) shows similar proportions across all DTRS datasets.

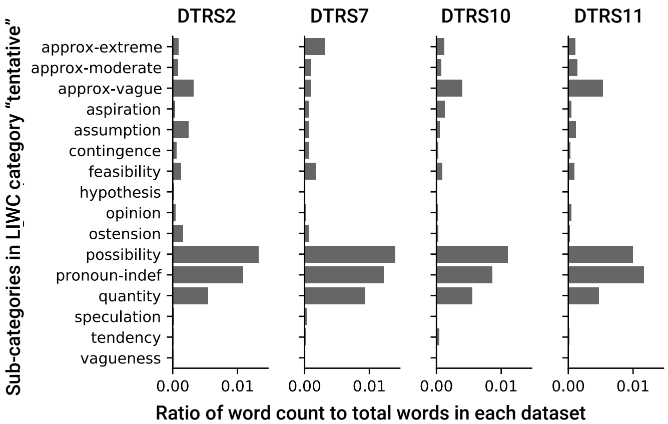

FIGURE 5. Ratio of words from the LIWC category *tentative*—divided into subcategories as explained in Table 3—occurring in each dataset, to the total words in the corresponding dataset. The figure shows a similarity in the distribution of words from each sub-category across all 4 datasets, with the sub-categories of *“possibility”*, *“indefinite pronoun”*, and *“quantity”* being the most dominant.

mity in expressions of tentative thinking across all four datasets, occurring in approximately 30% of speech turns.

Approximately 150 unique words from all four datasets were found to occur in the LIWC *tentative* category. These words included synonyms, inflected forms, and semantically-related terms. In order to make better sense of the words identified, we semantically grouped them through the use of word embeddings. Word embeddings have gained popularity in recent years as highly effective, lower-dimensional representations of word *meanings* [45]. They have been shown to capture semantic relationships ranging from the concrete (e.g. country–capital) to the more abstract (e.g. analogies). We used a pretrained word embedding [46] to measure semantic similarity between these 150 words and group them using hierarchical clustering [47] (see Table 3). Note that the names of the groups—column 1 in the table—are provided by us, and represent our attempt to describe each group of words in the context of this work. We then plot the distribution of each subcategory—or more precisely, the words in each subcategory—across each dataset, shown in Fig. 5. The figure shows that most of the discussion fall under three main sub-categories: *“possibility”*, *“indefinite pronouns”*, and *“quantity”*.

TABLE 3. Words from the DTRS corpora occurring in the LIWC “Tentative” category, semantically grouped.

Sub-Category	Words from LIWC “Tentative” Category
Approximation (extreme)	barely, hardly, almost, practically, virtually, fairly, pretty, quite
Approximation (moderate)	approximately, partly, most, often, occasionally, sometimes, mainly, mostly, generally, typically, usually
Approximation (vague)	vaguely, kinda, kind of, sorta, somewhat
Aspiration	hope, hoped, hopefully, hoping, may, someday, sometime, potentially, theoretically, likely, unlikely
Assumption	assume, assumed, assumes, assuming, guess, guessing, suppose, supposed, supposing
Contingence	contingent, depend, depending, depends, option, optional, optionally, options, random, randomly, variable, variables, varies, vary
Feasibility	chance, feasible, possibilities, possibility, possible, potential, try, trying
Hypothesis	assumption, assumptions, hypothesize, hypothetical, theory
Opinion	bet, betting, doubting, hesitant, luck, lucky, opinion, question, questioning, questions, undecided
Ostension	anyhow, apparently, appears, dunno, seem, seemed, seems, somehow, supposedly
Possibility	if, maybe, might, perhaps, possibly, probably
Pronoun (indefinite)	any, anybody, anyone, anything, anytime, anywhere, or, somebody, somebody’s, someone, someone’s, something, something’s, somewhere
Quantity	alot, lot, lots, some, sort, sorts
Speculation	doubt, wonder, wondered, wondering
Tendency	appear, appearing, border, borderline, indirect, indirectly, lotta, overall, temporally, temporary, temporarily
Vagueness	blurred, confuse, confused, confusing, confusion, confusions, mystery, puzzle, puzzling, unclear, unknown, vague

Note: The ‘sub-category’ names in the left column reflect our own characterisation of each group on the right, and are *not* LIWC categories.

The subcategory of “possibility” includes such terms 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 examining a condition. For instance:

> *“...is a mountain bike a problem when, **if** you have panniers hanging down the side, can people complain about that?”* (DTRS2, think-aloud session).

*"...did you see this as a space that **might** have its own small lectern in it or some altar-like feature inside it?"* (DTRS7, architect-client meeting).

The category of "indefinite pronoun" includes the terms "anybody", "anywhere", "somebody", "something" etc. (see Table 3). In the discussions, it is used as a way to suggest extending a theme, or to indicate ambiguity in the identity of people or objects in the speaker's speculation.

*"What if you—and you have some opportunities with colors and forms and we could do some kind of funky loopy handles or **something**, **someone** could pull things out."* (DTRS10, student design review).

In the example above, one of the design coaches is providing feedback and suggestions on a student's design concepts for a futuristic furniture-cum-storage system. The coach uses "funky loopy handles" as an instance of "forms", with the indefinite pronoun, i.e., "something", as a suggestion for the student to expand on the theme suggested. At the same time, "someone" is used as a placeholder for the potential end-user of the product being designed.

Finally, the category of "quantity" is characterised by terms such as *"alot"* (sic), *"lot"*, *"some"*, and so on. Closer examination of the occurrences of words from this category showed that most of the occurrences were of the word "some", used in most cases to convey non-specificity.

*"...we want them to confirm that this is still valid, that we did not, eh, push too hard in a direction or we are **some** kind of eh- or we are biasing them in **some** way that we didn't really realize."* (DTRS11, co-creation session).

Other terms, when found in context, also conveyed similar vagueness, shown below. The occurrences above and below seem to indicate that more than an expression of quantity, this category is similar to the "indefinite pronoun" category discussed earlier.

*"...it has to follow this articulating—like the suspension on a car as well. The car **sort** of has this inertia. It flies along in a straight line with suspension and all **sorts** of things as the wheels pump up and down"* (DTRS7, engineers' meeting).

We have furthermore observed some variations between individual sessions, which may indicate the limitations of using LIWC for discussion transcripts such as ours. For instance, Fig. 6 shows the distribution of the same categories as in Fig. 5, but in three sessions from DTRS10, involving the same student designer. The sub-category *"approximation (vague)"* immediately stands out as different from the dataset-level distribution.

Closer examination reveals the speaking style of the dominant speaker to be the cause of this variation. In the first session in

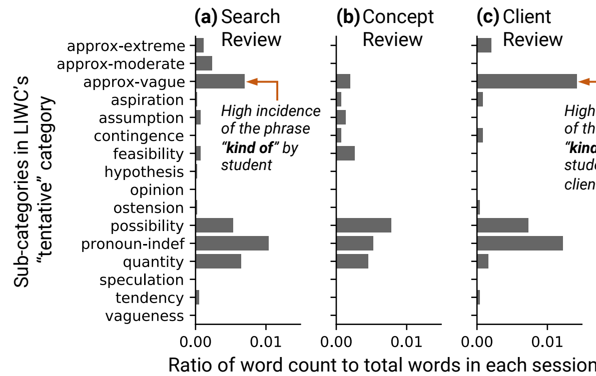

FIGURE 6. The similarity in distribution among the sub-categories identified within LIWC's "tentative" category is consistent across most sessions and match Fig. 5, with a few exceptions. The figure above shows three sessions involving the same student, at different stages of design, with a marked difference in the *"approximation (vague)"* sub-category. Examination of the transcript shows that the difference is more due to the dominant speaker's speaking style.

Fig. 6, the student was one of the two dominant speakers (94 of a total 279 turns), and had a tendency to use "kind of".

*"the first category is like just ram it all it, just do anything together, underwear and exercise clothes. they don't **kind of** care. Whereas this one they, they do it thoroughly and it's again that the lady—that's all of it, and on the video, um, when she's **kind of** filming going around and the guys, 'I don't do any laundry'. I was like so it's **kind of**..."*

A similar speech pattern is noted in the third of the displayed sessions for the same student—the high use of the term "kind of". In this session, the same student was the most dominant speaker (35 of a total of 71). In addition, another speaker—a client—also appears to use "kind of" frequently in this session, which results in the spike in the LIWC subcategory of "approximation (vague)" in Fig. 6(a) and (c).

This indicates that the context of word use in specific instances is an important consideration that should not be ignored when using word embeddings such as the one we used in this work [46] to identify LIWC sub-categories. Increasingly subtle embeddings such as BERT [48] are being developed to capture not only words but also their contexts of use. Applied to such corpora, these new approaches can help identify themes as well as connections made between concepts. In the future, we will explore the use of such models to identify and characterise terms appearing in designerly talk.

The detailed examination of subsets into the LIWC 'tentative' category shows a certain skew in the distribution of tentative terms in design discussions, mainly falling under expressions of exploring possibilities (if, might, possibly etc.) and expressions of non-specificity (something, someone, some etc.). Are these expressions of tentativeness sufficient in characterising a discussion as "designerly"? In the following section, we examine this question further by generating an artificial "transcript" using a language model to compare an instance of generated text with an actual excerpt from the DTRS dataset.

5 LANGUAGE MODELS OF DESIGN TALK

A third method that we have developed as a way of exploring the general nature of the DTRS corpus is to use machine learning to generate entirely new text in the style of the DTRS data. This allows us to compare very plausible generated talk with actual talk, again with the purpose of revealing the general characteristics of design talk. We used the online software Inferkit[2], an off-the-shelf text generation interface powered by NVIDIA's transformer model, Megatron-LM [49], trained on the data of 8 million webpages. Language models use large amounts of human-generated language as input to calculate probability distributions over sequences of words in a given language. This allows them to—given a sequence of words as an input or "seed text"—generate a likely sequence of words to follow the prompt based on these probability distributions. Crucially, Inferkit also allows "re-training" of the model with additional data, which generates text in the form and content of this additional data. We produced five custom text generators by retraining Inferkit's language model with the DTRS transcripts. One generator was produced for each DTRS dataset and a fifth generator was produced using the entire DTRS corpus as training data to explore what a more aggregated view of generated design talk would look like. Each custom generator took around 30 minutes to train and a further 5 minutes to deploy.

For the purposes of the present analysis the technical details of the language model are less important than how they function in practice. A text generator takes a seed text and generates what it thinks comes next up to a specified number of characters (in the case of this analysis, 300 characters). 'What it thinks comes next' hides, of course, the level of complexity in computing 'what comes next'. For humans, what comes next flows naturally in the context of what is being talked about and our ability (innate or otherwise) to make ourselves understood. For a machine, 'what comes next' can be any word or token based on the seed text and the training data. Using the DTRS dataset as training data means that 'what comes next' relates to the themes, content, and style of the DTRS dataset. For example, inputting the seed text 'I went and sat down' results in the following generated text from the language model trained on the DTRS7 dataset:

> *"**I went and sat down** and banged out the plans and we managed to get the studio space into the existing spaces which are two hospital wing facilities and we've had our meetings with them to look at how to move it forward so there's more space for meetings in it as well."*

This is a convincing text that is wholly artificial. No one *"banged out plans"* or *"managed to get studio space"* and there are no *"hospital wing facilities"*. The use of the word *"space"*, however, indicates there is an architectural dimension to the generated text relating to the training data. The language model seems to pick up on 'space' as a frequently mentioned word in the training data, thus increasing the probability of its use in generated text.

What this kind of analysis allows us to do is, on the one hand, explore design talk through what is not design talk (i.e. what is missing in the generated talk that would tell us something about our expectations of design talk). On the other hand, it allows us to take seed text that actually occurs in the DTRS dataset and compare what was said to what is generated, telling us something more contextual about the nature of design talk.

We carried out a number of explorations based on different seed texts and different DTRS language models, but we restrict ourselves to a discussion of one particular example to illustrate the method of inquiry. Table 4 shows two excerpts that derive from the same seed text *"It might be..."* based on the idea of exploring tentativeness in design conversation that was explored earlier in Section 4. One excerpt is a generated text from the language model trained on the DTRS2 data, the other is actual text from the DTRS2 data that was spoken. It is quite difficult to tell which is which, but attempting to do so can be revealing. Each excerpt contains a number of turns at talk with speakers and punctuation removed. No line breaks or text have been added so the generated text is presented as it appeared.

Spoken talk is obviously not grammatical text and that is the case for both excerpts illustrating the surprising sophistication of the language model. As you read the text the tendency is to try and impose meaning on what is said, to understand what is being talked about and what is being referred to. Excerpt 2 is an actual text, while excerpt 1 is the first example of generated text. Of course it is possible to generate an unlimited number of texts following the *"it might be"* seed but showing the first demonstrates that the generated text was not deliberately chosen. From the original data there were 12 instances of talk stemming from the words 'it might be'.

What is it about excerpt 1, if anything, that does not ring 'true'? Excerpt 2 represents a standard design exchange. The *"might be"* in line 1 relates to a possible problem about *"packing junk"*, a quick solution is suggested of *"two halves"* which leads to the identification of existing products that already have that functionality (*"people do that now"*, line 4). To this, a new zipper solution is proposed (*"you can... just pull it in half when you unzip"*, line 8) which is evaluated as *"good"* in line 9. Throughout the dialogue, each turn relates reasonably clearly to the previous turn, with a solution for the problem that is received positively. It is also clear that the *"it"* of line 2 refers to a proposed backpack

[2] https://inferkit.com/

TABLE 4. Two excerpts related to the DTRS2 data, one generated from the language model trained on the DTRS2 data, and the other actual text from DTRS2. Which is which?

Excerpt 1	
1	***It might be*** like the ring size maybe like
2	oh right I think that's em
3	we'll just err and er I'm gonna just just take a second and think about what are we using here
4	mm mm
5	it would be nice if it was on it
6	oh we can put it right there
7	it's way too wide so we're gonna stick it on the back here OK
Excerpt 2	
1	***It might be*** tough to get your junk to you'd have to make the
2	you'd have to pack it
3	backpack have two two halves
4	but that's but people do that now I mean you have the ones that have the big compartment on the bottom
5	mm mm separate compartments
6	maybe maybe a separate compartment's in there just a zipper
7	oh so like a middle (inaudible)
8	yeah yeah and you can prrp just pull it in half when you unzip
9	mm mm that'd be good

design that remains consistent in the following lines: *"it"* has two halves, separate compartments, a zipper, etc. The components of the backpack design object relate to the concept of the backpack.

In contrast, excerpt 1 is more difficult to understand, though a plausible attempt at understanding can be made. Line 3, for example, can be read as an attempt to take a step back and reframe what is being proposed, which appears to be an explanation of a problem (the *"ring size"* of line 1). The *"em"*s and *"err"*s in lines 2 and 3 indicate that other things could be happening besides talk—perhaps the interlocutors are sketching or using CAD? The combination of prepositions with the word 'it' is suggestive of the LIWC category of cognitive processing (*"on it"*, *"put it"*, *"stick it on"*) and give a sense of something being tried out and evaluated ('way too wide', line 7). It is difficult to determine what the 'it' refers to although we assume it is something that was mentioned earlier—an anaphoric reference. The difficulty, of course, also suggests that this might be the generated text. It is noteworthy that both excerpts use the word 'maybe' (excerpt 1, line 1; excerpt 2, line 6) part of the LIWC category of 'tentative'. This shows that the generated text has 'learned' that tentativeness is a feature of design talk. The definite final statement (*"so we're going to stick it on the back here"*, line 7) does not quite fit the tentative nature of the rest of the talk although it seems to suggest a solution. But it is the overall sense of not really knowing what is being talked about that comes through. The only concrete thing that is mentioned comes in line 1 (*"ring size"*) so we assume what is said is related to this, but there is no conceptual linkage as in excerpt 2. The common object of design does not appear in the talk, it has to be inferred and imagined.

The generated text clearly mimics the form and style of the actual text, a remarkable result in itself. The tentative suggestion and exploring of actions are similar, as is the closure of the excerpts with an evaluative statement. The text also captures a sense of an object being talked about as it is being designed, but lacks a conceptual structure when compared to excerpt 2.

To produce an artificial text that can reasonably convincingly be read as a design conversation not only illustrates the sophistication of language model, but reveals something about the general nature of that conversation. What we find in these two excerpts is a conceptual 'thing'—a 'virtual' design object—being developed through talk and action. The idea of a 'virtual design', a common imaginative object that everyone in a design project can discuss, is an idea that Medway [50] used to articulate the architectural design process, but using a language model takes this a step further by connecting this idea to specific patterns of talk. Comparing generated text to actual text in just this small example has shown how prepositions and anaphoric references indicate the concreteness of a virtual design under discussion, and thus leads to a hypothesis that can tell us something more general about the nature of 'virtual design objects' in design talk. Using the entire corpus of DTRS data to train the language models shows how a very different analytical method can be developed that is more akin to conversation analysis, but based on an entire dataset, not just through the use of selective examples.

6 DISCUSSION AND CONCLUSION

We have combined four large sets of data from the DTRS shared-data series to produce a singular corpus with which to explore the nature of design talk. With each dataset relating to a different form of design activity (thinking, conversing, dealing with clients, and reviewing), we made the argument that, taken together, they provide a composite view of the different types of talk that make up design activity. While most previous studies have focused on collecting and analysing a particular type of data, the combined corpus has enabled us to characterise designing more fully in terms of 'design talk'.

Three perspectives on characterising design talk are presented in this paper. All three approaches have one aspect in common: unlike prior studies of the DTRS datasets, these avoid looking at specific contexts of design, even though the kinds of design problems and the scenarios of the design discussions are different across each dataset, and indeed across different sessions in the same dataset.

Sec. 3 uses Linguistic Inquiry and Word Count (LIWC) categories to compare the DTRS corpus to other corpora as well as between datasets within the DTRS corpus. Using LIWC categories as a structural basis for a network map, we visualized all the design meetings as an aggregate, comparing this map to other written and conversational corpora. We observed that categories related to thinking and discussing, as well as time and space were more dominant and densely connected in design talk. These semantic dimensions are furthermore connected to a heightened use of function and verb forms. In contrast to this, categories of social relations, emotional content, and descriptions of risk, reward, and achievement take a more prominent role in written text, coming to the fore in design conversations only in certain design stages such as reflection and review of past design activity. When each DTRS dataset is visualised separately as a network, we see patterns such as the interlinking of cognitive processes to the LIWC subcategories of time, space, and motion, suggesting a mingling of abstract, conceptual thinking with concrete experiences. We posit that such a mingling is characteristic of conversations around designing, where cognitive processes such as hypothesizing are often derived from or applied to concrete entities such as objects, space, and movement.

We then take a magnifying glass to the cognitive processes in design discussions in Sec. 4. Specifically, we focus on one particular LIWC category, i.e. *'cognitive process (tentative)'* and identify patterns in word usage characteristic of designerly talk suggested in prior literature that examined specific excerpts from DTRS datasets. Semantically grouping the terms using word embeddings revealed subcategories such as the exploration of possibilities, the use of indefinite pronouns, and of words relating to quantity. We found that the most dominant terms related to talk that explored possibilities, and to talk that explored non-specifics. Our analysis also exposed the limitation of analysing such text using LIWC as the sole application, when some instances of tentative talk were revealed to simply represent the idiolect of dominant speakers. To an extent, such anomalies can be identified by comparing patterns in aggregated data to patterns in individual datasets/sessions. The existence of such "false positives"—non-tentative thinking that gets identified as tentative thinking—raises the question: could there be other indicators of tentative thinking that are not currently categorized as such in the LIWC dictionary? If so, how might such indicators be identified? On a more general level, are there other attributes of talk apart from tentativeness that signify a 'designerly' way of speaking?

While the former questions might require further investigation into machine learning approaches that will be the focus of future work, we investigated the latter question by training a language model with the DTRS datasets and examining the text generated by the model. In Sec. 5, we found that generating artificial text in the style of DTRS design talk and comparing it to actual talk meant that we could see what is 'learned' from the data by the machine learning software—the tentative nature of design talk, for example. We were also able to understand what was missing in the artificial text—the idea of a shared 'design object' with logical design components and relations. We presented only a brief example, but using text generation more systematically offers a radically different and creative way of engaging with data to reveal underlying patterns that might be present in a corpus. The option remains that we could close the loop and analyse generated text with LIWC categories, something that we might consider doing in the future. We posit that studying patterns such as tentativeness, the handling and transformation of virtual objects, and responses to ideas may help us understand what constitutes design talk.

In conclusion we have begun to tease out some interesting future areas of exploration and develop methods that might get us there. What we need is bigger data. Where could we get that from? Our future focus will be on identifying and adding datasets that we can consider as design activity consonant with the DTRS data to increase our general understanding of design talk. With many online platforms and repositories now providing open access to data there is a considerable opportunity to build a large corpus of design activity data. We have shown how comparing corpora can also reveal key differences between the nature of design discourse and other types of discourse. This opens the further possibility of comparing 'design' activity with 'design-like' activity, for example, political, scientific, or strategic discourse. Doing this will help provide a firmer evidence base for a more general theory of designing and its practice and use in a wider range of application areas.

REFERENCES

[1] Cross, N., 2018. "Developing design as a discipline". *Journal of Engineering Design,* ***29***(12), pp. 691–708.

[2] Lloyd, P., and Scott, P., 1994. "Discovering the design problem". *Design Studies,* ***15***(2), pp. 125–140.

[3] Suwa, M., and Tversky, B., 1997. "What do architects and students perceive in their design sketches? a protocol analysis". *Design Studies,* ***18***(4), pp. 385–403. Descriptive models of design.

[4] Cramer-Petersen, C. L., Christensen, B. T., and Ahmed-Kristensen, S., 2019. "Empirically analysing design reasoning patterns: Abductive-deductive reasoning patterns dominate design idea generation". *Design Studies,* ***60***, pp. 39–70.

[5] Cardella, M. E., Atman, C. J., and Adams, R. S., 2006. "Mapping between design activities and external representations for engineering student designers". *Design Studies,* ***27***(1), pp. 5–24.

[6] Menning, A., Grasnick, B. M., Ewald, B., Dobrigkeit, F., and Nicolai, C., 2017. "Combining computational and human analysis to study low coherence in design conversations". In *Analysing design thinking: Studies of cross-cultural Co-Creation*, B. T. Christensen, L. J. Ball, and K. Halskov, eds. CRC Press: Taylor and Francis Group, pp. 291–309.

[7] Archer, J., and Jockers, M. L., 2016. *The bestseller code: Anatomy of the blockbuster novel*. St. Martin's Press.
[8] Dorst, K., and Cross, N., 2001. "Creativity in the design process: co-evolution of problem–solution". *Design Studies, **22***(5), pp. 425–437.
[9] Gero, J. S., and Jiang, H., 2016. "Exploring the design cognition of concept design reviews using the fbs-based protocol analysis". In *Analyzing design review conversations*, R. Adams and A. Siddiqui, eds. Purdue University Press, pp. 177–194.
[10] Cash, P., Hicks, B., and Culley, S., 2015. "Activity theory as a means for multi-scale analysis of the engineering design process: A protocol study of design in practice". *Design Studies, **38***, pp. 1–32.
[11] Cross, N., 2018. "A brief history of the design thinking research symposium series". *Design Studies, **57***, pp. 160–164.
[12] Cross, N., Dorst, K., and Roozenburg, N., 1992. *Research in Design Thinking: Proceedings of a Workshop Meeting Held at the Faculty of Industrial Design Engineering, Delft University of Technology, the Netherlands, May 29-31, 1991*. Delft University Press.
[13] Cross, N., Christiaans, H., and Dorst, K., 1996. *Analysing design activity*. Wiley.
[14] McDonnell, J., and Lloyd, P., eds., 2009. *About: Designing—Analysing Design Meetings*. CRC Press: Taylor and Francis Group.
[15] Adams, R. S., and Siddiqui, J. A., 2016. *Analyzing design review conversations*. Purdue University Press.
[16] Christensen, B. T., Ball, L. J., and Halskov, K., eds., 2017. *Analysing Design Thinking: Studies of Cross-Cultural Co-Creation*. CRC Press.
[17] Dorst, K., 1995. "Analysing design activity: new directions in protocol analysis". *Design Studies, **16***(2), pp. 139–142. Analysing Design Activity.
[18] Lloyd, P., and McDonnell, J., 2009. "Editorial". *Design Studies, **30***(2), pp. 115–118. Values in the Design Process.
[19] McDonnell, J., and Lloyd, P., 2009. "Editorial". *CoDesign, **5***(1), pp. 1–4.
[20] Adams, R. S., Cardella, M., and Şenay Purzer, 2016. "Analyzing design review conversations: Connecting design knowing, being and coaching". *Design Studies, **45***, pp. 1–8. Special Issue: Design Review Conversations.
[21] Adams, R. S., McMullen, S., and Fosmire, M., 2016. "Codesigning review conversations – visual and material dimensions". *CoDesign, **12***(1-2), pp. 1–5.
[22] Halskov, K., and Christensen, B. T., 2018. "Designing across cultures". *CoDesign, **14***(2), pp. 75–78.
[23] Goldschmidt, G., 1995. "The designer as a team of one". *Design Studies, **16***(2), pp. 189–209. Analysing Design Activity.
[24] Goldschmidt, G., 2014. *Linkography: unfolding the design process*. MIT Press.
[25] Visser, W., 2009. "The function of gesture in an architectural design meeting". In *About: Designing—Analysing Design Meetings*, J. McDonnell and P. Lloyd, eds. CRC Press: Taylor and Francis Group, pp. 269–284.
[26] McMullen, S., Brightman, A., and Jaycox, H., 2016. "Making gestural frequencies". In *Analyzing Design Review Conversations*, R. Adams and J. Siddiqui, eds. Purdue University Press, pp. 31–57.
[27] Dong, A., MacDonald, E., Christensen, B., Ball, L., and Halskov, K., 2017. "From observations to insights: The hilly road to value creation". *Analysing Design Thinking: Studies of Cross-Cultural Co-Creation, Taylor & Francis, London*, pp. 465–482.
[28] Lloyd, P., and Oak, A., 2018. "Cracking open co-creation: Categories, stories, and value tension in a collaborative design process". *Design Studies, **57***, pp. 93–111.
[29] Glock, F., 2009. "Aspects of language use in design conversation". *CoDesign, **5***(1), pp. 5–19.
[30] Mabogunje, A., Eris, O., Sonalkar, N., Jung, M., Leifer, L., McDonnell, J., and Lloyd, P., 2009. "Spider webbing: A paradigm for engineering design conversations during concept generation". In *About: Designing—Analysing Design Meetings*, J. McDonnell and P. Lloyd, eds. CRC Press: Taylor and Francis Group, pp. 49–65.
[31] Lloyd, P., and McDonnell, J., 2009. *Introduction to about: designing-analysing design meetings*. Taylor and Francis Group.
[32] Adams, R. S., 2016. "Design review conversations: The dataset". In *Analyzing design review conversations*. Purdue University Press, pp. 23–30.
[33] Christensen, B. T., and Abildgaard, S. J. J., 2017. "Inside the dtrs11 dataset: Background, content, and methodological choices". *Analysing design thinking: Studies of cross-cultural co-creation*, pp. 19–37.
[34] Leech, G., Rayson, P., and Wilson, A., 2014. *Word frequencies in written and spoken English: Based on the British National Corpus*. Routledge.
[35] Pennebaker, J. W., Booth, R. J., Boyd, R. L., and Francis, M. E., 2015. *Linguistic inquiry and word count: LIWC 2015*. Pennebaker Conglomerates.
[36] Danescu-Niculescu-Mizil, C., and Lee, L., 2011. "Chameleons in imagined conversations: A new approach to understanding coordination of linguistic style in dialogs.". In Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics, ACL 2011.
[37] Schler, J., Koppel, M., Argamon, S., and Pennebaker, J. W., 2006. "Effects of age and gender on blogging.". In AAAI spring symposium: Computational approaches to analyzing weblogs, Vol. 6, pp. 199–205.
[38] Roozenburg, N. F., and Eekels, J., 1995. *Product design: fundamentals and methods*. Wiley.

[39] Schön, D. A., 1984. "Problems, frames and perspectives on designing". *Design Studies,* ***5***(3), pp. 132–136.

[40] Valkenburg, R., and Dorst, K., 1998. "The reflective practice of design teams". *Design studies,* ***19***(3), pp. 249–271.

[41] Darke, J., 1979. "The primary generator and the design process". *Design studies,* ***1***(1), pp. 36–44.

[42] Roozenburg, N. F., 2002. "Defining synthesis: on the senses and the logic of design synthesis". In *Engineering Design Synthesis*, A. Chakrabarti, ed. Springer, pp. 3–18.

[43] Kolko, J., 2010. "Abductive thinking and sensemaking: The drivers of design synthesis". *Design issues,* ***26***(1), pp. 15–28.

[44] Schön, D. A., 1992. "Designing as reflective conversation with the materials of a design situation". *Knowledge-based systems,* ***5***(1), pp. 3–14.

[45] Mikolov, T., Kombrink, S., Burget, L., Černockỳ, J., and Khudanpur, S., 2011. "Extensions of recurrent neural network language model". In IEEE international conference on acoustics, speech and signal processing, pp. 5528–5531.

[46] Speer, R., Chin, J., and Havasi, C., 2017. "Conceptnet 5.5: An open multilingual graph of general knowledge". In AAAI, pp. 4444–4451.

[47] Day, W. H., and Edelsbrunner, H., 1984. "Efficient algorithms for agglomerative hierarchical clustering methods". *Journal of classification,* ***1***(1), pp. 7–24.

[48] Devlin, J., Chang, M., Lee, K., and Toutanova, K., 2018. "BERT: pre-training of deep bidirectional transformers for language understanding". *CoRR,* ***abs/1810.04805***.

[49] Shoeybi, M., Patwary, M., Puri, R., LeGresley, P., Casper, J., and Catanzaro, B., 2019. "Megatron-LM: Training multi-billion parameter language models using model parallelism". *Computing Research Repository (CoRR)*.

[50] Medway, P., and Clark, B., 2003. "Imagining the building: architectural design as semiotic construction". *Design Studies,* ***24***(3), pp. 255–273. Designing in Context.

