

Multi-Dimensional Exploratory Factor Analysis of TED Talks

1. Introduction

This article conducts Multi-Dimensional (MD) analysis on a corpus of TED talks (TTC), aiming to characterise linguistic variation within this corpus in functional terms. TED talks are a contemporary digital genre via which a wide range of speakers communicate their ideas to a global audience. TED talks are typically 5 to 20 minutes long and are delivered to both live and online audiences. The conference originated in Silicon Valley in 1984 as a place where “technology, entertainment, and design converge” (ted.com, n.d.). Technology, entertainment and design comprise the original T-E-D of TED. However, since inception the brand has expanded and now has very few limitations in terms of included topics and target audience.

Prior research has identified TED talks as a spoken form of popular science, with certain similarities to academic texts. A popularisation discourse occurs when an expert “addresses not an expert group within the discipline but an audience of non-specialists” (Gotti, 2014, p. 16). TED talks achieve popularisation through reduced technicality in terms of content and vocabulary; an informal tone; and a preference for narrative text types (Mattiello, 2017). Some TED talks, can therefore be seen as a space where researchers communicate their findings to a public audience, with noteworthy linguistic adaptations.

Stance features have also been investigated in TED talks (Di Carlo, 2015), captured as “axiological evaluative adjectives” (Felices Lago, 1997), which imply emotional reaction or evaluative judgement, e.g. *important, interesting, wonderful*. It was argued that TED speakers use these adjectives to establish a connection with the audience, convey the relevance of their findings, and make judgements on the moral, political, and economic aspects of science (Di Carlo, 2015). These findings can also be linked to the perspective that TED talks are a popular form of science.

However, not all TED talks fall into this category, as talks are delivered by a broader range of speakers other than scientists and academics, including entrepreneurs, politicians, celebrities, writers, and artists (Cassidy et al., 2013). In fact, according to the curator of TED talks, a TED talk can be on any topic of “public interest” (Anderson, 2016, p. xii). This makes the TED talk register a broadly defined space open to a wide range of speakers and topics.

Wingrove, P. & Crosthwaite, P. (2022). Multi-Dimensional Exploratory Factor Analysis of TED Talks. *Register Studies* (authors' pre-publication manuscript, main manuscript DOI: <https://doi.org/10.1075/rs.21008.win>)

This is reflected in the content tags on the TED website which act as a guide to the content, such as *physics, economics, history, activism, environment, art, and culture*. Notably a number of these tags can be linked to academic disciplines, to extent that a full roster of academic disciplines are present. At the time of corpus creation, there were 427 tags on the TED website, which were collected as part of our metadata.

Due to the inclusion of academic content, a number of studies have approached TED from an English for academic purposes (EAP) perspective, noting their pedagogic potential. For instance, researching the use of TED talks for practising academic listening (Abdulrahman, 2017; Elk, 2014; Takaesu, 2014) and measuring the coverage of general English vocabulary (Nurmukhamedov, 2017) and academic vocabulary (Wingrove, 2017) to investigate their suitability as learning materials. These studies found that TED talks require a wider range of vocabulary for comprehension, closer to the average written text, rather than the average spoken text (Nurmukhamedov, 2017) and that TED talks have lower academic vocabulary coverage compared to academic lectures, although with notable overlap (Wingrove, 2017).

One interesting feature of TED talks is the fact that this register brings together many different speakers and topics within the same space, covering academic and non-academic content, simultaneously delivered to an online and live audience, and perhaps even with speakers having divergent purposes. This differentiates TED talks from other registers which are more restricted in terms of topics and participants.

It would be interesting investigate register variation in this unique contemporary digital genre that brings together aspects of situational context that are rarely brought together into one register. The current article aims to investigate register variation in TED talks by compiling a representative corpus and conducting a multi-Dimensional Analysis (MD; Biber, 1988). MD analysis is a key method in corpus linguistics for determining linguistic variation across genres and registers. This approach allows for the linguistic features of a given corpus to be quantified and subsequently reduced and grouped along functional 'dimensions'.

MD studies have been conducted on corpora comparable to TED talks. Studies on academic texts have identified dimensions which differentiate more informationally dense texts against more elaborated texts; for instance, in academic research articles (Gray, 2011), textbooks, (Egbert, 2014), and University classroom talk (Csomay, 2005). Moreover, stance dimensions have been discovered in academic texts (Biber, 2006; Csomay, 2005; Gardner, Nesi, & Biber, 2018), which coheres with the view that stance is a key feature of academic texts (Hyland, 2005).

Wingrove, P. & Crosthwaite, P. (2022). Multi-Dimensional Exploratory Factor Analysis of TED Talks. *Register Studies* (authors' pre-publication manuscript, main manuscript DOI: <https://doi.org/10.1075/rs.21008.win>)

The TED talk register is also defined by a broad range of topics and speakers. Studies on corpora which also include a broad range of topics have identified dimensions which differentiate texts by topic. These dimensions have been discovered in academic texts, such as in research articles, identified as "human vs. non-human" (Gray, 2011), but also outside academic texts, such as "idea-oriented discourse vs. action oriented discourse" in TIME magazine (Condi de Souza, 2014).

It can be expected that an MD analysis of TED talks may present information packaging, stance, and topic-associated dimensions, although the inclusion of talks on non-academic topics and the potential to either ad-lib or perform talks verbatim makes it difficult to predict exactly how these dimensions would be expressed. Moreover, there may be latent linguistic dimensions unique to the TED register, which can be uncovered by a robust MD analysis.

Therefore, the present study conducts an MD of a representative and large TED Talk Corpus (TTC, talks = 2483; words = 5,068,781), which includes 427 content tags. We aim to answer the following research questions:

RQ1: What linguistic dimensions can be found in the TTC by conducting exploratory factor analysis?

RQ2: How are the EFA (Exploratory Factor Analysis) dimensions represented across subregisters in the TTC?

We first provide the details of our TED Talk Corpus, before presenting the methodology and results of the MD model.

2. TED Talk Corpus

2.1 Compilation

The current study is based on a corpus of TED Talks extracted from the TED website. As of January 2018, all talks were collected from the TED website (date range: 2006 – 2017) except those without transcripts. Orthographic transcription was used. This captured spoken features such as contractions, but not filled pauses such as "um" and "er". Details of the TED Talk Corpus (TTC) are given in Table 1. The corpus includes 2483 talks across 427 content tags.

TED uses a system of content tags to act as a guide to the content and talks often have multiple tags. For instance, a talk on gravitational waves could have the tags *astronomy*, *science*, *exploration*, and *physics*. Note that because each talk can (and often does) have multiple tags, the content areas often overlap. For instance, a selection of talks with the tag *astronomy* can be expected to overlap with a selection of talks with the tag *physics*. These overlapping categories can be expected to affect the results by creating fuzziness in the data as the categories do not comprise distinct sets. However,

Wingrove, P. & Crosthwaite, P. (2022). Multi-Dimensional Exploratory Factor Analysis of TED Talks. *Register Studies* (authors' pre-publication manuscript, main manuscript DOI: <https://doi.org/10.1075/rs.21008.win>)

the decision to stick with TED's classification system, which frequently allocates multiple tags to individual talks, results in a more accurate description of the content, particularly in relation to talks which combine content areas. The complexity and diversity of the TED talk corpus is of particular interest to the study and maintaining TED's overlapping classification system enables our study to account for this complexity. For the sake of clarity, this article refers to TED talks in general as a *register* and the various areas delineated by the content tags as *subregisters*.

When the corpus data were collected, all of the content tags were recorded as part of our metadata and all talks had at least one tag. It is important to note that MD analysis was conducted on *all* talks in the corpus (RQ1). However, it would be impractical to try to investigate how the resultant dimensions are represented across all 427 tags. Therefore the current paper identifies the following 36 content tags as of particular interest to RQ2:

1. The 'main' tags that TED lists on their website (n = 6). These are: *technology*, *entertainment*, *design*, *business*, *science*, and *global issues*. These tags are offered first to the viewer when visiting the TED website and are particularly high frequency.
2. Academic disciplines associated with Biglan's (1973) disciplinary taxonomy (n = 21 disciplines, plus 5 aggregate categories) . Biglan's (1973) taxonomy divides academic disciplines into two continua: hard-soft and pure-applied. The "hard-soft" continuum characterises more paradigmatic areas, which have a "greater consensus about content and method", against less paradigmatic areas (Biglan 1973: 202). The "pure-applied" continuum is defined by "the concern of the [academic] area with application to practical problems" (Biglan, 1973, p.202). The continua are overlapping, resulting in four categories: hard pure, hard applied, soft pure, and soft applied. A further category, all academic tags, can also be included. The advantage of using Biglan's (1973) taxonomy is that it ensures a broad range of academic disciplines are represented and that these continua can be mapped onto the resultant dimensions.
3. Other high frequency tags (n = 4). These are: *health*, *TEDx*, *art*, and *culture*. The *TEDx* tag marks talks delivered at TEDx events: grassroots TED conferences, organised independently but under the guidance of the TED organisation.

Note that when the results are presented, we include five extra categories related to the academic tags. These are aggregate categories: *hard pure*, *hard applied*, *soft pure*, *soft applied*, and *all academic tags*. For instance, talks tagged *biology*, *chemistry*, *physics*, *math*, and *astronomy*, can be combined into one additional *hard pure* category. These aggregate categories will be marked by an

Wingrove, P. & Crosthwaite, P. (2022). Multi-Dimensional Exploratory Factor Analysis of TED Talks. *Register Studies* (authors' pre-publication manuscript, main manuscript DOI: <https://doi.org/10.1075/rs.21008.win>)

asterisk in the findings chapters. Note that the content tags were not included in the MD analysis, but were compared against the resulting dimensions following the MD analysis.

Table 1 Properties of the TED Talk Corpus

| | TTC |
|-------------------------------|--------------------------------|
| Number of Talks | 2,483 |
| Total Tokens | 5,068,781 |
| Average Tokens per transcript | 2,041 |
| Number of Speakers | 2,090 |
| Type | Spoken (US) |
| Date of transcripts | 2006-2017 |
| Transcription | Orthographic |
| Content | 427 content tags reduced to 36 |

2.2. Situational Characteristics of TED Talks

MDA is situated within the register perspective theoretical approach to corpus analysis (Biber & Conrad, 2009). This perspective assumes that linguistic features are functional and that certain features are used to meet the demands of the context. Therefore, an analysis of register involves: i. a description of the situational characteristics of the register; ii. a description of the pervasive linguistic features; and iii. an interpretation of the relationship between the situational characteristics and linguistic features in functional terms. As this perspective is adopted, this section will detail the situational characteristics of TED talks (Table 2).

Several characteristics are important to note. In particular, we can note that TED talks are delivered to a general audience often with an expert from the field communicating to people outside their field. TED talks are typically short, often between 5 to 20 minutes long, which means that talks have the potential to be scripted and performed verbatim. Moreover, the communicative purposes of TED talks can be linked to the desire to spread ideas and inspire change in the wider world. TED talks can also be delivered on a broad range of academic and non-academic topics, suggesting that the TED platform is less attached to topic than it is to other factors such as style and delivery.

Table 2 Situational Characteristics of the TTC (Adapted from Biber & Conrad, 2009)

| Situational Characteristics | TTC |
|-----------------------------|--|
| I. Participants | |
| A. Addressor(s) | |
| 1. who | TED speakers. Represents the TED brand. |

Wingrove, P. & Crosthwaite, P. (2022). Multi-Dimensional Exploratory Factor Analysis of TED Talks. *Register Studies* (authors' pre-publication manuscript, main manuscript DOI: <https://doi.org/10.1075/rs.21008.win>)

| | |
|---|---|
| 2. social characteristics | A range of professions: entertainers, artists, businessmen, engineers, and academics (Cassidy <i>et al.</i> 2013). |
| B. Addressees | |
| 1. who | Live and online audience. |
| 2. social characteristics | Live audience may be high status; online audience is restricted to those with internet access. |
| C. On-lookers | |
| 1. who | The online audience is arguably the primary audience in terms of views and promotion. Therefore they are termed <i>addressee</i> rather than <i>on-looker</i> . |
| 2. social characteristics | - |
| II. Relations among participants | |
| A. Interactiveness | Primarily a unidirectional monologue. Some instances of audience interaction, although not a dialogue. |
| B. Social roles | Speaker has an expert status. |
| C. Personal relationship | Global community. |
| D. Shared knowledge | Less shared knowledge compared to an academic community. Communicating expert knowledge to a lay audience. |
| III. Channel | |
| A. Mode | Speech; <i>limited</i> use of slides, images, other multimedia |
| B. Specific Medium | Taped & transcribed. Audio, video, text. |
| IV. Production circumstances | |
| A. Real time / planned / scripted / revised and edited | Performance: planned / scripted, real time, practiced. |
| V. Setting | |
| A. Is the time and place of communication shared by participants? | Yes by the speaker and live audience. But not by the online viewership. |
| B. Place of communication | |
| 1. Private / public | Public event. |
| 2. Specific setting | Conference. |
| C. Time: contemporary, historical time period | Contemporary: 2006-2017 |
| VI. Communicative purposes | |
| A. General purposes | From TED's mission statement: spread ideas, inspire change. |
| B. Specific purposes | Depending on speaker: inspire, narrate, explain, describe, persuade, entertain, reveal self, connect to audience, promote self, promote research. |
| VII. Topic | |
| A. General topical domain | Curator of TED: Anderson (2016: xii), TED talks can "cover any topic of public interest". Covers academic and non-academic areas. |
| B. Specific topic | Academic and non-academic topics. |

3. Method

3.1. Corpus annotation for MD

MD analyses require the automated parsing and tagging of corpus data for linguistic features, which are then reduced and grouped along functional lines through exploratory factor analysis (EFA).

Wingrove, P. & Crosthwaite, P. (2022). Multi-Dimensional Exploratory Factor Analysis of TED Talks. *Register Studies* (authors' pre-publication manuscript, main manuscript DOI: <https://doi.org/10.1075/rs.21008.win>)

When selecting the target features for tagging, it is best to start with a large number of features (Biber, 1988). This set (n = 90) was comprised of:

1. features identified by the multi-dimensional analysis tagger (Nini, 2019), such as past tense, predictive modals, nominalisations, type-token ratio, etc.;
2. semantic classes of content bearing words (Biber, Reppen, Byrd, Helt, Clark, Cortes, Csomay, & Urzua, 2004, p. 27-35), and;
3. additional features not separately counted in prior studies, such as comparatives, superlatives, first person pronouns plural, and first person pronouns singular.

For 2 and 3 above, *LancsBox* (Version 4, Brezina, Timperley, & McEnery, 2018) was used to calculate the frequency of linguistic features outside of the multi-dimensional analysis tagger. For semantic classes, each group was searched for using part-of-speech (POS) criteria. For instance, all animate noun types were input into *LancsBox* along with the POS criteria N*, restricting results to nouns only. *LancsBox* v.4 uses *TreeTagger* (Schmid, 1994) to tag the corpus as POS items. In testing, *TreeTagger* has achieved 96.36% accuracy on written texts (Schmid, 1994).

The set of linguistic features was then reduced by removing features which did not distinguish between texts, i.e. which did not load onto any factor; and features which overlapped excessively with each other. In terms of overlapping features, more general features were removed (e.g. total adverbs) and specific features were retained (e.g. concessive adverbial subordinators). However, overlap was inevitable in some cases, for instance where semantic verb classes overlap with grammatical verb categories. In such cases, all features were retained as the features did not consistently load together into one factor during the subsequent EFA.

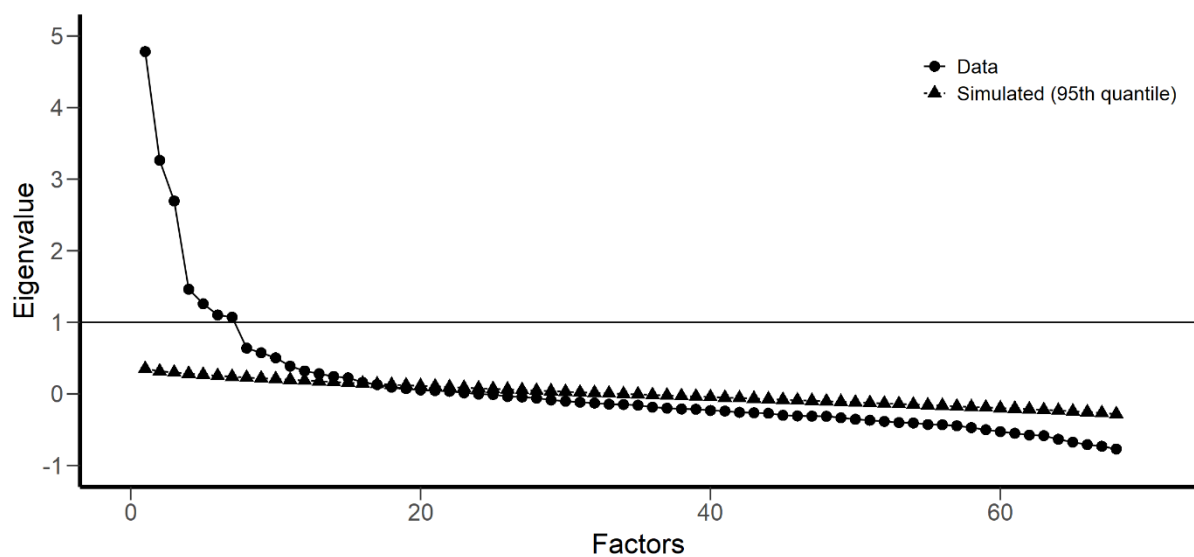
3.2. Exploratory Factor Analysis

EFA was conducted on the dataset using the *JASP* statistical package (JASP Team, 2020). An oblique promax rotation was used and 0.3 was set as the cut-off for variable loading. In prior studies, eigenvalues (greater than one) and visual analysis of scree plots have been used to help determine the number of factors (e.g. Thompson et al., 2017). In the current study, eigenvalues greater than one and the scree plot (Figure 1) both suggested a 7-factor model. The model was tested using RMSEA, a goodness of fit criterion which “estimates the discrepancy of the model and the data per degree of freedom for the model” (Crosthwaite & Cheung, 2019, p.38). As a rule of thumb, values below 0.05 suggest the model is a close fit; between 0.05-0.08 suggest a reasonable fit, and values

above 0.1 are inadequate (Browne & Cudeck, 1993). In the present model RMSEA was 0.049 with a 90% confidence interval of 0.48-0.50, which suggests the model is a close fit.

Once the final model was determined, the normalised frequencies for the linguistic features were converted to z-scores. This is an important step as it gives the variables within each factor an equal weighting. Following this, z-scores were converted to dimension scores. This was achieved by adding

Figure 1 Factor extraction scree plot



the positively loaded variables and subtracting the negatively loaded variables for each factor. This results in each text having a dimension score for each of the dimensions. Dimension scores for subregisters were calculated by finding the mean score for all talks within that subregister.

4. Results

4.1 EFA model

Table 3 shows the EFA output from JASP. Fifty-eight linguistic variables are spread out among seven factors. A range of linguistic features comprise the model, including grammatical features, semantic categories, clauses, and general properties of the text (e.g. type-token ratio). The following sections (4.2 – 4.8) concern the interpretation of the factors as linguistic dimensions.

Table 3 Factorial Structure

| Dimension 1 | | Dimension 4 | |
|--|-------|--|-------|
| Present tense | 0.63 | Private verbs | 0.97 |
| Demonstrative pronouns | 0.63 | Mental Verbs | 0.76 |
| Pro-verb <i>do</i> | 0.40 | Subordinator that deletion | 0.43 |
| Independent Clause Coordination | 0.39 | Verb + <i>wh</i> -clause | 0.36 |
| Pronoun <i>it</i> | 0.37 | <i>That</i> verb complements | 0.34 |
| Tokens | 0.37 | | |
| Existential <i>there</i> | 0.37 | | |
| Second person pronouns | 0.37 | Dimension 5 | |
| Conditional adverbial subordinators | 0.32 | Stance adverbials | 0.96 |
| Communication verbs | 0.32 | Emphatics | 0.88 |
| Causative adverbial subordinators | 0.31 | Hedges | 0.30 |
| Past participial <i>whiz</i> -deletion relatives | -0.31 | Comparatives | 0.30 |
| Present participial clauses | -0.38 | | |
| Attributive adjectives | -0.44 | | |
| Phrasal coordination | -0.46 | Dimension 6 | |
| Type-token ratio | -0.59 | <i>That</i> relative clauses on object position | 0.52 |
| | | Demonstratives | 0.39 |
| Dimension 2 | | <i>That</i> verb complements | 0.35 |
| Abstract/Process Noun | 0.58 | <i>That</i> relative clauses on subject position | 0.33 |
| Nominalizations | 0.55 | First Person Pronouns Plural | 0.33 |
| Attributive adjectives | 0.52 | Stance adverbials | 0.32 |
| Cognitive Noun | 0.46 | Split auxiliaries | 0.31 |
| Stance noun | 0.42 | Second person pronouns | -0.31 |
| Conjuncts | 0.38 | | |
| Present tense | 0.30 | Dimension 7 | |
| Independent Clause Coordination | -0.44 | Infinitives | 0.70 |
| Past tense | -0.59 | Desire Verb + <i>to</i> clause | 0.45 |
| | | Progressive / continuous tense | 0.34 |
| Dimension 3 | | Causation Verbs | 0.33 |
| Animate noun | 0.63 | Suasive verbs | 0.30 |
| Past tense | 0.58 | | |
| Third person pronouns | 0.53 | | |
| <i>Wh</i> -relative clauses on subject position | 0.44 | | |
| Public verbs | 0.40 | | |
| Perfect aspect | 0.35 | | |
| Group / institution Noun | 0.34 | | |
| Pronoun <i>it</i> | -0.30 | | |
| Present tense | -0.35 | | |

Wingrove, P. & Crosthwaite, P. (2022). Multi-Dimensional Exploratory Factor Analysis of TED Talks. *Register Studies* (authors' pre-publication manuscript, main manuscript DOI: <https://doi.org/10.1075/rs.21008.win>)

| | |
|----------------|-------|
| Technical noun | -0.37 |
| Concrete Nouns | -0.51 |

4.2 Dimension 1 Spontaneous involved versus edited informational

Dimension 1 is the largest dimension in terms of linguistic features. The positively loaded features with the highest loadings are the present tense, demonstrative pronouns (*those, this, these*), pro-verb *do*, independent clause coordination, pronoun *it*, total tokens, existential *there*, and second person pronouns. When these features are taken together they demonstrate spontaneous involved discourse. For instance:

1. We find patterns, **and** we represent them. **And** I think **this** is a nice day-to-day definition. But today I want to go a little bit deeper, **and** think about what the nature of **this** is. What makes **it** possible? **There's** one thing that's a little bit deeper, and that has to **do** with the ability to change your perspective. **And** I claim that when **you** change **your** perspective, **and** if **you** take another point of view, **you** learn something new about what **you** are watching or looking at or hearing. (Roger Antonsen: *Math is the hidden secret to understanding the world*)
2. I'm not quite sure why **you** decide not to **do it**. You're too lazy to **do it**. **It's** too hard. **You're** afraid **if you** look for **your** passion and **don't** find it, **you'll** feel like **you're** an idiot, so then **you** make excuses about why **you're** not going to look for **your** passion. They are excuses, ladies and gentlemen. We're going to go through a whole long list -- **your** creativity in thinking of excuses not to **do** what **you** really need to **do if you** want to have a great career. (Larry Smith: *Why you will fail to have a great career*)

Extracts 1 and 2 both display discourse which is more involved. There is evidence of directly referencing the audience in an interactive manner and of spontaneous elaboration through the use of demonstrative pronouns, subordinators, existential *there*, and independent clause coordination.

The key negatively loaded features are type-token ratio, phrasal coordination, and attributive adjectives. Taken together these features comprise condensed descriptive informational texts. For instance:

3. For 12 years, I have produced **collaborative** portraits, still lifes, landscapes and **aerial** views in order to build a **visual** archive to address the intersection of the steel industry, the environment, **and** the health care system's impact on the bodies of my family **and** community. The tradition **and grand** narrative of Braddock is mostly comprised of stories of industrialists **and trade** unions. Currently, the **new** narrative about Braddock, a poster child

for Rust Belt revitalization, is a story of **urban** pioneers discovering a **new** frontier. (Latoya Ruby Frazier: *A visual history of inequality in industrial America*)

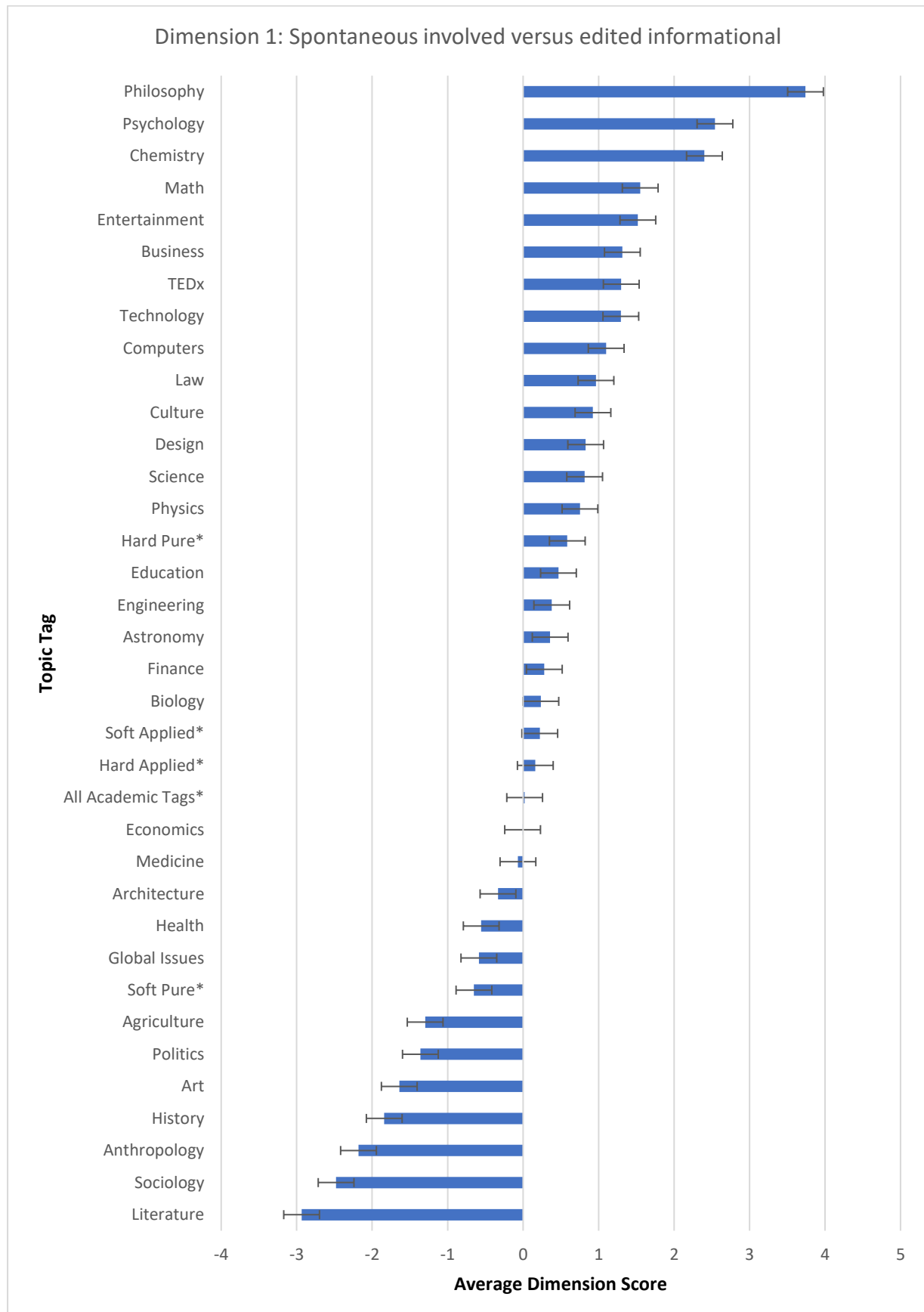
4. Jitish Kallat successfully practices across photography, sculpture, painting **and** installation. As you can see, he's heavily influenced by graffiti **and** street art, and his home city of Mumbai is an **ever-present** element in his work. He really captures that sense of density **and** energy which really characterizes **modern urban** Bombay. He also creates **phantasmagoric** sculptures made of bones from **cast** resin. (Ravin Agrawal: *10 young Indian artists to watch*)

Within extracts 3 and 4 the high type-token ratio measure is evident through the use of infrequent descriptive language. Information is condensed through the use of attributive adjectives and in contrast with independent clause coordination, these extracts use phrasal coordination instead. These texts are also shorter, as evidenced by the low tokens score, implying that information may be packaged in this way to meet a shorter time limit or may have been scripted and performed.

It appears that this dimension is identifying the tendency of speakers to lean towards more interactive, dispersed, involved texts in comparison to more condensed, descriptive, edited, informational texts. This dimension is very similar to Biber's (1988) first dimension in his original MDA model, "involved versus informational discourse". Whilst Biber's (1988) dimension broadly characterises variation across spoken and written registers, the current dimension finds a similar, more subtle expression of this same dimension within a single register. This can be connected to the situational characteristics of the corpus, which suggested that TED talks fall on a spectrum of planning, either strictly written, edited, and performed verbatim, resulting in a negative weighting, or planned but ad-libbed in a more spontaneous manner, resulting in a positive weighting. Note that total tokens loads positively on this dimension suggesting that longer talks are less likely to be written and performed compared to shorter talks.

When the various subregisters are mapped onto Dimension 1 (Figure 2) we can see that certain subregisters have a stronger preference for either side of the dimension. In particular it is interesting to note that many soft pure related subjects appear with a negative weighting with the exception of *philosophy* which has the highest score. This may be due to the inclusion of concessive and conditional adverbial subordinators used to structure arguments, or *philosophy* may wish to present itself in a more involved manner due to the conceptual subject matter.

Figure 2 TTC Subregisters mapped onto Dimension 1: Spontaneous involved versus edited informational



Wingrove, P. & Crosthwaite, P. (2022). Multi-Dimensional Exploratory Factor Analysis of TED Talks. *Register Studies* (authors' pre-publication manuscript, main manuscript DOI: <https://doi.org/10.1075/rs.21008.win>)

* Topic tags marked with an asterisk represent aggregate categories.

4.3 Dimension 2 Abstract informational versus narrative discourse

Dimension 2 identifies abstract informational discourse against narrative discourse. The use of abstract/process nouns, cognitive nouns, and stance nouns shows this discourse to be clearly abstract and the use of nominalisations and attributive adjectives demonstrate informational discourse. When texts with high Dimension 2 scores are investigated this tendency is apparent:

5. The **separation** between **science** and **human values** is an **illusion** -- and actually quite a dangerous one at this point in human **history**. Now, it's often said that **science** cannot give us a foundation for **morality** and **human values**, because **science** deals with **facts**, and **facts** and **values** seem to belong to **different** spheres. It's often thought that there's no **description** of the way the world is that can tell us how the world ought to be. But I think this is quite clearly untrue. **Values** are a certain kind of **fact**. They are **facts** about the well-being of conscious creatures. (Sam Harris: *Science can answer moral questions*)

6. I want a **scientific theory** of **consciousness** that works, and for a long time, I banged my head against the wall looking for a **theory** of **consciousness** in purely **physical** terms that would work. But I eventually came to the conclusion that that just didn't work for **systematic reasons**. It's a **long** story, but the **core idea** is just that what you get from **purely reductionist explanations** in **physical terms**, in **brain-based terms**, is stories about the **functioning** of a **system**, its **structure**, its **dynamics**, the **behavior** it produces, great for solving the **easy problems** ... (David Chalmers: *How do you explain consciousness?*)

Both these texts grapple with abstract concepts related to science, morality, consciousness, and philosophy. The use of abstract/process, cognitive, and stance nouns helps the speakers identify and define the concepts they aim to discuss. These texts also rely on the use of present tense rather than the past tense, showing a lack of narrative text types and a preference for declarative statements.

When texts on the opposite side of Dimension 2 are investigated the use of the past tense shows a tendency towards narrative text types. Independent clause coordination links components of the narrative together. These texts are notable for the lack of references to abstract concepts and processes. Consider the following extract:

7. Well, that chance encounter **inspired** my imagination, **and** I **created** the Lunch Lady graphic novel series, a series of comics about a lunch lady who uses her fish stick nunchucks to fight off evil cyborg substitutes, a school bus monster, and mutant mathletes, **and** the end of

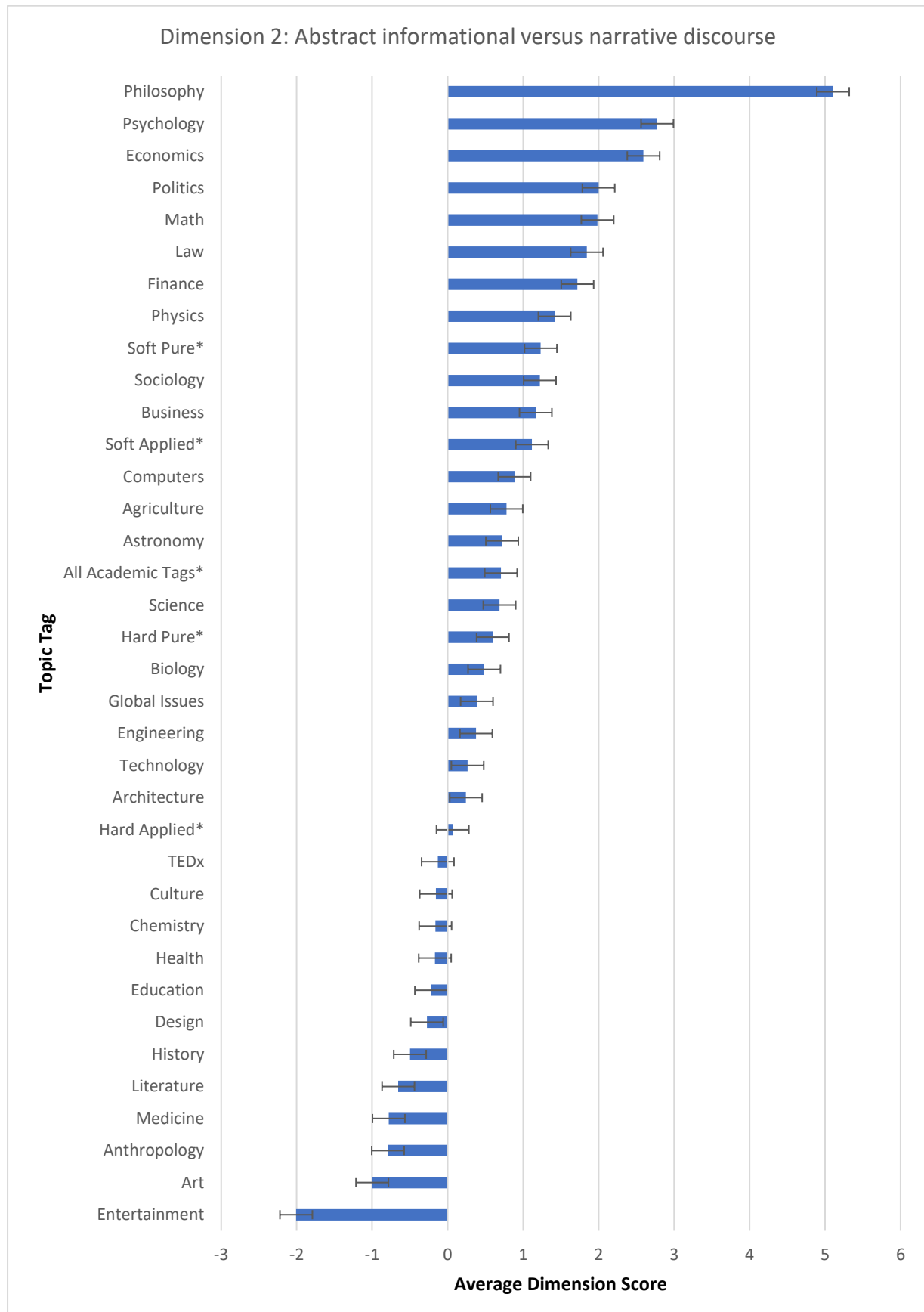
Wingrove, P. & Crosthwaite, P. (2022). Multi-Dimensional Exploratory Factor Analysis of TED Talks. *Register Studies* (authors' pre-publication manuscript, main manuscript DOI: <https://doi.org/10.1075/rs.21008.win>)

every book, they **get** the bad guy with their hairnet, **and** they proclaim, "Justice is served!"
(Jarrett J Krosoczka: *Why lunch ladies are heroes*)

In this extract the use of the past tense and independent clause coordination can be seen to build a narrative. The lack of references to abstract concepts and processes can be seen in the grounded nature of the topic. This means that texts on this side of the dimension are not particularly limited in terms of topic, except for being outside of abstract discourses. It may be the tendency of abstract discourses to lean towards the present tense which highlights narrative structures as an opposing dimension.

When the TTC subregisters are mapped onto Dimension 2, it is interesting to note that *philosophy* has the highest weighted score, which is unsurprising as philosophy can be expected to deal with abstract content. Conversely, *entertainment* talks have the lowest weighting, likely due to the inclusion of narratives and storytelling. Dimension 2 in particular seems to favour soft disciplines, with *psychology*, *economics*, and *politics* also having high weightings. Notably, *art* and *literature* have negative scores which may be due to these talks discussing more concrete entities such as artworks, works of literature, and their creators, rather than abstract concepts and processes.

Figure 3 TTC Subregisters mapped onto Dimension 2: Abstract informational versus narrative discourse



Wingrove, P. & Crosthwaite, P. (2022). Multi-Dimensional Exploratory Factor Analysis of TED Talks. *Register Studies* (authors' pre-publication manuscript, main manuscript DOI: <https://doi.org/10.1075/rs.21008.win>)

* Topic tags marked with an asterisk represent aggregate categories.

4.4 Dimension 3 Human-world oriented versus object oriented

Dimension 3 identifies human-world oriented discourse versus object oriented discourse. The positive pole is associated with verb forms often used in narration; people, groups, and institutions; and *wh*-relative clauses which elaborate on them. Taken together, these features are associated with texts which tell human stories, for example:

8. I research **students who** have overcome immense obstacles related to their background. **Students** from low-income, often single-**parent** households, **students who** have been homeless, incarcerated or perhaps undocumented, or some **who** have **struggled** with substance abuse or **lived** through violent or sexual trauma. So let me tell you about two of the grittiest **people** I've met. Tyrique was **raised** by a single **mother**, and then after high **school**, he **fell** in with the wrong **crowd**. He got **arrested** for armed robbery. But in **prison**, he **started** to work hard ... (Anindya Kundu: *The boost students need to overcome obstacles*)

9. I am the **daughter** of a **forgery**, not just any **forgery** ... When you hear the word "**forgery**," you often understand "**mercenary**." You understand "forged currency," "forged pictures." My **father** is no such **man**. For 30 years of his life, **he made** false papers -- never for **himself**, always for other **people**, and to come to the aid of the persecuted and the oppressed. Let me introduce **him**. Here is my **father** at age 19. It all **began** for him during World War II, when at age 17 **he found himself** thrust into a forged documents workshop. He quickly **became** the false papers **expert** of the Resistance. (Sarah Kaminsky: *My father the forgery*)

In extract 8 the co-occurrence of positively loaded Dimension 3 features can be seen to work together to help the speaker tell a human-oriented story, which references students, parents, school, and prison. In extract 9 the speaker is telling the story of her father who forged documents throughout World War II. In the first extract, the talk is aiming to shine a light on a social issue to spread awareness. In the second extract the speaker is using the TED platform to tell an inspirational story.

On the other end of the spectrum, the negative pole is associated with technical and concrete nouns, pronoun *it*, and the present tense. These features are associated with object oriented discourse and the use of the present tense here shows greater use of factual and declarative statements over narratives. Ostensively, on the TED platform this often takes the form of a demonstration, presenting artwork (extract 10), technology (extract 11), naturally occurring phenomenon (extract 12), or even an abstract concept such as compassion (extract 13):

10. This is a **sculpture** I made, which is a way of, kind of, freeing a **form** into an **object** that has different degrees of freedom. So, **it** can balance on a **point**. This is a bronze **ball**, an aluminium **arm** here, and then this wooden **disk**. And the wooden **disk** was really thought about as something that you'd want to hold on to, and that would glide easily through your **hands**. (Tom Shannon: *Anti-gravity sculpture*)

11. A solar **cell** absorbs **light** and converts it into electrical **energy**. This is why we can use a solar **cell** to charge our mobile **phone**. But now we need to remember that the **data** is encoded in subtle changes of the brightness of the **LED**, so if the incoming **light** fluctuates, so does the **energy** harvested from the solar **cell**. (Harald Haas: *Forget wi-fi meet the new li-fi internet*)

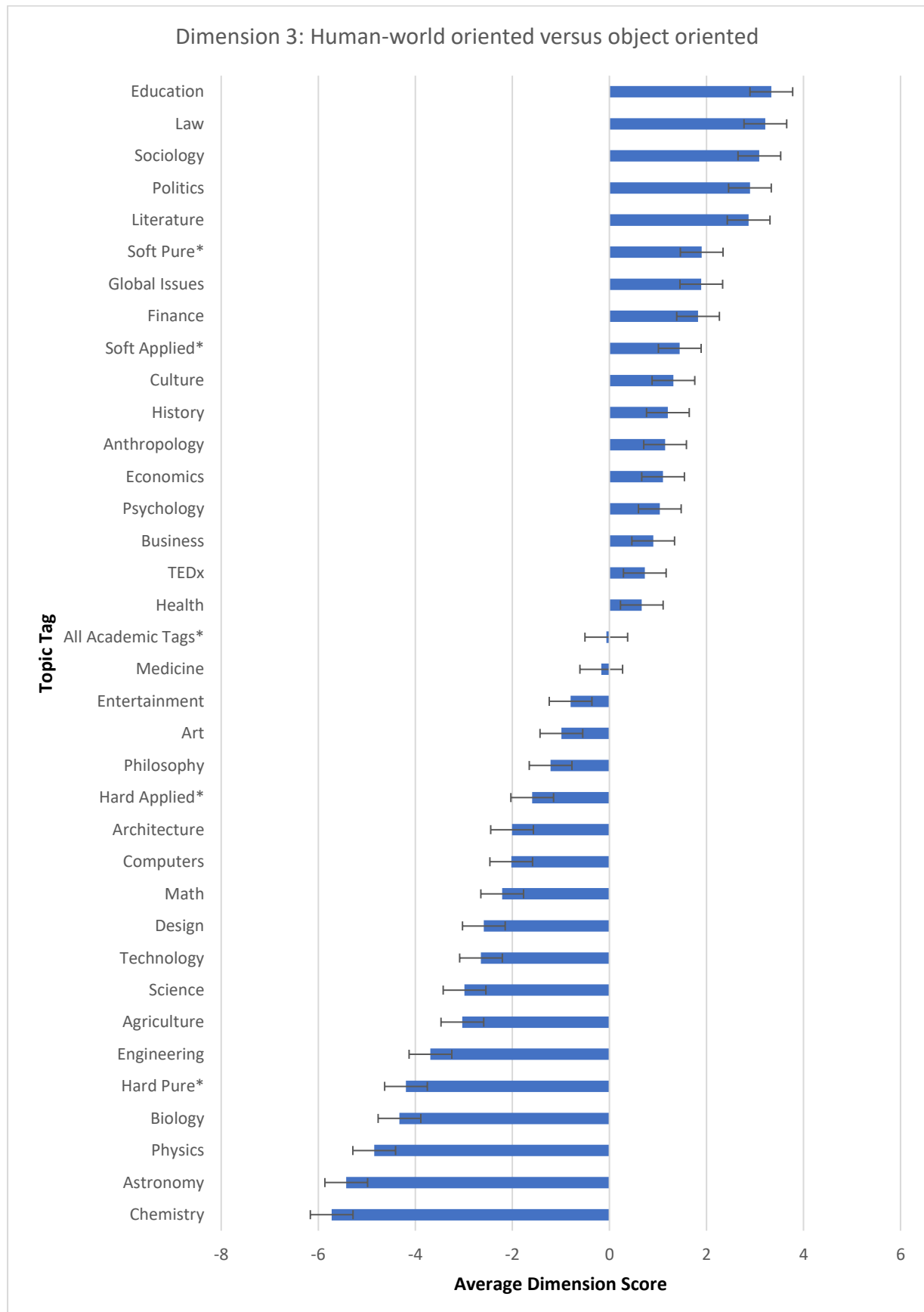
12. So let's take a deeper look. Let's look at **neurons**. So as I mentioned, there are 86 billion **neurons**. There are also these smaller **cells** as you'll see. These are support **cells** -- astrocytes glia. And the **nerves** themselves are the ones who are receiving input. They're storing **it**, they're processing **it**. Each **neuron** is connected via **synapses** to up to 10,000 other **neurons** in your **brain**. And each **neuron itself** is largely unique. (Allan Jones: *A map of the brain*)

13. Compassion is not defined in one form. There's no Indian compassion. There's no American compassion. **It** transcends nation, the gender, the age. Why? Because **it** is there in everybody. **It's** experienced by people occasionally. (Dayananda Saraswati: *The profound journey of compassion*)

In these extracts, the artwork, technology, biological phenomena, or concept is the focus of the talk, which the speakers refer to using pronoun *it*, concrete nouns, and technical nouns. The present tense is also used throughout these extracts to give factual information. The situational characteristics noted the broad range of topics in the TED talk corpus. This explains the presence of the current dimension, which differentiates talks by topic.

When the TTC's subregisters are mapped onto Dimension 3 (Figure 4), the human-world oriented topics appear at the top of the dimension, such as *education, law, and sociology*. At the opposite end we find more object oriented topics, such as *chemistry, astronomy, and physics*. Interestingly, the aggregated categories are clearly ordered: soft pure (1.9), soft applied (1.5), hard applied (-1.5), and hard pure (-4.2). This shows that the soft categories lean positive, the hard categories lean negative, the applied categories are closer to the mean, and the pure categories lean towards the extremes of the dimension.

Figure 4 TTC Subregisters mapped onto Dimension 3: Human-world oriented versus object oriented



Wingrove, P. & Crosthwaite, P. (2022). Multi-Dimensional Exploratory Factor Analysis of TED Talks. *Register Studies* (authors' pre-publication manuscript, main manuscript DOI: <https://doi.org/10.1075/rs.21008.win>)

* Topic tags marked with an asterisk represent aggregate categories.

4.5 Dimension 4 Subjective perspectives

Dimension 4 loads positively on private and mental verbs, subordinator *that*-deletion, verb + *wh*-clause and *that*-verb complements. There are no negatively loaded features, therefore this dimension identifies one type of discourse against all others. Taken together, the linguistic features function to state subjective perspectives and viewpoints. This is not to be confused with the speakers *only* giving their own perspective, as an analysis of the texts showed that the speakers often used these features to state the viewpoints of others as well as their own. One talk demonstrates this function in a particularly interesting way, as the speaker performs multiple different characters, beginning by taking on the personality of her Chinese-American friend:

14. Um, well, I, I want to start by **saying that** this is a very difficult experience for a Chinese-American. I don't **know what** to call myself now, because I have really my Chinese identity, but my kids, they are American-Chinese, but it's difficult to try to **express** myself in front of audience of people like this. But if had to give my opinion about meat, I **think** first, the most important thing is to **say that** we don't have to have perfect food, but maybe it can also not be poison. (Sarah Jones: *What does the future hold 11 characters offer quirky answers*)

In extract 14 the speaker takes on the personality of her friend and uses private verbs and mental verbs to give their perspective. The clauses help the speaker elaboration on information related to different perspectives. This includes an instance of *that*-deletion in the first sentence, “based on people (that) I really know”.

Dimension 4 can also be used to express subjective experiences. In the following example the speaker expresses their experience suffering with depression:

15. Well, I survived, and **that** just leaves me with my story, and my story is this: In four simple words, I **suffer** from depression. I **suffer** from depression, and for a long time, I **think**, I was living two totally different lives, **where** one person was always **afraid** of the other. I was **afraid that** people would **see** me for who I really was, **that** I wasn't the perfect, popular kid in high school everyone **thought** I was, **that** beneath my smile, there was struggle, and beneath my light, there was dark, and beneath my big personality just hid even bigger pain. (Kevin Brael: *Confessions of a depressed comic*)

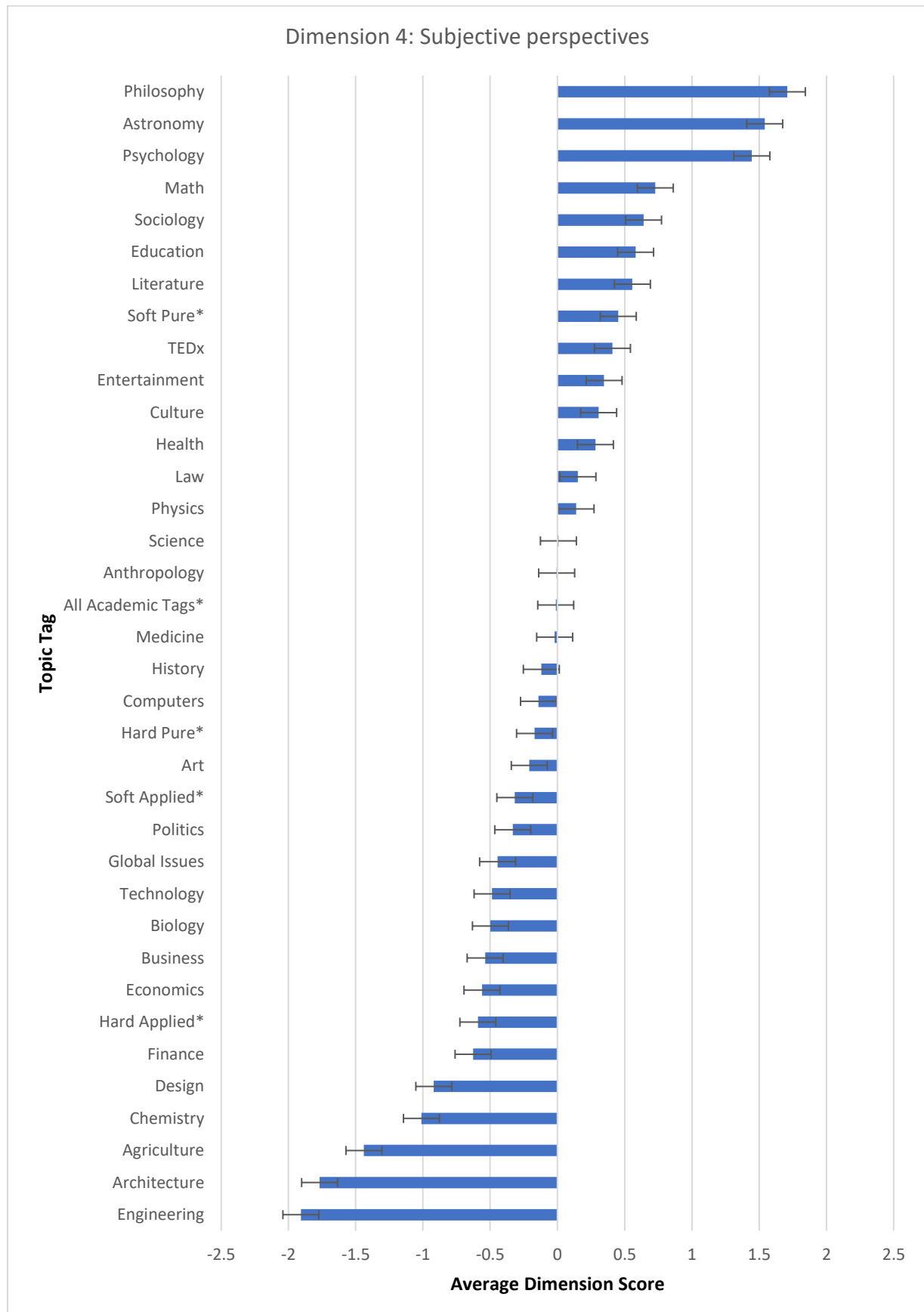
Again, mental verbs and private verbs are used to express thoughts and feelings and *that*-clauses enable them to elaborate. Note that in extract 15 their own thoughts are contrasted with the subjective experiences of others: “I was afraid that people would see me for who I really was .. I

Wingrove, P. & Crosthwaite, P. (2022). Multi-Dimensional Exploratory Factor Analysis of TED Talks. *Register Studies* (authors' pre-publication manuscript, main manuscript DOI: <https://doi.org/10.1075/rs.21008.win>)

wasn't the perfect, popular kid in high school everyone thought I was". It is interesting that the TED platform may be one of the few places where speakers feel comfortable expressing very personal experiences to an unfamiliar audience. This is likely due to the inclusive open-minded culture that TED has managed to cultivate and results in the identification of the current dimension.

When the TTC subregisters are mapped onto Dimension 4, philosophy stands out with the highest weighting. This suggests that expressions of perspectives are important in *philosophy* TED talks. Conversely, a number of hard applied disciplines appear at the bottom end of the spectrum, suggesting that expressions of subjective perspectives is uncommon in these areas. Arguably, this reflects the underlying epistemologies of the disciplines, where knowledge can be more open to interpretation in philosophy, compared to hard applied areas which aim for objectivity.

Figure 5 TTC Subregisters mapped onto Dimension 4: Subjective perspectives



Wingrove, P. & Crosthwaite, P. (2022). Multi-Dimensional Exploratory Factor Analysis of TED Talks. *Register Studies* (authors' pre-publication manuscript, main manuscript DOI: <https://doi.org/10.1075/rs.21008.win>)

* Topic tags marked with an asterisk represent aggregate categories.

Wingrove, P. & Crosthwaite, P. (2022). Multi-Dimensional Exploratory Factor Analysis of TED Talks. *Register Studies* (authors' pre-publication manuscript, main manuscript DOI: <https://doi.org/10.1075/rs.21008.win>)

4.6 Dimension 5 Persuasive Stance

Dimension 5 loads positively on stance adverbials, emphatics, hedges, and comparatives. Notably stance adverbials (+0.88) and emphatics (+0.96) are particularly highly weighted compared to hedges (+0.3) and comparatives (+0.3), which just meet the cut-off at +0.3 loading. When these features are taken together they function to develop stance in a particularly persuasive way.

This dimension appears to identify the tension between the speaker developing a nuanced stance and the desire to emphatically 'sell' the content of their talk in order to implicitly persuade the audience. For instance, consider the following talk where Sergey Brin presents Google Glass to the audience. This example contains two extracts from a single talk, where the speaker describes the audio and camera functions:

16. And also we wanted to free up the ears, so the sound **actually** goes through, conducts straight to the bones in your cranium, which is a little bit freaky at first, but you get used to it. And **ironically**, if you want to hear it **better**, you **actually just** cover your ear, which is **kind of** surprising, but that's how it works.

...

And the other **really** unexpected surprise was the camera. Our original prototypes didn't have cameras at all, but it's been **really** magical to be able to capture moments spent with my family, my kids. I **just never** would have dug out a camera or a phone or something else to take that moment. (Sergey Brin: *Why Google Glass*)

In the first part of example 16 the speaker balances overselling the technology with an attempt to sound more like an unbiased user. For instance, commenting on the experience being "kind of surprising". In the second part of example 16 the speaker veers more towards a sales pitch with the phrase "it's been really magical to be able to capture moments spent with my family, my kids". The reference to family is reminiscent of advertisements which aim to sell products by associating them with family, friendships, and relationships. In essence this talk functions as a tech demo where the speaker aims to develop a persuasive stance. In order to do this the speaker has to 'sell' the product but also maintain a position which appears unbiased and sincere.

Whilst this is quite common in technology demonstrations, the corpus also presents examples of persuasive stance being used to sell ideas. Consider the following extract from the talk entitled *Less stuff, more happiness*:

Wingrove, P. & Crosthwaite, P. (2022). Multi-Dimensional Exploratory Factor Analysis of TED Talks. *Register Studies* (authors' pre-publication manuscript, main manuscript DOI: <https://doi.org/10.1075/rs.21008.win>)

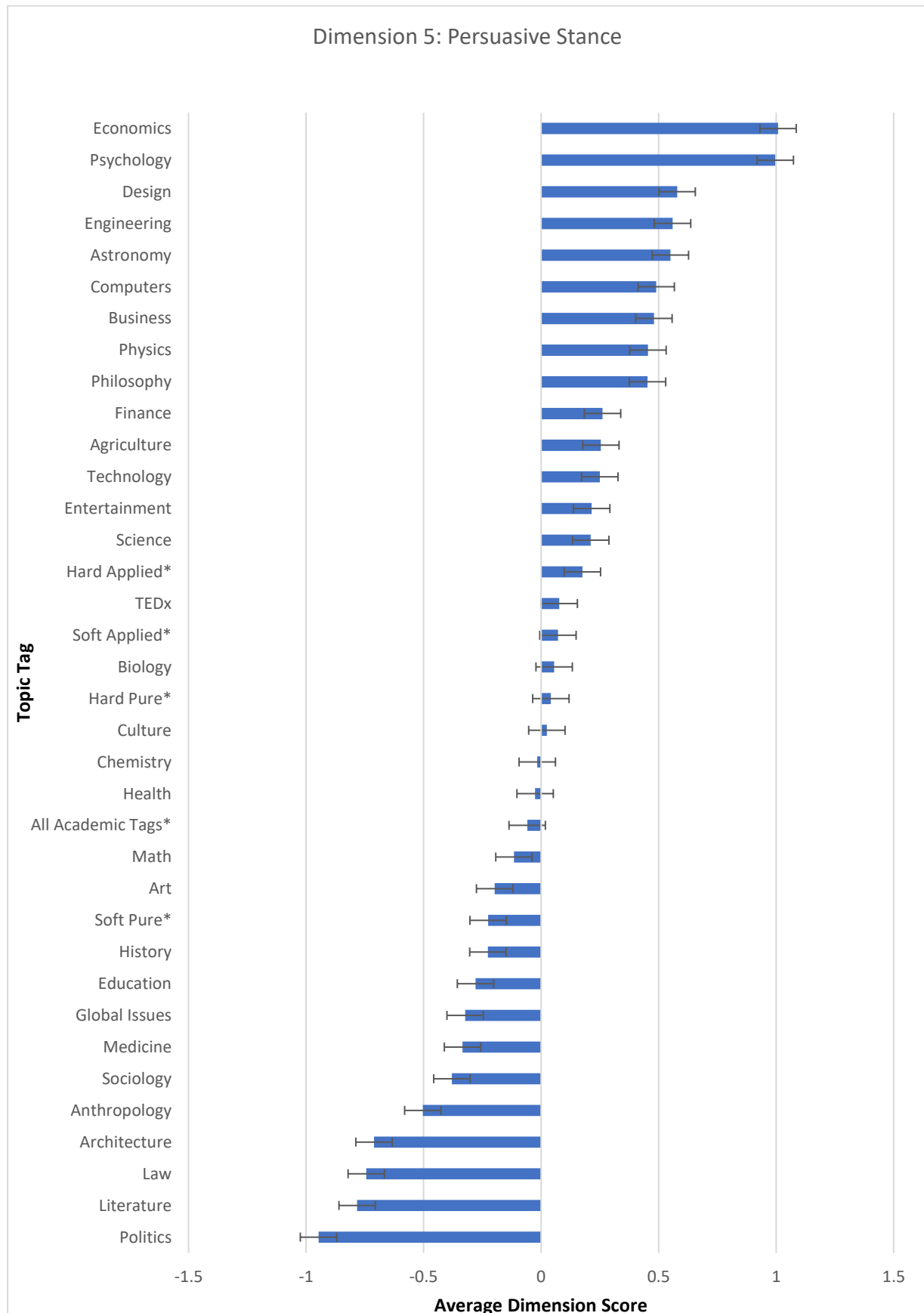
17. Well I'm here to suggest there's a better way, that less might **actually** equal **more**. I bet most of us have experienced at some point the joys of less: college -- in your dorm, traveling -- in a hotel room, camping -- rig up basically nothing, **maybe** a boat. Whatever it was for you, I bet that, among other things, this gave you a little **more** freedom, a little **more** time. So I'm going to suggest that less stuff and less space are going to equal a **smaller** footprint. It's **actually** a great way to save you some money. And it's going to give you a little **more** ease in your life. (Graham Hill: *Less stuff more happiness*)

In this talk the speaker aims to persuade the audience of the benefits of a minimalist lifestyle. They pre-empt the audience's reaction through the use of actually, "less might actually equal more ... it's actually a great way to save you some money", and uses comparatives to show the benefits of such as lifestyle, "more freedom ... more time ... a smaller footprint". This shows one of the ways in which idea oriented talks overlap with technology demonstrations, through the use of persuasive stance to sell the idea or technology.

The use of persuasive stance can be linked to the situational characteristics of TED talks. TED talks offer a platform with a general audience to speakers from a broad range of contexts. This gives TED speakers a unique opportunity to connect with a wide ranging audience. However, as TED talks are not infomercials, TED speakers have to present objects and ideas they wish to sell in a more subtle way, which includes stance adverbials and hedges as well as emphatics. This results in the use of persuasive stance to sell ideas without an overt sales pitch.

When the TTC subregisters are mapped onto Dimension 5, *design* and *engineering* talks have high loadings, associated with technology demonstrations. *Economics* and *psychology* also have high loadings which can be associated with presenting idea oriented discourse. On the other end of the spectrum we find humanities related subjects such as *politics*, *law*, and *literature*, which shows that talks associated with these disciplines are less likely to use persuasive stance, perhaps as they are less likely to have an idea or technology they wish to sell to the audience.

Figure 6 TTC Subregisters mapped onto Dimension 5: Persuasive Stance



Wingrove, P. & Crosthwaite, P. (2022). Multi-Dimensional Exploratory Factor Analysis of TED Talks. *Register Studies* (authors' pre-publication manuscript, main manuscript DOI: <https://doi.org/10.1075/rs.21008.win>)

* Topic tags marked with an asterisk represent aggregate categories.

4.7 Dimension 6 Expert elaboration

Dimension 6 loads positively on *that*-relative clauses on the objective and subject position, *that*-verb complements, demonstratives, first person pronouns plural, stance adverbials, and split auxiliaries. When these features are taken together, they appear to be used by speakers to establish an authorial expert self when elaborating on their research. Take the following example, which concerns multiple instances of this function within a single talk:

18. Two years ago here at TED I reported **that we** had discovered at Saturn, with the Cassini Spacecraft, an anomalously warm and geologically active region.

...

And I also reported **that we'd** made this mind-blowing discovery -- **this** once-in-a-lifetime discovery of towering jets erupting from **those** fractures at the south pole.

...

And at **that** time two years ago I mentioned **that we** were speculating **that these** jets might in fact be geysers.

...

So **we** are very encouraged by these results. And **we** are much more confident now than **we** were two years ago **that we** might indeed have on this moon, under the south pole, an environment or a zone that is hospitable to living organisms. (Carolyn Porco: *Could a Saturn moon harbor life?*)

In extract 18 the speaker uses 'we' to refer to herself and the expert community she belongs to. *That*-clauses are used to elaborate on key information that she is reporting to the audience. Demonstratives are also used to highlight key findings, such as "this once in a lifetime discovery", "those fractures", and "these jets".

Expert elaboration is evident in the following example:

19. But here **we** are on a planet, and **there** are so many things **that** are amazing about Earth **that we're** searching far and wide to find things **that** are like **that**. And when **we're** searching, **we're** finding amazing things.

...

Also, **we** see **that** it's the "red planet," and atmospheric escape plays a role in Mars being red. **That's** because **we** think Mars used to have a wetter past, and when water had enough energy, it broke up into hydrogen and oxygen, and hydrogen being so light, it escaped into

Wingrove, P. & Crosthwaite, P. (2022). Multi-Dimensional Exploratory Factor Analysis of TED Talks. *Register Studies* (authors' pre-publication manuscript, main manuscript DOI: <https://doi.org/10.1075/rs.21008.win>)

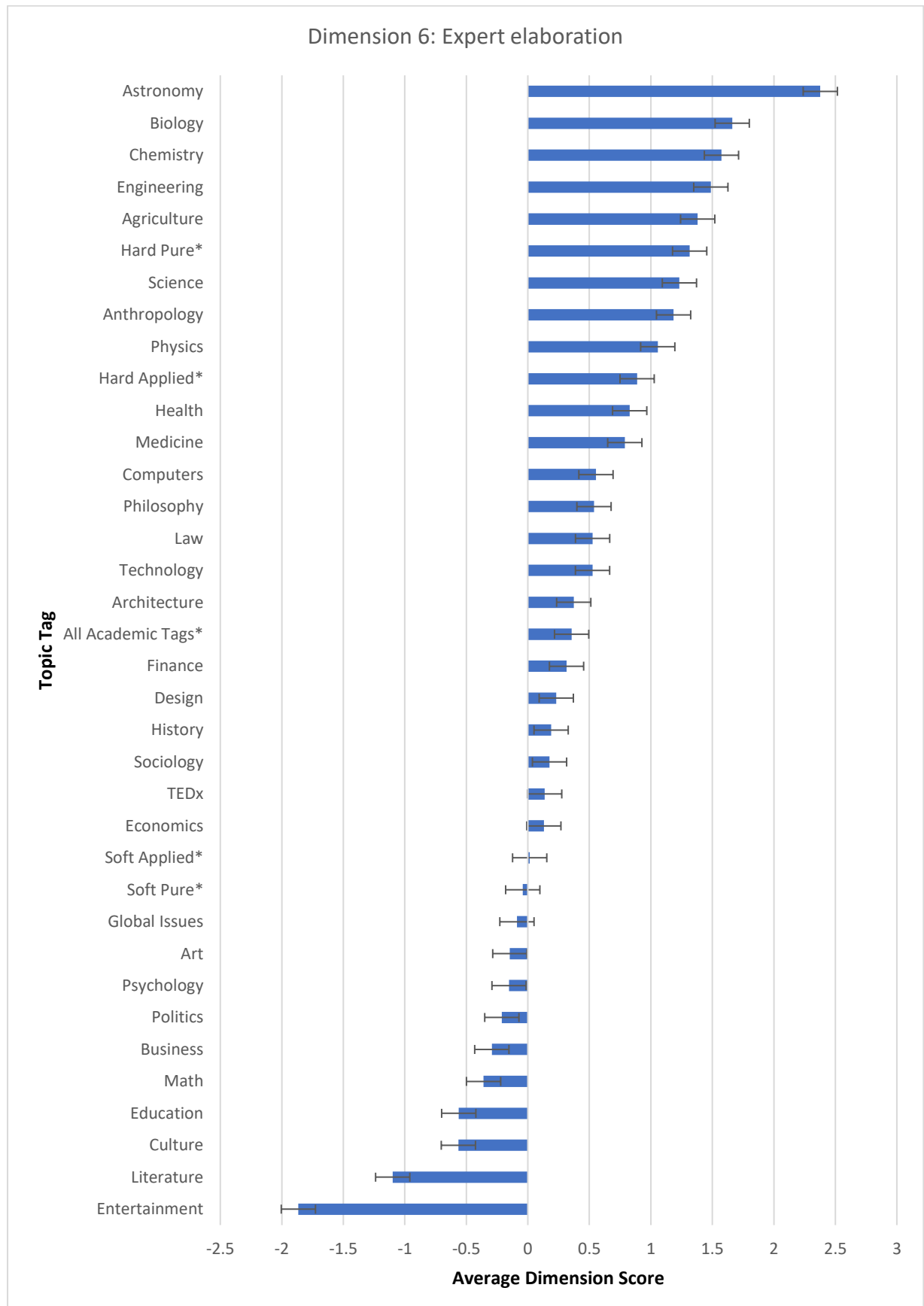
space, and the oxygen **that** was left oxidized or rusted the ground, making that familiar rusty red color **that we** see. (Anjali Tripathi: *Why Earth may someday look like Mars*)

In both these extracts the speakers are informing the public on the kinds of work they are doing in their field and communicating key findings. The use of 'we' is interesting in these extracts as it shows evidence of the speakers aiming to establish an authorial self to the audience. This function can be linked to the situational characteristics of the corpus concerning relationships between participants. As TED is a prestigious platform, whereby expert speakers communicate information to the general audience, the presence of expert elaboration appears to be appropriate in this context.

The opposite end of Dimension 6 concerns only one linguistic feature: second person pronouns. This is not substantial enough to warrant a contrasting dimension on its own, but suggests that talks high in expert elaboration refer to the audience much less directly, instead, preferring to refer themselves and their expert community; and talks without expert elaboration are more likely to directly reference the audience.

When the TTC subregisters are mapped onto Dimension 6, the hard sciences are at the top of the dimension. This suggests that hard science topics have the strongest tendency for this type of communication. This is likely due to the fact that these disciplines are more technical in nature and require the speaker to communicate key findings in this way. Conversely, more accessible topics appear lower down, such as within *entertainment, literature, and culture*.

Figure 7 TTC Subregisters mapped onto Dimension 6: Expert elaboration



Wingrove, P. & Crosthwaite, P. (2022). Multi-Dimensional Exploratory Factor Analysis of TED Talks. *Register Studies* (authors' pre-publication manuscript, main manuscript DOI: <https://doi.org/10.1075/rs.21008.win>)

* Topic tags marked with an asterisk represent aggregate categories.

4.8 Dimension 7 Change and Inspiration

Dimension 7 loads positively on infinitives, desire verbs + *to*-clauses, the continuous aspect, causation verbs, and suasive verbs. Taken together, these features are used to encourage change and inspire the audience. In some instances Dimension 7 is used to direct change in specific fields, such as in business management:

20. In the face of relentless competition, it may well be that you will **have to take** the painful decision to downsize the organization, just as you may **have to lose** some weight in order **to run** a marathon. But **losing** weight alone will not get you across the finish line with a winning time. **To win** you **need to go** all in. You **need to go** all in. Rather than just **cutting** costs, you **need to think** about initiatives that will enable you **to win** in the medium term, initiatives **to drive** growth, actions that will fundamentally change the way the company operates, and very importantly, investments **to develop** the leadership and the talent. (Jim Hemerling: *5 ways to lead in an era of constant change*).

In this extract the speaker is advising the audience how to lead a business in an era of constant change. In multiple instances they use infinitives and desire verbs + *to*-clauses in the pattern “you need to [verb]” to direct change in the audience. The continuous aspect is used to reference “losing weight” and “cutting costs” and a causation verb is used in “drive growth”. Together, all these features build a persuasive passage which directs change in the audience.

In other cases the aims are much broader, such as encouraging a global response to climate change:

21. We're **operating** outside the bounds of CO₂ concentrations that the planet has seen for hundreds of thousands of years. The larger point I'm trying to make is this. It's up to us **to look** at our homes and our communities, our vulnerabilities and our exposures to risk, **and to find** ways to not just survive, but **to thrive**, and it's up to us **to plan** and **to prepare** and **to call** on our government leaders and require them **to do** the same, even while they address the underlying causes of climate change. There are no quick fixes. There are no one-size-fits-all solutions. We're all **learning** by **doing**. (Vicki Arroyo: *Let's prepare for our new climate*)

In this passage the appeal is much broader as climate change is relatable to everyone. Notice how the speaker uses the continuous aspect to set up the current circumstances (“we’re operating”) and then uses infinitives to state the actions we should take, “to thrive ... to plan .. to prepare and to call”.

Wingrove, P. & Crosthwaite, P. (2022). Multi-Dimensional Exploratory Factor Analysis of TED Talks. *Register Studies* (authors' pre-publication manuscript, main manuscript DOI: <https://doi.org/10.1075/rs.21008.win>)

In this final passage the call to change is inspirational. Consider the following example where the speaker aims to inspire individuals in the audience with their own struggles:

22. The Arab states **are going through** tremendous change, and the struggles women face are overwhelming. Just like the women I photographed, I had **to overcome** many barriers to **becoming** the photographer I am today, many people along the way **telling** me what I can and cannot do. Umm El-Saad, Asma and Fayza, and many women across the Arab world, show that it is possible **to overcome** barriers to education, which they know is the best means to a better future. And here I would **like to** end with a quote by Yasmine, one of the four activist women I interviewed in Tunisia. Yasmine wrote, "Question your convictions. Be who you **want to** be, not who they want **you to** be. Don't accept their enslavement, for your mother birthed you free." (Laura Boushnak: *For these women reading is a daring act*)

In this passage the speaker uses infinitives and the continuous tense to describe how they overcame barriers and became the person they are. They end the passage with an inspirational quote which draws upon the desire verb + *to*-clause pattern.

The use of the current dimension can be linked to the situational characteristics of the corpus. Interestingly, the element of change and inspiration is clearly stated in the TED mission statement:

We believe passionately in the power of ideas to change attitudes, lives and, ultimately, the world. On TED.com, we're building a clearinghouse of free knowledge from the world's most inspired thinkers — and a community of curious souls to engage with ideas and each other.

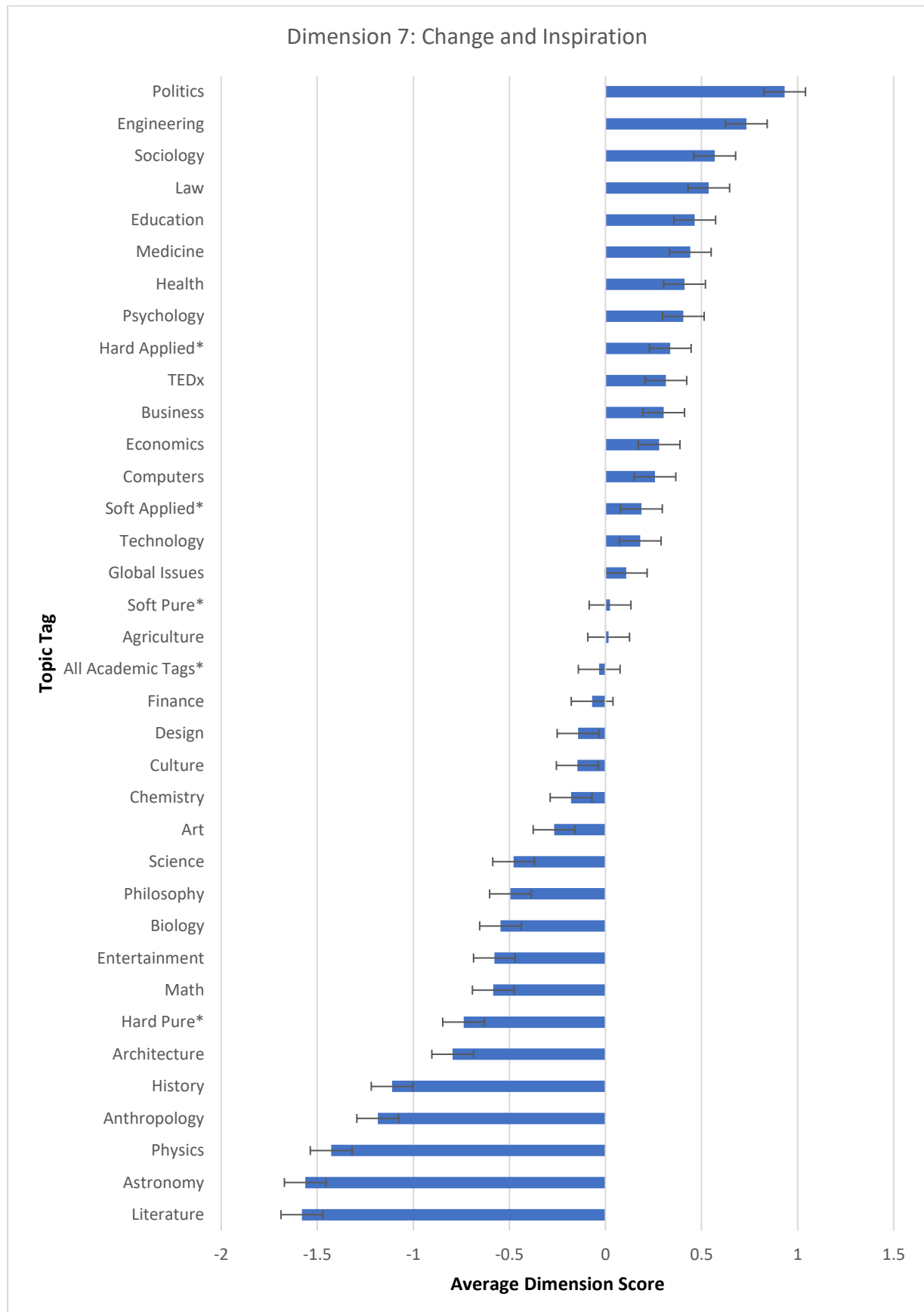
(<https://www.ted.com/about/our-organization>)

Dimension 7 captures talks which align themselves with this mission statement and in this sense is related to the communicative purposes of TED talks as a whole.

When the TTC subregisters are mapped onto Dimension 7, some social science disciplines can be seen at the top of the spectrum, such as *politics*, *sociology*, and *education*. These topics may be capturing talks related to encouraging change in the society. Interestingly, hard applied and soft applied are higher ranked than their pure counterparts. This may be due to the practical nature of this dimension, encouraging change in the real world. This explains why a discipline such as *engineering*, with a practical component, appears highly weighted, compared to literature, with the lowest score. *Engineering* may be using this dimension in a more specific functional manner to describe overcoming problems, whereas *literature* is less concerned about real world implications.

Wingrove, P. & Crosthwaite, P. (2022). Multi-Dimensional Exploratory Factor Analysis of TED Talks. *Register Studies* (authors' pre-publication manuscript, main manuscript DOI: <https://doi.org/10.1075/rs.21008.win>)

Figure 8 TTC Subregisters mapped onto Dimension 7: Change and Inspiration



Wingrove, P. & Crosthwaite, P. (2022). Multi-Dimensional Exploratory Factor Analysis of TED Talks. *Register Studies* (authors' pre-publication manuscript, main manuscript DOI: <https://doi.org/10.1075/rs.21008.win>)

* Topic tags marked with an asterisk represent aggregate categories.

5. Discussion and Conclusion

This paper aimed to investigate linguistic variation in a corpus of TED talks using MD analysis. The results reveal the latent functional dimensions of TED talks while identifying their distribution across the different topics comprising our corpus. These results are the first to characterise the linguistic variation of TED talks across the wide range of topics and subject areas that feature in such talks.

Regarding RQ1 (the linguistic dimensions of the TTC), our model revealed:

1. Spontaneous involved versus edited informational
2. Abstract informational versus narrative discourse
3. Human-world oriented versus object oriented
4. Subjective perspectives
5. Persuasive stance
6. Expert elaboration
7. Change and inspiration

These dimensions can be linked to the situational characteristics of TED talks and can be compared to MD models identified in prior research. In terms of situational characteristics, it was suggested that the production circumstances of TED talks resulted in the unique expression of Dimension 1, which captures a spectrum of spontaneous involved discourse versus more planned and edited texts. Moreover, the broad range of topics resulted in the human-world oriented versus object oriented dimension. The relationship between the expert speaker and the general audience can be connected to Dimensions 5 and 6, which capture the persuasive stance of the speaker and the use of expert elaboration. Finally, the presence of change and inspiration is intimately linked to TED's mission statement.

Compared to prior research, the current findings identify some dimensions reminiscent of dimensions in prior research, some dimensions which reinforce perspectives on TED talks, and some dimensions which are unique to the TED register. Beginning with familiar dimensions, spontaneous involved versus edited informational (Dimension 1), bears a close resemblance to Biber's (1988) Dimension 1, "involved versus informational discourse". Manifestations of this dimension have been captured in a variety of studies across different languages and times. In studies which capture a broad range of spoken and written registers this is sometimes framed as an oral vs literate divide, such as in a corpus of 18th century texts (Biber, 2001) and in Brazilian Portuguese (Sardinha, Kauffmann, & Acunzo, 2014), suggesting that this is typical of linguistic registers regardless of

Wingrove, P. & Crosthwaite, P. (2022). Multi-Dimensional Exploratory Factor Analysis of TED Talks. *Register Studies* (authors' pre-publication manuscript, main manuscript DOI: <https://doi.org/10.1075/rs.21008.win>)

language. Biber (2014), commenting on the persistence of this dimension across studies, characterises this kind of dimension as clausal/'oral' versus phrasal/'literate'. Interestingly, the current study identifies a more subtle version of this dimension, concerning variation within a single register, rather than across a broad range of spoken and written registers.

Two other familiar dimensions are subjective perspectives (Dimension 4) and persuasive stance (Dimension 5). Prior research has identified a number of dimensions related to developing opinions or perspectives (Hardy & Friginal, 2016; Bu, Connor-Linton, & Wang, 2020) and stances (Iberri-Shea, 2011; Garnder, Nesi, & Biber, 2018; Condi de Souza, 2014). It may be that stance is an integral (and perhaps even inseparable) component of registers which deal with informational content. In this sense, identifying perspectives and stance dimensions in themselves is not particularly surprising. However, it is interesting to note the specific expression of these dimensions within the TTC. In the case of subjective perspectives, this dimension was notably personal, with speakers sharing their thoughts and feelings with the audience; and in the case of stance, the persuasive nature was reflective of the speakers aiming to sell their idea or technology to the audience.

The current paper reinforces findings from prior research on TED talks. For instance, expert elaboration (Dimension 6), reinforces and expands some of our knowledge about TED talks as a means of science communication. Prior research has identified some TED talks as a form of popularisation discourse, where experts communicate their findings to the public (Mattiello, 2017). Expert elaboration identifies a way in which TED speakers establish an authorial expert self when elaborating on research. This contrasts with prior research on TED talks which investigated the use of the pronoun "we", which found that 57.90% of *we* forms referred to the speaker and audience simultaneously, to build a sense of commonality; and 39.90% referred to the speaker and other non-present agents, to reference their belonging to an academic or research community (Di Carlo, 2018). Taken together, there is evidence that TED speakers aim to establish an authorial expert self *and* create a sense of closeness with the audience. It appears that there may be some tension between these goals in terms of social positioning, which could be of interest to further research.

Whilst the dimensions identified in the current study do in many cases signal the similarity of TED talks to other genres and registers, some dimensions reflect the unique nature of TED talks. For instance, change and Inspiration (Dimension 7), is not comparable to any dimension from any prior study and appears to be identifying a unique aspect of the TED talk register. This dimension can be linked to TED's mission statement which believes in the power of ideas to change the world. This is only possible due to the presence of an online audience on TED's globally reaching platform. It

Wingrove, P. & Crosthwaite, P. (2022). Multi-Dimensional Exploratory Factor Analysis of TED Talks. *Register Studies* (authors' pre-publication manuscript, main manuscript DOI: <https://doi.org/10.1075/rs.21008.win>)

would be interesting to see whether similar dimensions are also present in registers which reach a expansive audience, such as political speeches. Moreover, abstract informational discourse versus narrative discourse (Dimension 2), bears resemblance to some dimensions in prior studies, which have identified narrative dimensions (Grieve, 2014) and abstract dimensions (Biber, 1988). However, the current finding is unique in identifying them both across a single dimension, differentiating more static abstract informational discourse against more dynamic action-oriented narration. It is worth noting that like the clausal/'oral' versus phrasal/'literate' divide, Biber (2014) also considers narrative versus non-narrative discourse a near universal dimension of register variation.

Regarding RQ2 (concerning the representation of the EFA dimensions across topics), many of the dimensions differentiated topics in interesting ways. Most notably, human-world oriented versus object oriented (Dimension 3) differentiated texts into familiar academic divisions, with soft categories leaning positive, hard categories leaning negative, applied categories closer to the mean, and pure categories at the extremes. This similarity with academic texts supports the position that TED talks are an academic-adjacent register. Other dimensions identified familiar disciplinary groupings, such as Dimension 2 grouping soft disciplines such as *psychology*, *economics*, and *politics* as particularly abstract, Dimension 6 grouping hard science disciplines as high in expert elaboration, and Dimension 7 finding social science disciplines such as *politics*, *sociology*, and *education* as highest in change and inspiration.

One of the strengths of the current model is that subregisters can be characterised in terms of the dimensions. This is shown visually in Figures 9 – 12, which show the profile of individual disciplinary areas in terms of their dimension scores. For instance, in Figure 9, we can see that all the hard-pure disciplines have a similar profile by their shape: typically object oriented (D3), high in subjective perspectives (D4), high in persuasive stance (D5) , high in expert elaboration (D6) and low in change and inspiration (D7). These shapes are comparable to hard applied areas shown in Figure 10, showing similarity between all hard areas, although with some applied areas being less object oriented (D3).

Figures 11 and 12 show soft pure and soft applied disciplines. In Figure 11, the outliers within the soft pure category are more clearly visible, for instance, *philosophy* appears as an outlier with a high D1 and D2 scores and a low D3 score, actually appearing more similar to typical hard pure disciplines, interestingly. This begins to present some evidence that philosophy may be an outlier in the humanities at least linguistically. Other outliers are also observable, such as *medicine* in Figure

Wingrove, P. & Crosthwaite, P. (2022). Multi-Dimensional Exploratory Factor Analysis of TED Talks. *Register Studies* (authors' pre-publication manuscript, main manuscript DOI: <https://doi.org/10.1075/rs.21008.win>)

10, with a higher D3 score and lower D2 score; and *architecture* in Figure 12 with much lower D2, D3, and D4 scores.

Figure 9 Hard Pure Disciplines Mapped onto Dimensions

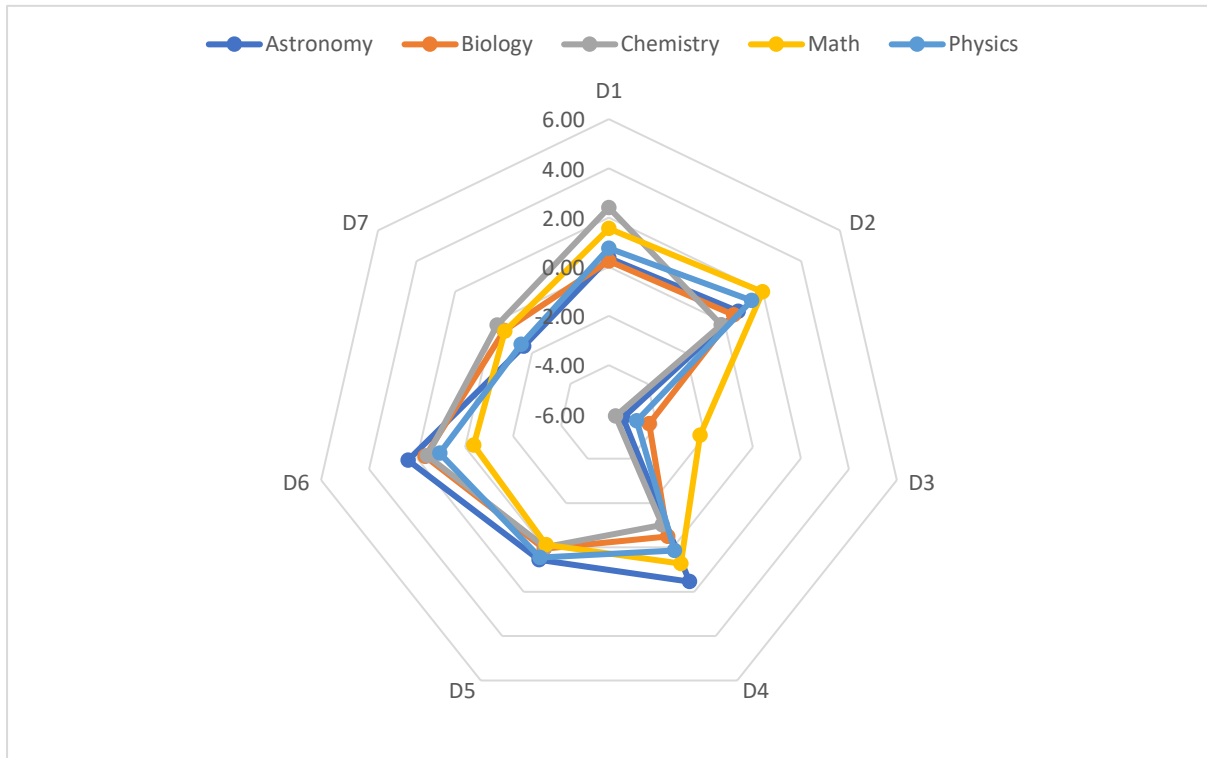


Figure 10 Hard Applied Disciplines mapped onto Dimensions

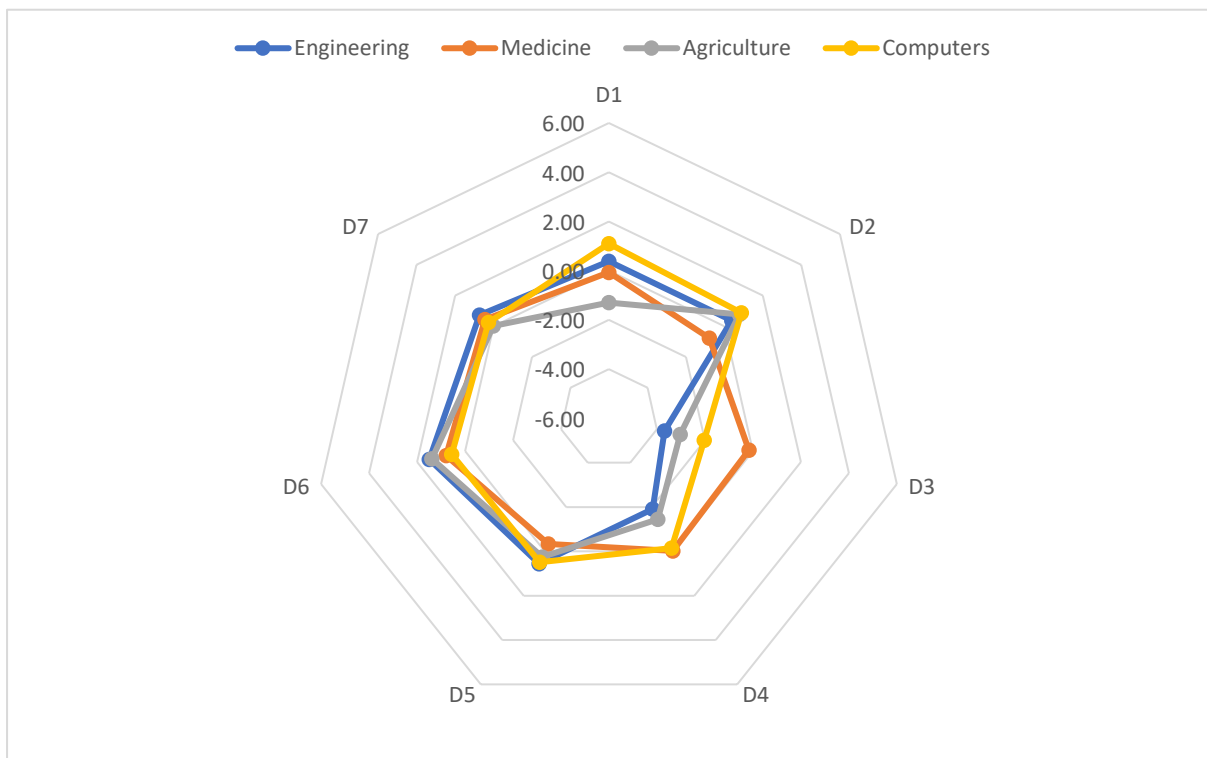


Figure 11 Soft Pure Disciplines Mapped onto Dimensions

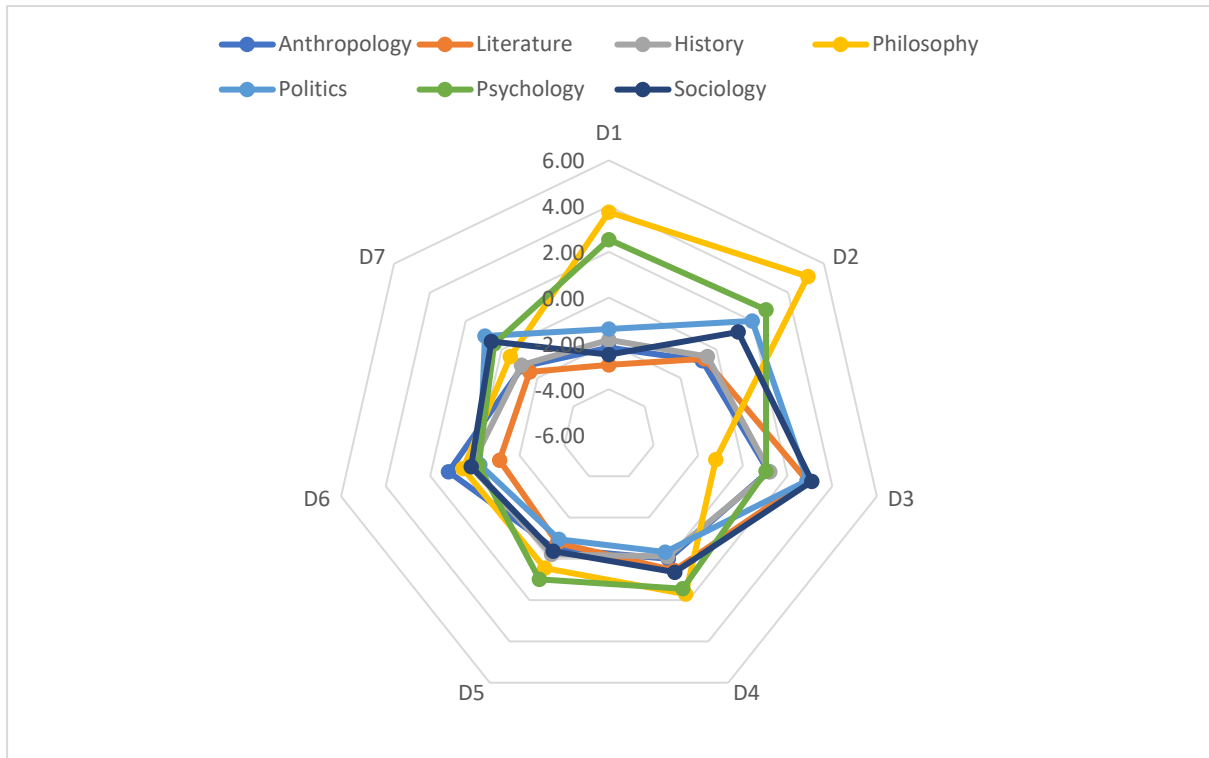
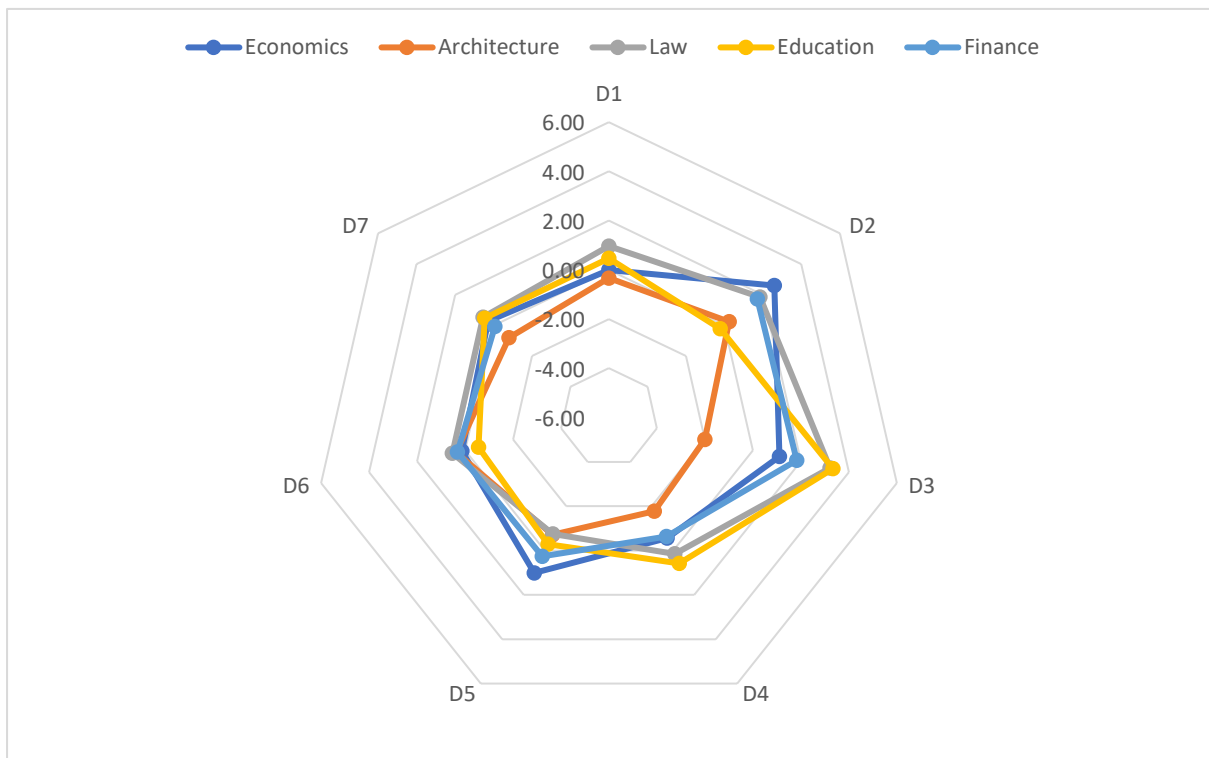


Figure 12 Soft Applied Disciplines Mapped onto Dimensions



Wingrove, P. & Crosthwaite, P. (2022). Multi-Dimensional Exploratory Factor Analysis of TED Talks. *Register Studies* (authors' pre-publication manuscript, main manuscript DOI: <https://doi.org/10.1075/rs.21008.win>)

In terms of the study's limitations, while investigating variation across a large number of topics and subject areas was one of the study's main aims and strengths, the issue of overlapping categories naturally presents some fuzziness within the data. This means that the clustering of similar registers on Figures 2 – 7 may be influenced by overlapping talks within these categories. However, this fuzziness is reflective of complexity within the data and the use of multiple tags provides a more accurate description of the content. The alternative would be to allocate talks to distinct sets, which would require removing tags and diverging from TED's classification system. Moreover, the use of multiple tags can be used to identify talks which combine topic areas. For instance, *Animations of Unseeable Biology* by Drew Barry has *art*, *science*, and *biology* tags and concerns scientifically accurate animations of cell biology.

A further limitation concerns the application of corpus methods to multimodal genres. The current study investigated linguistic variables only and therefore multimodal and paralinguistic features such as speech rate, intonation, gesture, accent, and so on, were not considered. This also includes filled pauses such as "um" and "er", which were not included in the transcripts. Future studies may wish to investigate TED talks in terms of a broader range of multimodal, paralinguistic, and linguistic features not afforded by orthographic transcription in the current study. It would be interesting to see how speakers draw upon multimodal and paralinguistic affordances to deliver information, engage the audience, and frame the content.

The current study has implications for the study of TED talks. Linguistic variation within TED talks was identified based upon statistical association. Notably, the subregisters varied in ways reminiscent of academic texts, which are often viewed as a separate disciplinary discourses (Hyland, 2007). As such, perhaps it would be better to view TED talks as windows into an array of discourses, rather than a monolithic whole. This emphasises the importance of accounting for topic variation when analysing TED talks in future research.

The current study contributes to the field of register studies by furthering our knowledge of a unique contemporary digital register, which brings together many topics and speakers into a single space. It is interesting to note that a disciplinary continuum (Dimension 3) and persuasive stance (Dimension 5) both appear within an academic-adjacent corpus, suggesting that these features are integral to academic disciplines, even outside an academic context. Moreover, the unique speaker-audience relationship results in persuasive stance (Dimension 5), expert elaboration (Dimension 6) and change and inspiration (Dimension 7). It would be interesting to see the extent to which, if at all, these

Wingrove, P. & Crosthwaite, P. (2022). Multi-Dimensional Exploratory Factor Analysis of TED Talks. *Register Studies* (authors' pre-publication manuscript, main manuscript DOI: <https://doi.org/10.1075/rs.21008.win>)

dimensions are present in comparable registers, such as sales pitches, political speeches, and new digital genres which will inevitably continue to emerge and evolve.

References

- Abdulrahman, T. (2017). TED talks as listening teaching strategy in EAP classroom. *Asian EFL Journal*, 1, 72–93.
- Anderson, C. (2016). *TED Talks: The Official TED Guide to Public Speaking*. London: Nicholas Brealy Publishing.
- Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/cbo9780511621024>
- Biber, D. (2001). Dimensions of variation among eighteenth-century speech-based and written registers. In Conrad, S. & D. Biber (Eds.), *Variation in English: multi-dimensional studies*. Harlow: Longman. <https://doi.org/10.4324/9781315840888>
- Biber, D. (2006). *Studies in corpus linguistics: University language* (Vol. 23). Netherlands: John Benjamins. <https://doi.org/10.1075/scl.23>
- Biber, D. (2014). Using multi-dimensional analysis to explore cross-linguistic universals of register variation. *Languages in contrast*, 14(1), 7-34. <https://doi.org/10.1075/lic.14.1.02bib>
- Biber, D., & Conrad, S. (2009). *Register, Genre, and Style*. Cambridge University Press. <http://doi.org/10.1017/CBO9780511814358>
- Biber, D., Reppen, R., Byrd, P., Helt, M., Clark, V., Cortes, V., Csomay, E., & Urzua, A. (2004). *Representing Language Use in the University: Analysis of the TOEFL 2000 Spoken and Written Academic Language Corpus*. TOEFL Monograph Series.
- Biglan, A. (1973). Relationships between subject matter characteristics and the structure and output of university departments. *Journal of Applied Psychology*, 57(3), 204–13. American Psychological Association. <https://doi.org/10.1037/h0034699>
- Brezina, V., Timperley, M., & McEnery, T. (2018). #LancsBox v. 4.x [software]. Available at: <http://corpora.lancs.ac.uk/lancsbox>.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing Structural Equation Models* (pp. 136-162). Newbury Park, CA: Sage. <https://doi.org/10.1177/0049124192021002005>
- Bu, H., Connor-Linton, J., & Wang, L. (2020). Linguistic variation in the discourse of corporate annual reports: A multi-dimensional analysis. *Discourse studies*, 22(6), 647-677. <https://doi.org/10.1177/1461445620928231>
- Cassidy, R. S., Mike, T., Vincent, L., Andrew, T., Philippe, M., & Benoit, M. (2013). Scientists popularizing science: characteristics and impact of TED talk presenters. *PLoS ONE*, 8(4), e62403. <https://doi.org/10.1371/journal.pone.0062403>

Wingrove, P. & Crosthwaite, P. (2022). Multi-Dimensional Exploratory Factor Analysis of TED Talks. *Register Studies* (authors' pre-publication manuscript, main manuscript DOI: <https://doi.org/10.1075/rs.21008.win>)

- Condi de Souza, R. (2014). Dimensions of variation in *TIME* magazine. In T.B. Sardina, M.V. Pinto, & D. Biber (Eds.) *Multi-dimensional analysis, 25 years on: a tribute to Douglas Biber. Multi-dimensional analysis, twenty-five years on*. Amsterdam: John Benjamins Publishing Company.
<https://doi.org/10.1075/scl.60>
- Crosthwaite, P., & Cheung, L. (2019). *Learning the language of dentistry: disciplinary corpora in the teaching of English for specific academic purposes*. Amsterdam: John Benjamins Publishing Company.
<https://doi.org/10.1075/scl.93>
- Di Carlo, G. S. (2015). Stance in TED talks: Strategic use of subjective adjectives in online popularisation. *Ibérica: Revista de la Asociación Europea de Lenguas para Fines Específicos (AELFE)*, 29, 201-222.
- Di Carlo, G. S. (2018). Patterns of clusivity in TED Talks: When 'you' and 'I' become 'we.' *Ibérica*, 35, 119–144. <https://doaj.org/article/969b60f26c5c416aafae72d2c644133e>
- Egbert, J. (2014). Student perceptions of stylistic variation in introductory university textbooks. *Linguistics and Education*, 25, 64–77. Elsevier Inc. <https://doi.org/10.1016/j.linged.2013.09.007>
- Elk, C. (2014). Beyond mere listening comprehension: Using TED talks and metacognitive activities to encourage awareness of errors. *International Journal of Innovation in English Language Teaching and Research*, 3(2), 215-230,246. <http://search.proquest.com/docview/1655287119/>
- Felices Lago, Á. (1997). The integration of the axiological classeme in an adjectival lexicon based on functional-lexematic principles. In C. S. Butler, J. H. Connolly, R. A. Gatward & R. M. Vismans (Eds.), *A Fund of Ideas: Recent Developments in Functional Grammar*. Amsterdam: IFOT
- Gallo, C. (2015). *Talk like Ted*. St. Martin's Griffin.
- Gardner, S., Nesi, H., & Biber, D. (2018). Discipline, Level, Genre: Integrating Situational Perspectives in a New MD Analysis of University Student Writing. *Applied linguistics*, 40(4), 646–74. Oxford University Press (OUP). <https://doi.org/10.1093/applin/amy005>
- Gotti, M. (2014). Reformulation and recontextualization in popularization discourse. *Ibérica*. 27, 15-34.
- Grieve, J. (2014). A Multi-Dimensional analysis of regional variation in American English. In T.B. Sardina, M.V. Pinto, & D. Biber (Eds.), *Multi-dimensional analysis, 25 years on: a tribute to Douglas Biber. Multi-dimensional analysis, twenty-five years on*. Amsterdam : John Benjamins Publishing Company. <https://doi.org/10.1075/scl.60>
- Groom, N. (2005). Pattern and meaning across genres and disciplines: An exploratory study. *Journal of English for Academic Purposes*, 4(3), 257-277. <https://doi.org/10.1016/j.jeap.2005.03.002>
- Hardy, J. A., & Friginal, E. (2016). Genre variation in student writing: A multi-dimensional analysis. *Journal of English for Academic Purposes*, 22, 119–31. Elsevier Ltd.
<https://doi.org/10.1016/j.jeap.2016.03.002>
- Hyland, K. (2005). Stance and engagement: a model of interaction in academic discourse. *Discourse Studies*, 7(2), 173–92. London, Thousand Oaks, CA and New Delhi.
<https://doi.org/10.1177/1461445605050365>

Wingrove, P. & Crosthwaite, P. (2022). Multi-Dimensional Exploratory Factor Analysis of TED Talks. *Register Studies* (authors' pre-publication manuscript, main manuscript DOI: <https://doi.org/10.1075/rs.21008.win>)

- Hyland, K. (2007). *Disciplinary Discourses: Social Interactions in Academic Writing*. Michigan: University of Michigan Press.
- Iberri-Shea, G. (2011). Speaking in front of the class: a multi-dimensional comparison of university student public speech and university language. *Classroom Discourse*, 2(2), 251–267. <http://doi.org/10.1080/19463014.2011.614061>
- JASP Team (2020). JASP (Version 0.14.1)[Computer software].
- Karia, A. (2015). *TED Talks Storytelling: 23 Storytelling Techniques from the BEST TED Talks*. AkashKaria.com.
- Kelly, P. (2016). *This is That* [TV Programme] CBC, Available online: https://www.youtube.com/watch?v=_ZBKX-6Gz6A
- Mattiello, E. (2017). The popularisation of science via TED talks. *International Journal of Language Studies*, 11(4), 77–106.
- Nini, A. (2019). The Multi-Dimensional Analysis Tagger. In Berber Sardinha, T. & Veirano Pinto M. (eds), *Multi-Dimensional Analysis: Research Methods and Current Issues* (pp. 67-94). London; New York: Bloomsbury Academic. <https://doi.org/10.5040/9781350023857.0012>
- Nurmukhamedov, U. (2017). Lexical Coverage of TED Talks: Implications for Vocabulary Instruction. *TESOL Journal*, 8(4), 768-790. <https://doi.org/10.1002/tesj.323>
- Oliver, J. (2016). *Last week tonight*. S03E11 [TV programme]. HBO.
- Reppen, R. (2001). Register variation in student and adult speech and writing. In S. Conrad, & D. Biber (Eds.), *Variation in English: multi-dimensional studies*. Harlow: Longman. <https://doi.org/10.4324/9781315840888>
- Sardinha, T. B., Kauffmann, C., & Acunzo, C. M. (2014) Dimensions of register variation in Brazilian Portuguese. In T.B. Sardinha & M.V. Pinto, (Eds), *Multi-Dimensional Analysis, 25 years on: A tribute to Douglas Biber*, John Benjamins Publishing Company. <https://doi.org/10.1075/scl.60>
- Schmid, H. (1994). Probabilistic Part-of-Speech Tagging Using Decision Trees. *Proceedings of International Conference on New Methods in Language Processing*, Manchester, UK.
- Takaesu, A. (2014). TED Talks as an Extensive Listening Resource for EAP Students. *Language Education in Asia*, 4. <https://doi.org/10.5746/LEiA/13/V4/I2/A05/Takaesu>
- Ted.com. (n.d). Our Organisation: History of TED . Available at: <https://www.ted.com/about/our-organization/history-of-ted> Accessed 1 Jan. 2021.
- Thompson, P., Hunston, S., Murakami, A., & Vajn, D. (2017). Multi-Dimensional Analysis, text constellations, and interdisciplinary discourse. *International Journal Of Corpus Linguistics*, 22(2), 153–86. John Benjamins Publishing Co. <https://doi.org/10.1075/ijcl.22.2.01tho>
- Wingrove, P. (2017). How suitable are TED talks for academic listening?. *Journal of English for academic purposes*, 30, 79–95. Elsevier Ltd. <https://doi.org/10.1016/j.jeap.2017.10.010>