Yeah, but how? Operationalising the functions of the discourse-pragmatic marker *yeah*

Ben Gibb-Reid[1] & Chloé Diskin-Holdaway[2]

[1]University of York, [2]University of Melbourne

ben.gibb-reid@york.ac.uk | chloe.diskinholdaway@unimelb.edu.au

Abstract

Discourse-pragmatic markers (DPMs) such as *yeah* are known for their multifunctionality, leading to challenges in coding for function in quantitative analyses. We provide an account of the incremental creation of a functional taxonomy for the DPM *yeah* using a decision tree and multi-level inter-rater reliability testing. Author 1 and rater 1 initially achieved a low agreement rate of 51%, but upon consultation and adaptation of the taxonomy, came to an agreement of 95%. Rater 2 used the revised taxonomy and achieved an agreement rate of 81% (for primary [P] functions) and 91% (for P and secondary [S] functions). Rater 3 achieved an agreement rate of 56% (P) and 67% (P+S) with author 1+rater 1, but a higher agreement rate with rater 2 of 74% (P) and 83% (P+S). We draw on top-down and bottom-up, data-driven approaches, and discuss the future of DPM operationalisation.

Keywords discourse-pragmatic variation, yeah, inter-rater reliability, decision trees

1. Introduction

Progress has been made in describing how discourse pragmatic variables (DPVs) can be analysed quantitatively, and there are accepted methods of how to include them within the envelope of variation (see Pichler [2010; 2013]). However, operationalising the specific functions of a DPV remains an issue. Studies usually provide a taxonomy of a DPV's functions (see e.g. Tagliamonte 2016; Pichler 2016), but rarely discuss how function was coded beyond references to previous literature, suggesting they have been top-down (see e.g. Diskin, [2017: 145]). To conduct systematic and comparable analyses of DPVs, it is imperative to operationalise their functions via a transparent process, which we offer here by way of a decision tree that can be replicated by other researchers. Data quantification of DPVs (where the DPV is a predictor variable) can allow for investigations into how their functions vary and change synchronically and diachronically, as well as across varieties – an enterprise broadly in line with a variationist sociolinguistic agenda.

Creating a taxonomy is a process, but with a few notable exceptions (Eiswirth 2022a, 2022b; Wagner et al. 2015), this is one that remains undisclosed, making studies hard to replicate. This paper provides an account of the incremental creation of a taxonomy for the discourse-pragmatic marker (DPM) *yeah* utilising a decision tree and multi-level inter-rater reliability testing. In doing so, we encourage future research to resist relying on top-down approaches but instead create bottom-up taxonomies which are validated by inter-rater reliability testing.

2. Previous literature

2.1 Previous literature on yeah

Drummond and Hopper (1993a) highlight three salient functions of *yeah*: firstly, *yeah* as a recipient continuer or backchannel (1); all examples taken from our dataset)[1]. Secondly, *yeah* as an acknowledgement or assessment token (2). Thirdly, *yeah* as a response to a yes/no question (3).

(1) letter_from_the_courts – minimal turn-final backchannel *yeah* (derbyFYN01-FYN02-02.trs)

```
     FYN01: oh I had another letter from the courts again
     FYN02: what for?
     FYN01: the same letter you know about Ron
   → FYN02: oh yeah?
     FYN01: I've had the same letter six times now
```

(2) cabin_bed – full turn-initial assessment *yeah* (derbyFYP01-FYP02-06.trs)

```
     FYP01: well I dreamt that my kid had got this [ cabin bed  ]
   → FYP02:                                        [yeah but you]
            haven't got a kid Poppy <laugh>
     FYP01: no <sigh> right I dreamt that I'd got a kid
```

(3) bride_and_groom – full turn-initial response *yeah* (derbyFEN01-FEN02-MEN01-03.trs)

```
     FEN01: do you remember when me and Elsie Wheeler were
            bride and groom?
     FEN02: y- yeah (.) in what was it?
```

Many prior studies into DPM *yeah* have utilised its structural (turn) position as an indicator of its function. Drummond and Hopper (1993b: 166–167), for example, propose three levels of incipiency for *yeah*, which they infer from its turn position: i) full (entirely incipient), ii) minimal (recipient, where a token is within a brief turn e.g. *oh yeah*), iii) freestanding (entirely recipient; where a speaker follows a token by "relinquish[ing] the floor without further utterance"). They propose that full and minimal tokens can occur turn-initially, medially or finally, and freestanding tokens occupy a full turn. In Drummond and Hopper (1993a), full-turn tokens of *yeah* are split into functions which are either primarily topic-changing (4), elaborating, or responding to next-turn repair initiators. However, the authors highlight the ambiguity around whether freestanding *yeah* is acknowledging, agreeing, or both (1993b: 207); (returned to in Section 4.1). This separates cases where *yeah* is used in a bid for speakership (incipient *yeah*) from cases where it is part of a passively recipient turn (recipient *yeah*). Since turn position cannot be used as a single indicator of a function (i.e. turn types can be ambiguous, or multifunctional), other work has moved away from focusing exclusively on these aspects. Turn position itself can also be difficult to define, as it is determined somewhat by speaker

[1]As Eiswirth (2022b) points out, there are many different labels for the action of showing listenership or acknowledging a prior turn. As we rely on Drummond and Hopper (1993a; 1993b), we use the term 'backchannel'.

prosody. As a consequence, turn position is not the primary focus of this study, but we do draw on work that considers this crossover to make our initial top-down functional taxonomy.

(4) packed_up_smoking – full turn-initial new topic *yeah* (derbyFYN01-FYN02-02.trs)

```
    W:     have you packed up smoking?
    FYN01: mm hmm
    W:     yeah(1.3)
    FYN02: y[e::s]
    FYN01:   [got] to I've been told
    W:     why? (1.5)
    FYN01: reasons
    W:     oh <laughs> oh (.) [very <breath>]
    FYN02:                    [cos what?]
    W:     <laugh> [reasons <laugh>]
    FYN01:         [I said  reasons]
    FYN02: oh reasons (0.8)
    W:     <laugh>
    FYN02: I though you said <a recents>
    W:     <laughs> (2.0)
→   FYN01: yeah my dad's not talking to me I came up tonight to turn
           round to my mum <and> says I want to pack night school in cos
           I can't cope with it
```

Fuller (2003: 37–38) finds that yeah is used as a confirmation marker, to introduce a new topic (4) and as a backchanneling device that "doesn't necessarily mark agreement" (1). Jucker and Smith (1998: 179) find that yeah is used to "acknowledge the receipt of information that is new to the discourse", as well as to bring attention to information, echo acknowledgement and confirm questioned information. Furthermore, there are analyses of yeah as an invariant tag: Stenström et al. (2002: 173–174) find that yeah checks whether immediately preceding referents are understood and Pichler (2021: 386) proposes that utterance-final tag yeah "requests that listeners accept […] information as the common ground". Note that the studies which found invariant tag yeah are based on London English, whereas the rest (Drummond and Hopper 1993a; Fuller 2003; Jucker and Smith 1998) are based on American English data. Additionally, many of these studies used face-to-face interview and conversation data (Fuller, 2003; Jucker and Smith, 1998; Pichler, 2021; Stenström et al., 2002), whereas Drummond and Hopper's (1993a; 1993b) studies were based on telephone calls. These discrepancies may account for differences in the functional taxonomies, but also highlight a need for taxonomies that can be used for any variety or speech type; something we propose in the present paper.

2.2 Previous literature on creating taxonomies

Analysing DPVs quantitatively relies on what Lavandera (1978: 174–175) terms "the requirement of sameness" – that is, variants of a variable such as *yeah* need to be different ways of saying the same thing. The difficulty of applying this concept is discussed at length by Pichler (2010; 2013), who describes the need for methodological coherence across studies to allow discourse-level variation to be analysed in a comparable way to phonological and morpho-syntactic variables. Pichler (2010: 588–590) distinguishes between **function-based** operationalization approaches and those which align with **form or structure-based** approaches. A function-based analysis might focus on, for example, "all strategies used to introduce reported speech, sounds, gesture and thought by self or other" (Buchstaller 2006: 5) and seek out all the DPMs that perform this function (e.g. *be like* or *say*). Whereas a form-based analysis would extract all tokens of a form (e.g. Pichler [2013] who looked at the constructions I DON'T KNOW and I DON'T THINK) and then identify each token's pragmatic function. Pichler (2010: 597–598) highlights the importance of function, arguing that the use of DPVs is "motivated solely, or at least primarily, by their functionality" and favours a bottom-up, "theoretically flexible" taxonomy creation process, driven by data as opposed to top-down approaches driven by analytical frameworks (see also Waters [2016]). However, this paper takes a form-based approach: it exclusively investigates the form *yeah*. While it is relatively easy to extract all tokens of *yeah* (or any other structure/form), it is not straightforward to reliably and consistently identify its functions (as shown in Section 4). Thus, although many studies have utilised Pichler's (2010; 2013) recommendations, there remains a lack of transparency in how function-based analyses are carried out. There are, however, three notable exceptions, which we describe below.

Firstly, Eiswirth (2022b) presents a process of making and assessing a qualitative coding scheme of listener responses for use in quantitative discourse-pragmatic research, thus representing an attempt to consistently isolate a specific function, regardless of form. She highlights the following steps: (i) reviewing relevant literature to develop an initial coding scheme, (ii) applying the coding scheme to data and creating a second revision, (iii) asking a second uninformed coder to code part of the data and then making a third revision to the scheme based on any mismatches. Secondly, Wagner et al. (2015) begin their analysis of general extenders by identifying two subcategories based on previous literature (set-extending and non-set-extending), and then discuss the creation of a hierarchically-structured decision tree with binary yes/no options discriminating between these categories. Thus, the approach of Wagner et al. (2015) isolates a form/structure and identifies one of its functions. Thirdly, Childs (2021: 544) adopts a similar approach and uses a decision tree to determine between three contexts of *never* in British English. She isolates two forms (*never*, *didn't*) and uses the decision tree to identify multiple functions. These three studies show that replicable taxonomies can be both form and function-based. Influenced by this work, in this paper we create a coding scheme and decision tree to categorise the functions of a single form, *yeah*, with the aim of creating an effective, replicable, and transparent taxonomy.

3. Methodology

3.1. Data and creation of a decision tree

A corpus recorded in Derby (described by Docherty and Foulkes [1999: 47–48]) was utilised as a searchable database of L1 British English speakers in a conversational setting consisting of pairs or trios. There were 581 tokens of *yeah* across 16 male and 17 female speakers aged 14–67. While Pichler (2010) recommends identifying functions bottom-up, our scheme is initially top-down. A preliminary scan of the data found four previously-identified categories of *yeah* from the literature (see Section 2.1) to be applicable – we found the functions highlighted by Fuller (2003) and Drummond and Hopper (1993a) to be most relevant. Thus creating an initial, top-down, taxonomy:

i. Response *yeah*: its canonical function; an affirmative reply to a yes/no question.

ii. Assessment *yeah*: offers opinion or comment on the prior turn (either agreement/affiliation or disagreement/disaffiliation) and may be incipient. If agreeing, the comment will align with a prior turn (e.g. *yeah, nice*) but if disaffiliating it will create some distance (e.g. *yeah but that's not what I said*).

iii. Topic change *yeah*: marks a departure in topic from the prior turn; incipient by nature as it is followed by a full turn. It can mark a new topic or a return-to-prior topic (synonymous with *anyway*).

iv. Backchannel *yeah*: entirely recipient; shows listenership or encourages the prior speaker to continue their turn.

We also included a function of *yeah* which we did not find reference to in previous literature (this followed a more data-driven, bottom-up approach, following Pichler (2010)'s advice to be theoretically flexible):

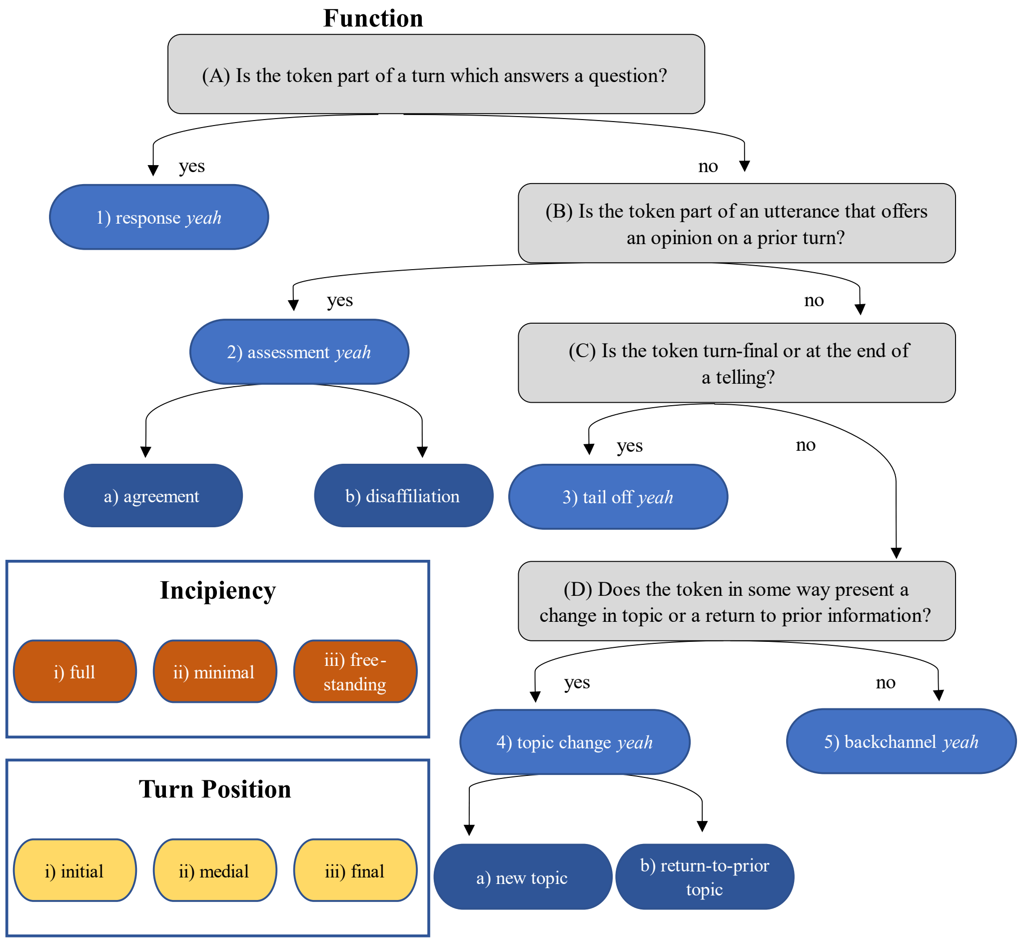
v. Tail-off *yeah*: closes off a speaker's discourse, indicating completeness to their turn or telling (a story or account of events, Mandelbaum, [2013: 492]) (5).

(5) Waynes_world – full turn-final tail-off *yeah* (derbyMYN03-MYN04-02.trs)

```
1    MYN04: so what do you reckon of Wayne's World then?
2    MYN03: it's okay I liked it uh I liked when they were in the car (0.4)
3  →          yeah (0.6)
4    MYN04: singing the Queen song?
```

A decision tree, influenced by Wagner et al. (2015), was created with five main functions (Figure 1). The tree nests the functions ordered by frequency of occurrence in the data (based on an initial scan). The exception to this is backchannel *yeah*, which was interpreted as being semantically bleached and therefore placed at the end; reflecting its distance from canonical response *yeah.*

Figure 1: Initial decision tree created to categorise *yeah* functions, incipiency and turn position.



3.2. Coding and inter-rater reliability

To test the effectiveness of the decision tree, a small sub-corpus was created from 15 different speakers across seven conversations by author 1, who was most familiar with the corpus. From these conversations, a subset of 102 tokens was selected to be somewhat representative of the five major function categories, so the selection was not entirely random. Form was utilised as the starting point for identifying eligible tokens, and these were selected across any clause or turn position. This was given to three raters to check for inter-rater reliability at different intervals. Written transcripts were provided[2] of the turn containing each token, along with two prior and following turns for context. All three raters were linguistically trained with some expertise in discourse-pragmatics. Raters were instructed to consult the decision tree, select functions in an Excel sheet, and, if necessary, provide a secondary function. The option of secondary function was added to capture the polysemy of *yeah*, which, like many DPMs, is inherently multifunctional. We omitted any definition for 'secondary function' however, leaving it open to interpretation by raters; this is returned to in Section 5. They also had a column in which to make notes about potential ambiguity in the data or uncertainty in their coding decisions.

[2] Author 1 had access to the audio, but raters 1–3 only had access to the written transcripts.

4. Results

4.1 Stage 1: coding by author 1 and rater 1

A summary of the inter-rater reliability between author 1 and rater 1 is presented in Table 1. Although given the option, rater 1 did not assign any secondary functions to *yeah.*

Table 1: Proportion of agreement across author 1 and rater 1, and the categories chosen by rater 1 when there was a mismatch. Numbers per category are shown in brackets.

Author 1 category	**Count**	**Rater 1 agreement (N) %**	**Rater 1 categories in cases of mismatch**
Response (1)	31	(27) 87%	Agreement (4)
Agreement (2a)	21	(20) 95%	Response (1)
Disaffiliation (2b)	7	(5) 71%	Response (2)
Tail-off (3)	5	(0) 0%	Agreement (3), Disaffiliation (1), Response (1)
New topic (4a)	3	(0) 0%	Agreement (2), Return-to-prior (1)
Return-to-prior (4b)	4	(0) 0%	Agreement (3), Response (1)
Backchannel (5)	31	(0) 0%	Agreement (22), Response (7), Not coded (2)
TOTAL	102	(52) 51%	Agreement (34), Response (12), Not coded (2), Return-to-prior (1), Disaffiliation (1)

The overall low agreement rate of 51% prompted author 1+rater 1 to consult and make necessary changes to the decision tree, particularly in section (A) and (B), which distinguish functions (1) and (2). This accounted for 92% of the divergence[3]. There were also seven mismatches for incipiency which were due largely to different categorisations of minimal incipiency. We do not analyse this further but focus instead on functions.

As part of their consultation, one example from section (A) that was discussed in detail was (6), where *yeah* was coded as a backchannel by author 1, but as a response by rater 1.

(6) only_five – full turn-final tail-off *yeah* (derbyFYP03-FYP04-08.trs)

```
1      FYP03: so who's coming round to you?
2             I mean you [obviously]
3    → W:                [yeah mum] and Derrick and
4             Missy and Paul that's it. It's only five anyway
```

Based on the decision tree, rater 1 ascertained that *yeah* was "part of a turn which answers a question", because line 3 indeed answers a question ("so who's coming round to you?").

[3] The total number of times agreement or response was offered by rater 1 in mismatch cases was 34+12=46. There were 102-52=50 cases of mismatch, resulting in 46/50=92% of cases involving agreement or response.

However, author 1 focused on the fact that the individual *yeah* token in line 3 did not constitute the actual response to the question in line 1. Instead, *yeah* seemed to be an acknowledgement of the question or assessment of the prior turn (which is how Drummond and Hopper [1993b] analysed *yeah* alongside *mm-hm* and *uh-huh*). Author 1 and rater 1 agreed to mark the token as an *agreement* token (2a). It was then agreed that if section (A) were rephrased to ask more specifically “is the token **an answer to** a question?” then it would be clearer to ascertain which *yeah* tokens were responses.

In consultation about section (B), example (7) was discussed. Here, *yeah* had been coded as backchannel by author 1, but agreement by rater 1.

(7) Costa_Blanca (derbyMON03-MON04-03.trs)

```
1      MON03: here’s Costa [Blanca]
2  →   MON04:              [ yeah  ] (0.4)
3           <noise of a page turning in a book>
4      MON03: now there’s your temperatures
```

The *yeah* in line 2 is responding to the prior turn without a change in topic which could be interpreted as MON04 agreeing with MON03 that he is indeed seeing Costa Blanca. However, if *yeah* functions as a backchannel here, then it merely shows MON04 is listening and allows MON03 to continue with his turn. It was agreed in consultation that this presented an ambiguity in the decision tree. Section B (“Is the token part of an utterance that offers an opinion on a prior turn?”) was worded somewhat openly; *yeah* only had to be “part” of an utterance offering opinion. It was therefore felt there was a lack of clarity in how do deal with freestanding tokens. This could lead to an over-reliance on assessment categories, without the opportunity to move down to the category of backchannel. Section (B) was thus changed to: “Is the token part of **a longer** utterance that offers **a comment** on a prior turn?”. *Yeah* in example (7) was therefore coded as a backchannel. A further discussion about section B led to *agreement* being renamed *affiliation* to create distance from the *response* category (which relates to questions), and to mirror the label *disaffiliation*.

The consultation between author 1+rater 1 resulted in re-analysis and a subsequent agreement rate of 95%. Of the remaining five cases, one token appeared in reported speech, and was subsequently excluded from the functional analysis. The remaining four cases were standalone tokens of *yeah* followed by long pauses, rendering it difficult to ascertain whether they were followed by a turn, and decide whether *yeah* should be coded as tail-off or backchannel. Rather than arbitrarily agree on a length of pause that would satisfy a turn boundary, the revised schema was shared with rater 2 to see if the same areas of disagreement or ambiguity would arise.

4.2 Stage 2: coding by rater 2

Figure 2 shows the updated decision tree, which was shared with rater 2, who was also given the same Excel coding sheet and transcripts as rater 1.

Figure 2: Revised decision tree for *yeah* functions, incipiency and turn positions. The altered parts have been highlighted in green.

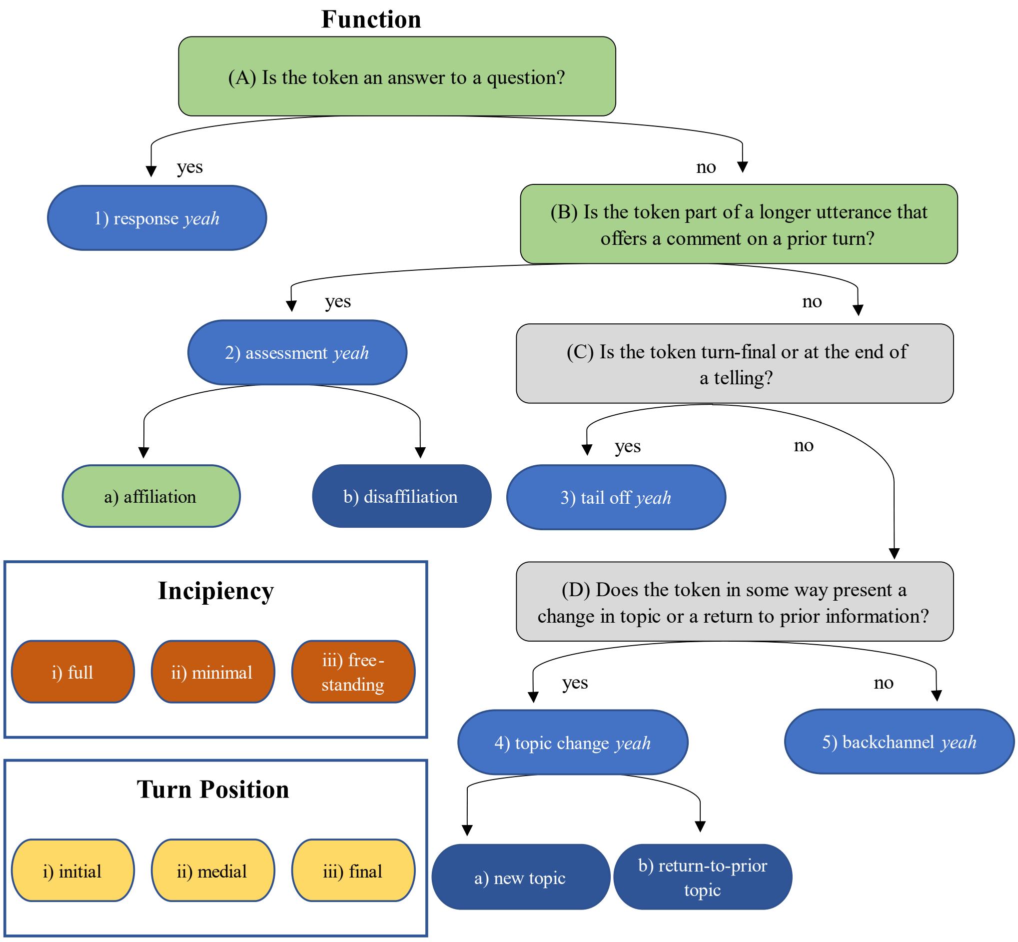


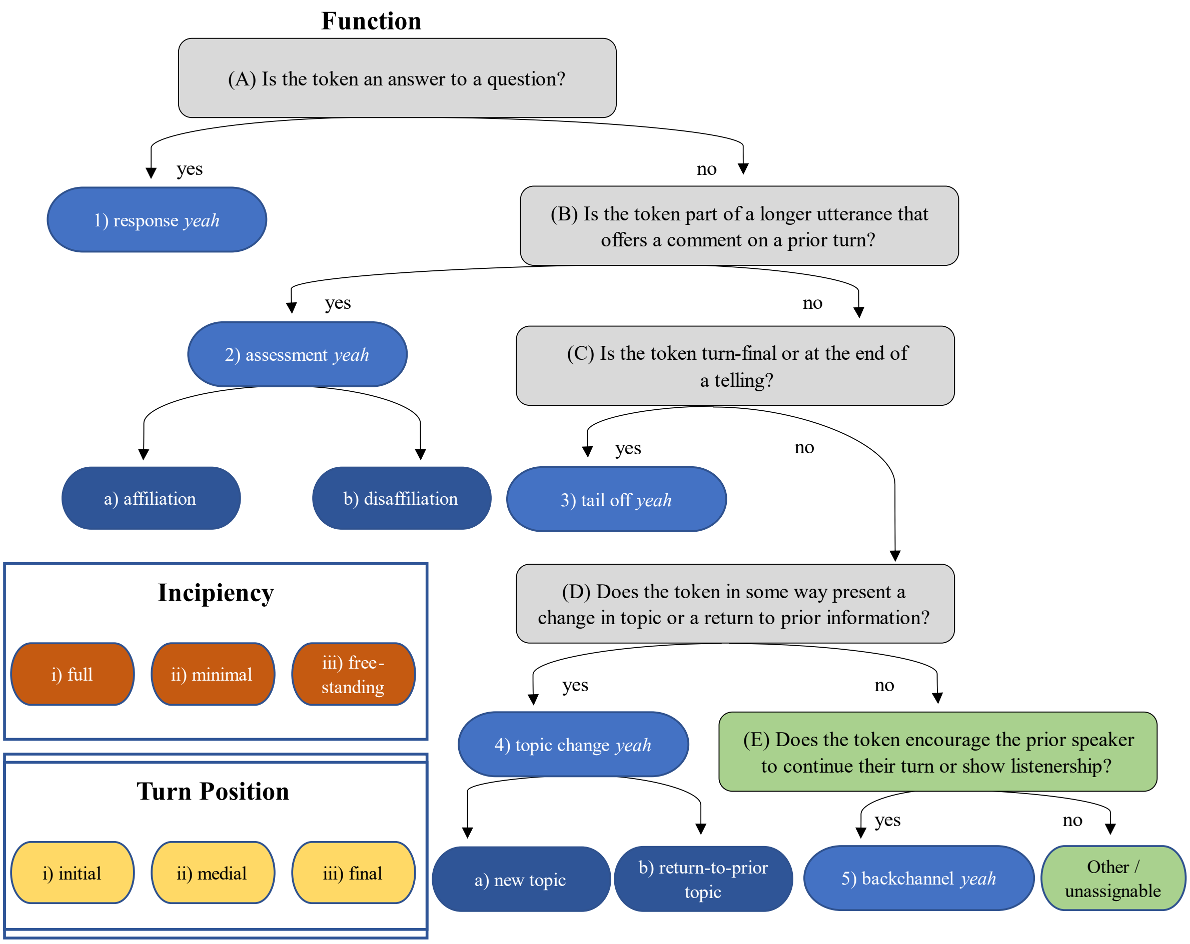
Table 2 presents a comparison between the consensus reached by author 1+rater 1, and rater 2. There was a more satisfactory agreement rate of 81% with 19 mismatches. Further, rater 2 offered six secondary functions (not shown in the tables) which were matches with the primary functions of author 1+rater 1, and an additional four of author 1's secondary function tokens were primary functions of rater 2. Taking these secondary functions into account, the agreement rate increases to 91%.

Table 2: Proportion of agreement across the consensus reached between author 1+rater 1 compared to rater 2 and the categories chosen by rater 2 when there was a mismatch. Numbers per category are shown in brackets.

Author 1+rater 1 category	**#**	**Rater 2 agreement (N) %**	**Rater 2 categories in cases of mismatch**
Response (1)	26	(26) 100%	-
Affiliation (2a)	22	(16) 73%	Backchannel (3), Response (1), Tail-off (1), Return-to-prior (1)
Disaffiliation (2b)	6	(5) 83%	Affiliation (1)
Tail-off (3)	5	(4) 80%	Disaffiliation (1)
Return-to-prior (4a)	7	(4) 57%	New topic (1), Affiliation (1), Backchannel (1)
New topic (4b)	1	(1) 100%	-
Backchannel (5)	30	(26) 87%	Response (3), New topic (1)
Not coded	5	(1) 20%	Backchannel (4)
TOTAL	102	(83) 81%	Backchannel (8), Response (4), Affiliation (2), New topic (2), Tail-off/return-to-prior/disaffiliation (3)

One area of notable disagreement was backchannel, representing the largest category of mismatches (8/19; 42%). The initial rationale for including backchannel last in the decision tree was that as a semantically bleached token with the least-incipient function, it would be easier to filter out other functions before assigning a token to backchannel. However, rater 2 categorised four of author 1+rater 1's unresolved ('not coded') tokens as backchannels, suggesting that the tree could be leading raters to the 'backchannel' function after all options are exhausted, even if it is still not the best fit. It also came to our attention that backchannel had no definition in the decision tree, which could lead it to be categorised without sufficient reflection. It was thus decided to add a new question, (E) which defines the backchannel function: "Does the token encourage the prior speaker to continue their turn or show listenership?". We also added an 'other/unassignable' category and shared the revised tree (Figure 3) with rater 3.

Figure 3: Third decision tree for *yeah* functions, incipiency and turn position revised after inter-reliability coding stage 2. The altered parts have been highlighted in green.



4.3 Stage 3 coding by rater 3

Table 3 presents a comparison between the consensus reached by author 1+rater 1, and the coding of rater 3, with a less satisfactory rate of 56%. Rater 3 offered five secondary functions, all of which fell within the mismatches, and were matches with the primary functions of author 1+rater 1. For six further mismatches, author 1 offered the function chosen by rater 3 as a secondary function. Taking these secondary agreements into account, the agreement rate increases to 67%.

Table 3: Proportion of agreement across the consensus reached between author 1+rater 1 and rater 3, and the categories chosen by rater 3 when there was a mismatch. Numbers per category are shown in brackets.

Author 1+rater 1 category	**#**	**Rater 3 agreement (N) %**	**Rater 3 categories in case of mismatch**
Response (1)	26	(20) 77%	Affiliation (5), Disaffiliation (1)
Affiliation (2a)	22	(15) 68%	Tail-off (2) Response (1), Disaffiliation (1), New topic (1) Backchannel (1) Other (1)
Disaffiliation (2b)	6	(3) 50%	Affiliation (2), New topic (1)
Tail-off (3)	5	(1) 20%	Affiliation (3), Backchannel (1)
Return-to-prior (4a)	7	(3) 43%	Affiliation (3), New topic (1)
New topic (4b)	1	(0) 0%	Affiliation (1)
Backchannel (5)	30	(15) 50%	Affiliation (13), Response (1), Tail-off (1)
Other/unassignable	5	(0) 0%	Affiliation (2), Backchannel (2), Response (1)
TOTAL	102	(57) 56%	Affiliation (29), Backchannel (4), Response (3), Tail-off (3), New topic (3), Disaffiliation (2), Other (1)

Rater 3 coded 51 tokens as ‘affiliation’, far more than any previous rater. This propensity to answer ‘yes’ to question (B) accounts for 29 mismatches against stage 1. A total of 17 mismatches occurred where rater 3 marked a token as affiliation even though it was not part of a longer turn (e.g. “yeah <laugh>”), accounting for 12/15 backchannel misalignments. Tail-off also had low inter-rater reliability: in four cases, tokens were coded as tail-off at stage 1 and either affiliation or backchannel by rater 3. For a further three cases, rater 3 offered tail-off but author 1+rater 1 had affiliation or backchannel. Future iterations of the taxonomy could be more explicit about defining utterance lengths.

However, Table 4 shows that there is a much higher rate of agreement between rater 2 and rater 3 of 74%. This increases to 83% when taking secondary agreements into account. Similarly to the comparison with author1+rater 1, a large proportion of mismatches (55% in this case) occurred where rater 3 marked a token as affiliation, but rater 2 had different categories. Rater 3 matched with every token that rater 2 coded as backchannel, but they had four extra backchannel tokens that were coded differently by rater 2. Four further mismatches occurred for tokens coded as affiliation/disaffiliation by rater 2 but as one of the two topic change functions by rater 3.

Table 4: Proportion of agreement across rater 3 compared to rater 2 with numbers per category in brackets and the categories chosen by rater 3 when there was a mismatch.

Rater 2 category	**#**	**Rater 3 agreement (N) %**	**Rater 3 categories in case of mismatch**
Response (1)	32	(22) 69%	Affiliation (7), Backchannel (2), Disaffiliation (1)
Affiliation (2a)	33	(29) 88%	Disaffiliation (1), Tail off (1), New topic (1), Other (1)
Disaffiliation (2b)	6	(3) 50%	Affiliation (1), Tail off (1), New topic (1)
Tail-off (3)	6	(2) 33%	Affiliation (3), Backchannel (1)
Return-to-prior (4a)	6	(3) 50%	Affiliation (3)
New topic (4b)	3	(1) 33%	Affiliation (1), Backchannel (1)
Backchannel (5)	15	(15) 100%	-
Other / unassignable	1	(0) %	Response (1)
TOTAL	102	(75) 74%	Affiliation (15), Backchannel (4), Disaffiliation (2), Tail off (2), New topic (2), Response (1), Other (1)

5. *Discussion and conclusion*

In this paper, we have outlined the process of creating a taxonomy of functions for the DPM *yeah,* eventually reaching an agreement rate between 67–95% when accounting for primary and secondary functions. We would like this agreement rate to be higher and more consistent (≥95%); however, we believe that this process improved our approach by moving beyond a reliance on a top-down functional taxonomy based on a single rater's interpretations (see e.g. Aijmer [2013]; Childs [2021]; Columbus [2010: 292–293]; Tagliamonte [2016]). We recognise that this taxonomy was not entirely bottom-up, however, it has been flexible and data-sensitive as Pichler (2010) recommends.

We made improvements to our decision tree by providing more explicit instructions to raters, in line with the recommendations of Eiswirth (2022b). We also achieved more agreement when taking secondary functions into account, which is not always considered in studies of discourse-pragmatic variation (but see Cheshire [2007]). However, we acknowledge that our schema does not provide a specific definition of what constitutes a secondary function. Indeed, rater 2 noted that they were unsure whether to utilise the secondary function to indicate uncertainty as to a specific token, or to indicate that a token definitively fulfilled two functions at once. Furthermore, by making secondary function coding optional, we ended up with significant variability, with author 1 providing 30 secondary functions, rater 2 providing 40, rater 3 providing 12, but rater 1 providing none. Future work on DPM operationalisation should instruct raters more explicitly on secondary functions, and perhaps make secondary coding compulsory where tokens definitively fulfil multiple functions. This would permit for an analysis of DPMs which accounts for multifunctionality (for a discussion of polysemy in DPM research see Denis and Tagliamonte [2016: 95–96]). We also think that raters could be more explicitly guided to think about how incipiency and turn position interact with function as there was some minor mismatch for the data in this study. The tree already guides raters to consider turn position in question (C) for *tail-off*, but turn position is not included in any other function descriptions. In future, other criteria would be useful to inform raters about these ambiguities. For example, ensuring that freestanding tokens are only marked as either *backchannel* or *response*, or that *new/return-to-prior topic* tokens are only ever fully incipient. Furthermore, grouping categories together remains a viable option in the refinement of decision trees. For instance, there remains the possibility of uniting *affiliation* and *disaffiliation* under the broader *assessment* label where there is uncertainty. Similarly, in cases where it is unclear whether *yeah* functions to return to a topic or present a new one, it could be aligned to the broader *topic change* category.

We note at this point that the application of decision trees to DPM operationalisation can be immensely helpful for replicable, robust, and transparent research. However, we recognise that it doesn't necessarily reflect how individuals produce or process DPMs in conversation. We propose our decision tree as a methodological tool allowing for DPMs to be operationalised by their functions for a range of applications. We recognise, though, that different approaches are needed to model how speakers themselves interpret DPM functions.

Another issue to consider for the future is testing the accuracy of DPM coding when all raters are provided with audio data. The phonetics of *yeah* has been examined by many (Freeman et al. 2015; Grivičić and Nilep 2004; Trouvain and Truong 2012; Truong and Heylen 2010), finding that assessment *yeah* generally has a longer duration (Truong and Heylen 2010; Freeman et al. 2015), 'reluctant stance' *yeah* is highest in intensity and pitch (Freeman et al. 2015), and freestanding recipient *yeah* has a lower intensity than turn-initial fully incipient *yeah* (Trouvain and Truong 2012). Conversely, the most semantically bleached function category we have highlighted is backchannel *yeah*, and this may show some phonetic reduction, along with recipient tokens of *yeah* (as Gibb-Reid et al. [2022] find). This is a ripe avenue for future research and DPM studies generally, which we intend to pursue.

We encourage any future research into DPMs to resist relying on previously-made taxonomies (top-down) but instead work to adapt and create flexible, bottom-up taxonomies. We also encourage further inter-rater reliability in discourse-pragmatic research. By reporting the process of operationalisation, we invite future research that creates more effective, transparent, and replicable taxonomies which are less subjective, can be used for other research, and are adapted and applied to different contexts.

6. References

Aijmer, Karin 2013. *Understanding pragmatic markers*. Edinburgh University Press.

Buchstaller, Isabelle 2006. Diagnostics of age-graded linguistic behaviour: The case of the quotative system. *Journal of Sociolinguistics*. Wiley Online Library. 10(1). 3–30.

Cheshire, Jenny 2007. Discourse variation, grammaticalisation and stuff like that. *Journal of Sociolinguistics* 11(2). 155–193.

Childs, Claire 2021. The grammaticalisation of never in British English dialects: Quantifying syntactic and functional change. *Journal of Linguistics* 57(3). 531–568.

Columbus, Georgie 2010. A comparative analysis of invariant tags in three varieties of English. *English World-Wide*. John Benjamins. 31(3). 288–310.

Denis, Derek & Sali Tagliamonte 2016. Innovation, right? Change, you know? Utterance-final tags in Canadian English. In Heike Pichler (Ed.), *Discourse-pragmatic variation and change in English: New methods and insights*, 86–112. Cambridge: Cambridge University Press.

Diskin, Chloé. 2017. The use of the discourse-pragmatic marker "like" by native and non-native speakers of English in Ireland. *Journal of Pragmatics 120*: 144–157.

Docherty, Gerard & Paul Foulkes 1999. Derby and Newcastle: instrumental phonetics and variationist studies. In Paul Foulkes & Gerard Docherty (Eds.), *Urban voices: Accent studies in the British Isles*, 47–71. Routledge.

Drummond, Kent & Robert Hopper 1993a. Some uses of yeah. *Research on language and social interaction* 26(2). 203–212.

Drummond, Kent & Robert Hopper 1993b. Back channels revisited: Acknowledgment tokens and speakership incipiency. *Research on language and Social Interaction* 26(2). 157–177.

Eiswirth, Mirjam Elisabeth 2022a. Modeling listener responses. In Elizabeth Peterson, Joseph Kern & Turo Hiltunen (Eds.), *Discourse-Pragmatic Variation and Change: Theory, Innovations, Contact*, 173–189. Cambridge: Cambridge University Press.

Eiswirth, Mirjam Elisabeth 2022b. Developing and testing interaction-based coding schemes for the analysis of sociolinguistic variation. *Language & Communication* 87. 11–28.

Freeman, Valerie, Gina-Anne Levow, Richard A Wright & Mari Ostendorf 2015. Investigating the role of "yeah" in stance-dense conversation. *Interspeech 2015*, 3076–3080.

Fuller, Janet M. 2003. The influence of speaker roles on discourse marker use. *Journal of Pragmatics* 35(1). 23–45.

Gibb-Reid, Ben, Paul Foulkes & Vincent Hughes 2022. Exploring the phonetic variation of 'yeah' and 'like'. *York Papers in Linguistics 2,* 18. 1–27, December 2022.

Grivičić, Tamara & Chad Nilep 2004. When phonation matters: The use and function of yeah and creaky voice. *Colorado Research in Linguistics* 17(2). 1–11.

Jucker, Andreas H. & Sarah W. Smith 1998. And people just you know like, wow: Discourse markers as negotiating strategies. In Andreas H. Jucker & Yael Ziv (Eds.), *Discourse markers. description and theory* (Pragmatics & Beyond New Series 57), 171-201. Amsterdam/Philadelphia: John Benjamins.

Lavandera, Beatriz R. 1978. Where does the sociolinguistic variable stop? *Language in society*. Cambridge University Press. 7(2). 171–182.

Mandelbaum, Jenny 2013. Storytelling in conversation. In Jack Sidnell & Tanya Stivers (Eds.), *The handbook of conversation analysis*, 492–507. 1st edn. Wiley-Blackwell.

Pichler, Heike 2010. Methods in discourse variation analysis: Reflections on the way forward 1. *Journal of Sociolinguistics* 14(5). 581–608.

Pichler, Heike 2013. *The structure of discourse-pragmatic variation* (Studies in Language Variation). Bd. 13. John Benjamins Publishing.

Pichler, Heike 2016. Uncovering discourse-pragmatic innovations: innit in Multicultural London English. In Heike Pichler (Ed.), *Discourse-pragmatic variation and change in English: New methods and insights*. Cambridge: Cambridge University Press.

Pichler, Heike 2021. Tagging monologic narratives of personal experience: utterance-final tags and the construction of adolescent masculinity. In Karen V Beaman, Isabelle Buchstaller, Susan Fox & James A Walker (Eds.), *Advancing Socio-Grammatical Variation and Change: Sociolinguistic research in honour of Jenny Cheshire* (Routledge Studies in Sociolinguistics), 377–398. Routledge.

Stenström, Anna-Brita, Gisle Andersen & Ingrid Kristine Hasund 2002. Trends in teenage talk. *Amsterdam: Benjamins*.

Tagliamonte, Sali 2016. *Teen talk: The language of adolescents*. Cambridge University Press.

Trouvain, Jürgen. & Khiet Phuong Truong 2012. Acoustic, morphological, and functional aspects of `yeah/ja' in Dutch, English and German. *Proceedings of the Interdisciplinary Workshop on Feedback Behaviors*, 77–80. El Paso: University of Texas.

Truong, Khiet Phuong & Dirk Heylen 2010. Disambiguating the functions of conversational sounds with prosody: the case of 'yeah'. *Interspeech 2010*, 2554–2557. ISCA.

Wagner, Suzanne Evans, Ashley Hesson, Kali Bybel & Heidi Little 2015. Quantifying the referential function of general extenders in North American English. *Language in Society*. Cambridge University Press. 44(5). 705–731.