Cardiff University Press
Gwasg Prifysgol Caerdydd

JCaDS
JOURNAL OF CORPORA AND DISCOURSE STUDIES 2023, 6(1):61–90
E-ISSN 2515-0251

'HER DREADFUL PLIGHT': A CORPUS-BASED ANALYSIS OF THE INDEXICAL AND AFFECTIVE STANCE PROPERTIES OF *POOR THING*

SEAN NONNENMACHER
UNIVERSITY OF PITTSBURGH

BEN NAISMITH
DUOLINGO

CITATION

Nonnenmacher, S. & Naismith, B. (2023). 'Her dreadful plight': A corpus-based analysis of the indexical and affective stance properties of *poor thing*. *Journal of Corpora and Discourse Studies*, *6*(1):61–90

ABSTRACT

Curiously, the English multiword expression *poor thing* (*PT*) often refers to entities that are neither economically impoverished nor inanimate objects. By using a mixed-methods corpus linguistics and critical discourse analysis (CDA) approach, we demonstrate that *PT* functions as an expression of affective stance based on evidence from two American English corpora. In cases where the social identities of speakers and referents can be determined, *PT* is frequently used by women and in reference to women, intimates (including children), or animals. Additionally, the expression may refer to entities of low vitality due to illness or death. Our results indicate that *PT* indexes a speaker's compassionate stance alongside a referent's misfortune, bundling together a set of (stereo)typically 'disempowered' personae. This study demonstrates the potential of corpus-based CDA investigations for the analysis of the indexical associations and stance properties of relatively infrequent lexical expressions that, nonetheless, have recognizable meaning for speakers.

KEYWORDS

Corpus analysis; critical discourse analysis; *poor thing*; affective stance; indexicality

CONTACT

Sean Nonnenmacher, Department of Linguistics, University of Pittsburgh, 2816 Cathedral of Learning, Pittsburgh, PA 15260, USA. sen40@pitt.edu

DOI

10.18573/jcads.90

ISSUE DOI

doi:10.18573/jcads.v6

ORCID

0000-0002-5246-200X, 0000-0001-8347-3142

Manuscript accepted 2022-12-19

'Her dreadful plight': A corpus-based analysis of the indexical and affective stance properties of *poor thing*

Sean Nonnenmacher
University of Pittsburgh

Ben Naismith
Duolingo

1. Introduction

In the 2007 film *Sweeney Todd: The Demon Barber of Fleet Street* (Burton, 2007), the song 'Poor Thing' addresses the dreadful plight and eventual demise of Sweeney Todd's wife, Lucy Barker. Given that linguistic expressions have the potential to index other information in the sociocultural realm (Eckert, 2008; Labov, 1972; Ochs, 1992; Silverstein, 2003), it is possible that the use of 'poor thing' in this context, more than being merely incidental, is suggestive of a broader set of social-indexical and discursive properties for the expression. In this paper, we explore instances of the English expression *poor thing* (henceforth *PT*) using a mixed-methods and computationally-aided approach that combines a corpus-based analysis of quantitative trends with a critical discourse analysis of representative corpus samples (e.g., Baker, 2006; Baker *et al.*, 2008). *PT* serves as a case study, being selected out of many possible expressions due to its semantic bleaching and idiomaticity (cf., *poor boy* or *poor me*), its higher frequency than other *poor* + (pro)noun types in the selected corpus data, and because there are no sociolinguistic studies of this multiword expression to date.

In their analysis of British media representations of refugees, asylum seekers, and migrants, Baker *et al.* (2008) model how practitioners of critical discourse analysis (CDA) might productively collaborate with scholars in corpus linguistics (CL) to use language patterns like collocations to bolster CDA arguments. Rather than seeing the union of CDA and CL as primarily enhancing one or the other approach, Baker and colleagues seek to identify the sweet spot where 'each [approach] contributes equally and distinctly to a methodological synergy' (p. 274) of the two. Additionally, scholars of corpus-based discourse analysis have recently devoted greater attention to the study of stancetaking, or how speakers and writers evaluate an object of joint attention and take up positions based on these evaluations (Kiesling *et al.*, 2018; Poole & Hayes, 2022). Yet it remains true that more socioculturally-oriented notions of stancetaking, such as those that emerge from interactional approaches to the study of discourse (Du Bois, 2007; Jaffe, 2009; Kiesling, 2022), have yet to be adequately integrated into CL-CDA scholarship.

Our analysis addresses this gap by considering what the instances of *PT* in two American English corpora indicate about the social and linguistic properties of this multiword expression as well as its stance-relevant discourse functions. These two corpora are the Corpus of Contemporary American English (COCA; Davies, 2008–), a contemporary American English corpus, and the Corpus of Historical American English (COHA; Davies, 2010), a historical American English corpus, respectively. Following Baker *et al.* (2008), we first use a corpus analysis to discover and summarize the properties of *PT*

with automated computational methods and trained human annotators. CDA (Fairclough, 2013; van Dijk, 2015) then elaborates on the discourse functions of *PT* through an exploration of representative concordance lines. This choice allowed for a robust CL-CDA analysis, with each approach enhancing the other to yield patterns, observations, and conclusions that either alone would have been unable to achieve. Our discussion frames the observed corpus patterns and supporting CDA with an eye toward the theories of indexicality and affective stancetaking.

2. Background

2.1. The social indexicalities of linguistic forms

Recent work on the relationship between lexicogrammatical choices and discourse meaning has demonstrated that corpus-based investigations can greatly enhance our understanding of culture-specific words and their associations (Gladkova & Romero-Trillo, 2021b). For instance, Gladkova and Romero-Trillo (2021a) use a corpus methodology to investigate collocations and pragmatic uses of the term `ugly` in English, finding a difference in the cognitive salience of concepts associated with *ugly* and its antonym *beautiful*, such that 'people' and 'nature' figure into the use of *beautiful* but 'human actions' are attended to with *ugly*. Theoretical concepts of interest to linguistic anthropologists, such as the enregisterment of social personae within speech communities (Agha, 2003), are useful for investigations of previously unexamined identity terms. For example, Cole and Tieken-Boon van Ostade (2021) analyze stereotyped working-class personae in the Netherlands and England and find overlapping and distinct linguistic features related to local patterns of variation (regional dialects and class-based dialects) and other semiotic properties (e.g., attire).

Scholars working at the intersection of sociolinguistics and linguistic anthropology (or sociocultural linguistics, as per Bucholtz & Hall, 2008) have increasingly sought to explain the ways in which linguistic forms index (Silverstein, 2003) or point to the social life of speakers. Phonological variables are commonly indexical of the social identities of a speaker because they often sit below the level of conscious awareness (e.g., Eckert, 2019). However, virtually all dimensions of linguistic structure can be incorporated into indexical meaning-making. Kiesling's (2004) research on American English *dude* and Bucholtz's (2009) later work on Mexican Spanish *güey* ('dude') demonstrate that single-word slang expressions achieve a variety of immediate interactional or discursive functions, while co-occurring with higher-order indexical associations. The terms *dude* and *güey* may be used as referentially null discourse markers of exclamation, agreement, or conflict mitigation. Alternatively, they can appear as referentially full address terms, for instance when one fraternity member calls a fellow member 'dude'. As Kiesling notes, regardless of the specific function being activated in a particular utterance of *dude*, the lexical item helps to construct a speaker's stance of 'cool solidarity' (Kiesling, 2004, p. 286) which, in turn, is also indexical of masculinity. The range of meanings for any sociolin-

guistic variable exist within what Eckert (2008, p. 453) has called the indexical field, which contains 'a constellation of ideologically related meanings, any one of which can be activated in the situated use of a variable'. It is within the indexical or, more broadly, semiotic field (Babel, 2018) of a given linguistic form that discourse functions like agreement become linked with speaker qualities like coolness and masculinity. While Kiesling and Bucholtz both attend to the discursive patterns of these expressions, neither author uses corpus data in their work. Nonetheless, these findings about the indexical meanings of broadly circulating referential and address terms might be further bolstered with the help of CL.

Our own view of the social indexicality of linguistic forms like *dude, güey,* and *PT* draws from both sociolinguistic and linguistic anthropological thinking about how semiotic processes sustain indexical connections over time (Irvine & Gal, 2000; Ochs, 1992; Silverstein, 2003). Important for our analysis is the distinction between direct and indirect forms of indexicality first articulated by Ochs in her work on the caregiving practices of American and Samoan mothers. Ochs (1992) argues that 'the relation between language and gender is not a simple straightforward mapping of linguistic form to social meaning of gender … [but] is constituted and mediated by the relation of language to stances, social acts, social activities, and other social constructs' (pp. 336–337). Ochs (1992) discusses how the mapping between linguistic forms and social orders like gender is (i) *non-exclusive*, meaning a linguistic feature may be used by anyone and may also point to multiple possible meanings simultaneously, (ii) *constitutive*, so that a given linguistic pattern may directly index interactionally-relevant phenomena like stances, acts, or activities, and (iii) *temporally transcendent*, with indexes drawing from and contributing to more broadly circulating sociocultural meaning outside the immediate context (pp. 340–346). Our analysis of *PT* finds that the three indexical mapping mechanisms identified by Ochs are likewise important for understanding this multi-word expression's (MWE) social meaning. We also build on the above tradition by analyzing the indexical associations of *PT* alongside affective stancetaking, as we define next.

2.2. Affective stance in sociocultural linguistics

Scholars working at the intersection of corpus analysis and discourse analysis have recently devoted greater attention to the notion of stancetaking. Poole and Hayes (2022), for instance, use an applied linguistic approach of stance analysis in their study of shifts in climate change discourse over time, focusing particularly on modal makers of epistemic stance (following Biber & Finegan, 1989; Hyland, 2005). In their work on interactional stancetaking in online Reddit forums, Kiesling *et al.* (2018) find that the stance dimensions of affect (feeling or emotionality), investment (commitment to one's position), and alignment (between interlocutors) are associated with lexical features and discoverable through qualitatively-informed computational analysis. Yet it remains the case, as Poole and Hayes (2022) contend, that relatively few scholars have approached the study of corpus-based discourse from the perspective of stance analysis.

Here, we are specifically interested in applying affective notions of stancetaking, as conceptualized in sociocultural linguistics, to corpus-based CDA work. Jaffe (2009) introduced the term 'sociolinguistic stance' as a means of bringing sociolinguistic insights to bear on the positionalities taken up by speakers engaged in talk. In its simplest formulation, 'stancetaking' refers to the linguistic action taken by a speaker in positioning themself with respect to the form of an utterance or its referential content (Jaffe, 2009). Jaffe builds her notion of sociolinguistic stance from earlier work by Du Bois (2007), an anthropologist who argues that stance acts can be understood as unfolding in a triangulated manner. Figure 1, which is Kiesling's (n.d.) adaptation of the Du Bois (2007) stance triangle, indicates that a first subject (or speaker) evaluates a stance object and thus positions themself in relation to it. A second subject (or interlocutor) then takes up a separate positionality with respect to the object based on their own evaluation and their alignment or disalignment with the first subject. The subjects' agreement or disagreement with each other's evaluation of the same stance object results in their alignment or disalignment, respectively.

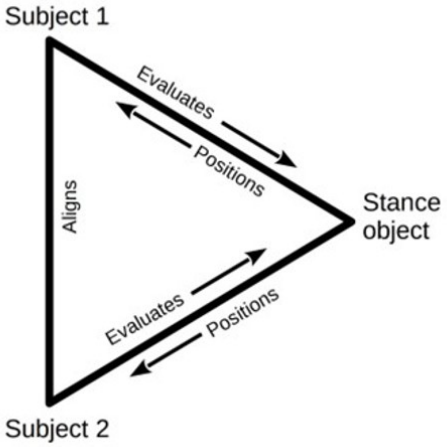

Figure 1*. Stance triangle indicating evaluation, positioning, and (dis)alignment (Kiesling, n.d.)*

Affective stance is characterized by three distinct properties. It conveys feelings or emotional intensity about an utterance, text, focus, or object of concern (Besnier, 1990; Martin, 2000). Further, affective stance involves displays of interlocutors' intersubjective feelings toward each other (Du Bois, 2007; Du Bois & Kärkkäinen, 2012; Goodwin, Cekaite, & Goodwin, 2012). Finally, affective stance is recognizable through the repeated circulation of affective practices (Ahmed, 2014; Kiesling, 2018; Milani & Richardson, 2020; Wetherell, 2013), whereby displays toward a focus of concern follow sociocultural scripts that render emotions like joy or anger legible (Ahmed, 2014; Lakoff, 1987).

Whether stance is epistemic or affective, it is always 'socially situated and socially consequential' (Jaffe, 2009, p. 3) and implicated in other indexical relations. Moreover, the same 1989 issue of *Text* that featured Biber and Finegan's seminal work on a corpus-based examination of stancetaking also included an article by anthropologists Elinor Ochs

and Bambi Schieffelin about affective frames for social actions. Ochs and Schieffelin (1989) argue that affect is encoded in language across multiple levels of structure (sound patterns, morphemes, lexical items, and broader discourse patterns) and can be understood through affective frames that trigger particular emotional responses in speakers' feelings, moods, dispositions, or attitudes. As already noted, Ochs (1992) would later describe how stance contributes to the development and circulation of indexical associations between linguistic forms and sociocultural meanings, including affective meanings. Based on our hunch that the discourse function of *PT* has something to do with affective stance, we move forward with an analysis that focuses primarily on affective stance, acknowledging that other kinds of stance (epistemic, instrumental, cooperative, or moral) are always co-present alongside affective displays (Goodwin, 2007).

3. Current study

The current study addresses the following research questions:

1. What are the sociolinguistic (i.e., social and linguistic) properties of *poor thing*?
2. To what extent is the function of *poor thing*, in its broader discourse context, related to affective stancetaking?

In our study, we employ a corpus-based analysis (Baker, 2006; Tongini-Bonelli, 2001), where the two corpora serve as sources of evidence that can confirm, refute, or expand our understanding of indexical associations of *PT* and its affective stance properties. Addressing the above two research questions positions us to better consider the all-important 'how' of discourse analysis (Mautner, 2019, p. 10): here, how affective stancetaking is the mediational link between the micro level of linguistic form to the macro level of sociocultural meaning of *PT*. As advocated by Hashemi (2012), qualitative and quantitative linguistic data are integrated within a single study, with the goal of creating an initial quantitative mapping of key patterns to guide a qualitative parsing of how the patterns emerge (Bryman & Burgess, 1994). We have adopted an explanatory sequential design in which the collection of quantitative data (i.e., the descriptive corpus statistics) is used to guide the resulting qualitative analysis and interpretation (Cresswell & Plano Clark, 2011). We additionally see our methodology as following a triangulation approach, which Baker (2006) describes as 'using multiple methods of analysis (or forms of data)' (p. 16, following McNeill 1990; Newby, 1977). The corpus instances of *PT* serve as data to better understand both the indexical associations of the term and its stance-relevant functions in the broader discourse context of the concordance line. To operationalize 'discourse context' we have constrained our focus to the words immediately preceding and following each *PT* token, in part to determine if such discourse allows for claims to be made about stancetaking in accordance with the stance triangle (Du Bois, 2007).

In line with recent recommendations by Mautner (2019) about Corpus-assisted discourse studies (CADS) and calls for greater methodological fluidity in the critical study of discourse (e.g., Ehrlich & Romaniuk, 2013; Lakoff, 2015; Tannen, 1990; van Dijk, 2009; Wodak & Meyer, 2009), we used a systematic protocol in selecting and cleaning the cor-

pus data, appropriately quantifying linguistic and social patterns, and supplementing the computer-assisted coding of *PT* tokens with manual coding by human annotators. Our motivation was to allow a range of patterns to bubble to the surface, even if they did not reach the minimum threshold of significance with statistical testing. Following Baker *et al.* (2008), these larger quantitative patterns guided downsampling to examine representative concordance lines for a CDA-informed analysis of affective stancetaking. This integrated approach between CL and CDA (Partington, 2010, discussed in Mautner, 2019) allowed us to assemble suitable discourse evidence to model affective stancetaking in *PT* usage with the help of the stance triangle (Du Bois, 2007), while at the same time accounting for the host of common indexical associations across contexts. Our approach is particularly informed by Baker's (2006) discussion of analyzing concordances in discourse analysis with corpora. Most of what we have done closely mirrors the steps he describes for concordance analysis, and we diverge only in the latter steps related to sorting concordances and making sense of results. We use the exploratory analysis of part one to guide our selection and presentation of a smaller set of concordance lines for more detailed discourse analysis. These lines were chosen not only because they exemplify the quantitative trends uncovered in the first part of our analysis, but also because they demonstrate how *PT* is implicated in affective stancetaking.

4. Corpus analysis

4.1. Corpus methodology

4.1.1 Data collection

Two large corpora were initially queried online through the English-Corpora.org interface (Table 1). The resulting concordance lines were subsequently collated in a comma-separated values (csv) file and processed using Python programming language in a Jupyter notebook. In each concordance line, the node *PT* is in the middle of the line with 15 words on either side to provide discourse context. As previously stated, the two corpora selected were COCA (Davies, 2008–) and COHA (Davies, 2010). These corpora can be taken as representative of American English because of their size (Table 1) and because they encompass a wide variety of genres produced in the United States (described in Section 4.2.1). COCA, in particular, is possibly the most widely-used English corpus (Davies 2008–), appearing in many studies as representative of L1 English norms (e.g., Monteiro, Crossley, & Kyle, 2020; Naismith & Juffs, 2021). It was deemed necessary to use a combination of American English corpora to maximize the potential data because *PT* is a relatively low-frequency MWE and because each of the two corpora sample language from different time periods. Whereas COCA supplies modern American uses of *PT* from 1990 onwards, COHA includes older texts starting in 1810.

A cursory diachronic overview of *PT* usage indicates that the rate of occurrence has remained relatively constant across time in both corpora, with slight increases beginning in the mid–19th century and again in the early 21st century (see Figure 2 for a visualiza-

tion of usage across time in Google Ngram; note that the data in our analysis come from COCA/COHA, not Google). Although we do not perform a diachronic analysis, we use these two corpora to allow for a large enough sample of *PT* tokens for analysis. Additionally, the authors manually inspected tokens from both corpora and found them suitably comparable for an analysis of affective stancetaking and indexical meaning generally. All corpus data from this point forward refer to the combined COCA/COHA dataset.

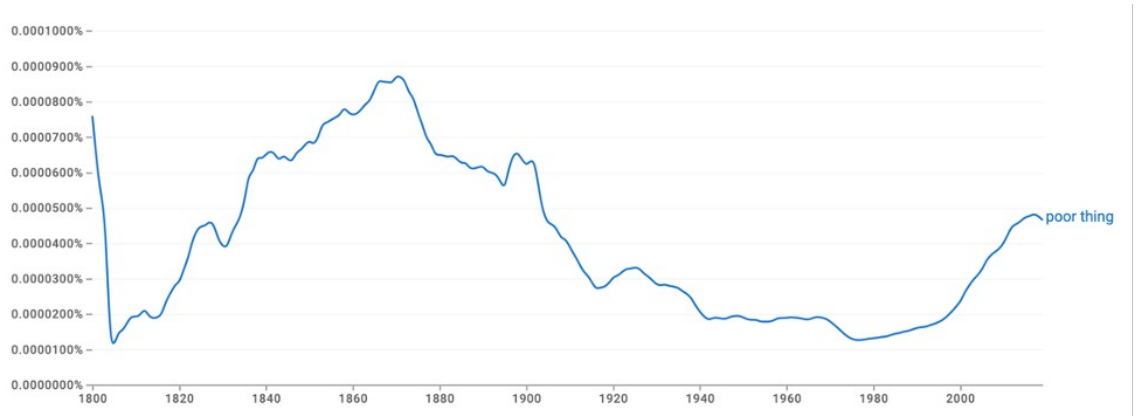

***Figure 2**. Rate of occurrence of PT in the Google Ngrams American English corpus 1800-2019* (Google Ngram, 2021)

Table 1 provides the total count of *PT* instances of 2,378 tokens (for a total discourse context of about 70,000 words, inclusive of words to the left and right of each *PT* token per concordance line). Importantly, this total refers to the forms of the lemma THING, i.e., both singular *thing* and plural *things*. However, after consideration, the other forms of the lemma POOR were excluded (comparative degree *poorer* and superlative degree *poorest*) due to their very low frequencies (5 tokens in COCA and 1 token in COHA).

Corpus	Year range	Corpus size	*PT* tokens	per M
COCA	1990–2019	1.0 billion	1,189	1.19
COHA	1810–2009	400 million	1,189	2.97
Total	1810–2019	1.4 billion	2,378	1.70

***Table 1.** Occurrences of PT in COCA and COHA (after data cleaning)*

4.1.2 Data processing

After compiling the concordance lines, a number of steps were taken to standardize the data, extract information through automated processes, and prepare the files for manual annotation. To first clean the data, duplicate concordance lines were dropped. Where possible, automated linguistic analyses were carried out to reduce the need for manual annotation. The majority of these linguistic categories were form-related rather than

meaning-related, as form-related characteristics rely on the surface forms of the data and not inferencing.

Category	Options
Grammatical number	Singular / Plural
Placement in sentence	Start / Middle / End
Capitalization	True / False
Exclamation mark	True / False
Reduplication ('poor thing poor thing')	True / False
Article	Definite (`the`), Indefinite (`a/an`), No article (`Ø`)
This/That	*This* / *That* / Neither
Collocations	Tokens occurring up to two words on either side of *PT*

Table 2. *Automated processing of PT concordance lines for eight categories*

After tokenizing the data using the Natural Language Toolkit (NLTK) word tokenizer (Bird, Loper, & Klein 2009), eight categories were identified for each concordance line (Table 2). Here we note that for the collocation analysis, raw frequencies were used rather than other common collocation association measures such as Mutual Information. This decision was made because we were interested in co-occurrences with grammatical words such as pronouns (in addition to more lexical collocations) and because the relatively small number of total word combinations allowed for manual inspection of all instances.

4.1.3 Data annotation

A team of three annotators was responsible for annotating the concordance lines described above: two undergraduate research assistants (principal annotators) and one of the authors (adjudicator). Initially, the first 100 lines from COCA, representing the time period of 1990 to 2019 and a variety of genres, were used as a pilot study and as training data to standardize annotator responses. This pilot revealed that, in many cases, the variables of interest (e.g., speaker gender) were not apparent, yet with sufficient annotations it was determined that a reasonable dataset could be compiled. Next, these 100 training lines from COCA were excluded from the resulting analysis and Annotators 1 and 2 independently annotated the remaining 2,278 lines each (1,089 lines from COCA, and 1,189 lines from COHA), resulting in two complete annotations for each concordance line in

the final dataset.[1] It is this figure of 2,278 which is used throughout the remainder of the paper. Annotator 3 then acted as an adjudicator, annotating any items for which the original annotators had provided different responses. In every case of disagreement, the adjudicator's annotation corresponded with that of one of the two initial annotators, and therefore all data was maintained for analysis. Although the annotation process was labor-intensive and time-consuming, relying on a small number of trained annotators is considered preferable to a larger number of untrained ones (Bhardwaj *et al.*, 2010). Simple agreement rates were used to assess inter-annotator reliability (as in Hovy *et al.* 2006).

In total, there were 12 separate judgements that annotators made for each concordance line, for a variety of meaning-related aspects. These judgements pertained to the speaker (the person saying *PT*) and the referent (the entity being described as *PT*). Often, this information was not present, in which case annotators were instructed not to guess, but to use an 'unsure' option. Table 3 presents these 12 categories and the possible options for each (in addition to 'unsure'). Speaker and referent are presented together within a single cell for the categories of gender, occupation, and age. A sample of four concordance lines is provided in Figure 3. Additional methodological decisions are discussed where relevant throughout the analysis.

Categories	Options
Referent	1st / 2nd / 3rd person
Direct speech (e.g., quotations)	True / False
Speaker gender / Referent gender	Male / Female / Other
Speaker occupation / Referent occupation	Open category (e.g., *Teacher*)
Speaker age / Referent age	Infant, Young, Teen, Adult, Senior
Referent is human	Human / Animal / Object
Relationship (between Speaker and Referent)	Open category (e.g., *Family*)
Sarcasm (is the speaker being sarcastic?)	True / False
Vocative (e.g., 'Oh you poor thing')	True / False

Table 3. *Manual annotation of PT concordance lines for 12 categories*

1 In hindsight, we acknowledge that a better training set would have randomly sampled from both corpora and the complete range of years. Nonetheless, the final dataset is still well balanced, being 47.8% COCA and 52.2% COHA.

1	. - Do n't be ridiculous . It 's wriggling all over the place .	poor thing	, it 's fighting for its life . I do n't know what
2	some hometown girl . " The captain took a drink of his cola . "	poor thing	. She grew up in a place that sells tractors , then Stefan
3	Like would have been ejected if it had been a " real " game .	poor thing	did n't even know that was n't allowed . He also made two
4	Taking Lillian up the stairs to her apartment , Frieda said to Lillian , "	poor thing	, " and she gestured over her shoulder to a small room filled
5	. Fatima ? Fatima ? You let that	poor thing	out of there . That p**r th*ng 's got distemper . Fatima ! She does

Figure 3. *Example of concordance lines for poor thing*

4.2. Corpus findings

4.2.1 Genres

As evidenced in Table 4, the majority of tokens are found in fictional written texts. Although only 2% of tokens are from the spoken genre, many of the tokens are from speech or facsimiles of speech, such as dialogue in movies or novels.

Genre	Code	Tokens	% of tokens
Fiction	FIC	1476	64.8%
Movies	MOV	226	9.9%
Television	TV	219	9.6%
Magazines	MAG	120	5.3%
Blogs	BLOG	66	2.9%
Online	WEB	60	2.6%
Spoken	SPOK	45	2.0%
Non-fiction	NF	39	1.7%
News	NEWS	16	0.7%
Academic	ACAD	11	0.5%

Table 4. *PT tokens by genre*

4.2.2 Linguistic variables

The first group of form-related linguistic variables paint a consistent picture of the syntactic distribution of *PT* (Table 5). That the majority of *PT* referents are third person singular (3SG; 65.4%) indicates that referents are most often being discussed by others, whether they are present in the conversation or not. At a sentence level, *PT* shows a fairly even distribution in terms of placement, occurring sentence-initially (23.5%), -medially

(42.5%) and -finally (33.9%). However, nearly two-thirds of the time (65.9%) *PT* is set apart from the rest of the sentence as a clause-independent interjection or vocative (Zwicky, 1974), shown in (Excerpt 1), compared to non-vocative use as a noun phrase (NP) with some syntactic function within the clause (34.1%), shown in (Excerpt 2). Typically, vocative utterances do not use a determiner (67.4%), which is also clear in (Excerpt 1). Both examples come from COCA.

(1) her head to her shoulder. Her skin was so hot it frightened me. *Poor thing*, she was miserable. When she looked at me through those red (1996, FIC)

(2) him away? The matted coat? The dirt, maybe?" "The *poor thing* could use a bath," Paige admitted. She 'd always had (2010, FIC)

	1SG	2SG	2PL	3SG	3PL	Unknown
Person	8	395	21	1490	349	15
	0.4%	17.3%	0.9%	65.4%	15.3%	0.7%
	The	**A/An**	**This**	**That**	**None**	**Unknown**
Determiner	612	77	17	36	1536	0
	26.9%	3.4%	0.7%	1.6%	67.4%	0%
	Start	**Middle**	**End**			**Unknown**
Placement	536	969	773			0
	23.5%	42.5%	33.9%			0%
	Yes	**No**				**Unknown**
Vocative	1502	776				0
	65.9%	34.1%				0%

Table 5. *Form-related linguistic variables showing syntactic distribution of PT*

The next three categories relate to pragmatics and provide further evidence of the situational use of *PT* (Table 6). The majority of tokens were judged by annotators to be used in direct speech contexts, consistent with the typical conventions of spoken language and dialogue in written genres (82.9%). As might be expected of speech-like language, emphatic markers like *oh* were sometimes employed (18.6%), though the majority of tokens did not have emphatic markers (81.4%). The most frequent *PT* collocations in fact include two such markers: *oh* (16th most frequent) and repeated or reduplicated *PT* (23rd most frequent).

	Yes		No		Unknown	
Direct speech	1911	83.9%	316	13.9%	51	2.3%
Emphasis	430	18.9%	1848	81.1%	0	0%
Sarcasm	20	0.9%	2253	98.9%	5	0.2%

Table 6. *Pragmatic variables showing the situational use of PT*

A concordance with *oh* is shown in (Excerpt 3).

(3) my littlest has had nightmares ever since. She cries every night. Oh, *poor thing.* Ah, what are you gonna do? You got ta take (COCA, 1994, TV)

In terms of sarcasm, we found that the vast majority of tokens were used non-sarcastically or when the *PT* being referred to had actually suffered some misfortune (98.9%), compared to sarcastically (0.8%). A sarcastic use of *PT* is presented in (Excerpt 4) below, displayed in a vertical format to capture the dialogic nature of this *PT* instance (with shifts between speakers indicated by dashes).

(4) ? Amen to that. My husband's on Viagra.
— Oh, you poor thing!
— Every minute, he wants it. He has no right to (COCA, 2004, MOV)

4.2.3 Social variables

In shifting to social variables, it is important to note that the agreement rate among Annotators 1 and 2 was quite low for some variables. Before adjudication it was determined that 'uncertain' annotations, which are those in which one annotator noticed something but the second annotator left the cell blank, would be collapsed into the 'unknown' category. On the one hand, this decision limited the dataset overall. However, we judged a conservative approach to be most suitable for making claims about the indexical properties of *PT*. Therefore, the discussion reports on cases in which (a) Annotators 1 and 2 noticed something in the concordance line related to the category of interest and (b) the adjudicator (Annotator 3) agreed with either Annotator 1 or 2.

Despite the above limitations, certain key trends emerge from the data in Table 7. With respect to gender, there appears to be an imbalance in male and female speakers and referents, with *PT* being said by female speakers 14.4% of the time (compared to 5.2% for male speakers) and in reference to female entities 35.3% of the time (compared to 8.0% male). A chi-square test of significance was performed which confirmed the significance of these differences, using an expected baseline of 50% male and 50% female (Speaker: χ^2 (1, n = 893) = 51.7, p < .001, d = 0.5; Referent: χ^2 (1, n = 1,972) = 216.5, p < .001, d = 0.7). The most frequent *PT* collocations again provide supporting evidence: the feminine subject pronoun *she* and object pronoun *her* are the third and 14th most frequent collocates respectively, whereas *he* is the 24th and *him* is the 76th.

Although annotators also assessed concordance lines for information about speaker/referent age and relationship, in the vast majority of cases there was not enough information to make a clear determination (for relationship: 98.8% unknown; for speaker age 99.9% unknown; for referent age 98.9% unknown). Annotators were also instructed to code for information about speaker and referent roles if such information appeared in the concordance line.

	Female	Male				Unknown
Speaker gender	329	118				1831
	14.4%	5.2%				80.4%
Referent gender	804	182				1292
	35.3%	8.0%				56.7%
	Infant	**Young**	**Teen**	**Adult**	**Senior**	**Unknown**
Speaker age	0	0	0	0	3	2275
	0%	0%	0%	0.0%	0.1%	99.9%
Referent age	7	9	2	2	5	2253
	0.3%	0.4%	>0.1%	>0.1%	0.2%	98.9%
	Family	**Owner/Pet**	**Caretaker /Patient**	**Friend**		**Unknown**
Relationship	16	6	4	1		2251
	0.7%	0.3%	0.2%	>0.1%		98.8%
Humanness	**Human (alive)**	**Human (dead)**	**Animal (alive)**	**Animal (dead)**	**Object**	**Unknown**
	1851	65	165	25	72	100
	81.3%	2.9%	7.2%	1.1%	3.2%	4.2%

Table 7. *Social variables of PT relating to gender, age, relationship, and humanness*

In processing the annotation data, the authors collapsed the role terms provided by each annotator into a closed set of 25 categories, such as animal, child, concept, family member, or partner. We discovered that role categories provided useful information about the relationship between speaker and referent as well as the relative age of each, with such information being tied directly to details from the concordance line. Table 8 presents a summary of the five most common speaker and referent roles when known. It should be

noted that for the majority of tokens, there was not enough information to determine role category (89.6% unknown). The following observations therefore relate to only the 10.4% of cases in which a determination about role category could be made.

Speaker		Referent	
Family member	19	Animal	91
Job:Government	5	Partner	37
Partner	5	Child	26
Job:Healthcare	3	Family member	23
Job:Misc	3	Nature	8
Total known	**39**	**Total known**	**241**

***Table 8**. Five most common speaker and referent role categories*

Focusing on speaker, the largest determinable role category was family member (19), which included traditional kinship items like *mother* (11) and *grandmother* (5). Apart from the non-family speaker role categories presented in Table 8, the remaining determinable speaker role categories occurred infrequently (fewer than 10 times) and are not considered further. More useful patterns about frequency emerge for role categories of *PT* referents, with a varied set of role categories for *PT* referents occurring infrequently for entities like body part, concept, occupation, food items, or vehicle. More frequent role categories for *PT* referents include animal (91), and familiar and romantic relationships (partner, $n = 37$; child, $n = 27$; family member, $n = 23$).

4.2.4 Correlations between variables

All social and linguistic variables were checked for possible correlations, and those with significant correlations ($p < .05$) were included in a correlation matrix (Figure 4). To compare the categorical variables, each possible choice for each of the categorical variables was transformed using dummy variables into numerical values, i.e., True = 1, False = 0 (Levshina, 2015). Circle size and color intensity represent magnitude (i.e., the value of correlation coefficients), and the scale is from -1 (negative correlations in red) to 1 (positive correlations in blue).

In Figure 4, some strong correlations are self-explanatory,[2] for example absolute negative correlations for mutually exclusive categories like the definite determiner the (Determiner_Definite) and the absence of a determiner (Determiner_None). Other, less ob-

2 Information about year was included for each *PT* instance in the dataset, but because we do not do a diachronic analysis, we do not consider year as a variable here. Note that in Figure 4, Genre_TV.Film is positively correlated with COCA due to the recency of this corpus (1990–2019), whereas Genre_Fiction is positively correlated with COHA due to the historical nature of this corpus (1820–2019).

vious correlations strengthen our earlier observations about the syntactic context of *PT* at the sentence level. For vocatives, emphasis, and 2SG context, there is a significant positive correlation with the start and end of sentences, and a significant negative correlation for the middle of the sentence.

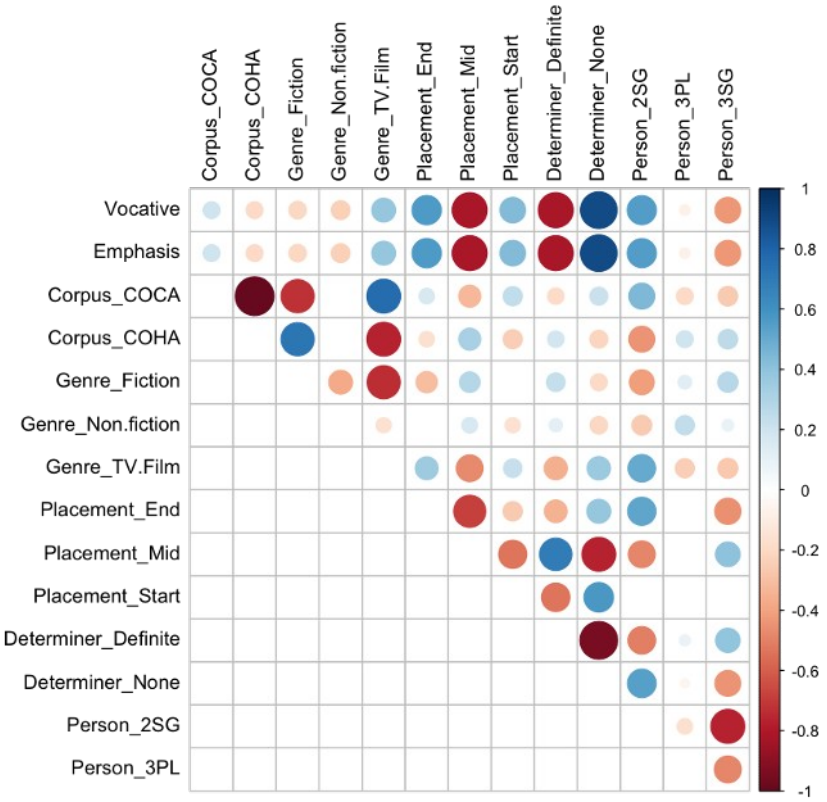

Figure 4. *Correlations of social and linguistic variables for PT*

Notably, no significant correlations were found involving the social variables. Taken together, these findings reflect that we can expect speakers to use *PT* when speaking directly to another person, often with emphatic markers and prominent placement at the start or end of utterances.

4.2.5 Summary of corpus findings

In response to our first research question, the preceding corpus analysis revealed that *PT*, as it appears in two American English corpora, has a number of identifiable linguistic and social properties. The MWE often appears as a vocative, independent from the clause structure of the sentence, suggesting that it is more likely to occur in informal registers of speech or speech-like writing. Additionally, it appears most frequently in conversation about third singular referents, suggesting that a speaker who says *PT* is often talking about a referent that is not immediately present or, at the very least, not addressed.

In terms of social variables, the corpus data indicate that in the vast number of cases a 30-word concordance line does not offer enough information to determine the social

identities of speaker or referent. However, in the sample of instances where clear determinations can be made, the expression is often used by women, in reference to women, and with a large number of references to children and animals. While these global observations offer insight into the indirect indexical meanings of the expression (Ochs, 1992), they tell us little about the possible source of these associations, which are constructed and reconstructed constantly in discourse. We now shift to the second part of our analysis, a critical discourse analysis of *PT* corpus tokens, to better understand the expression's stance-relevant functions.

5. Critical discourse analysis (CDA)

5.1. CDA approach

For the second part of our analysis, we use CDA to delve deeper into the stance function of *PT* in discourse. Through CDA, we also hope to develop a better sense of how the (indirect) indexical properties of *PT* suggested through the corpus analysis might somehow be consequential of the MWE's functions in discourse, with an eye toward the identification of power relations, as is customary in CDA work (van Dijk, 2008). Our approach to CDA mirrors the approach used by Baker (2006, 2008) in his analysis of the English words bachelor and spinster. We use the corpus instances of *PT* (described in Section 4) as a way into understanding key dimensions of affective stancetaking (Kiesling, 2022): evaluations of a *PT* referent or addressee, the positioning of a speaker as compassionate, and the possibility for disalignment between an evaluating speaker and an evaluated *PT* through the latter's meta-linguistic awareness. CDA also helps situate a hinted-at indexical association in the corpus analysis between *PT*s and low vitality due to illness or death, as we will show. While we must view decontextualized corpus excerpts with some caution, and we recognize that the wider co-text of each *PT* instance may supply useful contextual information for further analysis,[3] our goal is to start local by attending to the immediate information within each concordance line and gradually build a set of observations that are shared across concordance lines. Because CDA is well suited for uncovering the workings of power and ideology in texts across a variety of written or spoken genres (Fairclough, 2013; van Dijk, 2015), it is ideal for our dataset.

5.2. CDA findings

5.2.1 Disadvantage and misfortune in referential context

Excerpt 5 indicates that speakers who refer to another entity as *PT* are evaluating a referent's conditions of existence, which may lead to the inference that a referent is disadvantaged or unfortunate.

(5) I wanted to show you my mama. She's an old lady. *Poor thing* ... I'm taking care of her. She's sleeping. Very (COCA, 2002, MOV)

3 We thank the editor for this helpful comment.

In Excerpt 5, the speaker uses the *PT* expression to describe their mother, whom they are speaking about with an unspecified addressee. Additional information about the film's plot, this particular scene, and characters involved would offer even more context for understanding the use of *PT* in this concordance line. However, the immediate discourse context provides enough information to ascertain not only that a *PT* referent is disadvantaged but also the source of their misfortune: being 'an old lady' who the speaker is 'taking care of'. Similarly, Excerpt 6 presents an instance of the immediate context offering insight into the discourse function of *PT*. The narrator describes how Lady Maccon 'pursed her lips' and 'looked down at her daughter', who is considered a *PT* for inheriting 'her mother's complexion and curly hair'.

(6) # Lady Maccon pursed her lips and looked down at her daughter. The *poor thing* had inherited her mother's complexion and curly hair. Alexia hoped the nose (COCA, 2012, FIC)

The preceding two excerpts, though brief, demonstrate that the immediate discourse context may provide enough information to establish a referent's misfortune. Additional examples provided in Concordance 1 suggest that misfortune is at the heart of *PT* reference across concordance lines. Aspects of discourse indicating misfortune and surrounding the *PT* token are underlined.

1	get up at all, anymore. Monahan *had to shoot him* to put the	poor thing	*out of his misery*. A whole decade later, he still felt
2	better (around 10 months). But it got *ugly* last night. #	poor thing	, she just *kept pushing and then crying* and them screaming. So
3	so rude. You think your accents don't sound funny to other people?	poor thing	. *Terrible, terrible tragedy*. She was supposed to have been married
4	at your campaign HQ licking envelopes. She 'd do it herself, but the	poor thing	*can't produce saliva*. -Hello. -You agreed to keep
5	Me? What'd I do? I didn't do anything. Oh,	poor thing	. Did Dr. Fleischman *frighten* you? Huh? Come on.
6	re exhausted. And one of them needs to see a vet immediately. The	poor thing	has a *bleeding* nose. ""Ms Bradshawe ... " Johannsen began
7	) (m) Take her! Unhand her! Let her be! The	poor thing	has *gone mad*. Let her be! Get her out of my

***Concordance 1**: Sample of occurrences indicating misfortune*

The sample of occurrences in Concordance 1 indicate that the misfortune implied by *PT* can have a variety of sources (being shot in 1.1, not producing saliva in 1.4, or having a bleeding nose in 1.6) and assumes different forms tied to affective expression (crying in 1.2, being frightened in 1.5, or going mad in 1.7). In all cases, speakers mobilize context to determine that an entity meets the requirements to be called 'poor thing'.

5.2.2 *Vitality*: a special case of referential context

Next, we discuss a special case of referential context we refer to as *vitality*, which encompasses two related continua: living vs. dead and, if living, healthy vs. sick. Vitality was not a variable we specifically coded for in the human annotation portion of the corpus analysis. Instead, it was hinted at through annotator comments about a referent's sickness or roles like 'hospital patient'. Although in no case did the two principal annotators independently agree upon the vitality of a referent, prior to adjudication the annotators noted more than 100 instances where a referent's vitality was relevant in the concordance line.

We take this as infrequently occurring but nonetheless useful corpus evidence that *PT* may have an indexical association related to a referent's vitality. Excerpt 7 looks at one instance of a referent's vitality serving as the context for establishing misfortune.

(7) Details fill the letters: 'four deluges of vomiting she has had today -*poor thing* --'; 'you could not in the least recognize her with' (COCA, 1995, ACAD)

In Excerpt 7, a female referent is called *PT* with a syntactically-independent vocative after a description of her illness ('four deluges of vomiting'). Here the MWE appears in speech-like written dialogue offset by single quotation marks. This is evidence that physical illness is one form of low vitality that may render one a *PT*.

(8) lost. At nightfall I found your horses and ponchos -- the horses was dead, *poor things*. I slept on the desert that night, and the next mornin (COHA, 1898, FIC)

In Excerpt 8, horses are 'poor things' because they are dead. Additionally, Table 7 showed that the proportion of dead to alive animals is greater than the proportion of dead to alive humans (25:165 or 13% for animals vs. 65:1851 or 3% for humans), suggesting that when *PT* refers to an animal, vitality is more likely to be a factor. Additional examples of *PT* being associated with low vitality are provided in Concordance 2.

1	, Daddy ... Emmeline. Oh, my goodness, Dr. Hart. You	poor thing	, you *look so tired*. Sort of *puffy-like*. Thank you.
2	facts ... .do our job. Thanks. Good luck with everything. You	poor thing	, sweetheart. Just sit tight. I'll get you some *aspirin*
3	got *rheumatoid arthritis and kidney failure*. Has to get *dialysis twice a week,*	poor thing	. Oh , yeah. Give it to me, Earl. Give
4	na be all right. -It does look *pretty infected* though. -You	poor thing	. You don't understand. There were things out there that ...
5	a beet. And hot like an oven. You 're *burning with fever,*	poor thing	. Why didn't you say something? Sit, sit on the
6	Nasyrova shrugged. #The dog, too, *hadn't been feeling well,*	poor thing	. # `*It was sick*,` she told The Post.
7	unless they are fricasseed. He shot at our little runaway pullet, and the	poor thing	came home *dragging a broken and useless leg*. Now, if any

***Concordance 2**: Sample of occurrences indicating low vitality*

Low vitality may be generalized, having to do with looking tired (2.1), needing aspirin (2.2), having an infection (2.4), or not feeling well (2.6). Or it may be specific, due to having rheumatoid arthritis and kidney failure (2.3), having a fever (2.5), or dragging a broken and useless leg (2.7).

5.2.3 Compassionate positioning by the speaker

Leveraging discourse context to establish an entity's misfortune in service of *PT* reference has the simultaneous effect of putting the speaker in a complementary position as someone able to evaluate another as disadvantaged. Excerpt 9 offers a particularly explicit instance of a speaker's positionality as an evaluating subject being established.

(9) her in a detached way. She felt sorry, dreadfully sorry, for the *poor things*; but as she could not help them she dismissed them from her thoughts (COHA, 1920, FIC)

In this example, the author provides information about the inner thoughts and feelings of the unnamed female character ('she'), who is described as feeling 'sorry, dreadfully

sorry' for the objects of her sympathy. Other examples of compassionate positioning toward *PT* referents are provided in Concordance 3.

1	if you know him, this is his daughter. Oh my God! You	poor thing	. I'm so *sorry* about your dad. Thank you. We
2	. It's about *helping him*, since he can't possibly win alone,	poor thing	. *After all, how can the Democratic Party survive -- and win*
3	girl, Stuart 's mother said, whom *she felt the most sorry for.*	Poor thing	, raised by hippies, not even given the comfort of a Christian
4	be laying around in front of the TV any more than she already does.	poor thing	, *bless her heart,* she's gonna have a rough g
5	# Oh how I love a Denny's breakfast. # Anonymous # Oh you	poor thing	, *I feel for you.* # Anonymous # i have sobered up

***Concordance 3**: Sample of occurrences indicating compassionate stancetaking*

In effect, because a *PT* utterance is both the result of a speaker's evaluation and the means of positioning the same speaker as capable of expressing sympathy, it comes to index a speaker's compassionate stance, a concept we return to in our discussion. For now, Excerpt 8 and the examples in Concordance 3 all indicate that *PT* helps to establish a speaker's compassion, sympathy, or pity for a disadvantaged referent. *PT* may co-occur with other indexes of compassion, such as apologies (3.1, 3.3), expressions of desire to help or support (3.2, 3.5), or conventionalized expressions like 'bless her heart' (3.4).

5.2.4 Metalinguistic awareness

As a final example, Excerpt 10 provides evidence that individuals may possess metalinguistic awareness about *PT* as an index of a speaker's compassionate positioning and a referent's misfortune.

(10) , some men can be so mean." # "Don't you '*poor thing*' me, you stupid bimbo!" Elli barked. "If it (COCA, 2011, FIC)

Elli 'barks' at their interlocutor for calling them 'poor thing', a statement punctuated with the pejorative insult 'you stupid bimbo'. The immediate discourse context of this concordance line is evidence enough that Elli recognizes the social-indexical meaning being conveyed when one speaker refers to another as *PT* and goes on to resist their classification. Metalinguistic awareness of *PT* is likewise encountered in the OED definition for 'poor-thing' (i.e., 'to "poor thing" someone' as a verb form, Oxford University Press, 2022).

5.2.5 Summary of CDA findings

In answer to our second research question, *PT* signals a speaker's mobilization of discourse context to establish a referent's misfortune. The qualitative analysis additionally supported an insight first discovered through the corpus analysis: a special instance of referential context related to vitality, with *PT* occasionally being used in reference to sick or dead entities.

6. Discussion

We now return to the notions of affective stancetaking (Jaffe, 2009; Kiesling, 2022) and indexicality (Ochs, 1992; Silverstein, 2003) introduced previously to model affective stancetaking by way of the stance triangle (Du Bois, 2007). Combining the insights from our corpus analysis with the discourse analysis leads to a view of the social meaning of *PT* being discursively established in the following way. First, a speaker mobilizes discourse context to evaluate an entity (the stance object) as unfortunate or disadvantaged. Our discourse analysis indicates that the immediate context of the concordance line, consisting of 15 words before and after the *PT* token, may provide enough information to establish a referent's misfortune and justify *PT* reference by a compassionate speaker. Mapped onto the stance triangle (Du Bois, 2007) in Figure 5, Excerpts 5–9 (as well as Concordances 1–3) provide enough information to make claims about the axis connecting speaker 1 to the *PT* referent. As per the figure, speaker 1 ascribes misfortune to the *PT* referent (1), which in turn allows the speaker to take up a compassionate stance position (1a).

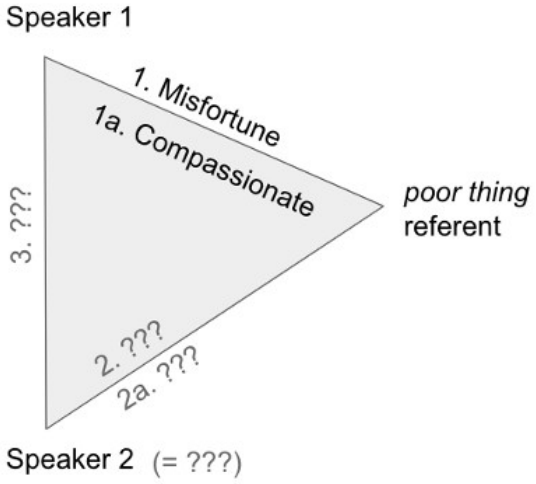

Figure 5. *Compassionate stancetaking connecting Speaker 1 to PT, as shown in Excerpts 5-9, Concordances 1-3*

However, the concordance lines in these examples do not provide any information about the evaluation and positioning of a second speaker. The example in Excerpt 10 allows for a fuller conception of the stance triangle, as depicted in Figure 6, where referential misfortune and compassionate stance are still being established by the speaker. However, aided by metalinguistic awareness, the *PT* referent, who is also speaker 2, resists the evaluation of speaker 1 and their resulting classification (2a in Figure 6). In negatively evaluating their own positioning as a disadvantaged *PT* referent (2), speaker 2 produces a stance disalignment with speaker 1 (3).

Following Kulick (2003), we note that the enunciation (or not) of particular utterances by particular kinds of speakers in specific discursive environments materialize performative subject positions. Speaker 1's affective stancetaking in relation to a *PT* referent produces two related subjectivities in discourse: the compassionate speaker and the dis-

advantaged referent. As discussed previously, Ochs (1992) has argued that the mapping between linguistic forms like *PT* and various kinds of social meaning is non-exclusive, constitutive, and temporally transcendent. Our analysis finds that compassionate stancetaking is an emergent phenomenon, one that is specific to an immediate discourse context and which entails responsiveness to context by a speaker (i.e., it is constitutive). However, our quantitative analysis of two corpora also finds that several indexical associations have become temporally transcendent, appearing across different time periods.

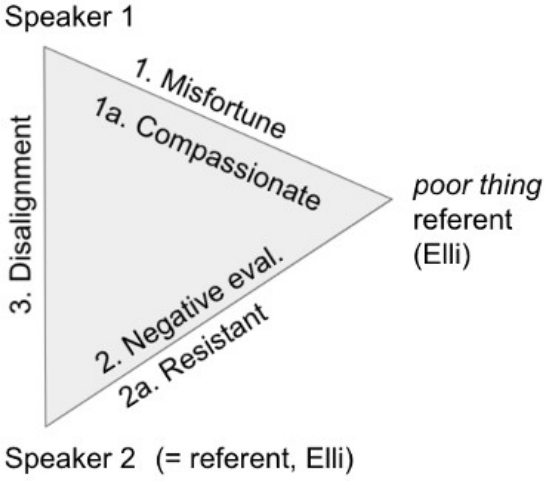

Figure 6. *Resistance by a PT referent in Excerpt 10*

Although it is true that any (once-)living or (once-)sentient being can be selected through *PT* reference, our analysis finds that, when it can be inferred, 'poor things' are typically women, intimate relations (including children), animals, or entities of low vitality (see Table 7). These speaker and referent positionalities emerge through stancetaking and index information about the sociocultural world, helping to constitute the indexical meanings of this MWE.

In discussing what it means to analyze discourse with the help of corpus methods, Baker (2008, following Stubbs 2001), contends that lexical expressions may invoke cultural stereotypes formed through the accumulation of similar such encounters with language across time and texts.[4] While it is beyond the scope of the current investigation to offer a detailed explanation of why women, intimates/children, animals, and entities of low vitality are bundled together in the indexical field for *PT*, we suggest that ideologies of disempowerment are at play in these associations. Decades of scholarship demonstrate that women and children in particular are viewed as disempowered (Cameron, 2007; Eckert & McConnell-Ginet, 2013; Hadodo & Kanwit, 2020; Lakoff, 2004 [1975]; Ochs & Schieffelin, 2011). As critical discourse analysis has long demonstrated, power is a key di-

4 A related point is made by Fairclough (1989): 'A single text on its own is quite insignificant: the effects of media power are cumulative, working through the repetition of particular ways of handling causality and agency, particular ways of positioning the reader' (p. 54). Thank you to the editor for this helpful suggestion.

mension of the material and symbolic production of text and talk (van Dijk, 2008), being ubiquitous and operating at various scales of social life while always entailing forms of resistance (Cameron & Kulick, 2003, p. 112). Thankfully, our analysis is a hopeful reminder that in corpus data, reflective as it is of everyday life, evidence is likely to be found that (stereotypically) 'disempowered' subjects contest their categorization through language and, in so doing, speak back to local articulations of power.

7. Conclusion

We conclude first by noting the limitations of our analysis. As previously mentioned, we did not conduct a diachronic analysis, but future work on historical changes in *PT* may well find shifts in affective stance or indexical associations over time. Furthermore, many of the *PT* tokens in this dataset are underspecified for social information related to our main variables of interest (gender, age, humanness, and vitality). Were we to look in closer detail at the actual textual sources of the 2,278 examples, it is possible that a fuller picture of the social meaning of *PT* would begin to emerge. However, our conservative approach to data processing and annotation produced a selection of evidence we can be confident about in basing our assertions, namely that *PT* indexes a speaker's compassionate stance and referent's misfortune. Further, as stancetaking is essential to the social meaning of language (Kiesling, 2022), our analysis finds that it is through the mechanism of stancetaking that a set of 'usual suspects' (i.e., typical referents) come to be indexically linked, possibly through the help of still-dominant ideologies about disempowerment. Our analysis serves as another instance of mixed-methods synergy between CL and CDA, which is quite suitable for investigating patterns that occur relatively infrequently in discourse but nonetheless have clear and identifiable social meaning to speakers.

Acknowledgements

We thank undergraduate research assistants Vanessa Todd and Nathan Andrud for their hard work annotating the *poor thing* dataset. We also thank Nonnenmacher's Spring 2020 Introduction to Linguistics course for their initial 'poor thing' judgements, the University of Pittsburgh Department of Linguistics Graduate Student Writing Workshop members, and the Social Meaning in Language (SMiL) Reading Group members for helpful feedback on an earlier draft of this paper. We thank the editor and two anonymous reviewers of the *Journal of Corpora and Discourse Studies* for their helpful comments. All errors are, of course, our own.

Funding

This research was supported in part by funding from the Social Sciences and Humanities Research Council of Canada doctoral fellowship (award number 752-2019-0006).

Competing interests

The authors have no competing interests to declare.

References

Agha, A. (2003). The social life of cultural value. *Language & Communication, 23*(3–4), 231–273. doi:10.1016/S0271-5309(03)00012-0

Ahmed, S. (2014). *The cultural politics of emotion* (2nd ed.). Edinburgh: Edinburgh University Press.

Babel, A. (2018). *Between the Andes and the Amazon: Language and social meaning in Bolivia.* Tucson: The University of Arizona Press. doi:10.2307/j.ctt1zxsmkv

Baker, P. (2006). *Using corpora in discourse analysis.* London: Continuum. doi:10.5040/9781350933996

Baker, P. (2008). 'Eligible' bachelors and 'frustrated' spinsters: Corpus linguistics, gender and language. In K. Harrington, L. Litosseliti, H. Sauntson, & J. Sunderland (Eds.), *Gender and Language Research Methodologies* (pp. 73–84). New York: Palgrave Macmillan.

Baker, P., Gabrielatos, C., KhosraviNik, M., Krzyżanowski, M., McEnery, T., & Wodak, R. (2008). A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. *Discourse & Society, 19*(3), 273–306. doi:10.1177/0957926508088962

Besnier, N. (1990). Language and affect. *Annual Review of Anthropology, 19*(1), 419–451. doi:10.1146/annurev.an.19.100190.002223

Bhardwaj, V., Passonneau, R. J., Salleb-Aouissi, A., & Ide, N. (2010, July). Anveshan: A framework for analysis of multiple annotators' labeling behavior. *Proceedings of the Fourth Linguistic Annotation Workshop, ACL 2010* (pp. 47–55). Uppsala, Sweden.

Biber, D., & Finegan, E. (1989). Styles of stance in English: Lexical and grammatical marking of evidentiality and affect. *Text, 9*(1), 93–124. doi:10.1515/text.1.1989.9.1.93

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python.* Sebastopol, CA: O'Reilly Media Inc.

Bryman, A., & Burgess, R. G. (Eds.). (1994). *Analyzing qualitative data.* London: Routledge. doi:10.4324/9780203413081

Bucholtz, M. (2009). From stance to style: Gender, interaction, and indexicality in Mexican immigrant youth slang. In A. Jaffe (Ed.), *Stance: Sociolinguistic perspectives*

(pp. 146–170). Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780195331646.003.0007

Bucholtz, M., & Hall, K. (2008). All of the above: New coalitions in sociocultural linguistics. *Journal of Sociolinguistics, 12*(4), 401–431. doi:10.1111/j.1467-9841.2008.00382.x

Burton, T. (2007). *Sweeney Todd: The demon barber of Fleet Street* [Film]. DreamWorks Pictures.

Cameron, D. (2007). *The myth of Mars and Venus*. Oxford: Oxford University Press.

Cameron, D., & Kulick, D. (2003). *Language and Sexuality*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511791178

Cole, A., & Tieken-Boon van Ostade, I. (2021). *Haagse Harry*, a Dutch chav from The Hague? The enregisterment of similar social personas in different speech communities. *International Journal of Language and Culture, 9*(1), 72–96. doi:10.1075/ijolc.21040.col

Cresswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2nd ed.). London: Sage.

Davies, M. (2008–). *The Corpus of Contemporary American English (COCA): One billion words, 1990–2019*. Retrieved from https://www.english-corpora.org/coca/

Davies, M. (2010). *The Corpus of Historical American English (COHA): 400 million words, 1810–2009.* Retrieved from https://www.english-corpora.org/coha/

Du Bois, J. W. (2007). The stance triangle. In R. Englebretson (Ed.), *Stancetaking in discourse: Subjectivity, evaluation, interaction* (pp. 139–182). Amsterdam: John Benjamins. doi:10.1075/pbns.164.07du

Du Bois, J. W., & Kärkkäinen, E. (2012). Taking a stance on emotion: Affect, sequence, and intersubjectivity in dialogic interaction. *Text & Talk, 32*(4), 433–451. doi:10.1515/text-2012-0021

Eckert, P. (2008). Variation and the indexical field. *Journal of Sociolinguistics, 12*(4), 453–476. doi:10.1111/j.1467-9841.2008.00374.x

Eckert, P. (2019). The limits of meaning: Social indexicality, variation, and the cline of interiority. *Language, 95*(4), 751–776. doi:10.1353/lan.2019.0072

Eckert, P., & McConnell-Ginet, S. (2013). *Language and Gender* (2nd ed.). Cambridge: Cambridge University Press. doi:10.1017/CBO9781139245883

Egbert, J., & Baker, P. (2020). Synthesis and conclusion. In J. Egbert & P. Baker (Eds.), *Using corpus methods to triangulate linguistic analysis* (pp. 268–282). New York: Routledge. doi:10.4324/9781315112466-11

Ehrlich, S., & Romaniuk, T. (2013). Discourse analysis. In R. J. Podesva & D. Sharma (Eds.), *Research methods in linguistics* (pp. 460–493). Cambridge: Cambridge University Press. doi:10.1017/CBO9781139013734.024

Fairclough, N. (1989). *Language and power*. London: Longman.

Fairclough, N. (2013). *Critical discourse analysis: The critical study of language* (2nd ed.). London: Routledge. doi:10.4324/9781315834368

Gladkova, A., & Romero-Trillo, J. (2021a). Is ugliness in the mind of the beholder?: The conceptualization of 'ugly'in English. *International Journal of Language and Culture, 8*(1), 106–127. doi:10.1075/ijolc.00037.gla

Gladkova, A., & Romero-Trillo, J. (2021b). The linguistic conceptualization in folk aesthetics: Past, present and future. *International Journal of Language and Culture, 8*(1), 1–13. doi:10.1075/ijolc.00032.int

Goodwin, C. (2007). Participation, stance and affect in the organization of activities. *Discourse & Society, 18*(1), 53–73. doi:10.1177/0957926507069457

Goodwin, M. H., Cekaite, A., & Goodwin, C. (2012). Emotion as stance. In A. Peräkylä & M. Sorjonen (Eds.), *Emotion in Interaction* (pp. 16–41). Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199730735.003.0002PMid:22590847

Google Ngram. (2021). Search results for "poor thing". Retrieved from https://www-oed-com.pitt.idm.oclc.org/view/Entry/237584?redirectedFrom=poor+thing&

Hadodo, M. J., & Kanwit, M. (2020). *Menos masculino, demasiado infantil*: How age and gender ideologies index sexuality in diminutive perception in Madrid Spanish. *Journal of Language and Sexuality, 9*(2), 127–151. doi:10.1075/jls.19013.had

Hall, K. (2019). Middle class timelines: Ethnic humor and sexual modernity in Delhi. *Language in Society, 48*(4), 491–517. doi:10.1017/S0047404519000435

Hashemi, M. R. (2012). Reflections on mixing methods in applied linguistics research. *Applied Linguistics, 33*(2), 206–212. doi:10.1093/applin/ams008

Hovy, E. H., Marcus, M., Palmer, M., Ramshaw, L., & Weischedel, R. (2006). OntoNotes: The 90% solution. *Proceedings of the Human Language Technology Conference, NAACL 2006* (pp. 57–60). New York, NY. doi:10.3115/1614049.1614064

Hyland, K. (2005). *Metadiscourse: Exploring interaction in writing.* New York: Bloomsbury.

Irvine, J., & Gal, S. (2000). Language ideology and linguistic differentiation. In P. V. Kroskrity (Ed.), *Regimes of language: Ideologies, polities, and identities* (pp. 35–84). Santa Fe, NM: School of American Research Press.

Jackendoff, R. (1995). The boundaries of the lexicon. In M. Everaert, E. van der Linden, A. Schenk, & R. Schreuder (Eds.), *Idioms: Structural and psychological perspectives* (pp. 133–165). New York: Lawrence Erlbaum.

Jaffe, A. (Ed.). (2009). *Stance: Sociolinguistic perspectives.* Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780195331646.001.0001

Kiesling, S. F. (n.d.). *Scott F. Kiesling* [Figures]. Retrieved from https://sfkiesling.com/figures/creative-commons-figures

Kiesling, S. F. (2004). Dude. *American Speech, 79*(3), 281–305. doi:10.1215/00031283-79-3-281

Kiesling, S. F. (2018). Masculine stances and the linguistics of affect: On masculine ease. *Norma, 13*(3–4), 191–212. doi:10.1080/18902138.2018.1431756

Kiesling, S. F. (2022). Stance and stancetaking. *Annual Review of Linguistics, 8*, 409–426. doi:10.1146/annurev-linguistics-031120-121256

Kiesling S. F., Pavalanathan, U., Fitzpatrick, J., Han, X., & Eisenstein, J. (2018). Interactional stancetaking in online forums. *Computational Linguistics, 44*(4), 683–718. doi:10.1162/coli_a_00334

Kulick, D. (2003). No. *Language and Communication, 23*(2), 139–151. doi:10.1016/S0271-5309(02)00043-5

Labov, W. (1972). *Sociolinguistic patterns.* Philadelphia, PA: University of Pennsylvania Press.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind.* Chicago, IL: The University of Chicago Press. doi:10.7208/chicago/9780226471013.001.0001

Lakoff, R. T. (2004 [1975]). Language and woman's place: The original text with annotations by the author. In M. Bucholtz (Ed.), *Language and woman's place: Text and commentaries* (pp. 39–103). Oxford: Oxford University Press.

Lakoff, R. T. (2015). Nine ways of looking at apologies: The necessity for interdisciplinary theory and method in discourse analysis. In D. Tannen, H. E. Hamilton, & D. Schiffrin (Eds.), *The handbook of discourse analysis* (2nd ed., Vol. 1–2, pp. 293–308). Malden, MA: Wiley-Blackwell. doi:10.1002/9781118584194.ch13

Langacker, R. W. (1987). *Foundations of cognitive grammar: Theoretical prerequisites* (Vol. 1). Stanford: Stanford University Press.

Levshina, N. (2015). *How to do linguistics with R: Data exploration and statistical analysis.* Amsterdam: John Benjamins. doi:10.1075/z.195

Martin, J. (2000). Beyond exchange: Appraisal systems in English. In S. Hunston & G. Thompson (Eds.), *Evaluation in text: Authorial stance and the construction of discourse* (pp. 142–175). Oxford: Oxford University Press.

Mautner, G. (2019). A research note on corpora and discourse: Points to ponder in research design. *Journal of Corpora and Discourse Studies, 2*, 1–13. doi:10.18573/jcads.32

McNeill, P. (1990). *Research methods* (2nd ed.). London: Routledge. doi:10.4324/9780203133729

Milani, T. M., & Richardson, J. E. (2020). Discourse and affect. *Social Semiotics, 31*(5), 671–676. doi:10.1080/10350330.2020.1810553

Monteiro, K. R., Crossley, S. A., & Kyle, K. (2020). In search of new benchmarks: Using L2 lexical frequency and contextual diversity indices to assess second language writing. *Applied Linguistics, 41*(2), 280–300. doi:10.1093/applin/amy056

Naismith, B., & Juffs, A. (2021). Finding the sweet spot: Learners' productive knowledge of mid-frequency lexical items. *Language Teaching Research.* doi:10.1177/13621688211020412

Newby, H. (1977). In the field: Reflections on the study of Suffolk farm workers. In C. Bell & H. Newby (Eds.), *Doing Sociological Research* (pp. 108–129). New York: The Free Press.

Ochs, E. (1992). Indexing gender. In A. Duranti & C. Goodwin (Eds.), *Rethinking context: Language as an interactive phenomenon* (pp. 335–358). Cambridge: Cambridge University Press.

Ochs, E. & Schieffelin, B. (1989). Language has a heart. *Text, 9*(1), 7–25. doi:10.1515/text.1.1989.9.1.7

Ochs, E. & Schieffelin, B. B. (2011). The theory of language socialization. In A. Duranti, E. Ochs, & B. B. Schieffelin (Eds.), *The handbook of language socialization* (pp. 1–21). Malden, MA: Wiley-Blackwell. doi:10.1002/9781444342901.ch1

Oxford University Press (2022). 'Poor-thing, v.' In *OED Online*. Retrieved from http://www.oed.com/view/Entry/237584

Partington, A. (2010). Modern diachronic corpus-assisted discourse studies (MD-CADS) on UK newspapers: An overview of the project. *Corpora, 5*(2), 83–108. doi:10.3366/cor.2010.0101

Poole, R., & Hayes, N. (2022). Stance in climate science: A diachronic analysis of epistemic stance features in IPCC physical science reports. *Journal of Corpora and Discourse Studies, 5*, 37–60. doi:10.18573/jcads.100

Silverstein, M. (2003). Indexical order and the dialectics of sociolinguistic life. *Language & Communication, 23*(3–4), 193–229. doi:10.1016/S0271-5309(03)00013-2

Stubbs, M. (2001). *Words and phrases: Corpus studies of lexical semantics*. Oxford: Blackwell.

Tagliamonte, S., & D'Arcy, A. (2005). When people say, "I Was Like...": The quotative system in Canadian youth. *University of Pennsylvania Working Papers in Linguistics, 10*(2), 20. Retrieved from https://repository.upenn.edu/pwpl/vol10/iss2/20

Tannen, D. (1990). Discourse analysis: The excitement of diversity. *Text, 10*(1–2), 109–111. doi:10.1515/text.1.1990.10.1-2.109

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition.* Cambridge, MA: Harvard University Press.

Tognini-Bonelli, E. (2001). *Corpus linguistics at work.* Amsterdam: John Benjamins. doi:10.1075/scl.6

van Dijk, T. (2009). Multidisciplinary CDA: A plea for diversity. In R. Wodak & M. Meyer (Eds.), *Methods of critical discourse analysis* (2nd ed., pp. 95–120). London: Sage. doi:10.4135/9780857028020.n5

van Dijk, T. A. (2008). *Discourse and Power*. New York: Palgrave Macmillan. doi:10.1007/978-1-137-07299-3

van Dijk, T. A. (2015). Critical discourse analysis. In D. Tannen, H. E. Hamilton, & D. Schiffrin (Eds.), *The handbook of discourse analysis* (2nd ed., Vol. 1–2, pp. 466–485). Malden, MA: Wiley-Blackwell. doi:10.1002/9781118584194.ch22

Villarreal, D., Clark, L., Hay, J., & Watson, K. (2020). From categories to gradience: Auto-coding sociophonetic variation with random forests. *Laboratory Phonology: Journal of the Association for Laboratory Phonology, 11*(1), 1–31. doi:10.5334/labphon.216

Wetherell, M. (2013). Affect and discourse – What's the problem? From affect as excess to affective/discursive practice. *Subjectivity, 6*(4), 349–368. doi:10.1057/sub.2013.13

Wodak, R., & Meyer, M. (2009). Critical discourse analysis: History, agenda, theory and methodology. In R. Wodak & M. Meyer (Eds.), *Methods of critical discourse analysis* (2nd ed., pp. 1–33). London: Sage. doi:10.1080/17405900903180954

Zwicky, A. M. (1974). Hey, Whatsyourname! In M. W. La Galy, R. A. Fox, & A. Bruck (Eds.), *Papers from the Tenth Regional Meeting of the Chicago Linguistics Society* (pp. 787–801). Chicago, IL: Chicago Linguistics Society. Retrieved from http://chicagolinguisticsociety.org/list.html