*Language and Cognition (2022), 14: 2, 185–207*
doi:10.1017/langcog.2021.28

ARTICLE

# Understanding demonstrative reference in text: a new taxonomy based on a new corpus

Alfons Maes*, Emiel Krahmer and David Peeters

Department of Communication and Cognition (TiCC), Tilburg University, Tilburg, Netherlands
*Corresponding author. Email: maes@tilburguniversity.edu

(Received 09 April 2021; Revised 20 December 2021; Accepted 21 December 2021)

**Abstract**

Endophoric demonstratives such as *this* and *that* are among the most frequently used words in written texts. Nevertheless, it remains unclear how exactly they should be subdivided and classified in terms of their different types of use. Here, we develop a new taxonomy of endophoric demonstratives based on a large-scale corpus including three written genres: news items, encyclopedic texts, and book reviews. The taxonomy enables analysts to reliably code endophoric demonstratives based on objectively applicable criteria, while at the same time making them aware of many subtle borderline cases. We consider the taxonomy as a theoretical foundation for future theoretical and empirical work into endophoric demonstratives, and as an analytical tool allowing researchers to unify and compare the results of studies on endophoric demonstratives coming from different genres and languages.

**Keywords:** demonstratives; endophoric reference; discourse genre; referential communication; referent accessibility; corpus linguistics; pragmatics

## 1. Endophoric demonstratives: distinctions and definitions

Demonstratives, such as 'proximal' *this* and 'distal' *that* in English, take up an important position in the system of referential devices that languages offer their users. On the one hand, demonstratives contribute to establishing successful reference to entities within a joint speaker-hearer space, in close coordination with other devices such as deictic bodily gestures (Cooperrider, 2016; Levinson et al., 2018; Peeters & Özyürek, 2016). On the other hand, they are part of a larger set of referential expressions (e.g., including pronouns and definite noun phrases) that speakers or writers may use to activate or reactivate a *discourse* referent in the mind of one's addressee (Ariel, 1990; Cornish, 2011; Gundel, Hedberg, & Zacharski, 1993). Traditionally, a clear distinction is hence made between the exophoric vs. endophoric use of demonstratives respectively, with exophoric demonstratives referring to actual entities in the speech situation, predominantly studied in spoken interaction, and

© The Author(s), 2022. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

https://doi.org/10.1017/langcog.2021.28 Published online by Cambridge University Press

endophoric demonstratives covering all other uses (e.g., Diessel, 1999; Halliday & Hasan, 1976).

The exophoric use of demonstratives is largely considered the ontogenetic, phylogenetic, and grammatical precursor from which other types of use have derived (e.g., Bühler, 1934; Diessel, 1999; Lyons, 1977; Tomasello, 2008) and it has been studied in a wide variety of languages (e.g., Diessel, 1999; Levinson et al., 2018). The study of exophoric demonstratives relies on research tools that enable researchers to unify and compare results coming from different languages and communicative situations. For example, Wilkins (2018) developed an analytical tool that can be used to elicit the basic use of exophoric demonstratives in virtually any spoken language based on a fixed set of interactional configurations between speaker, hearer, and object referent (see Levinson et al., 2018). Experimental laboratory studies into exophoric demonstratives also often use comparable spatial (game) setups, likewise enabling refined parametrization of physical and interactional variables, and study their effect on the demonstrative variant speakers decide to use (e.g., Coventry, Valdés, Castillo, & Guijarro-Fuentes, 2008; Diessel & Coventry, 2020; Reile, Plado, Gudde, & Coventry, 2020).

The empirical study of *endophoric* demonstratives shows a more divergent picture. Endophoric demonstratives are commonly considered as part of the larger set of referential expressions, either expressing different degrees of mental activation of the underlying referent (e.g., Ariel, 1990; Gundel et al., 1993) or different degrees of deictic force to mentally construct the intended referent (e.g., Cornish, 1999). A large number of studies discuss pragmatic functions of endophoric demonstratives going beyond simple reference, and relate demonstratives to the assumed psychological and social relations between writer and addressee (see for an overview, Maes, Krahmer, & Peeters, in review; Peeters, Krahmer, & Maes, 2021). However, in the endophoric domain, helpful tools enabling a unified and comparative analysis including all classes of endophoric demonstratives are lacking.

The aim of this paper is therefore to develop a taxonomy of endophoric demonstrative use, based on the analysis of demonstratives in three representative genres of written texts (narrative, expository, and evaluative). We consider all demonstratives occurring in written discourse endophoric, and here restrict our scope to *written non-interactive* '*one-to-many*' *discourse aimed at a generic audience*, with typical examples being everyday newspaper articles, Wikipedia texts, and written product reviews. The exclusion of one-to-one personal discourse (as in many instances of spoken interaction, and conversational and personal written genres) enables us to focus on situations in which communication partners cannot rely on direct interactional feedback, nor on 'private' common ground, two conditions expected to affect the use of demonstratives, in particular, when they are used to access a new referent (e.g., 'recognitional' *that*; Gundel et al., 1993; Himmelmann, 1996), or to construct a referent on the basis of non-nominal antecedents (e.g., discourse deixis; Webber, 1988).

Within this scope, we aim at creating an analytical tool enabling analysts to *reliably code* and *exhaustively classify* demonstratives in written corpora based on observable surface variables, thus allowing researchers to unify and compare the results of studies on endophoric demonstratives coming from different genres and languages. The taxonomy should provide them with a sound empirical basis for future experimental and analytic work into endophoric demonstratives, for example, when contrasting and testing different theoretical proposals on demonstrative variance (Diessel & Coventry, 2020; Peeters et al., 2021). In testing these proposals, clear and reliable taxonomic

https://doi.org/10.1017/langcog.2021.28 Published online by Cambridge University Press

distinctions are needed to develop claims about demonstrative variance in different types of demonstrative use in written discourse (Maes et al., in review).

Figure 1 presents our taxonomy and Table 1 provides typical examples for each of the classes we propose. We globally distinguish two main classes of endophoric demonstratives: *text-based* and *situation-based* demonstratives. We consider endophoric demonstratives *text-based* when the discourse context contains explicit linguistic 'antecedent' elements: (i) direct or indirect nominal elements (Noun Phrases and pronouns) or non-nominal elements (a Verb Phrase, clause, sentence, and parts or combinations thereof) in the case of *anaphoric* or *cataphoric* demonstratives as in examples (1) to (5) and (ii) elements included in the noun phrase (NP) following the demonstrative determiner in cases of *first mention* demonstratives as in examples (6) to (8).

*Situation-based* endophoric demonstratives instead find their interpretation outside the text proper, but in the writing situation: for instance, in relation to the communicative situation of the text (*origo*), in the container of the text (*self-reference*), in elements typically activated in specific genres (*displaced exophoric*), or even in non-linguistic visible objects in multimodal texts (*exophoric*). Their interpretation is triggered by the absence of explicit linguistic antecedent cues, and by the inability (at least in English) to change the obligatory proximal variant into a distal one (see examples (9)–(12)).

Note that the two basic classes we distinguish roughly represent what traditionally has been termed *anaphoric* vs. *deictic*, respectively (e.g., Diessel, 1999; Levinson, 2004). Yet, we will not make use of these notions in this sense, as we consider all endophoric demonstratives deictic. *Situation-based* demonstratives can be seen as pointing directly to referents outside the written text proper, relatively comparable to the typical deictic function of exophoric demonstratives. *Text-based* demonstratives are used when a discourse referent needs a deictic device (as compared to a regular pronoun or definite NP) to be accessed properly, as explained in Section 3.1.1.

The definitions of demonstrative classes in our taxonomy are based as much as possible on observable linguistic cues. Still, we fully realize that surface cues represent only a first step in mentally constructing a discourse referent, accessing semantic representations, and making pragmatic inferences associated with demonstrative

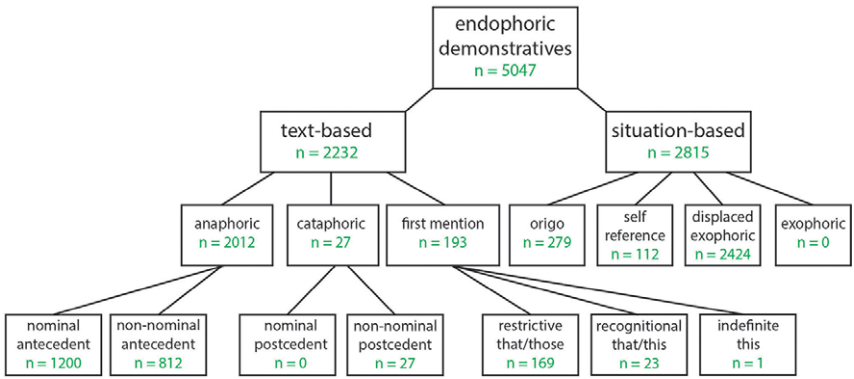


**Fig. 1.** Proposed taxonomy of endophoric demonstrative reference, and the number of demonstratives observed in our corpus per class, after data exclusion (see Section 2)

https://doi.org/10.1017/langcog.2021.28 Published online by Cambridge University Press

**Table 1.** Typical examples for each class in our taxonomy In these and following examples, we have underlined the proposed antecedent or postcedent, presented the critical demonstrative in boldface, and added the source (News, Wikipedia, and Reviews) and the corpus ID number.

| | |
|---|---|
| **Text-based** | |
| anaphoric nominal antecedent | (1) The Canadian maker of telecommunications equipment said it had a net loss of US$105 million (euro 87.6 million), or 2 cents a share, in the period ended Sept. 30. **That** compares with a loss of US$259 million, or 6 cents a share, in the prior third quarter. (N246) |
| | (2) Seeing Jacob young and old, and hearing how time affected him, and how he still remained connected to his history after all those years was great. It really added to the book and really tied the story together for me. It was very dramatic for me to see how **this** once active and interesting man was effected by the years, and what age does to a person. (R4252) |
| anaphoric non-nominal antecedent | (3) By law, non-U.S. investors may own no more than 25 percent of a U.S. airline's voting stock. Shane said the Bush administration isn't asking Congress to change **that**. (N336) |
| cataphoric non-nominal postcedent | (4) All you need to know for now is **this**, it is a fresh well-written book that you will love and ends in a good way so, have no fear of it and enjoy. (R3853) |
| cataphoric nominal postcedent | (5) Imagine **this**: a nice warm bath. |
| first mention restrictive *that/those* | (6) For **those** who don't know Gibran, get to know his work. (R3332) |
| first mention recognitional *that/this* | (7) Maybe I'm just nostalgic for the grandpa I never had. It's like **that** scene from The Royal Tenebaums, when Royal reads the gravestone of a man who died heroically on a sinking ship. (R5184) |
| first mention indefinite *this* | (8) My father has **this** huge book collection and I remember I was around 12 when I found this book amidst a sea of books. (R3471) |
| **Situation-based** | |
| origo | (9) In Vietnam, more than 3,000 poultry died or were culled **this** week in three villages in Bac Giang province, … (N627) |
| self-reference | (10) Associated Press writer Parfait Kouassi contributed to **this** report. (N649) |
| displaced exophoric | (11) If there were more than five stars, **this** book would surely deserve it. This book is the paradigm of wisdom, humility, and grace. (R3434) |
| exophoric | (12) It is shown in **this** diagram. |

reference (e.g., Lambrecht, 1994; Verhagen, 2005). We consider the taxonomy as a sound basis for more in-depth analyses of subclasses and borderline cases, in which surface cues are weighed against and complemented with less objective cues, such as assumptions on mental representations of discourse referents.

The remainder of this paper is organized as follows: In Section 2, we discuss the selection and coding criteria of our corpus, which includes text fragments taken from three well-known written genres. In Section 3, we further discuss our taxonomy of endophoric demonstrative classes in detail, based on the coding results and in light of existing taxonomies. Section 4 concludes the paper.

## 2. A corpus approach to endophoric demonstratives

In this section, we discuss the details of the corpus used to develop our taxonomy. We compiled a corpus of 6,884 demonstratives coming from three different, common

https://doi.org/10.1017/langcog.2021.28 Published online by Cambridge University Press

genres of written discourse: news articles, Wikipedia items, and book reviews. We used different clear-cut genres for reasons of representativity, but also because of the assumed theoretical relevance of genre for endophoric demonstrative variance (Peeters et al., 2021). The perspective we take is that a taxonomy of endophoric demonstratives should enable us to distinguish and code all demonstratives in a written corpus, and that the classification of demonstratives should be based on analytical variables that can be coded objectively and reliably.

Over the years, different types of corpora have been analyzed for referential expressions in general (e.g., Toole, 1996; Uryupina et al., 2020) or demonstratives in particular (e.g., Botley & McEnery, 2001a, 2001b; Byron & Allen, 1998; Maes, 1996; Petch-Tyson, 2000). Although these studies offer valuable distributional information, to the best of our knowledge, no existing corpus presents an in-depth discussion of endophoric classes of demonstratives based on a sizeable and balanced set of well-defined, different written genres. Early proposals often tested their claims on demonstrative variation in small-scale, unbalanced, or genre-unspecific corpora (Ariel, 1988; Gundel, Hedberg, & Zacharski, 1988; Himmelmann, 1996; Kirsner, 1979; Oh, 2001). Other studies restricted themselves to studying either only anaphoric demonstrative NPs (Maes, 1996), abstract (Dipper & Zinsmeister, 2012), proximal (Poesio & Modjeska, 2005), or distal (Byron & Allen, 1998; Passonneau, 1989) demonstratives, to a single genre (Acton & Potts, 2014; Botley & McEnery, 2001b; Gundel, Hedberg, & Zacharski, 2004; Hedberg, Gundel, & Zacharski, 2007; Potts & Schwarz, 2010), or they did not explicitly define the categories used (Botley & McEnery, 2001a, 2001b).

The corpus together with the annotations is publicly available via the Open Science Framework (OSF; https://osf.io/b32xz/).

### 2.1. Selection of the corpus

We selected English texts coming from three written genres: narrative (news articles), expository (Wikipedia texts), and evaluative (book reviews). The *news* texts consisted of Associated Press news articles ($n = 2{,}021$) on national and international news from the period 2004–2006, taken from the AQUAINT-2 Information-Retrieval Text Research Collection (Voorhees & Graff, 2008). *Wikipedia* entries ($n = 1{,}755$) were taken from the GREC corpus, and consisted of topics on persons, mountains, rivers, cities, and countries (Belz, Kow, Viethen, & Gatt, 2010). The *book reviews* ($n = 1{,}904$) were taken from the Amazon product data corpus (He & McAuley, 2016; http://jmcauley.ucsd.edu/data/amazon/), and were written in a more informal style by assumedly less professional writers.

### 2.2. Data exclusion and coding

We used the Stanford CoreNLP system (https://stanfordnlp.github.io) to tokenize text in the corpus, and to extract lemmas and parts-of-speech. Paragraphs were the input in the news and Wikipedia texts, complete reviews were the input in the review texts. We selected the complete set of Wikipedia texts in the GREC corpus and the first 3,000 paragraphs from the news and the reviews corpus. We automatically retrieved all sentences that contained at least one demonstrative, plus the three sentences that preceded the retrieved sentence (or as many preceding sentences as

https://doi.org/10.1017/langcog.2021.28 Published online by Cambridge University Press

**Table 2.** Across the three genres (All), and per genre (News, Wiki, and Reviews), the overall number of words, the number of demonstratives in total (and per 1,000 words between brackets), and the number (and percentages between brackets) of text-based, situation-based, and excluded demonstratives

| | All | News | Wiki | Reviews |
|---|---|---|---|---|
| number of words | 1,382,876 | 855,359 | 239,470 | 288,047 |
| number of demonstratives | 6884 (4.98) | 2,625 (3.07) | 667 (2.79) | 3,592 (12.47) |
| text-based | 2232 (32.4) | 825 (31.4) | 609 (91.3) | 798 (22.2) |
| situation-based | 2815 (40.9) | 348 (13.3) | 4 (0.6) | 2,424 (67.5) |
| excluded | 1837 (26.7) | 1,452 (55.3) | 54 (8.1) | 370 (10.3) |

possible when the retrieved sentence was the first, second, or third in the text). Only demonstrative determiners and demonstrative pronouns were included, thus excluding other uses of *that* (conjunction, complementizer, and relative pronoun). This resulted in 6,156 fragments (news $n = 2{,}503$; Wikipedia $n = 660$; and reviews $n = 2{,}993$) with one or more demonstratives, some of them partially overlapping with other fragments. For each of these fragments, we first manually coded all demonstratives ($n = 6{,}884$) according to their main endophoric class (text-based vs. situation-based). Table 2 shows substantial differences in the total number of observed demonstratives (in particular per 1,000 words) across the three genres.

In total, from this first selection, 1,837 demonstratives were excluded prior to analysis. As one of our aims was for the taxonomy to be useful in contrasting and testing theoretical proposals on demonstrative variance (Maes et al., in review; Peeters et al., 2021), we excluded 1,283 demonstratives (832 proximal), almost all from the news texts, because they were part of quoted text. A quotation makes the demonstrative form (e.g., proximal vs. distal) unreliable, in particular, when it is impossible to determine whether the chosen demonstrative variant is literally taken from an original context, or adapted when written up by the journalist. Furthermore, 354 demonstratives (225 proximal), most probably anaphors, were excluded due to the automatic selection procedure, with the (demonstrative) anaphor sentence being selected in the absence of the sentence containing the antecedent. Furthermore, a total of 169 duplicates were excluded. Finally, we excluded a small number of degree modifier demonstratives (e.g., *this/that much;* $n = 31$, 10 proximal), apparently also considered a determiner by the automatic parser. Although these examples show demonstrative variation and can be seen as an extension of regular deictic reference (Klipple & Gurney, 2002; König & Umbach, 2018; Labrador, 2011), we did not include them, as they do not access a discourse referent; they are adverbials followed by an adjective, and thus do not fit the coding variables we used to analyze the demonstrative pronouns and determiners.

As a starting point for our taxonomy, we first divided the remaining endophoric demonstratives ($n = 5{,}047$) into two overarching categories based on the source of interpretation, as shown in Fig. 1: *text-based* demonstratives having explicit antecedent or postcedent triggers in the surrounding text; *situation-based* demonstratives, obligatorily proximal in English and finding their interpretation triggers outside the text proper but in the writing situation.

We coded the remaining text-based ($n = 2{,}232$) and situation-based ($n = 2{,}815$) demonstratives on their *demonstrative type* (proximal *this/these* vs. distal *that/those*) and *demonstrative number* (singular *this/that* vs. plural *these/those*). Furthermore, we coded all text-based demonstratives with regard to variables that could be assigned

https://doi.org/10.1017/langcog.2021.28 Published online by Cambridge University Press

objectively and which were deemed relevant to explain demonstrative functions and demonstrative variance, thus enabling us to test different theoretical proposals on demonstrative variance elsewhere (Maes et al., in review). These were *demonstrative form* (pronominal vs. unmodified vs. modified NP with and without head noun, for example, *those* vs. *those people* vs. *those great people* vs. *those who are great*), *syntactic function* (subject vs. non-subject), *sentence position* (initial vs. non-initial), *type of referent* (abstract vs. concrete vs. human vs. named human). In addition, for anaphoric and cataphoric demonstratives ($n = 2{,}039$), we coded the following variables: *antecedent type* (nominal vs. non-nominal), *referential distance* (same sentence vs. previous sentence(s)). Finally, for unmodified anaphoric NPs with a nominal antecedent ($n = 692$), we coded the *lexical relation* between the anaphor noun and the head noun of their antecedent (same vs. different noun).

### 2.3. Descriptive observations

Table 3 presents the distribution of demonstratives across the coded variables, overall and separately for the three genres. Situation-based demonstratives in our corpus, by definition exclusively proximal, were almost all singular (99.2%). As to text-based

**Table 3.** Across the three genres (All), and per genre (News, Wiki, and Reviews) (i) the proportion of situation-based demonstratives for the variables demonstrative type: proximal (vs. distal) and number: singular (vs. plural), (ii) the proportion of text-based demonstratives for the variables demonstrative type: proximal (vs. distal), number: singular (vs. plural); form: pronouns, unmodified NPs, modified NPs, modified elliptic NPs; syntactic function: subject (vs. non-subject); sentence position: initial (vs. non-initial); and type of referent: abstract, concrete, human vs. named human; (iii) the proportion of anaphoric and cataphoric demonstratives for the variables antecedent type: nominal (vs. non-nominal); and referential distance: same sentence (vs. earlier sentence); (iv) the proportion of unmodified NP anaphora with nominal antecedent with either a lexically same (vs. different) head noun.

| | | All | News | Wiki | Reviews |
|---|---|---|---|---|---|
| **(i) situation-based demonstratives** | | ***n = 2,815*** | ***n = 348*** | ***n = 4*** | ***n = 2463*** |
| demonstrative type | % proximal | 100 | 100 | 100 | 100 |
| demonstrative number | % singular | 99.2 | 99.4 | 100 | 99.1 |
| **(ii) text-based demonstratives** | | ***n = 2,232*** | ***n = 825*** | ***n = 609*** | ***n = 798*** |
| demonstrative type | % proximal | 43.1 | 20.1 | 77.8 | 40.4 |
| demonstrative number | % singular | 70.0 | 60.3 | 79.6 | 72.6 |
| demonstrative form | % pronoun | 32.3 | 29.2 | 23.3 | 42.5 |
| | % unmodified NP | 42.5 | 42.2 | 57.5 | 31.5 |
| | % modified with head N | 12.1 | 8.7 | 9.9 | 17.4 |
| | % modified without head N | 13.0 | 19.9 | 9.3 | 8.6 |
| syntactic function | % subject | 47.2 | 48.6 | 53.4 | 41.0 |
| sentence position | % initial | 48.9 | 50.7 | 59.3 | 39.1 |
| type of referent | % abstract | 78.6 | 78.5 | 77.3 | 79.7 |
| | % concrete | 9.9 | 5.1 | 19.5 | 7.6 |
| | % human | 9.4 | 16.0 | 3.0 | 7.4 |
| | % named human | 2.1 | 0.4 | 0.2 | 5.3 |
| **(iii) anaphoric–cataphoric demonstratives** | | ***n = 2,039*** | ***n = 749*** | ***n = 596*** | ***n = 694*** |
| antecedent type | % nominal | 58.9 | 59.5 | 73.5 | 45.5 |
| referential distance | % same sentence | 33.0 | 34.7 | 25.5 | 37.5 |
| **(iv) unmodified NP anaphors with nominal antecedent** | | ***n = 692*** | ***n = 239*** | ***n = 286*** | ***n = 167*** |
| lexical relation | % same head noun | 35.3 | 35.1 | 37.8 | 31.1 |

https://doi.org/10.1017/langcog.2021.28 Published online by Cambridge University Press

demonstratives, stable across the three genres, a strong preference was observed for *singular* (70%) vs. plural demonstratives, for *antecedents in previous sentence(s)* (67.0%) rather than in the same sentence, and for *different* (64.7%) vs. the same *head nouns*. There was a preference for *demonstrative NPs* (67.7%) vs. pronouns, which resonates well with a long tradition of stylistic guidelines and prescriptions in favor of the use of NP over pronominal demonstratives (Finn, 1995; Geisler, Kaufer, & Steinberg, 1985; Moskovit, 1983; Rustipa, 2015; Wulff, Römer, & Swales, 2012), and with studies concluding that essays with a higher proportion of pronominal demonstratives are more difficult for readers, although at the same time receiving higher marks from trained essay raters (Crossley, Rose, Danekes, Rose, & McNamara, 2017). Also, there was a preference for higher order, *abstract* (78.6%) vs. concrete or human referents and a relatively high proportion of demonstratives with *non-nominal* antecedents (41.1%), pointing at a majority of demonstratives accessing abstract referents. Finally, modified NPs with and without head noun (25.1%, $n = 561$) were partly first mention demonstratives ($n = 193$), partly anaphoric ($n = 368$). In the latter case, they typically predicated new information on an activated referent, as in example (2) (e.g., Cornish, 2011; Maes & Noordman, 1995).

Other variables showed considerable variation over the three genres. There was a huge difference in *demonstrative type* across genres, with the news and reviews texts having a (strong vs. weak) preference for distal demonstratives (79.9% and 59.6% respectively), while the Wikipedia texts displayed a strong proximal preference (77.8%), highlighting the importance of discourse genre in explaining demonstrative variance as we argued for elsewhere (Maes et al., in review; Peeters et al., 2021). Also, demonstratives in the book reviews predominantly took up a non-subject (59%) and non-initial (60.9%) sentence position.

### 2.4. Reliability of the coding

Although the coding of the demonstratives for the various variables is fairly objective and formal in nature, about 10% of the data ($n = 689$; 222 situation-based and 467 text-based demonstratives) was independently coded by a second coder. For each of the coding variables, the agreement between the two coders was between 96.3% and 98.7%. In total, 69 coding differences were found in 60 out of the 689 fragments (7 situation-based; 62 text-based: 14 demonstrative form, 17 antecedent type; 14 distance, 11 syntactic function, 6 sentence position). Most of these (78%) were simple errors made by one of the coders. Some other coding differences were borderline cases or were resolved after discussion, in particular, concerning the exact type or extension of the antecedent, or the difference between a nominal or non-nominal antecedent, as in example (13). The coded corpus, as well as details of the second coding procedure, are available online via the OSF entry for this study.

## 3. A new taxonomy of endophoric demonstratives

In this section, we will discuss our taxonomy class by class and relate it to the properties of the demonstratives observed in the corpus and to choices made in alternative, existing approaches. Some earlier distinctions originate from the study of exophoric demonstratives in interactional discourse (Diessel, 1999; Himmelmann, 1996; Levinson, 2004), while others are developed within an endophoric tradition

https://doi.org/10.1017/langcog.2021.28 Published online by Cambridge University Press

(e.g., Ariel, 1990; Cornish, 2001; Doran & Ward, 2019; Gundel et al., 1993; Halliday & Hasan, 1976). We will also include ideas and proposals coming from studies of particular uses of demonstratives, such as bridging demonstratives (Lücking, 2018) or demonstratives with non-nominal antecedents (e.g., Kolhatkar, Roussel, Dipper, & Zinsmeister, 2018). The discussion of our taxonomy below will follow the classes distinguished in Fig. 1 and discuss these one by one.

### 3.1. Text-based endophoric demonstratives

In our taxonomy, we distinguish three classes of text-based endophoric demonstratives: anaphoric, cataphoric, and first mention demonstratives. *Anaphoric* demonstratives ($n$ = 2,012) find their interpretation trigger in an antecedent that by definition precedes it, as in examples (1), (2), and (3). They can be further subdivided as a function of whether the antecedent has a nominal (examples (1) and (2)) or non-nominal form (example (3)). *Cataphoric* demonstratives were observed substantially less often ($n$ = 27) and all had a non-nominal 'postcedent', as in example (4). Finally, *first mention* demonstratives ($n$ = 193) were observed to flexibly create a new discourse referent on the spot, with a productive class formally restricted to a distal demonstrative followed by a post-modification ($n$ = 169), for example, a relative clause as in example (6), and a smaller category introducing a new referent using a modified NP ($n$ = 24), as in examples (7) and (8). We will now discuss these three text-based classes in more detail, relate them to earlier work, and discuss several borderline cases to illustrate our coding and classification procedures.

#### 3.1.1. Anaphoric demonstratives

Our classification of anaphoric demonstratives differs from existing taxonomies that consider anaphoric and discourse deictic demonstratives as conceptually distinct classes (e.g., Diessel, 1999; Himmelmann, 1996; Levinson, 2004). In these taxonomies, *anaphoric* demonstratives are proposed to track discourse entities with a past and a future. They refer to 'pre-existing' or 'pre-activated' discourse entities, established in previous discourse via a nominal antecedent, which then often persist in subsequent discourse. Discourse deictic demonstratives, on the other hand, are typically argued to 'point' at a variety of non-nominal stretches of preceding discourse, thereby connecting this information to the current sentence. As such, they are usually considered 'impure' deictic devices: they do not refer to clearly definable pre-existing entities, but create entities on the spot, which rarely persist in subsequent discourse (e.g., Diessel, 1999; Hauenschild, 1982; Himmelmann, 1996; Lyons, 1977). Although we acknowledge this distinction, and consider the type of antecedent (nominal vs. non-nominal) indeed as a crucial analytical variable, we propose to attenuate the conceptual distinction between nominal/anaphoric and non-nominal/discourse deictic. This position is inspired by the behavior of demonstratives in our corpus and congruent with ideas proposed elsewhere (see e.g., Cornish, 2001; Ehlich, 1982; Kolhatkar et al., 2018; Piwek, Beun, & Cremers, 2008).

First, we consider demonstratives with nominal and non-nominal antecedents both *deictic*, as their referents need the deictic force of a demonstrative (rather than a regular pronoun or definite NP) to be accessed properly. In case of nominal antecedents as in example (1), demonstratives more often have a non-subject antecedent than regular pronouns (e.g., Brown-Schmidt, Byron, & Tanenhaus,

https://doi.org/10.1017/langcog.2021.28 Published online by Cambridge University Press

2005; Çokal, Sturt, & Ferreira, 2014; 2018; Fossard, Garnham, & Cowles, 2012; Kaiser & Trueswell, 2008), and bring less accessible entities into focus (e.g., Ariel, 1990; Gundel et al., 1993, 2004; Hauenschild, 1982; Linde, 1979). In their function as determiner in a (modified) demonstrative NP accessing highly activated referents, as in example (2), they create additional inferences compared to definite determiners (e.g., 'predicating' demonstrative NPs, Cornish, 2011; de Mulder, 1998; Doran & Ward, 2015; Gaillat, 2016; Goethals, 2013; Kleiber, 1991; Maes & Noordman, 1995; Schnedecker, 2006). In the case of non-nominal antecedents and higher order abstract referents, as in example (3), demonstratives are found to be used more frequently than pronouns (e.g., Gundel et al., 2004; Kolhatkar et al., 2018; Maes, 1997). Demonstratives also enable access to new referents (e.g., recognitional *that* or indefinite *this* as in examples (7) and (8)) more easily than definite NPs, and more easily allow cataphoric relations, as in example (4), than pronouns. Finally, they are also more powerful in creating mental representations of referents based on indirect cues, as we will see below: both nominal cues (e.g., in deferred/bridging or generic reference, Doran & Ward, 2019; Lücking, 2018) and non-nominal cues (e.g., discourse deixis, Cornish, 2007; referent coercion, Kolhatkar et al., 2018; Webber, 1988). All these observations justify them to be considered a combination of deixis and anaphora, captured adequately by the notion of *anadeixis* (e.g., Cornish, 2001, 2007, 2011; Ehlich, 1982).

Second, we consider demonstratives with nominal and non-nominal antecedents both *anaphoric*, in the sense that they always require some explicit antecedent trigger being present in the text. This condition is widely accepted for anaphors with nominal antecedents, but the 'required presence' of a non-nominal trigger should also be seen as a necessary condition for the construction of abstract referents (e.g., Kolhatkar et al., 2018; Recasens, 2008; Webber, 1988).

Third, anaphoric demonstratives with *nominal* antecedents in our corpus do not typically enable access to discourse referents 'with a past and a future'. Most of these antecedents (78.6%) actually refer to *abstract* entities, they are often derived from verbs as in example (1), and rarely refer to persistent referents. This makes them conceptually similar to the referents accessed by non-nominal antecedents. In many cases, it is even unclear whether the antecedent is nominal or non-nominal, as we observed in our second coding exercise. For instance, in example (13), we may consider the whole NP (*the fact that…*) or only the *that*-clause as the nominal or non-nominal antecedent respectively. Whatever (subjective) interpretation we take, in either case, the demonstrative is roughly referring to the same conceptual entity.

(13) This huge amount of water is responsible for the fact that the Amazon has no clouds above its channel near its mouth, as it is very easy to see in the satellite image. The reason for **this** is that satellite images are almost always taken during morning hours, when there are fewer clouds. (W2902)

Finally, we include in the two anaphoric classes regular cases as well as borderline cases, as we see this as the best starting point for more in-depth analyses of these cases. Scholars from different backgrounds use different labels for different types of borderline cases. Nominal antecedents are said to enable bridging, deferred, indirect, and/or inferrable reference (e.g., Gundel et al., 2004; Prince, 1981a), while non-nominal antecedents have been claimed to allow coercion, ostension, or indirect reference (e.g., Hedberg et al., 2007; Kolhatkar et al., 2018; Webber, 1991). Here, we will review some borderline cases as we found them in the corpus.

https://doi.org/10.1017/langcog.2021.28 Published online by Cambridge University Press

*Borderline cases of demonstratives with nominal antecedents* are mostly characterized by anaphor referents having a different semantic interpretation or denotation compared to their antecedent referent. Typical cases involving quantificational shifts, often discussed in semantic accounts of demonstratives, were not found in our corpus (e.g., "*Every dog in the neighborhood, even the meanest, has an owner who thinks that* ***that*** *dog is a sweetie*"; Roberts, 2002, p. 93). Shifts between a generic and a specific referent interpretation (Bowdle & Ward, 1995; Doran & Ward, 2019) are illustrated in the shift from a specific (*a story)* to a generic (*these stories*) referent in example (14), or the other way around (*elephants → this one [elephant]*) in example (15).

(14) It's a story about coming home. It's a story about dignity. Gruen frames the story so that you are hit by the fact that this old man, left to drool over Jello his last few years, is the twenty-three-year-old Jacob in the circus. **These** stories aren't just stories. (R5170)

(15) It was good. About elephants for the most part. They have tusks sometimes, but **this** one didn't. (R5318)

We coded a small number of demonstratives ($n = 16$) as bridging inferences (e.g., Apothéloz & Reichler-Béguelin, 1999; Doran & Ward, 2019), most of them having distal forms ($n = 12$) and/or coming from the book reviews ($n = 13$). Example (16) presents a fairly standard case based on a general knowledge inference (*the "Our Father" in Polish → those words*). Most other cases, however, partly also rely on genre-specific knowledge, in particular, on assumed knowledge of the book reviewed, and thus also have a displaced exophoric flavor, as in example (17) where *all those chapters* refer to the substantial part of the story that is situated in a retirement home.

(16) Being of Polish descent, I enjoyed the fact that the protagonist was Polish, and I was taken back to my grade school days when a character began reciting the "Our Father" in Polish! It amazed me that, after all of these years, I could still read and pronounce **those** words correctly. This is simply a wonderful novel, and I highly recommend it! (R5058)

(17) I didn't go in expecting literary greatness, but even so this book fell flat. For starters, the "old man reflecting from a retirement home" is a tired technique and all **those** chapters bored me to tears. (R4000)

As noted in previous work, demonstratives can furthermore sometimes defer reference ("*A car drove by. The horn was honking. Then another car drove by.* ***That*** *horn was honking even louder,*" e.g., Doran & Ward, 2019; Lücking, 2018; Wolter, 2006). In the corpus, we did not observe such cases, but we did find a productive class of deferred *elliptic those/that* anaphors ($n = 156$), as in example (18), where the demonstrative picks out the head noun of the antecedent NP (*Cubans*) to create a new entity (*wet foot Cubans).*

(18) Under the government's wet foot/dry foot policy, Cubans who set foot on U.S. soil are generally allowed to stay, while **those** intercepted at sea are usually returned to Cuba. (N623)

https://doi.org/10.1017/langcog.2021.28 Published online by Cambridge University Press

Finally, some demonstratives rely on information present in the previous discourse, but without representing a clear reference, as in example (19), where the demonstrative NP (*that point*) refers to a point in the story that is implied in the writer's strategy of *foreshadowing*. These cases are often distal, like in semi-fixed expressions (e.g., *at that, that being said, for that matter*), and their semantic borderline type is hard to determine.

(19) Gruen writes an engaging story that kept me turning the pages. She knows how to build suspense; she uses foreshadowing in the first few pages that makes the reader want to get to **that** point in the story. (R4035)

In sum, borderline cases of demonstratives with nominal antecedents show a varied picture. Yet, for all of them, the initial interpretation trigger can be found in the text, which is why we classified them all as anaphoric demonstratives.

For *demonstratives with non-nominal antecedents*, distinguishing between regular and borderline cases is much more complex. Apart from discussions on the nominal vs. non-nominal status of the antecedent, as in example (13), analysts have to find the exact syntactic extension of the antecedent, the semantic type of the anaphor, and the semantic type of the antecedent. In particular, the semantic interpretation turns out to be notoriously difficult, as both the antecedent and the anaphor can have a different semantic interpretation, to be selected from a fluid list of object types with increasing degrees of abstractness, including events, states, situations/facts, or propositions/speech acts (Asher, 1993; Gundel et al., 2004; Kolhatkar et al., 2018). Because of these complexities, early annotation attempts restricted themselves to distinguishing between direct and indirect cases of non-nominal demonstrative reference (Gundel et al., 2004; Hedberg et al., 2007).

More recently, Dipper & Zinsmeister (2012) developed an annotation system based on three linguistic tests, applied here to example (13). The *namely test* identifies the antecedent in a *namely* construction following the demonstrative (*this → this, namely (the fact) that the Amazon has no clouds above its channel near its mouth*). The *NP-replacement test* adds a noun to the demonstrative, taken from a fixed list, and identifies the semantic type of the demonstrative anaphor (*this → this state of affairs/situation/fact*). The *colon test* adds a writer's statement in front of the antecedent, and so identifies the semantic type of the antecedent (*I state the fact: "the Amazon has…"*). The combination of these tests allows these authors to detect semantic shifts between anaphor and antecedent with a fair level of intersubjective reliability, by using expert annotators and lists of nouns and statements connected to different abstractness levels.

Discussions remain however, as the linguistic tests do not prevent that several nouns or statements are acceptable at the same time. In example (13), the antecedent is clearly presented as a fact, but it remains unclear whether the anaphor has to be interpreted as a state, a situation, or a fact. Example (3) can be given a regular interpretation with the antecedent and the anaphor being both state-of-affairs or situations (i.e., *Congress being asked to change the situation or state-of-affairs described in the antecedent sentence*) or a slightly shifted one with an anaphor representing an action-event inferred from the antecedent (i.e., *Congress being asked to carry out the action of changing the law which is responsible for the situation described in the antecedent).* Also, the tests are not always easily applicable, for

https://doi.org/10.1017/langcog.2021.28 Published online by Cambridge University Press

example in the semi-fixed construction in example (20), where the demonstrative refers to the container of the previous sentence or perhaps to the illocutionary act of saying, rather than to states, facts, or propositions expressed in the sentence.

(20) Maybe some of the folks in his homeland considered him a "prophet," but I see him as an other person with some ideas but not too many truths. **That** being said, I like many of his ideas and concepts. (R3347)

In sum, congruent with our view, non-nominal antecedents are considered a necessary condition for the linguistic tests to be applied properly. Although the tests do not guarantee unambiguous coding of all subtle semantic changes between non-nominal antecedents and demonstrative anaphors, they are very useful in distinguishing regular from borderline cases.

### *3.1.2. Cataphoric demonstratives*

Endophoric demonstratives also allow for cataphoric relations with nominal or non-nominal postcedents (see example (4)). Most taxonomies restrict these to proximal demonstratives with non-nominal postcedents (Chen, 1990; Diessel, 1999; Gundel et al., 1988; Himmelmann, 1996), although occasional distal cases have been reported as well (Danon-Boileau, 1984; Fraser & Joly, Fraser & Joly, 1980; Quirk, Greenbaum, Leech, & Svartvik, 1985), and cross-linguistic differences in whether the proximal or distal (or any other) form functions as the preferred or 'unmarked' cataphoric demonstrative are expected. In our corpus, we found a small number of cataphoric demonstratives ($n = 27$). Six of them include a distal demonstrative, but all these can be considered borderline cases between an anaphoric and cataphoric interpretation, as in example (21) where the distal demonstrative can also be said to vaguely refer to the preceding proposition. The absence of cataphoric demonstratives with nominal antecedents in the literature and in our written corpus does not radically exclude such cases, as in example (5), but it is clear that they more typically fit in an interactive (and spoken) context.

(21) This book was a bestseller, but in case you missed it, it's not too late. **That**'s the thing about a great book – it never gets old. (R4876)

### *3.1.3. First mention demonstratives*

Apart from anaphoric and cataphoric demonstratives, we distinguish a variety of demonstratives ($n = 193$) used to introduce a new entity. We consider them text-based because the NP information following the demonstrative includes necessary triggers for an acceptable interpretation. In that sense, it is reasonable to consider them cataphoric as well, as has been done elsewhere (Deichsel & von Heusinger, 2011; Gary-Prieur, 1998; Labrador, 2011, see also Kolhatkar et al., 2018 for a similar use of the notion of cataphoric). In the corpus, their characteristics deviated from the general picture presented in Table 3: they are modified, often take up a non-subject function (81.9%) and non-initial (75.6%) sentence position, and are predominantly used in reference to first order (human or concrete) entities (88.6%), thus very much similar to the 'predicating' anaphoric modified demonstrative NPs.

https://doi.org/10.1017/langcog.2021.28 Published online by Cambridge University Press

An example of a first mention demonstrative that is extensively discussed in existing literature is *recognitional that* (e.g., Coniglio, Murphy, Schlachter, & Veenstra, 2018; Cornish, 2001; Diessel, 1999; Himmelmann, 1996). The crucial characteristic of this class of first mention demonstratives is the assumption on the part of the speaker or writer that the addressee will be able to identify the intended referent on the basis of the newly provided information (Cornish, 2001; Gundel et al., 1993). Two additional modality-specific conditions are that the demonstrative is meant to signal to the addressee that a given referential expression may be elaborated on if necessary (Auer, 1981; Himmelmann, 1996; Schlegloff, 1996), and that the NP information is assumed to be privately shared between speaker and hearer (Diessel, 1999; Doran & Ward, 2019). These two latter conditions hence do not apply in written communication where direct feedback is not available or audiences may be unknown. Therefore, in our corpus, recognitional demonstratives tend to stay more on the safe side and suggest only a generic type of familiarity, often combined with an attitudinal or rhetorical stylistic effect, as illustrated in examples (7) and (22). Note that the distal demonstrative in example (7) remains acceptable, even for readers not acquainted with the scene of *the Royal Tenebaums,* and that the effect of the demonstrative remains the same when reference to the new entity is repeated, as in example (22), the latter being observed in Norwegian as well (Johannessen, 2008).

(22) And then there is **that** rare book you open to the first page and several hours later, your butt numb, your joints stiff from inactivity, your eyes misting, your vision blurred, you close the cover on the last page. **That** rare book that sucks you in so completely you lose track of time and place. (R4947–4948)

We should mention here that some scholars also discuss examples of *unmodified* recognitional or familiar *that* (*"I couldn't sleep last night. That dog (next door) kept me awake,"* Gundel et al., 1993, p. 277; "*[Sticker on rear window of car] Mind that child! He may be deaf,"* Cornish, 2001, p. 300). Although one may see familiarity indeed as part of the pragmatic interpretation of these demonstratives, their acceptability also depends on generic scenario knowledge (*dogs being part of neighborhoods; children being vulnerable in traffic*), rather than individual knowledge about one specific dog or child. Moreover, the two cases have an exophoric flavor: the location of the rear window sticker activates the relevant exophoric context, and addressees with private knowledge of *that annoying dog next door* may give *that dog* a displaced situational interpretation (Himmelmann, 1996, p. 220–221). So, in non-interactive written discourse, where such conditions do not apply, we consider the presence of NP information crucial for these first mention demonstratives.

For *indefinite this*, as in example (8), roughly the same story holds. Proximal demonstratives in English are able to introduce new 'indefinite' referents (MacLaran, 1980; Prince, 1981b), but only if they can assure the reader that enough information is provided in the modified NP, and again with the demonstrative as a signal marking the upcoming new information, together with the pragmatic 'proximal' inference that the speaker or writer will provide this information. A typical context for these proximal NPs is the presentational sentence (e.g., *She is this lawyer!* Doran & Ward, 2019, p. 250), but a similar effect can also be found in predicating anaphoric demonstratives as in example (23). In these instances, highlighting new information

https://doi.org/10.1017/langcog.2021.28 Published online by Cambridge University Press

is combined with a positive evaluation based on the degree modifying effect of demonstratives (e.g., *she is so good as a lawyer).*

(23) Since the book was told in first person from Jacob's point of view, the character of Marlena was even more under-developed than Jacob. All we know is that she's **this** beautiful girl in pink sequin that caught Jacob's eye the day he joined the circus and from then on he was lost. (R4703)

In our corpus, recognitional *that* and indefinite *this* were found almost exclusively in the book reviews ($n = 23$), attesting their informal nature. The distal cases ($n = 15$) represent typical cases of recognitional or familiar *thatN.* Only one of the remaining proximal cases could be replaced by an indefinite determiner, that is, in example (8). The others evoked similar inferences of familiarity based on assumed knowledge of the activated book, thus combining familiarity with a flavor of displaced exophoricity, as in example (24). All had a modified NP format, except example (25), where the reader of the book review is assumed to know that the referents of the distal and proximal unmodified demonstratives *that barn, those years,* and *this man,* as well as the referent of *these poor men and women* are to be found in the book, while *those summers with my Grand-dad,* a referent playing a part in the reviewer's private life, again is modified with sufficient private (although not privately shared) information.

(24) I haven't read novels in many many years and I decided to pick this up at Target. I had no idea it was going to be a movie. I have to say it was incredibly engrossing, and so much fun to be a part of **this** other world of train circuses! (R4366)

(25) Reading this book was like sitting in **that** barn all of **those** years ago listening to **this** man … Reading this book was like spending **those** summers with my Grand-dad. Sara Gruen made me feel the train swaying as it ran through the night and let me smell the smell of the animals as they stood in the heat of the day waiting to be tended to. She let me cry over the hardship of the lives that **these** poor men and women lived then let me feel the excitement of hearing the music start up for the shows. (R4765)

The most productive type of first mention demonstratives we observed are so-called restrictive *that/those* ($n = 167$) cases, found in all three text genres (news 75, wiki 13, and reviews 81): postmodified distal demonstratives without a head noun, as illustrated in example (6). Again, here the acceptability of the demonstrative type depends on the information provided by the subsequent relative clauses. These cases have a typical syntax and do not allow for replacement by a proximal variant. Semantically, they are somewhat related to quantificational demonstratives (Doran & Ward, 2019; MacLaran, 1980), and akin to deferred anaphoric demonstratives as in example (18).

#### *3.1.4. In sum*

We have proposed three main classes of *text-based* endophoric demonstratives. Anaphoric demonstratives, both in pronominal use and when part of a demonstrative NP, can have nominal and non-nominal antecedents, and include regular as well

https://doi.org/10.1017/langcog.2021.28 Published online by Cambridge University Press

as borderline cases. Cataphoric demonstratives have a postcedent that is most typically non-nominal. First mention demonstratives are used to introduce new referents on the spot, thereby relying on information in the demonstrative NP. Objective formal variables (e.g., the presence of antecedent information preceding or following the demonstrative, or included in the demonstrative NP) suffice to distinguish the three main classes, as well as some subcategories, such as demonstratives with non-nominal antecedents or restrictive *that* demonstratives.

### 3.2. Situation-based endophoric demonstratives

In sharp contrast with text-based demonstratives, situation-based endophoric demonstratives find their interpretation outside the ongoing text. We used two analytic criteria to define robust and objectively measurable classes of situation-based endophoric demonstratives. First, the impossibility, at least in English, for a proximal form to be reasonably replaced by a distal form, as situation-based endophoric demonstratives in English are exclusively proximal and almost always have a singular NP format. Second, the absence of explicit linguistic antecedent information. Apart from that, we assume for all situation-based demonstratives the mental presence of a referent being activated on the basis of either the standard coordinates of a communicative situation or genre-specific assumptions. Taking into account these criteria, we distinguish four classes of situation-based demonstratives (cf. Fig. 1), which are illustrated below.

#### 3.2.1. Origo demonstratives

Each communicative situation is intrinsically connected to a here-and-now (Bühler, 1934), which provides the opportunity to use situation-based demonstratives in reference to the *origo* of the ongoing discourse, as in example (9). This includes examples such as *this week, this era, this room,* or *this country,* which have previously been termed 'symbolically exophoric' (Levinson, 2004). These demonstratives are commonly observed in spoken discourse, but also occur in written text. In our corpus, the frequency of origo-based demonstratives differed substantially as a function of text genre: the news texts were responsible for 90% ($n = 246$) of all cases, consistent with the important role of space and time in news items.

Endophoric demonstratives can also refer in terms of a projected, transposed, imagined, or displaced origo (Bühler, 1934; Diessel, 1999; Levinson, 2004; Lyons, 1977). In our corpus, we however found them in these cases to be dependent on antecedent information needed to evoke a projected origo. Example (26) is taken out of the narrative part of a book review. The demonstrative has a clear antecedent, and replacement by a distal variant is a matter of stylistic effect rather than a drastic change of semantics (as would be in the case of origo deixis). In our taxonomy, this makes them text-based and therefore not situation-based.

(26) Sara Gruen does a phenomenal job of painting a vivid portrait of Depression era America and of the end of the golden age of circus life in the USA. The photos that head each chapter are fun little bonuses that let you see the actual performers of **this** time as well as some of the settings in which the book takes place. The characters come to life and are vivid portrayals of heroes and villains. (R5116)

https://doi.org/10.1017/langcog.2021.28 Published online by Cambridge University Press

### 3.2.2. Self-reference demonstratives

Each communicative situation also allows for *self-reference* via demonstratives, as in example (10). Such use of demonstratives includes deictic expressions referring to 'the container' rather than the content of the ongoing discourse, such as *this chapter*, *this conversation*, *this article*, *this manual*, etc. (e.g., Gundel et al., 1988; Hauenschild, 1982; Himmelmann, 1996; Paraboni & van Deemter, 2002). In our corpus, self-references ($n = 112$) were all NPs, almost all referring to the news reports ($n = 102$), as in example (10), and in exceptional cases also to the Wikipedia article ($n = 3$) or book review ($n = 7$).

Traditional taxonomies sometimes discuss borderline cases of self-reference when they distinguish between references to propositions (i.e., discourse deixis or 'impure' deixis) and references to the 'material side of language', i.e., [pure] text deixis (Cornish, 1999; Diessel, 1999; Himmelmann, 1996; Lyons, 1977; Webber, 1991). In example (27), *that word* does not refer to the meaning of *unsophisticate* but to the container of the meaning, similar to example (20). Although we consider this a valid distinction, adding to the other subtle semantic distinctions found in anaphoric demonstratives, the presence of clear antecedent information makes these latter cases, as in example (27), according to our taxonomy, text-based and not situation-based.

(27) I invented the word "hedon" … Let's say a hedon is a pleasure-seeker without the philosophy. A hedon is an unsophisticate. I probably made up **that** word too. (R3429)

### 3.2.3. Displaced exophoric demonstratives

Some demonstratives in the corpus enabled direct access to entities typically associated with specific written genres, and thus assumed to be activated on the basis of genre-specific knowledge. This class is particularly productive in the book reviews, as almost all reviews refer to *this book* as if it was physically present, as in example (11). These demonstratives can introduce the referent in a flexible way, using a pronoun only, as in example (30), a bridging inference as in example (31), or a generic demonstrative as in example (32).

(30) I usually don't read fiction, but **this** kept popping up as a suggestion for me. (R5115)

(31) **This** author is a master story teller. (R4181)

(32) I generally do not usually read **these** types of stories. (R4289)

We expect these displaced exophoric demonstratives to show up in many other genres as well. Manuals or product reviews, for instance, typically almost by default activate particular products or services. This results in a productive class of displaced exophoric demonstratives, reminiscent of displaced situational demonstratives (Himmelmann, 1996, p. 220–222).

https://doi.org/10.1017/langcog.2021.28 Published online by Cambridge University Press

#### 3.2.4. Exophoric demonstratives

Situation-based endophoric demonstratives can paradoxically also be *exophoric*, for instance, when they refer to non-linguistic objects physically present in multimodal written genres, as via the words *this diagram* or *this photograph,* see example (12). They cannot be considered text-based, as the crucial cue for the interpretation of these demonstratives is not a linguistic one. On the other hand, they do not point at a genuine exophoric object either, as the referent can be seen as integral to the interpretation of the text. Therefore, we consider these cases endophoric situation-based exophoric demonstratives. In the corpus, due to the nature of the selected texts, they were not observed.

#### 3.2.5. In sum

In the proposed taxonomy, situation-based demonstratives are demonstratives finding their interpretation in the origo of the discourse, in the container of discourse itself, in prominent (displaced) entities in the writing situation, or in non-linguistic objects present in multimodal written texts. None of them can reasonably be replaced by the distal variant in English and no linguistic antecedent is available. As a final note, it is worth mentioning that the usage most central in the study of exophoric demonstratives (*this here* vs. *that there*) did not show up in our endophoric corpus.

## 4. Conclusion

The corpus data presented in the current study allowed us to develop a new taxonomy of endophoric demonstrative reference. We conclude that the taxonomy enables a reliable coding of a large and varied corpus of endophoric demonstratives. Endophoric demonstratives are re-categorized on the basis of easily measurable criteria. Thus far, we do not have proof of the usability of our taxonomy across languages and across a larger variety of genres. Studying a wider variety of genres and languages may well result in more specific uses of endophoric demonstratives. Nevertheless, we expect analyses of other languages and other genres to fit in the classes we distinguished in Fig. 1, because in developing the taxonomy, a large variety of endophoric demonstratives from studies involving many different genres and languages were taken into account.

Although the present corpus analysis did not allow us to compare demonstratives with other types of referential expressions, many corpus distributions support the view of text-based demonstratives as giving access to relatively complex discourse referents, thus referents in need of an (ana)deictic device to be accessed properly. The proportion of anaphoric demonstratives with a non-nominal antecedent (41.1%), the huge proportion of abstract referents (78.6%), and the variety of (mostly distal) borderline cases points to demonstratives being devices flexibly creating complex referents on the basis of non-nominal, abstract, or vague antecedent information (e.g., Wittenberg, Momma, & Kaiser, 2021). Likewise, the relatively high proportion of modified demonstrative NP anaphors (25.1%), as in example (2) (compared to anaphoric definite NPs, e.g., Fraurud, 1990; Vieira & Poesio, 2000), supports the idea of demonstrative NPs often predicating new information about the referent (e.g., Cornish, 2001; Maes & Noordman, 1995).

The taxonomy introduced in this paper can be seen as a sound theoretical basis for future experimental or analytic work into endophoric demonstratives. In addition, it

https://doi.org/10.1017/langcog.2021.28 Published online by Cambridge University Press

allows researchers to unify and compare the results of corpus studies on demonstratives coming from different genres and languages, and it can be used to compare and refine the analysis in studies in which demonstratives are seen as instrumental, for example, in assessing the quality of translations (e.g., Goethals, 2007) or the proficiency of learners of a (foreign) language (e.g., Blagoeva, 2004; Petch-Tyson, 2000). Elsewhere, we have successfully used it to test hypotheses on demonstrative variance (*this* vs. *that*), and found that variance in English is not determined in the first place by the discourse-local activation status of discourse referents, but by discourse-global knowledge of genres, in particular, assumptions on subtle interactional inferences with respect to writer, addressee, and referent connected to specific discourse genres. Thus, we explained the dominance of distal vs. proximal demonstratives in the news vs. Wikipedia texts as a result of a reader- vs. writer-oriented interaction assumption connected to narrative vs. expository written texts, respectively (Maes et al., in review).

Finally, we hope that our taxonomy will provide the theoretical and empirically supported foundation for a coherent and cooperative future research agenda focusing on the analysis of endophoric demonstratives in text.

**Acknowledgments.** The authors thank three anonymous reviewers for valuable suggestions on an earlier draft of this work, and Thiago Castro Ferreira and Emiel van Miltenburg for help in the construction of the corpus.

**Funding statement.** D.P. was supported by a Veni grant (275-89-037) awarded by De Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO, the Dutch Research Council).

## References

Acton, E. K. & Potts, C. (2014). That straight talk: Sarah Palin and the sociolinguistics of demonstratives. *Journal of Sociolinguistics* 18(1), 3–31.

Apothéloz, D. & Reichler-Béguelin, M.-J. (1999). Interpretations and functions of demonstrative NPs in indirect anaphora. *Journal of Pragmatics* 31(3), 363–97.

Ariel, M. (1988). Referring and accessibility. *Journal of Linguistics* 24(1), 65–87.

Ariel, M. (1990). *Accessing antecedents*. London: Routledge.

Asher, N. (1993). *Reference to abstract objects in discourse*. Berlin: Springer Science & Business Media.

Auer, P. (1981). Zur indexikalitätsmarkierenden Funktion der demonstrativen Artikelform in deutschen Konversationen. In G. Hindelang & W. Zillig (eds), *Sprache: Verstehen und handeln*, 301–11. Berlin: Walter de Gruyter.

Belz, A., Kow, E., Viethen, J. & Gatt, A. (2010). Generating referring expressions in context: The GREC task evaluation challenges. In E. Krahmer & M. Theune (eds), *Empirical methods in natural language generation: Data-oriented methods and empirical evaluation*, 294–327. New York: Springer.

Blagoeva, R. (2004). Demonstrative reference as a cohesive device in advanced learner writing: A corpus-based study. *Advances in Corpus Linguistics* 49, 297–307.

Botley, S. & McEnery, T. (2001a). Demonstratives in English: A corpus-based study. *Journal of English Linguistics* 29(1), 7–33.

Botley, S. & McEnery, T. (2001b). Proximal and distal demonstratives: A corpus-based study. *Journal of English Linguistics* 29(3), 214–33.

Bowdle, B. F. & Ward, G. (1995). Generic demonstratives. In J. Ahlers, L. Bilmes, J. S. Guenter, B. A. Kaiser, & J. Namkung (Eds.), *Proceedings of the twenty-first annual meeting of the Berkeley Linguistics Society* (pp. 32–43). Berkeley, CA: Berkeley Linguistics Society.

Brown-Schmidt, S., Byron, D. K. & Tanenhaus, M. K. (2005). Beyond salience: Interpretation of personal and demonstrative pronouns. *Journal of Memory and Language* 53(2), 292–313.

Bühler, K. (1934). *Sprachtheorie*. Berlin: Fischer.

https://doi.org/10.1017/langcog.2021.28 Published online by Cambridge University Press

Byron, D. & Allen, J. (1998). Resolving demonstrative anaphora in the TRAINS93 corpus. In S. Botley & T. McEnery (eds), *New approaches to discourse anaphora: Proceedings of the second colloquium on discourse anaphora and anaphor resolution (DAARC2)*, Lancaster. 68–81.

Chen, R. (1990). English demonstratives: A case of semantic expansion. *Language Sciences* 12(2), 139–53.

Çokal, D., Sturt, P. & Ferreira, F. (2014). Deixis: *This* and *That* in written narrative discourse. *Discourse Processes* 51(3), 201–29.

Coniglio, M., Murphy, A., Schlachter, E. & Veenstra, T. (2018). It's not all just about this and that. In M. Coniglio, A. Murphy, E. Schlachter & T. Veenstra (eds), *Atypical demonstratives: Syntax, Semantics and pragmatics*, 1–19. Berlin: Walter De Gruyter.

Cooperrider, K. (2016). The Co-organization of demonstratives and pointing gestures. *Discourse Processes* 53 (8), 632–56.

Cornish, F. (1999). *Anaphora, discourse, and understanding*. Oxford: Oxford University Press.

Cornish, F. (2001). 'Modal' *that* as determiner and pronoun: The primacy of the cognitive-interactive dimension. *English Language and Linguistics* 5(2), 297–315.

Cornish, F. (2007). English demonstratives: Discourse deixis and anaphora. A discourse-pragmatic ac-count. In R. A. Nilsen, N. A. A. Amfo & K. Borthen (eds), *Interpreting utterances: Pragmatics and its interfaces. Essays in honour of Thorstein Fretheim*, 137–56. Oslo: Novus Press.

Cornish, F. (2011). 'Strict' anadeixis, discourse deixis and text structuring. *Language Sciences* 33(5), 753–67.

Coventry, K. R., Valdés, B., Castillo, A. & Guijarro-Fuentes, P. (2008). Language within your reach: Near–far perceptual space and spatial demonstratives. *Cognition* 108(3), 889–95.

Crossley, S. A., Rose, D. F., Danekes, C., Rose, C. W. & McNamara, D. S. (2017). That noun phrase may be beneficial and this may not be: Discourse cohesion in reading and writing. *Reading and Writing* 30(3), 569–89.

Danon-Boileau, L. (1984). That is the question. In A. Grésillon & J.-L. Lebrave (eds), *La langue au ras du texte*, 31–55. Stanford: Presses Universitaires de Lille.

De Mulder, W. (1998). Du sens des démonstratifs à la construction d'univers. *Langue Française* 120, 21–32.

Deichsel, A. & von Heusinger, K. (2011). Cataphoric potential of indefinites in German. In I. Hendrickx, S. L. Devi, A. Branco & R. Mitkov (eds), *Anaphora processing and applications: 8th discourse anaphora and anaphor resolution colloquium, DAARC 2011, Faro Portugal, October 6–7, 2011. Revised selected papers*, 144–56. Berlin: Springer Science & Business Media.

Diessel, H. (1999). *Demonstratives: Form, function and grammaticalization*. Amsterdam: John Benjamins Publishing.

Diessel, H. & Coventry, K. R. (2020). Demonstratives in spatial language and social interaction: An interdisciplinary review. *Frontiers in Psychology* 11, 3158.

Dipper, S. & Zinsmeister, H. (2012). Annotating abstract anaphora. *Language Resources and Evaluation* 46 (1), 37–52.

Doran, R. B. & Ward, G. (2015). Proximal demonstratives in predicate NPs. *Proceedings of the Annual Meeting of the Berkeley Linguistics Society* 41, 61–70.

Doran, R. B. & Ward, G. (2019). A taxonomy of uses of demonstratives. In J. Gundel & B. Abbott (eds), *The Oxford handbook of reference*, 236–59. Oxford: Oxford University Press.

Ehlich, K. (1982). Anaphora and deixis: Same, similar, or different? In R. J. Jarvella & W. Klein (eds), *Speech, place, and action. Studies in deixis and related topics*, 315–38. New York: Wiley.

Finn, S. (1995). Measuring effective writing: Cloze procedure and anaphoric "This". *Written Communication* 12(2), 240–66.

Fossard, M., Garnham, A. & Cowles, H. W. (2012). Between anaphora and deixis … The resolution of the demonstrative noun phrase "that N". *Language and Cognitive Processes* 27(9), 1385–404.

Fraser, T. & Joly, A. (1980). Le système de la deixis (2): Esquisse d'une théorie d'expression en anglais. *Modèles Linguistiques* 2, 22–49.

Fraurud, K. (1990). Definiteness and the processing of noun phrases in natural discourse. *Journal of Semantics* 7(4), 395–433.

Gaillat, T. (2016). *Reference in interlanguage: The case of this and that. From linguistic annotation to corpus interoperability* [Theses, Université Paris Diderot (Paris 7) Sorbonne Paris Cité]. https://hal.archives-ouvertes.fr/tel-01705743

Gary-Prieur, M.-N. (1998). La dimension cataphorique du démonstratif. Étude de constructions à relative. *Langue Française* 120, 44–50.

https://doi.org/10.1017/langcog.2021.28 Published online by Cambridge University Press

Geisler, C., Kaufer, D. S. & Steinberg, E. R. (1985). The unattended anaphoric "This": When should writers use it? *Written Communication* 2(2), 129–55.

Goethals, P. (2007). Corpus-driven hypothesis generation in translation studies, contrastive linguistics and text linguistics: A case study of demonstratives in Spanish and Dutch parallel texts. *Belgian Journal of Linguistics* 21(1), 87–103.

Goethals, P. (2013). Demonstratives on the move: What translational shifts tell us about demonstrative determiners and definite articles in Spanish and Dutch. *Linguistics* 51(3), 517–53.

Gundel, J. K., Hedberg, N. & Zacharski, R. (1988). On the generation and interpretation of demonstrative expressions. In D. Vargha & E. Hajičová (eds), *Proceedings of the 12th conference on computational linguistics*, 216–21. Melbourne, VIC: Association for Computational Linguistics.

Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive Status and the Form of Referring Expressions in Discourse. *Language* 69(2), 274–307.

Gundel, J. K., Hedberg, N., & Zacharski, R. (2004). Demonstrative pronouns in natural discourse. In: *Proceedings of the 5th Discourse Anaphora and Anaphora Resolution Colloquium(DAARC-2004)*, pp. 81–86.

Halliday, M. A. K. & Hasan, R. (1976). *Cohesion in English*. London: Routledge.

Hauenschild, C. (1982). Demonstrative pronouns in Russian and Czech—Deixis and anaphora. In J. Weissenborn & W. Klein (eds), *Here and there: Cross-linguistic studies on deixis and demonstration*, 167–186. Amsterdam: John Benjamins Publishing.

He, R. & McAuley, J. (2016). Ups *and* downs*:* Modeling *the* visual evolution *of* fashion trends *with o*ne-class collaborative filtering. In *Proceedings of the 25th international conference on World Wide Web*, New York: ACM Press. 507–17.

Hedberg, N., Gundel, J. K. & Zacharski, R. (2007). Directly and indirectly anaphoric demonstrative and personal pronouns in newspaper articles. In *Proceedings of the 6th discourse anaphora and anaphora resolution colloquium (DAARC)*, Daarc-2007, Lagos. 31–6.

Himmelmann, N. P. (1996). Demonstratives in narrative discourse: A taxonomy of universal uses. In B. A. Fox (ed), *Studies in anaphora*, 205–54). Amsterdam: John Benjamins Publishing.

Johannessen, J. B. (2008). The pronominal psychological demonstrative in Scandinavian: Its syntax, semantics and pragmatics. *Nordic Journal of Linguistics* 31(2), 161–92.

Kaiser, E. & Trueswell, J. C. (2008). Interpreting pronouns and demonstratives in Finnish: Evidence for a form-specific approach to reference resolution. *Language and Cognitive Processes* 23(5), 709–48.

Kirsner, R. S. (1979). Deixis in discourse: An exploratory quantitative study of the Modern Dutch demonstrative adjectives. *Discourse and Syntax*, 355–375.

Kleiber, G. (1991). Celui-ci/-là ou comment montrer du nouveau avec du déjà connu. *Revue Québécoise de Linguistique* 21(1), 123–69.

Klipple, E. & Gurney, J. (2002). Some observations on deixis to properties. In K. van Deemter & R. Kibble (Eds.), *Information sharing. Reference and presupposition in language generation and interpretation*, 355–89. Stanford: CSLI Press.

Kolhatkar, V., Roussel, A., Dipper, S., & Zinsmeister, H. (2018). Anaphora with non-nominal antecedents in computational linguistics: A survey. *Computational Linguistics* 44(3), 547–612.

König, E. & Umbach, C. (2018). Demonstratives of manner, of quality and of degree. In M. Coniglio, A. Murphy, E. Schlachter & T. Veenstra (eds), *Atypical demonstratives: Syntax, semantics and pragmatics*, 285–327. Berlin: Walter De Gruyter.

Labrador, B. (2011). A corpus-based study of the use of Spanish demonstratives as translation equivalents of English demonstratives. *Perspectives* 19(1), 71–87.

Lambrecht, K. (1994). *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents*. Cambridge: Cambridge University Press.

Levinson, S. C. (2004). Deixis. In L. Horn (ed), *The handbook of pragmatics*, 97–121. Oxford: Blackwell.

Levinson, S. C., Cutfield, S., Dunn, M., Enfield, N., Meira, S. & Wilkins, D. (eds) (2018). *Demonstratives in cross-linguistic perspective*. Cambridge: Cambridge University Press.

Linde, C. (1979). Focus of attention and the choice of pronouns in discourse. In T. Givón (ed), *Discourse and syntax*, 337–54. New York: Academic Press.

Lücking, A. (2018). Witness-loaded and witness-free demonstratives. In M. Coniglio, A. Murphy, E. Schlachter & T. Veenstra (eds), *Atypical demonstratives: Syntax, semantics and pragmatics*, 255–84. Berlin: Walter De Gruyter.

https://doi.org/10.1017/langcog.2021.28 Published online by Cambridge University Press

Lyons, J. (1977). *Semantics*. Cambridge: Cambridge University Press.
MacLaran, R. (1980). On two asymmetrical uses of the demonstrative determiners in English. *Linguistics* 18 (9–10), 803–20.
Maes, A. (1996). *Nominal anaphors, markedness and coherence of discourse*. Leuven: Peeters.
Maes, A. (1997). Referent ontology and centering in discourse. *Journal of Semantics* 14(3), 207–35.
Maes, A., Krahmer, E. & Peeters, D. (in review). Explaining variance in writers' use of demonstratives: A corpus study demonstrating the importance of discourse genre.
Maes, A. & Noordman, L. G. M. (1995). Demonstrative nominal anaphors: A case of nonidentificational markedness. *Linguistics* 33(2), 255–82.
Moskovit, L. (1983). When Is broad reference clear? *College Composition and Communication* 34(4), 454–69.
Oh, S.-Y. (2001). A focus-based study of English demonstrative reference: With special reference to the genre of written advertisements. *Journal of English Linguistics* 29(2), 124–48.
Paraboni, I. & van Deemter, K. (2002). Towards the generation of document-deictic references. In K. van Deemter & R. Kibble (eds), *Information sharing: Reference and presupposition in language generation and interpretation* (329–54). Stanford: CSLI Press.
Passonneau, R. J. (1989). Getting at discourse referents. In *Proceedings of the 27th Annual Meeting on Association for Computational Linguistics*, 51–9. Canada: Association for Computational Linguistics.
Peeters, D., Krahmer, E. & Maes, A. (2021). A conceptual framework for the study of demonstrative reference. *Psychonomic Bulletin & Review* 28(2), 409–33.
Peeters, D. & Özyürek, A. (2016). This and that revisited: A social and multimodal approach to spatial demonstratives. *Frontiers in Psychology* 7, 222.
Petch-Tyson, S. (2000). Demonstrative expressions in argumentative discourse: A computer corpus-based comparison of non-native and native English. In S. P. Botley & T. McEnery (eds), *Corpus-based and computational approaches to discourse anaphora*, 43–64. Amsterdam: John Benjamins Publishing.
Piwek, P., Beun, R.-J. & Cremers, A. (2008). 'Proximal' and 'distal' in language and cognition: Evidence from deictic demonstratives in Dutch. *Journal of Pragmatics* 40(4), 694–718.
Poesio, M. & Modjeska, N. N. (2005). Focus, activation, and this-noun phrases: An empirical study. In A. Branco, T. McEnery & R. Mitkov (eds), *Anaphora processing: Linguistic, cognitive and computational modelling* (429–42). Amsterdam: John Benjamins Publishing.
Potts, C. & Schwarz, F. (2010). Affective "this". *Linguistic Issues in Language Technology* 3(5), 1–30.
Prince, E. F. (1981a). On the interfacing of indefinite-this NPs. In A. K. Joshi, B. L. Webber & I. A. Sag (eds), *Elements of discourse understanding*, 231–50. Cambridge: Cambridge University Press.
Prince, E. F. (1981b). Towards a taxonomy of given-new information. In P. Cole (ed), *Radical pragmatics*, 223–55. New York: Academic Press.
Quirk, R., Greenbaum, S., Leech, G. & Svartvik, J. (1985). *A comprehensive grammar of the English language*. London: Longman Inc.
Recasens, M. (2008). Discourse deixis and coreference: Evidence from AnCora. In *Proceedings of the 2nd workshop on anaphora resolution (WAR) II*, Bergen. 73–82. Bergen.
Reile, M., Plado, H., Gudde, H. B. & Coventry, K. R. (2020). Demonstratives as spatial deictics or something more? Evidence from common estonian and võro. *Folia Linguistica* 54(1), 167–95.
Roberts, C. (2002). Demonstratives as definites. In K. van Deemter & R. Kibble (eds), *Information sharing: Reference and presupposition in language generation and interpretation*, Stanford. 89–136. CSLI Press.
Rustipa, K. (2015). The use of demonstrative pronoun and demonstrative determiner "This" in upper-level student writing: A case study. *English Language Teaching* 8(5), 158–167.
Schlegloff, E. A. (1996). Some practices for referring to persons in talk-in-interaction: A partial sketch of a systematics. In B. A. Fox (ed), *Studies in anaphora*, 437–86. Amsterdam: John Benjamins Publishing.
Schnedecker, C. (2006). SN démonstratifs "prédicatifs" Qu'est-ce qui limite leur apport informatif ? *Langue Francaise* 152(4), 39–55.
Tomasello, M. (2008). *Origins of human communication*. Cambridge, MA: MIT Press.
Toole, J. (1996). The effect of genre on referential choice. In T. Fretheim & J. K. Gundel (eds), *Reference and referent accessibility*, 263–90. Amsterdam: John Benjamins Publishing.
Uryupina, O., Artstein, R., Bristot, A., Cavicchio, F., Delogu, F., Rodriguez, K. J. & Poesio, M. (2020). Annotating a broad range of anaphoric phenomena, in a variety of genres: The ARRAU corpus. *Natural Language Engineering* 26(1), 95–128.

https://doi.org/10.1017/langcog.2021.28 Published online by Cambridge University Press

Verhagen, A. (2005). *Constructions of intersubjectivity. Discourse, syntax, and cognition*. Oxford: Oxford University Press.
Vieira, R. & Poesio, M. (2000). An empirically based system for processing definite descriptions. *Computational Linguistics* 26(4), 539–93.
Voorhees, E. & Graff, D. (2008). *AQUAINT-2 information-retrieval text research collection*. Philadelphia: Linguistic Data Consortium.
Webber, B. L. (1988). Discourse deixis: Reference to discourse segments. In *Proceedings of the 26th annual meeting of the association for computational linguistics*, 113–22. New York: Association for Computational Linguistics.
Webber, B. L. (1991). Structure and ostension in the interpretation of discourse deixis. *Language and Cognitive Processes* 6(2), 107–35.
Wilkins, D. (2018). The demonstrative questionnaire: "THIS" and "THAT" in comparative perspective. In S. C. Levinson, S. Cutfield, M. Dunn, N. Enfield, S. Meira & D. Wilkins (eds), *Demonstratives in cross-linguistic perspective*, 43–71. Cambridge: Cambridge University Press.
Wittenberg, E., Momma, S. & Kaiser, E. (2021). Demonstratives as bundlers of conceptual structure. *Glossa: A Journal of General Linguistics* 6(1), 33.
Wolter, L. (2006). That's that: The semantics and pragmatics of demonstrative noun phrases. PhD thesis, University of California, Santa Cruz.
Wulff, S., Römer, U. & Swales, J. (2012). Attended/unattended this in academic student writing: Quantitative and qualitative perspectives. *Corpus Linguistics and Linguistic Theory* 8(1), 129–57.

**Cite this article:** Maes, A., Krahmer, E. & Peeters, D. (2022). Understanding demonstrative reference in text: a new taxonomy based on a new corpus *Language and Cognition* 14: 185–207. https://doi.org/10.1017/langcog.2021.28

https://doi.org/10.1017/langcog.2021.28 Published online by Cambridge University Press