10 Representing Social Actors

10.1 Introduction

We can think of healthcare as a constellation of social practices: as the repeated performance of actions determined by the co-constitutive relationship between structure and individual agency (Giddens, 1984). In other words, individuals can make decisions that impact their health, such as going to see the doctor for advice and treatment and manging their diet and exercise, but healthcare outcomes equally depend on structural effects, such as the availability of treatments, nutritious food, or facilities that enable active lifestyles. Structural aspects and individual (health) choices are co-constitutive in that one determines the other (Maller, 2015). By conceptualising healthcare in terms of social practices, we can posit the various stakeholders (i.e., health practitioners, patients, governments, business owners, news writers etc.) as social actors – agents that play a key role in creating our social world, including how healthcare is delivered.

Studies of social actor representation have typically drawn on Van Leeuwen's (1996) influential social actor network, in which he describes the various strategies through which participants are foregrounded or backgrounded when they are identified in the text or otherwise excluded entirely. For example, social actors can be described according to their functional role or physical characteristics, and individuals can be marked out independently or presented as part of a collective. In short, the key question driving social actor analysis can be expressed as Darics and Koller (2019: 224) posit in their framework: 'How are social actors represented: as active or passive, as more or less agentive, and in personal or impersonal ways?'

Relatedly, we can consider what participants are represented as *doing* as indicative of their role and their capacity to contribute to the delivery of healthcare. One of the concepts through which researchers have investigated the actions and behaviours that social actors are involved in is 'transitivity', and transitivity analysis offers a framework for classifying processes that assists us in pointing to different levels of agency (Halliday and Matthiessen, 2014). Furthermore, if we posit that representations of social actors are choices,

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 21:47:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009477680.010

consistent with the principles of the wider conceptual framework for systemic functional linguistics, we can consider the selection of particular naming strategies and reported actions in relation to the institutional and social contexts in which they appear. Or, as Van Leeuwen (2008: 33) asks, 'What interests are served by them, and what purposes achieved?'

In this chapter, we demonstrate how corpus approaches assisted us in examining the representation of social actors: first, in UK media coverage of people with obesity and, second, in interview responses provided by participants with psychosis, leading them to hear voices that others cannot. Through discussion of these case studies, we will show how recurrent naming strategies contribute to the stigmatisation of people with obesity and how the actions ascribed to voices demonstrates how voice-hearers navigate experiences that are often distressing. Ultimately, we will show that there are steps in the analysis of social actors that can be supported by the computationally driven procedures of corpus analysis, while other aspects are better suited to the more contextualised and discerning reading provided by the informed analyst.

10.2 Representing Obesity in the British National Press

In this section we describe how members of the CASS team identified representations of people with obesity in a 36-million-word corpus of British newspaper articles about obesity published between 2008 and 2017 (see Brookes and Baker, 2021, and Sections 2.3, 3.2, and 7.2 of this book). The articles were collected using the LexisNexis online news database, with the stipulation that they had to contain at least one mention of the word *obese* or *obesity*. The researchers had conducted similar types of research on newspaper discourse in the past. For example, Baker and co-authors (2013) examined a 140-million-word corpus of news articles about Muslims and Islam. They examined representations by conducting collocation and concordance analyses around a small set of relevant words like *Muslim*, *Muslims*, *Islam*, and *Islamic*. However, when examining the corpus of articles about obesity, it was not as easy to focus the analysis on a few words, as initial examination of samples of articles indicated that people with obesity were referred to in a wide number of ways, ranging from positive to neutral to euphemistic to explicitly stigmatising.

10.2.1 Social Actors in Obesity Coverage

In their analysis, Brookes and Baker (2021) incorporated Reisgl and Wodak's discourse historical approach (2001), which like social actor analysis considers referential/nomination and predication strategies. The former pertains to how individuals are named and referred to, while the latter deals with how they are described and what qualities or characteristics are attributed to them.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 21:47:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009477680.010

A nomination strategy might involve using a noun like *fatso*, while a predication strategy could include adjectives like *pathetic* or verbs describing certain actions (e.g., *wolfing down food*) or being the recipient of actions (e.g., *labelled as obese*).

Additionally, Brookes and Baker (2021) utilised Van Leeuwen's (1996) social actor representation framework, which offers a system for categorizing the ways that social actors are portrayed in English discourse. For instance, terms like *the obese* involve physical identification, uniquely defining individuals based on their physical characteristics within a given context. This framework helped researchers identify how references to people with obesity can either include or exclude, personalise or impersonalise, assimilate or differentiate them in the discourse. The researchers uploaded the corpus into the online analysis tool CQPweb (Hardie, 2012), so the analysis described in the following section uses the search facilities associated with that tool, although other tools should allow for similar kinds of searches to be carried out.

10.2.2 Nomination Strategies

The analysis began with the identification of nomination strategies. Once a set of those were found, Brookes and Baker (2021) could then more easily find the predication strategies through the consideration of collocates and concordance lines. Identifying nomination strategies was not a simple task, however, and the researchers employed several non-corpus-assisted and corpus-assisted tactics in order to produce a list.

The two main non-corpus-assisted methods involved using introspection and reading samples of articles from the corpus. The researchers asked friends, family members, and colleagues to think of possible terms that might appear in the corpus, in order to triangulate from different perspectives. Subsequent to the completion of the study, the researchers decided to see whether a tool like ChatGPT would have been useful for identifying further terms. ChatGPT initially offered a list of terms that it described as 'neutral and descriptive', like *person with a higher body mass*. Its initial response was lightly chastising, stating: 'Remember to use language that is respectful and non-stigmatizing when discussing individuals' weight or body size.' However, when the purpose of the study was explained in more detail, ChatGPT produced a wider list of words which included stigmatising ones like *whale* and *blob*.

The corpus-assisted methods of eliciting nominations involved the examination of frequency and keyword lists. The researchers identified keywords for each newspaper by using 'the remainder method' (i.e., comparing a frequency list of articles from a single newspaper against a reference corpus consisting of the remaining newspaper articles they had collected; see Chapter 7). This helped identify terms that tended to be popular in a single paper. For instance,

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 21:47:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009477680.010

nouns like *fatties* and *flab* were prominent in *The Star* compared to the other newspapers in the corpus. The researchers also tried to identify relatively infrequent nouns in each newspaper by looking at wordlists. For example, when they obtained a frequency list of *The Times*, they looked at plural nouns that occurred between 4 and 10 times in the corpus. This produced words like *bloaters*, *heavyweights*, and *overeaters*, which had not been identified by any other means.

10.2.3 Collocation Analysis

Once they had an initial set of words, Brookes and Baker (2021) identified collocates, focussing particularly on adjectival and verb collocates, as these were most likely to reveal predication strategies. These collocates also provided an additional route for finding new nominations. For example, when examining collocates of *slob*, the researchers found adjectives like *fat*, *boozy*, *lazy*, and *out-of-shape*. Conducting searches of these adjectives and examining words which collocated with them produced additional nomination labels like *lard-arse*, *hog*, *porker*, *lump*, and *fatty*. As they conducted their analyses, the researchers kept an open eye when reading concordance lines, in order to identify further terms. This meant that they had to regularly update frequency tables as new terms were elicited during the analysis.

The researchers were working with a grammatically tagged version of the corpus, which allowed them to easily expand the analysis to account for both singular and plural forms of nouns (e.g., *slob*, *slobs*), as well as comparative and superlative forms of adjectives (e.g., *chunky*, *chunkier*, *chunkiest*) and related verb tenses (e.g., *guzzle*, *guzzles*, *guzzled*, *guzzling*). Within CQPweb, putting a search word within curly brackets retrieves all its related grammatical forms. For instance, searching for {waddle/V} would elicit cases of *waddle*, *waddles*, *waddling*, and *waddled* as verbs.

For some terms, searches elicited a large number of false positives (i.e., unwanted cases). For example, the word *fat* often referred to the amount of fat in food. The researchers could search just for adjectival cases (tagged JJ), using the search query fat_JJ, although this still produced a high number of noun cases that were erroneously tagged as adjectives. Other cases used *fat* to modify nouns that did not refer to human beings (e.g., *fat crisis*, *fat gene*, *fat camp*). There were 22,578 cases of *fat* tagged as an adjective, so in order to obtain a reasonable estimation of the number of times the word *fat* appeared as an adjective which referred to a human being, the researchers carried out a qualitative examination of 10 per cent of concordance lines of fat_JJ, presented in random order (2,258 cases). They then multiplied the number of cases that referred to people by 10. Another set of unwanted cases involved the phrase *the obese*, which sometimes occurred as a collective noun but also often

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 21:47:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009477680.010

occurred as part of a longer noun phrase, such as *the obese singer*. A similar strategy to *fat* was used to isolate only the cases where *the obese* was used as a collective. This term only occurred 1,703 times in the corpus, so it was possible to carry out a concordance scan of all cases and remove cases like *the obese singer*, which didn't apply.

The researchers also obtained the collocates of a set of nouns that more generally referred to human beings (e.g., *people*, *men*, *women*, *children*, *kids*, *boys*, and *girls*), as these words sometimes occurred as parts of noun phrases relating to obesity (e.g., *obese people*, *supersize kids*). Different collocational statistics are likely to produce different kinds of findings. For example, the log ratio and mutual information statistics tended to produce low-frequency collocates which would point to rare nominations (e.g., *super-obese* is a top log-ratio collocate of *people*, occurring 11 times in the corpus). On the other hand, log likelihood foregrounded more frequent collocates like *obese*, which co-occurred with *people* 3,714 times.

In order to consider representation beyond simple noun labels, the researchers aimed to find a list of adjectives and verbs that tended to be used to refer to people with obesity in the corpus of news articles. Once they had identified a list of nominations, it was possible to obtain collocates for each one. As there were a large number of terms, comprising both singular and plural forms, a short cut to obtaining lists of adjective and verb collocates was to combine searches on multiple words in CQPweb. For example, researchers can put a number of similar terms within brackets, separating them with the | symbol, to consider them together:

(pig_NN1|hog_NN1|fatty_NN1|slob_NN1|porky_NN1|overweight people)

Researchers can then obtain the collocates of these collated terms, which would save time compared with having to conduct separate searches on each word. The researchers checked the resulting concordance lines of collocates to remove false positives. For example, based on the search query provided, the top-10 collocates (using log ratio) are *hydroxyl*, *Peppa*, *'Oi*, *suckling*, *guinea*, *Percy*, *Daddy*, *salty*, and *metabolise*. Most of these collocates are modifying non-human cases such as *hydroxyl fatty acids* and *Peppa Pig*, and thus would need to be removed completely or at least have their frequencies adjusted.

10.2.4 Predication Strategies

Once Brookes and Baker (2021) had removed the false positives, they were able to obtain a list of adjectives used to refer to people with obesity, which was illuminating in revealing typical qualities that are associated with them. The researchers grouped the adjectives into categories which related to attractiveness (e.g., *ugly*, *beautiful*), health (*ill*, *unfit*), emotional state (*depressed*, *happy*), level of activity (*lazy*), intelligence (*stupid*), or other qualities (*funny*,

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 21:47:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009477680.010

proud). However, the researchers also wanted to identify adjectives that refer more directly to obesity as opposed to other qualities, such as *tubby*, *solid*, and *expansive*. Adjectives such as these can appear as predicative (e.g., *he is tubby*), and when attributive, they might not necessarily modify nouns relating to obesity but rather occur with general nouns like *person* or *man*. Again, a number of retrieval strategies were employed in order to obtain a list of these adjectives: introspection, reading samples, concordance lines, and through prospective collocation searches on relevant nouns or pronouns.

Once they had obtained a list of these adjectives, the researchers categorised them into those which were positive, euphemistic, or directly insulting towards people with obesity. One additional aspect of these adjectives that emerged as the researchers collected them was that some of them seemed to be gender-specific (e.g., *voluptuous* and *curvy* were used to refer to women with obesity, while *beefy* and *portly* referred to men with obesity), so the researchers noted such cases accordingly. Further, they noted that many of these words were often used alliteratively. For example, of the 125 references to *porky*, 63 (50.4 per cent) were followed by another word beginning with *p* (e.g., *pooch*, *pets*, *PCs*, *pupils*, etc.). Similar alliterative patterns are found with the other euphemistic adjectives (e.g., *chubby children*, *tubby toddlers*, *lardy lags*, *hefty hounds*, *flabby felines*, *voluptuous vixens*). Such cases are more typical of tabloid newspapers, which often use literary devices to make headlines more eye-catching. However, in terms of representing people, it could be argued that this alliterative language can result in a fictionalising effect, encouraging readers to think about the subjects as caricatures and distancing them from thinking about them as actual human beings.

Using collocates, the researchers also obtained lists of verbs that were used to modify references to people with obesity. Verbs were classified into groups which referred to

- food being consumed quickly and in large quantities (e.g., *gorge*, *scoff*, *cram*, *shovel*);
- ungainly movement (*lumber*, *waddle*, *jiggle*);
- discomfort (*sweat*, *wheeze*);
- weight gain (*balloon*, *pile on*); and
- difficulty fitting into space (*cram*, *wedge*, *clog*, *squeeze*).

These verbs were considered further in terms of the metaphors that they implied – for example, the verbs related to eating sometimes represented people with obesity as animal-like (e.g., *wolf*, *swill*, *pig out*). The researchers also considered verbs that positioned people with obesity as patients of actions, as opposed to agents. (The Sketch Engine analysis tool has a Word Sketch function that can make this distinction automatically; see Chapter 9.) This led the researchers to a small set of verb collocates like *brand*, *dub*, and *shame*, which were often used in sympathetic articles about people with obesity, where

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 21:47:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009477680.010

they were sometimes presented as objects of pity or were the subject of redemptive weight-loss narratives. For example,

> Tubby teen to trolley dolly: Junk food addict **branded** "**fat** friend" on girls' holiday sheds 6st and lands job as air stewardess. (*Mirror*, 29 March 2017)

This mention of narratives points to a further, more qualitative stage of the analysis, which extends beyond identifying collocates to consider the wider contexts in which people with obesity are involved. This can be a less systematic form of analysis to engage in, and even a concordance line analysis may not be especially productive, requiring analysts to expand concordance lines to read entire texts to obtain a sense of the narrative structure of a particular article. Continuing with the redemptive weight-loss narratives, for example, the researchers found a pattern whereby numerous news articles (particularly in tabloids) told a real-life story about an individual (often an ordinary member of the public, although sometimes a celebrity) who had been bullied in childhood for being overweight and had gone on to lose a lot of weight as a result of a particular diet or exercise regime. The individual was described as having undergone an amazing and positive transformation, both physically and in terms of achieving success in other areas of their life. Such stories were somewhat formulaic, although the researchers also noted what they did *not* include, such as criticism of the bullies or acknowledgement that many people who lose a lot of weight quickly are likely to regain it after a few years. These narratives are therefore perhaps both inspiring and misleading to readers.

Alongside these narratives, the researchers' more detailed readings of concordance lines helped them identify two other ways that people with obesity were represented – as subject to ridicule and criminals. For example, they found a number of cases involving stories of convicted criminals who were described as having obesity:

> BEHIND BARS; Evil Huntley clinically obese after guzzling chocolates. SOHAM killer Ian Huntley has ballooned to almost 18 stone after bingeing on Wispa and Toffee Crisp chocolate bars in his cell. The 35-year-old monster is now clinically obese and is nicknamed "Blobby" by fellow lags. (*Star*, 31 May 2009)

This article observes Huntley's weight gain, portraying it as deserving public attention. However, reporting on his weight gain is tinged with a sense of delight, which, considering the overall media stance on weight gain, may be interpreted as a form of fitting retribution (with the article labelling Huntley as *evil* and *a monster*). The *Sun* article also implicitly links weight gain and crime, and the researchers also found articles that referred to people whose criminal behaviour was described as being directly related to their obesity:

> Obese woman who went on TV to complain that she was too fat to get a job caught stealing cakes just hours after This Morning appearance. (*Mail*, 29 March 2013)

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 21:47:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009477680.010

Such stories collectively contribute to the portrayal of individuals with obesity as clumsy, petty wrongdoers, associating negative traits like greed and laziness with their weight. Consequently, these narratives contribute to depicting people with obesity in a comedic role. While the crimes of individuals like Huntley are not presented as humorous, stories about their weight gain are portrayed more frivolously, demonstrating the scope within which obesity is presented as subject for ridicule. Although such stories are not as prevalent as the weight-loss narratives mentioned earlier in this section, their frequency is sufficient to identify a pattern. The casual association of crime with obesity arguably represents one of the most unfavourable forms of representation in the corpus.

In summary, the case study described in this section indicates how representations can be found in corpora by beginning with the idea of nomination and nouns, and gradually expanding the remit to include predicative and attributive adjectives, as well as verbs which position social actors as agents or patients. We also saw how the analysis was expanded further to consider metaphors, gendered language, and stylistic features like alliteration, and then finally went beyond the lexis to explore narratives. In doing so, the analysis moved from the quantitative to the qualitative. While it cannot be said that every form of representation of people with obesity in the corpus has been identified, what the approach taken here does offer is a representative picture in order to draw meaningful and reasonable conclusions, while also allowing the researchers to provide information about the frequencies of different kinds of representations in comparison to one another.

10.3 Investigating the Agency of Voices in Psychosis

Our second case study concerns the experience of auditory verbal hallucinations (AVHs), commonly referred to as 'voices', as a kind of psychosis that is associated with mental health conditions such as schizophrenia, bipolar disorder, and borderline personality disorder (Woods et al., 2022). Such hallucinations are defined as 'sensory perceptions in the absence of any externally generated stimulus' (Lindenmayer and Khan, 2006: 198) or, to put it another way, an individual's perception of voices that others cannot hear. There are issues with the terms 'voices' and auditory verbal hallucinations, since these experiences can manifest in ways that are not 'heard' (Wilkinson and Bell, 2016). Nevertheless, in the absence of a suitable alternative, we adhere to the convention among researchers in the field, and indeed generally among those with lived experience, to use the term 'voices' to capture stimuli that might be otherwise described as ambient rather than communicative (e.g., *buzz*) or as some other sensation (e.g., *flashing*, *shaking*).

Voices in psychotic disorders are typically distressing and cause disruption in the lives of those who experience them, though interactions between the

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 21:47:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009477680.010

voice-hearer and their voices can range in their complexity and affect (Woods et al., 2022). Subsequently, there is growing interest in the ways that voices are personified and represented as social agents (see Wilkinson and Bell, 2016). It was on this basis that CASS researchers saw a way for language-based models of agency and social actor representation to contribute to mapping out the interpersonal mechanisms of voice-hearing. In what follows, we discuss transitivity as a system through which analysts can document the agency of social actors as it is demonstrated through processes. We reflect on aspects of existing frameworks that require some decision-making on the part of the analyst when applied to language data in context and consider how corpus procedures can assist in such investigations.

10.3.1 Hearing the Voice: Interviews with People Experiencing Psychosis

The data for this study were collected by the Hearing the Voice (HtV) team, operating in northeast England and carrying out work to better understand how voice-hearing experiences change over time (Woods et al., 2022). The HtV team interviewed volunteers who had sought support from local early intervention in psychosis services for their voices, and those interviews were transcribed and made available to CASS researchers for (corpus) linguistic analysis. Participants were given pseudonyms, and other personal information was anonymised. Forty individuals took part in semi-structured interviews with the HtV team, describing what their voices are like, when they started, what they do and say, and how they have changed over time. This amounted to 205,941 words of participant data that the CASS team compiled and analysed as a corpus.

One of the objectives of the HtV team was to investigate the personification of the reported voices and how this varies according to degrees of complexity. This aim was predicated on the understanding that voices can variously be attributed attitudes, intentions, and different kinds of identities (even names). They can also manifest as other non-human (*demon, birds*, *bomb*) or abstract entities (*thoughts*, *scenario*, *sensation*). The HtV team established a binary classification for manually coding the reported voices in terms of the complexity of personification, according to the following definitions:

Minimal personification: The voice has few person-like qualities; is attributed to a person or described as being ‘like a person’ but without further elaboration. Person-like characteristics tend to remain stable over time and follow a single theme (e.g., the voice is ‘mean’, or a ‘nasty man’).

Complex personification: The voice is described as having more than one kind of person-like quality. These may include elaborate descriptions of intentional states (the voice wants/thinks/feels), agency (the voice will ‘make something happen’), or identity

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 21:47:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009477680.010

(the voice 'comes' from somewhere or has a specific and idiosyncratic ontological status). Complexity . . . will typically involve a voice being attributed multiple, qualitatively different person-like states. (Alderson-Day et al., 2021: 233)

Based on these definitions, the CASS research team drew on concepts from literary linguistic theory to develop a characterisation model through which analysts could formalise the documentation of personness in the descriptions of voices by voice-hearers (see Semino et al., 2021).

In this chapter, we focus specifically on agency as a component of the HtV team's definitions for personification and our own view of recording personness (Semino et al., 2021). Agency has also been shown to be a key part of other taxonomies developed among clinical psychologists for describing personification (see Wilkinson and Bell, 2016). From a linguistic perspective, the CASS team was interested in a semantic view that allowed them to discuss *degrees* of agency, as opposed to a strictly binary grammatical view (i.e., who is active and who is passive in a clausal structure; Darics and Koller, 2019). As Darics and Koller (2019: 219) argue, 'Clearly, it is more agentive to effect a material change in the world . . . than to merely become or be something'. To this end, we can refer to the transitivity model and the classification of process types to help us discuss the different ways in which social actors are reported to have agency.

10.3.2 Transitivity

Transitivity is one element of the complex system of linguistic analysis known as systemic functional linguistics (SFL), most commonly associated with Halliday (e.g., Halliday, 1994; Halliday and Matthiessen, 2014). Transitivity is oriented around meaning insofar as it represents the happenings around us, the states of affairs of the world, and our responses to them. The transitivity framework gives us the resources to describe and evaluate this flow of events 'as a configuration of elements centred on a process' (Halliday and Matthiessen, 2014: 213). These processes involve participants (who enact and are affected by these processes) and circumstances (how, when, and where these processes occur). Each component can broadly be mapped onto clausal elements – that is, the participants are typically realised in the nominal group (*it*, *the voices*), circumstances by the adverbial group (*today*) or prepositional phrase (*from next door*), and processes in the verbal group (*started*, *will be going*).

The various processes through which we can refer to the many different aspects of our experience have been classified into the following process types (Halliday and Matthiessen, 2014: 214–15):

- Material: construing the outer experience of physical actions
- Mental: representing the inner experience of thoughts and emotions
- Relational: referring to processes of identifying and classifying

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 21:47:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009477680.010

- Behavioural: representing the outer manifestations of inner workings such as laughing or sleeping
- Verbal: concerned with saying and expressing meaning
- Existential: recognising the existence or happening of various phenomena

However, Halliday and Matthiessen (2014: 218) acknowledge that these are 'fuzzy categories'; in other words, the boundaries between them are not clear-cut. In response to ambiguities in the classification, alternative taxonomies have been developed and the prevailing schools in this area are those derived from Halliday (1994), known as the Sydney model (SM) and the Cardiff grammar (CG) model, best described in Fawcett (2000) and Neale (2002). The CG model proposes an alternative taxonomy for process types, namely action, mental, relational, influential, environmental, and event-relating. One of the key differences between the models is indicated in the influential category, which captures verbs such *start*, *try*, *continue*, or *stop*, as well as verbs that denote success and failure (Bartley, 2018). In our discussion, we will take the SM taxonomy as our starting point and refer to pertinent aspects of the CG model as they correspond with the challenges identified through the CASS research team's investigation of processes described in their interviews with voice-hearers.

10.3.3 Querying References to Voices

Since the CASS team were interested in the variety of labelling terms used to refer to voices, a preliminary task in their analysis was to locate references to voices in the participant responses. A member of the research team read through the interview transcripts and manually tagged references to voices in the corresponding corpus files. This process demonstrated how identifying a referent is based on contextual information, particularly in the case of pronouns (i.e., differentiating when 'it' referred to a voice and when it did not) and when the referent is introduced in a preceding turn. For example,

> Interviewer: you mentioned that two are male and one was female
> Participant: Female, yeah.

Manual identification of references to voices confirmed that the range of terms used was beyond the team's introspection and showed that voices could be invoked according to different parts of speech:

- Pronouns: *it, she, they*
- Nouns: *voices, shadow, Roxy*
- Determiners: *this, some, which*
- Verbs as gerunds: *commenting, whispering*
- Adjectives: for example, 'there's one *good* and one *bad'*

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 21:47:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009477680.010

As such, the procedure of identifying references to voices in the first instance was not amenable to (a simple) corpus query. However, once the researchers had manually tagged references to voices in the data, they were able to run corpus queries for frequency and distribution of voice label types (see Collins et al., 2023). This allowed them to determine that there were 9,030 instances of references to voices in the data, expressed as 392 different types. The most common types were *it* (1952), *they* (1518), *he* (732), *she* (673), *them* (481), and *voices* (383).

Furthermore, the tagging of voice referents allowed the CASS team to perform additional queries that highlighted associated qualities and actions through adjective and verb collocates. In particular, the voice labels (i.e., the referents that we tagged) and their collocates directed the researchers to identification strategies that correspond with Van Leeuwen's (2008) social actor network. For example,

- Classification: *girl*, *man*, *feminine*, *childlike*, *old*
- Relational: *dad*, *ex-girlfriend*
- Physical: *big, angular*
- Appraisement: *evil*, *clever*, *useless*

These are discussed in some detail by Collins and colleagues (2023), including which kinds of labelling strategies are favoured in cases of minimal and/or complex personification. Here, we consider the reporting of actions, positing the voices as social actors with various degrees of agency.

10.3.4 Classifying Process Types

The investigation of processes associated with voices was guided by the identification of verb collocates of the voice labels that the researchers tagged. The proximity of a referent and a verb does not entail a definite clausal relationship (i.e., we cannot assume that the verb is carried out by or acts upon our node as a participant in the clause). Nevertheless, we can set our collocational window to maximally target subject-verb combinations. Manual checking of the outputs from alternative settings (i.e., three/four/five tokens to the left/right of the node) showed that three tokens to the right of the node were optimal for the precision and recall of processes directly attributable to a voice. The research team set a minimum frequency of one as an association measure in order to capture the full list of verb collocates. They identified 462 different lemmatised verb types, the most common of which were *be* (1569), *say* (433), *come* (280), *go* (270), *do* (265), *get* (252), and *tell* (230).

The researchers then considered automatically categorising the verb collocate types according to Halliday and Matthiessen's (2014) taxonomy; in other words, collating the verb types that correspond with each process type category

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 21:47:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009477680.010

(*talk* as verbal; *think* as mental, etc.). However, it was problematic to reduce each verb type to a singular meaning that corresponded with just one process type category. Other researchers have found that because it is a semantically oriented task, there is a range of contextual factors, along with different usage patterns, that make it difficult to develop algorithms for automatic process-type classification (Yan, 2014). As such, while the CASS team's observations began with a table of the verb collocates and their frequency, the researchers' interpretations were based on instances as they occurred in the original context of the interview.

One of the first steps in classifying process types is to identify the main process in a clause. Bartley (2018: 12) explains that there are differences in how the SM and CG model instruct analysts to approach phrasal elements with a catenative verb. These differences apply to verb phrases such as

- She *started talking* about them.
- He *wants to hurt* me.
- They're just *trying to distract* me.

In such cases, the SM would tend to focus on *talking*, *hurt* and *distract* as processes, whereas the CG model considers *started*, *wants* and *trying* to be the main processes. Furthermore, verbs such as *start* and *try* typify the influential process category in the CG model. Researchers must establish whether they will record one or both these elements of the process and decide which model they want to refer to, if they are looking to (quantitatively) describe patterns according to these top-level process types. Choosing either the SM or the CG model will generate very different views of the (same) data.

In the analysis of the voice-hearer interviews, while the researchers might collectively address the category 'material processes' based on the prevailing meaning of terms such as *control*, *fight*, and *hurt*, it is important to acknowledge the significance of the extended verbal group (*trying to hurt*), given its clear relevance to the discussion of capabilities and intentions – in other words, agency (see the definitions provided in Section 10.3.1). There are other instances in which the researchers focussed on the particular elements of a verbal phrase, based on its relevance to the inquiry into agency. For example, Collins and co-authors (2019: 48) discuss the qualitative differences in how verbs such as *stop* are used: the intransitive form (e.g., *the voices stopped*) demonstrates very limited agency; however, in the non-finite complementation clause 'they've *stopped me from doing* so much', the voice-hearer is explicitly attributing a negative outcome to the voice, which is shown to have the capacity to 'make something happen'.

The CASS research team found that the capacity of the voice(s) to communicate was of particular significance to their investigation, with communication verbs such as terms *say*, *tell*, and *talk* appearing as high-frequency collocates of

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 21:47:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009477680.010

references to voices. Bartley (2018: 13) is critical of the CG model for its classification of communicative processes (under the mental cognition category), as this seems to understate the significance of the action of deliberately transferring information to other sources – which has direct relevance to our discussion of the agency of voices. Verbal processes are treated as a distinct category in the SM and defined as covering 'any kind of symbolic exchange of meaning' (Halliday, 1994: 140). Among the terms that are included in this category, it is possible to further differentiate levels of agency and impact. For example, Collins and colleagues (2019: 48) discuss how *respond* and *answer* indicate the capacity for voice-hearers to participate in dialogue with their voices. This is particularly significant when the unidirectional nature of *threats* and *shouting* is often reported as a source of distress for voice-hearers, when they have limited options for affecting the interpersonal and communicative dynamic.

Similarly, there are certain mental processes that describe different levels of intent and subsequently – different degrees of surveillance or antagonism that can be a source of distress for voice-hearers. For instance, some voices are reported to *judge*, *embarrass*, *hate*, or *reassure* the voice-hearer, demonstrating their capacity to 'want/think/feel' (Alderson-Day et al., 2021). Bartley (2018: 5) explains that the subcategorisation of the CG model offers more detail – compared with the SM – for distinguishing 'the act of consciously perceiving something and doing so intuitively', which we can see in the comparison between *saw* versus *looked* and *hear* versus *listen*. This distinction can be particularly useful in capturing the perceived intent behind the voice *ignoring* the voice-hearer, for example.

10.3.5 Considerations for Investigating Transitivity

In this exploration of voice-hearer accounts, we have established that the concepts of agency and positing participants as social actors are of interest to researchers working in disciplines beyond linguistics, but that language-based theories offer frameworks for documenting dimensions of identification and agency in the labelling strategies and description of processes. While there are established frameworks that are predicated on shared principles for how agency is encoded in language, we have seen that there are different approaches to categorising processes. This means that researchers will need to critically reflect on which aspects they are particularly interested in capturing, particularly if they are looking to provide quantification, since fundamental steps such as identifying the main process in a clause will have implications for what is categorised and how. The different practices for documenting processes and related participants show how this is an analytical procedure that is difficult to automate; nevertheless, we have shown how wordlists and collocation analysis

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 21:47:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009477680.010

(following annotation) can help in identifying the most common terms used to denote participants and processes. Subsequently, we have considered some of the complexities in how processes are described. Given these complexities, processes that appear in the data warrant close, contextualised examination – particularly for the purposes of discerning degrees of agency attributed to different social actors in the text.

10.4 Conclusion

In this chapter we have demonstrated the value of investigating representations of social actors in two different health-related contexts. We have shown that documenting nomination strategies can involve different combinations of human introspection, corpus procedures such as generating wordlists and keywords, and outputs from large language model-based chatbot systems such as ChatGPT. Such combinations reiterate that while frequency-based approaches help establish prevailing patterns for language use, it is also useful to append these techniques with those that are less dependent on frequency, in order to try to capture the breadth of ways in which people talk about those experiencing various health challenges. Similarly, automated processes have their limitations in capturing the polysemy of lexical forms, and the human analyst has an important role to play, not only in choosing and implementing the framework for documenting social actors, but also in interpreting their position in the text and their reported contribution to wider social practices.

References

Alderson-Day, B., Woods, A., Moseley, P., Common, S., Deamer, F., Dodgson, G. and Fernyhough, C. (2021). Voice-Hearing and Personification: Characterizing Social Qualities of Auditory Verbal Hallucinations in Early Psychosis. *Schizophrenia Bulletin*, *47*(1), 228–36. https://doi.org/10.1093/schbul/sbaa095.

Baker, P., Gabrielatos, C. and McEnery. T. (2013). *Discourse Analysis and Media Attitudes: The Representation of Islam in the British Press*. Cambridge University Press.

Bartley, L. V. (2018). Putting Transitivity to the Test: A Review of the Sydney and Cardiff Models. *Functional Linguistics*, *5*(4), 1–21. https://doi.org/10.1186/s40554-018-0056-x.

Brookes, G. and Baker, P. (2021) *Obesity in the News: Language and Representation in the Press*. Cambridge University Press.

Collins, L. C., Brezina, V., Demjén, Z., Semino, E. and Woods, A. (2023). Corpus Linguistics and Clinical Psychology: Investigating Personification in First-Person Accounts of Voice-Hearing. *International Journal of Corpus Linguistics*, *28*(1), 28–59. https://doi.org/10.1075/ijcl.21019.col.

Darics, E. and Koller, V. (2019). Social Actors "to Go": An Analytical Toolkit to Explore Agency in Business Discourse and Communication. *Business and*

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 21:47:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009477680.010

Professional Communication Quarterly, *82*(2), 214–38. https://doi.org/10.1177/2329490619828367.

Fawcett, R. P. (2000). In Place of Halliday's Verbal Group Part 1: Evidence from the Problems of Halliday's Representations and the Relative Simplicity of the Proposed Alternative. *Word*, *51*(2), 157–203. https://doi.org/10.1080/00437956.2000.11432500.

Giddens, A. (1984). *The Constitution of Society: Outline of the Theory of Structuration*. Polity Press.

Halliday, M. A. K. (1994). *An Introduction to Functional Grammar*, 2nd ed. Edward Arnold.

Halliday, M. A. K. and Matthiessen, C. (2014). *Halliday's Introduction to Functional Grammar*, 4th ed. Routledge.

Hardie, A. (2012). CQPweb: Combining Power, Flexibility and Usability in a Corpus Analysis Tool. *International Journal of Corpus Linguistics*, *17*(3), 380–409. https://doi.org/10.1075/ijcl.17.3.04har.

Lindenmayer, M. D. and Khan, A. (2006). Psychopathology. In J. A. Lieberman, T. S. Stroup and D. O. Perkins (eds.), *Textbook of Schizophrenia* (pp. 187–222). American Psychiatric Publishing. https://doi.org/10.1016/j.psc.2007.04.005.

Maller, C. J. (2015). Understanding Health through Social Practices: Performance and Materiality in Everyday Life. *Sociology of Health & Illness*, *37*(1), 52–66. https://doi.org/10.1111/1467-9566.12178.

Neale, A. C. (2002). *More Delicate Transitivity: Extending the Process Type System Networks for English to Include Full Semantic Classifications*. Dissertation. Cardiff University. www.isfla.org/Systemics/Print/Theses/amy_neale_final_thesis.pdf.

Reisigl, M. and Wodak, R. (2001). *Discourse and Discrimination: Rhetorics of Racism and Antisemitism*. Routledge.

Semino, E., Demjén, Z. and Collins, L. (2021). Person-ness of Voices in Lived Experience Accounts of Psychosis: Combining Literary Linguistics and Clinical Psychology. *Medical Humanities*, *47*(3), 354–64. https://doi.org/10.1136/medhum-2020-011940.

Van Leeuwen, T. (1996). The Representation of Social Actors. In C. R. Caldas-Coulthard and M. Coulthard (eds.), *Texts and Practices: Readings in Critical Discourse Analysis* (pp. 32–70). Routledge. https://doi.org/10.4324/9780203431382.

Van Leeuwen, T. (2008). *Discourse and Practice: New Tools for Critical Analysis*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195323306.001.0001.

Wilkinson, S. and Bell, V. (2016). The Representation of Agents in Auditory Verbal Hallucinations. *Mind & Language*, *31*(1), 104–26. https://doi.org/10.1111/mila.12096.

Woods, A., Alderson-Day, B. and Fernyhough, C. (2022). Voices in Psychosis: Interdisciplinary Listening. In A. Woods, B. Alderson-Day and C. Fernyhough (eds.), *Voices in Psychosis: Interdisciplinary Perspectives* (pp. 3–16). Oxford University Press.

Yan, H. (2014). Automatic Labelling of Transitivity Functional Roles. *Journal of World Languages*, *1*(2), 157–70. https://doi.org/10.1080/21698252.2014.937563.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 17 Nov 2026 at 21:47:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009477680.010