Abstract: The focus of much sensory language research has been on vocabulary and codability, not how language is used in communication of sensory perceptions. We make a case for discourse-oriented research about sensory language as an alternative to the prevailing vocabulary orientation. To consider the language of sound in authentic textual data, we presented participants with 20 everyday sounds of unknown sources and asked them to describe the sounds in as much detail as possible, as if describing them to someone who could not hear them. We explored how the participants use language to describe these sounds. Do they describe their listening experiences (stressful), sound properties (intermittent beeping), and/or the events that caused the sounds (eating an apple)? The results show that out of these three soundscape elements, events are the most frequent and most indispensable element. We let the results from the study illustrate the need for more discursive data in studies of sensory language and argue that there is no designated language of sound. Our study highlights that in order to account for sensory language use, we need an analytical framework that accommodates discursive language in a non-trivial way beyond stable couplings between individual words and meanings.

Keywords: cognitive semantics; construal; soundscapes; everyday sounds

1 Introduction

We have seen an increase of interest in the interactions between sensory perceptions and language in recent years (e.g., Caballero et al. 2019; Winter 2019a), and yet there is clearly more to be said on this topic, in particular from a language science point of view. A number of studies have investigated how perceptions are coded in language, typically by means of naming experiments of different types (Majid et al. 2018;
Olofsson and Gottfried 2015a) or through interpretation tasks for sets of words that explore their polysemy range and metaphorical patterning (Hörberg et al. 2022; Ibarretxe-Antuñano 1999; Lynott and Connell 2009, 2013; Lynott et al. 2020; Strik Lievers and Winter 2018; Viberg 2015, 2019). The focus has primarily been on vocabulary, codability and naming, rather than how language *de facto* is used to communicate sensory perceptions. Investigating how speakers express their sensory experiences to others in actual language use requires both discourse data and a usage-based approach.

In this article, we make a case for research about sensory language in which the object of study is language use in authentic communication. We explore a discourse-oriented alternative to the vocabulary approach that currently defines research about language for sensory perceptions (e.g., Majid et al. 2018), and discuss some implications of using naming and discrimination tasks vis-à-vis descriptive tasks and connected discourse. We also make a call for an ecological shift in semantic modeling to allow greater emphasis on meaning-making in authentic communication, and raise the questions: What if actual communication about sensory perceptions entails more than drawing upon lists of dedicated lexicalized items? What would be the consequences for semantic modeling and, more precisely, the modeling of couplings between form and meaning? Our prediction is that when speakers communicate about sensory perceptions, in our case perceptions of sound, the possibilities are far more diverse and complex than lexicalized percepts based on the auditory domain (cf. Caballero and Paradis 2020, 2023). There is, of course, occasion in communication to label experiences by finding appropriate lexicalized terms for them, but when speakers describe their experiences to another person, it is often with the aim of inviting that person to become an immersed experiencer (Zwaan 2004) of their sensory world.

In order to explore the language of sound with the use of authentic textual data, we presented participants with 20 everyday sounds and asked them to describe each sound in as much detail as possible, as if describing it to someone who could not hear it. The participants did not receive any information about the sounds beyond what they could hear, as the aim of the study was to find out how people use language to describe everyday sounds of unknown sources. We set out to investigate: What aspects of the sounds do the participants describe and how do they go about it? Do they describe their experiences of the sounds (calming), properties of the sounds (high-pitched clink), and/or the events that may have caused the sounds (brushing teeth)? How do they construct couplings between wordings and meanings and what language resources are used? The findings we present in this article are based on close analyses of almost 4,000 descriptions totaling over 51,000 words.

In summary, we aim to (i) provide an overview of what speakers focus on when they describe everyday sounds of unknown sources, (ii) account for the conceptual
domains evoked by language in the descriptions and discuss how these are cued linguistically, (iii) discuss how methodological choices and assumptions about linguistic meaning may influence conclusions drawn about sensory language, specifically language about sound. The article is structured as follows: Section 2 reports on previous work on sensory perception and everyday sounds. In this section we also make explicit our assumptions about meaning in language and provide a cognitive semantic framework for the description and explanation of uses of language to describe sound. Section 3 presents the study, the results of which are outlined in Section 4. Section 5 concludes the article.

2 Background

In this section, we report on previous work on language and sensory perception of relevance to this study, and briefly present some research on everyday sounds from academic fields outside the language sciences. We also consider some foundational matters related to linguistic meaning and present a framework for analyzing meaning in language.

2.1 Sensory perception, cognition and language

Sensory perception has received more attention in the human sciences in recent years, but people’s re-contextualization of auditory experiences through language has not been in the spotlight of such research. In the language sciences, sound has been included in studies about sensory language more generally and/or considered in the context of cross-modality, metaphorization and typological comparisons across languages and cultures (e.g., Caballero and Paradis 2015; Dubois 2000; Howes and Classen 2014; Ibarretxe-Antuñano 1999; Koptjevskaja-Tamm and Nikolaev 2021; Majid et al. 2018; Paradis 2015b; Viberg 2015, 2019; Winter 2019a); sound talk in professional discourse (Barten 1998; Caballero 2006; Caballero et al. 2019; Porcello 2004); and sound symbolism and iconicity (Dingemanse 2012; Johansson et al. 2020a, 2020b; Speed et al. 2021; Svantesson 2017; Winter et al. 2017).

Many of these investigations of sensory language have focused on the codability (naming) of sensory perceptions in language. The aim of a study of codability is to ascertain which sensory experiences are encoded as lexical items in a given language, so that, for instance, there is agreement that a particular visual experience of color is called green and another is called blue. In this type of lexically oriented research, methods generally hone in on individual words outside of a context of use. Such methods can facilitate cross-linguistic comparisons of different languages’
vocabularies and provide insight into the diverse constellations of lexical resources for sensory meaning that languages offer. With a similar lexical focus, but with more emphasis on meaning flexibility, some studies have aimed to determine sensory intermodality profiles and perceptual strengths of individual adjectives such as hard, gigantic, bright, sharp (Lynott and Connell 2009), adjectives in antonymic relations (van de Weijer et al. 2023), metaphoric relations (Winter and Strik Lievers 2023), and later also words other than adjectives (Lynott and Connell 2013; Lynott et al. 2020). These studies investigate the extent to which words evoke meanings related to single or multiple sensory modalities.

Underlying assumptions—most often not explicitly stated, but taken for granted—in much of the current literature on the language of the senses appear to be that (i) there is a relatively stable one-to-one correspondence between words and meanings, and (ii) that lexicalized meanings constitute the most effective and/or natural way of communicating sensory experiences (e.g., Juhasz and Yap 2013; Majid and Burenhult 2014; Olofsson and Gottfried 2015a; Speed et al. 2019).

An implication of the first assumption—that there is a one-to-one correspondence between words and meanings—is a problematic disregard of the dynamics of language in use and the meaning multitasking of words. In a conceptually driven study of language for sound, Caballero and Paradis (2020) found that sound descriptions in narrative texts in English and Spanish draw upon a range of domains and domain matrices such as PERCEPTION (She heard the beat of the drums), MOTION (She emitted a hoot of derision), MANIPULATION (A blast from a conch shell cut through the murmuring voice), EMOTION-REACTION (Outside the wind bellowed and raged), and CONSUMPTION (The dank, steaming jungle closed around her, dark and dense, swallowing sound and light). Some wordings conflate meanings for SOUND and ACTION (The doors to her bedroom banged open) and SOUND and MOTION (The gust of wind cluttered to a stop). The main focus of their study was the verb phrase, hence the underscores in the sentences above, but from these examples it should be clear that the sounds are not described by the main verbs alone, but by the conceptually structured and contextually situated expressions (beyond individual words), and that the wordings multitask by evoking a wide range of meanings, drawing upon domains above and beyond SOUND. Also, the descriptions include wordings that would not come to mind as designated ‘sound words.’

An explication of the basis for the second assumption—that sensory words provide the most natural and/or effective way of communicating sensory

---

1 In these two typologically different languages, sound descriptions are instantiated in different domains in a similar way and sounds are described from different perspectives, but the languages differ from one another in that conflations such as SOUND FOR ACTION and SOUND FOR MOTION are not found in Spanish. For an explanation see Pedersen (2019) and Caballero and Paradis (2018, 2020).
experiences—is elusive in the literature. Descriptions that do not take the form of lexicalized percepts, such as source descriptions or experiential accounts, are sometimes referred to as *ad hoc* or as something speakers resort to when their language provides no appropriate sensory vocabulary (Majid and Levinson 2011; Majid et al. 2018). Given the tasks participants are presented with in many lexically oriented studies (to name a sensory experience), answers beyond a label in the form of a single word (*red, swish, pungent*) can be described as a make-shift solution to the task at hand. However, it is unclear to what extent conclusions can be drawn about communication of sensory experiences based on studies of isolated words (Zwaan 2014; Zwaan and Madden 2005).

Another issue in much sensory language research is that uses of fundamental notions such as metaphor, antonymy and abstract/concrete meaning are not explained, making it difficult to assess results and conclusions. There are no theoretical statements about what meaning in language is and how lexical resources are used to create meaning. As we see it, these are non-trivial matters that need to be expounded (see Olofsson and Gottfried 2015b for comments from a neurocognitive point of view). We return to this issue in Section 2.3 in which we present our explanatory framework.

While many lexically oriented studies contribute significantly to our knowledge of lexical meaning and referents of individual words in the sensory domain, they do not extend our knowledge of how meanings are conveyed in natural language production, simply because it is not without pitfalls to draw conclusions about communicative meaning based on the study of decontextualized words (Zwaan 2014: 230). If we consider sensory language outside of communication, we risk overlooking descriptions that come about in extended natural discourse such as similes or metaphors (e.g., Hartman 2017; Hartman and Paradis 2018, 2021; Winter 2019a). This study addresses communication of sound meanings in natural language use, a venture that requires experientially and conceptually driven research. The meaning domain of *sound* is the starting point as we explore how language is used to portray everyday sounds from unknown sources.

### 2.2 Everyday sounds

The influence of sound on human experiences of the world has not been investigated to the same extent as that of visual stimuli, even though there has been an upsurge of interest in recent years in some fields of research (O’Callaghan and Nudds 2009). Environmental sounds, in particular, have not been given much attention, in contrast to speech sounds and music (Carello et al. 2005; Guastavino 2021; Gygi et al. 2007). However, recent research on auditory perception has firmly established that sounds play a key role
in many areas of human experience and may guide behavior and shape our interactions with the world. Sounds influence people’s appreciation of food and beverages (Caballero et al. 2019; Dematté et al. 2014; Spence 2015; Zampini and Spence 2004), allow identification of unseen objects, including their shapes, sizes and materials, and help determine properties of sound-producing actions (Aglioti and Pazzaglia 2010; Lemaitre and Heller 2012, 2013; Lemaitre et al. 2010). Studies have found that sounds are experienced primarily as meaningful events (e.g., Carello et al. 2005; Gaver 1988, 1993a, 1993b; Gygi et al. 2007; Houix et al. 2012; Marcell et al. 2000; Nudds 2009; Steenson and Rodger 2015; VanDerveer 1979), the observation being that “[e]veryday listeners’ care about the source of the sound, not the sound itself. They want to know what happened and what it means for them” (Carello et al. 2005: 80). As cues for event meanings, sounds can provide listeners with information that may guide their actions (Steenson and Rodger 2015).

Models of auditory perception from ecological psychology and the philosophy of perception have proposed that a sound can be conceptualized in different ways. A sound can be characterized by its source (sound-producing event), its properties (qualities of the sound itself), or its influence on a listener (sensations or emotions) (Killin 2022; O’Callaghan and Nudds 2009). Models of auditory perception thus include a few elements regarded as constitutive of an auditory situation: an event, a sound, and a listener, elements that may or may not take center stage in a given experiential context. These essential elements of an auditory situation can also be conceived of as complementary deictic perspectives on a sound from the point of view of a person experiencing/describing the sound: distal, medial and proximal perspectives. In this study, we set out to investigate the ways in which the language used to describe sound reflects these different auditory elements and perspectives on sound. To what extent do the sound descriptions adopt a distal perspective that focuses on sound-generating events (someone eating an apple), a medial perspective that hones in on sound properties (high-pitched, very loud), or a proximal perspective that highlights listeners’ experiences of sounds (gross, happy)? How is language put to use in such descriptions?

We use soundscape to refer to the elements constitutive of an auditory situation. Soundscape are generally defined in environmental terms as the sounds that stem from and characterize a place (Guastavino 2021 for an overview). We likewise use the term for sounds that establish and reinforce auditory environments, but not in a way limited to what we may refer to as ‘sounds of place’. Our use of the term soundscape is conceptual in nature and is in agreement with that of Caballero and Paradis (2020) who use it to describe a Gestalt—a sound event—that includes information about how a sound came about, who/what caused it, from where it was emitted, in what direction it traveled and how it was perceived. A soundscape thus encompasses an event that generates sound and an acoustic signal that is actively perceived by a listener vis-à-vis an experiential context, previous experiences, and current goals.
Based on this auditory input, the listener may be able to simulate an event that may or may not correspond to the causal event.

### 2.3 A cognitive semantic framework

As noted, a limitation of many studies of sensory matters in the human sciences, including the language sciences, is that theoretical assumptions about the nature of linguistic meaning-making are not stated, which means that readers are left in the dark about fundamental matters. In the preface to *Language and Perception* (1976), Miller and Johnson-Laird point to a lack of a theory of meaning that can accommodate the language of the senses. Their work as cognitive psychologists can be said to be among the first attempts at raising awareness of the fact that “the nature of meaning is too important for psychology to set aside” (1976: V). This observation was certainly true in the 1970s, but it is fair to say that there is still a want of work that clearly states its basic assumptions about the nature of meaning and how meanings are created using the resources that languages offer. To overcome this limitation, we need discursive production data as well as a principled semantic framework that can accommodate findings beyond individual words. Such a framework has to account for both conceptual structures and conceptual structuring in communication.

The current study takes a grounded perspective on cognition. From this perspective, cognition emerges from sensory-motor experiences of the world, including their psychological and social contexts (e.g., Barsalou 2020; Goldin-Meadow 2003; Pecher and Zwaan 2005; Varela et al. 1991). Language in turn rests on a socio-sensory-cognitive foundation that is realized in and through communication, so that knowledge representation and meaning-making in language are seen as grounded in three systems (social, sensory, cognitive) that are communicating vessels in the act of communication (Gärdenfors 2014; Paradis 2005, 2015a). This means that we take meanings to be grounded in people’s experiences of the world; the way we perceive the world is the way we conceive of it and talk about it (e.g., Gibbs 2006; Santos et al. 2011; Talmy 2000; Zwaan 2004). Moreover, we see word forms as cues to structures in regions of conceptual space. These spatial regions constitute the entire use potential of a word. The region of the use potential of a word form is not static but constantly changing as an effect of new experiences, new input and/or lack of input from language use (Paradis 2015b). Following Cruse (2001, 2002), we may even say that there is no reason to assume that our intuitive interpretations of isolated words are particularly illuminating or reliable. Every time we use a word, it makes distinct contributions to what we want to communicate, so that what we experience are particularized messages in specific contexts. Outside of such communicative contexts, words have only potential for meaning. This meaning potential (in its entirety)
is greater than what can be realized in a given communicative situation. For example, outside of context a *crack* has potential for meanings that include sound properties (duration, loudness), actions and motion (breaking, explosions, fast movement), objects (branches, whips), and a range of associated experiences (being startled, enjoying a walk in the woods, sitting by a fire). Outside of context, the meaning of a *crack* is by its very nature diffuse (Zwaan 2004); its meaning is both very rich (in potential) and unconstrained, in need of direction.

In the following, we outline a cognitive semantic framework, *Lexical meaning as Ontologies and Construals* (LOC for short) that can facilitate and scaffold investigations into auditory experiences, while also taking seriously the flexibility with which the resources of language are put to use in different communicative contexts. As noted, the dynamics of linguistic meaning-making has been and still is largely a non-issue in research on sensory language, which tends to focus on sensory words (*sticky, pungent, minty, raspy*), treating them as by-default expressions of sensory meanings. Even though there is ample evidence that language is not a naming game in which words represent atoms of meanings with fixed referents, the view of language as a collection of words that map onto entities, properties and activities in a relatively stable manner is still widespread.

Table 1 is a reduced adaption of the LOC framework (Paradis 2005), which we make use of in the analyses of our data.\(^2\) The framework includes conceptual structures (ontologies) as well as resources that account for the dynamics of portrayals of

| Table 1: The Ontologies and Construals framework adapted from Paradis (2005). |
|-----------------------------------------|---------------------------------|---------------------------------|
| **Ontologies (conceptual structures)** | **Configurations**              | **Cognitive processes**         |
| CONCRETE PHENOMENA                    | BOUNDARY                        | Comparison                     |
| EVENTS, PROCESSES, ACTIVITIES, STATES | THING, RELATION                 | Perspective                    |
| ABSTRACT PHENOMENA                   | DEXIS                           | Salience                       |
|                                        | SCALE                           |                                |
|                                        | FORCE                           |                                |
|                                        | PATH                            |                                |
|                                        | FIGURE/GROUND                    |                                |
|                                        |                                 |                                |

\(^{2}\) For a more extensive description and for different implementations of the framework and how it is used, see, for instance, Bianchi et al. (2017), Hartman and Paradis (2018), Jones et al. (2012), Paradis (2004, 2005, 2015a, 2015b), and Paradis and Willners (2011).
experiences and thoughts (construals). The ontologies are not ready lexical meanings, but pre-meaning structures that can be realized through language in use. The conceptual pre-meaning structures are of two types: contentful and configurational.

Table 1 includes only the most general categories of contentful pre-meanings, but these have more fine-grained subcategories (e.g., ENTITY > SOUND > CRUNCH; EVENT > ACTIVITY > CUTTING; STATE > EXPERIENCE > HEARING). Configurational pre-meanings take the form of schematic templates, such as FIGURE/GROUND, DIRECTION and FORCE. In the analyses of the participants’ responses, we are concerned both with contentful meaning (what they describe) and configurational meaning (how this content is structured). For instance, CRUNCH can be configured as THING or RELATION, exemplified by a crunch and crunching/crunchy respectively. We are also concerned with how the participants realize these pre-meanings in communication, which invariably includes the dynamic part of the framework, namely the construals that crystallize in their responses, such as comparison (e.g., in the form of similes or metaphorizations), salience (e.g., in the form of metonymizations), and perspective (e.g., in the form of foregrounding/backgrounding of aspects of meaning). Construal is imposed by language users in actual communication; all contextualized meanings involve construals effected on the occasion of use. On that occasion of use, a portion of a word’s entire pre-meaning is evoked as it is made relevant by context (e.g., linguistic, communicative, experiential, task-related). We can say that all language forms are associated with a body of conceptual structure, which Croft and Cruse (2004: 100) refer to as purport, which outside of a communicative context is not a constructed meaning, but just a body of meaning potential. In Section 4 we make clear how the framework can be used.

3 Data and method

In this study we explore how people use language to describe environmental sounds without known sources and investigate in what ways and to what extent participants describe these sounds in terms of (i) what caused them, (ii) properties of the sounds themselves, and (iii) subjective experiences of the sounds.

3.1 Participants

To connect with potential participants, we advertised the survey on social media and used Prolific to recruit research participants. The Prolific participants were compensated for their time (£8.16/h). The selection criteria used for all participants were: 18+ years old and English as native language. No names, IP-addresses, or other
identifying information were collected. A total of 214 participants took the survey, 150 of which were recruited through Prolific. The number of descriptions collected for each sound ranges from 189 to 199.

3.2 Task

The survey began with an introductory page and one page of instructions. The participants were informed that they would get to listen to 20 sounds, one at a time, that some sounds would be brief and other sounds would be longer, and that shorter sounds would be repeated 3 times automatically with the option of listening to them again. The participants were also given an overview of what they would be asked to do in response to each sound. Following the instructions, 20 sounds (Table 2) were presented in random order (one sound per page) along with questions. The participants were instructed to listen to the sound and to answer the questions after or while the sound was playing. They were then prompted to describe the sound. The

Table 2: Sound stimuli ordered by duration (longest to shortest), labels used to refer to the sounds in the article, and their original labels from the database.

<table>
<thead>
<tr>
<th>Sound labels used in examples</th>
<th>Duration min:sec</th>
<th>Sound labels from database</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fire</td>
<td>2:00</td>
<td>Fireplace crackling</td>
</tr>
<tr>
<td>Grocery</td>
<td>2:00</td>
<td>Grocery store</td>
</tr>
<tr>
<td>Traffic</td>
<td>2:00</td>
<td>Traffic sounds</td>
</tr>
<tr>
<td>Forest</td>
<td>1:29</td>
<td>Forest (bird sounds)</td>
</tr>
<tr>
<td>Rain</td>
<td>1:28</td>
<td>Thunderstorm</td>
</tr>
<tr>
<td>Digging</td>
<td>0:38 (×3)</td>
<td>Digging with shovel</td>
</tr>
<tr>
<td>Water</td>
<td>0:36 (×3)</td>
<td>Moving in water</td>
</tr>
<tr>
<td>Gravel</td>
<td>0:32 (×3)</td>
<td>Walking on gravel</td>
</tr>
<tr>
<td>Heels</td>
<td>0:31 (×3)</td>
<td>Walking down stairs with heels</td>
</tr>
<tr>
<td>Snore</td>
<td>0:27 (×3)</td>
<td>Man snoring</td>
</tr>
<tr>
<td>Dishes</td>
<td>0:23 (×3)</td>
<td>Doing dishes</td>
</tr>
<tr>
<td>Onion</td>
<td>0:23 (×3)</td>
<td>Cutting an onion</td>
</tr>
<tr>
<td>Baby</td>
<td>0:21 (×3)</td>
<td>Baby laughing and/or crying</td>
</tr>
<tr>
<td>Crowd</td>
<td>0:19 (×3)</td>
<td>Crowd cheering at outside event</td>
</tr>
<tr>
<td>Teeth</td>
<td>0:19 (×3)</td>
<td>Brushing teeth</td>
</tr>
<tr>
<td>Chip</td>
<td>0:17 (×3)</td>
<td>Eating crunchy chip</td>
</tr>
<tr>
<td>Apple</td>
<td>0:06 (×3)</td>
<td>Apple bite and chew</td>
</tr>
<tr>
<td>Textile</td>
<td>0:06 (×3)</td>
<td>Textile, tearing cloth</td>
</tr>
<tr>
<td>Tea</td>
<td>0:04 (×3)</td>
<td>Sip of hot tea</td>
</tr>
<tr>
<td>Cough</td>
<td>0:03 (×3)</td>
<td>Woman coughing</td>
</tr>
</tbody>
</table>
verbatim prompt was: *Describe the sound in as much detail as possible, as if you are describing it to someone who cannot hear it.* We included *as if describing it to someone who cannot hear it* to provide a communicative purpose for the description vis-à-vis an imagined recipient without experience of the sound. The study was designed to elicit detailed descriptions of sounds from a large number of participants. To the extent that previous studies have elicited descriptions of environmental sounds, they have imposed limitations on length of descriptions and/or time allotted to compose them resulting in rather brief accounts (e.g., Marcell et al. 2000; VanDerveer 1979). No such limitations were imposed on the participants in this study. Moreover, the task was not designed to orient the listeners to a particular form of description, such as by asking *What do you hear?* Instead, the participants were free to include any information that they deemed pertinent. This description task was the first task in a survey that also included additional subsequent tasks in the form of Likert-scale ratings.

At the end of the survey, the participants answered demographic questions about their gender, age, occupation, and any sensory impairments. Due to the online setup of the study, it was not possible to control the environment of the listening, but we recommended that the participants “sit somewhere quiet (preferably inside)” and told them that they could wear headphones if it made it easier to hear the sounds. At the end of the survey, we included questions about the environment in which they completed the survey and their use (or not) of headphones. The survey was set up so that it was only possible to take it one time from a given device. Participants could edit their answers until they had finished the survey but not after submission. Data are available at: osf.io/xv52g.

### 3.3 Sound stimuli

The sound stimuli (Table 2) were selected from Storyblocks, a subscription database of royalty-free stock sounds. We picked environmental sounds of varying complexity, including brief sounds (e.g., someone taking a sip of tea) as well as longer and more complex sounds (e.g., traffic) because we saw a value in situating our study vis-à-vis research on brief environmental sounds as well as building on the more limited research on complex environmental sounds (i.e., sounds that stem from multiple and often overlapping sound-causing interactions, as opposed to less complex sounds that stem from fewer and often sequential sound-causing interactions). Previous research “has focused on the identification of relatively brief, isolated environmental sounds” whereas “less is known about how people perceive and understand complex, coherent, and lengthy sequences of natural sounds” (Marcell et al. 2007: 561).
We included both sounds of human activities (e.g., taking a bite of an apple) and sounds of non-human events (e.g., a thunderstorm) because research has shown that people respond differently to sounds of living and non-living sources (Giordano et al. 2010). For human sounds, we selected activities that typically engage different parts of the body. Most of these sounds stem from activities that involve physical interaction with the environment (e.g., digging with a shovel). Some vocalizations (no recognizable speech) and machine sounds (e.g., cars and cash registers) were also included. We aimed to include sounds involving different types of matter (solids, gases, liquids) and different sound-producing interactions. Brief sounds (<60 s) were repeated three times in the survey, with 6 s between repetitions. Table 2 shows the names of the sounds we created for our report of the results, the duration of the sounds and whether they were repeated or not, as well as the original names from the database for convenience of reference.

### 3.4 Data

The 214 participants wrote 3,873 descriptions of sounds, totaling 51,089 words. Figure 1 shows the total number of words written for each of the 20 sounds.

**Figure 1:** Total number of words written in the descriptions of each sound.
The descriptions vary considerably in length, not just between the sounds, but also within the sounds. Table 3 details the ranges for number of words and gives the average length of descriptions for each sound. The sound that elicited the longest descriptions on average was Grocery and the shortest descriptions were written in response to Cough. However, all sounds elicited short descriptions (1–2 words) as well as long ones (39–261 words). Overall, longer and more complex sounds (Table 2) elicited longer descriptions on average, as shown in Table 3.

3.5 Analytical steps and considerations

All descriptions were analyzed manually in two steps by the two authors. In the first round, we exported all the descriptions into a database and scrutinized them in order to set up a coding scheme for the analysis (coding scheme with explanations is available at: osf.io/xv52g), (Hartman and Paradis 2023). In the second round, A1 analyzed and annotated all descriptions and A2 did the same for a random set of 20 descriptions for each sound (10 % of the data set). The analyses were carried out
independently by the researchers based on the agreed-upon coding scheme and were then compared. There were only minor differences between A1’s and A2’s respective annotations in the double coding of the smaller data set. The discrepancies were resolved through discussion, after which the analysis of the entire data set was adjusted.

As detailed in Table 4 and illustrated by examples 1–4, the sound descriptions were analyzed according to the three components of soundscapes (EVENT, SOUND, EXPERIENCE) and corresponding deictic perspectives (distal, medial, proximal) that we presented in Section 2.2. Please note that all the responses to the survey describe sound. EVENT meanings describe sounds via events in the sense that something is happening, or somebody is doing something (fire crackling, someone is eating an apple); SOUND meanings describe properties of sounds directly (high-pitched, uneven rhythm, bang); and EXPERIENCE meanings describe sounds via experiences (Super annoying!). Some responses include only one type of description whereas other ones include more than one type of description (see examples 1–4). In the analysis, each response was coded as either including or not including description of EVENT, SOUND and/or EXPERIENCE.

As mentioned, the meaning types (EVENT, SOUND, EXPERIENCE) in Table 4 effect complementary perspectives on sound that can be conceptualized in terms of distance from a listener as distal, medial and proximal, conceptualizing sound respectively as “sound sources (sounding objects or events),” “longitudinal compression waves,” or “sonic mental images or sensations” (Killin 2022: 445 italics in original; also O’Callaghan and Nudds 2009).

Table 4: Coding scheme for the analysis of the sound descriptions.

<table>
<thead>
<tr>
<th>EVENT meanings (distal perspective)</th>
<th>SOUND meanings (medial perspective)</th>
<th>EXPERIENCE meanings (proximal perspective)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGENT (someone, a baby, a man)</td>
<td>QUALITY (clunk, pop)</td>
<td>AUDITION (can hear)</td>
</tr>
<tr>
<td>SOLIDS (an apple, the ground, a brush)</td>
<td>PROPERTIES (high-pitched, loud)</td>
<td>VISION, TOUCH, SMELL, TASTE (can see, feels sharp)</td>
</tr>
<tr>
<td>LIQUIDS (juice, tea, water)</td>
<td>TEMPORAL STRUCTURE (intermittent, continuous)</td>
<td>PLEASANTNESS, EMOTION, AROUSAL (very gross, annoying, love, stressful)</td>
</tr>
<tr>
<td>GASSES (wind)</td>
<td>SPATIAL STRUCTURE (foreground, background)</td>
<td>FAMILIARITY (familiar, strange)</td>
</tr>
<tr>
<td>SETTING (a grocery store, the kitchen)</td>
<td>(INTER)ACTION (chewing, rubbing against)</td>
<td></td>
</tr>
<tr>
<td>VOCALIZATIONS (crying, howling)</td>
<td>VOCALIZATIONS (crying, howling)</td>
<td></td>
</tr>
<tr>
<td>MACHINERY (engine, vehicle)</td>
<td>MACHINERY (engine, vehicle)</td>
<td></td>
</tr>
<tr>
<td>MANNER (vigorously, slowly)</td>
<td>MANNER (vigorously, slowly)</td>
<td></td>
</tr>
</tbody>
</table>
Considering the meanings and perspectives in Table 4 from the point of view of our theoretical framework (Table 1), we can note that the meanings primarily draw upon the contentful pre-meanings of **CONCRETE PHENOMENA** (*baby, apple, engine*) and **PROCESSES, ACTIVITIES and STATES** (*crying, loud, bright*) that exist in **SPACE and TIME** (Table 1, left-most column). As is the case for all meanings, these pre-meanings are not only contentful; when they are used in discourse they are configured in various ways (Table 1, middle column), for instance in terms of **THING** (*baby, apple, engine*) or **RELATION** (all other descriptors that are predicates of other meanings, *crying, loud, bright*), **DEIXIS** (*can hear; in the distance*), or **FIGURE/GROUND** (*in the background*). On the occasion of use, meanings are always subject to construal operations (Table 1, right-most column), which operate on the conceptual structures (the combined contentful and configurational structures) in a dynamic manner in discourse, for instance by affording selective salience to aspects of a soundscape. Examples (1)–(4) illustrate the implementation of the coding scheme and the use of our framework.

(1) *Birds singing in the country side near a small village* (Forest P206)

(2) *This sound is very sharp, bouncy, jagged and rough. The flow is irregular.* (Onion P177)

(3) *It was annoying, unpleasant and aggravating. I would never listen to anything like this again.* (Chip P96)

(4) *A deep bass sound of traffic that makes you feel heavy interspersed with sounds of nature which serves as a nice contrast.* (Forest P170)

In example (1) the sounds of a forest are construed from a distal perspective and described as an **EVENT** (*birds singing*) taking place in a setting (*in the country side near a small village*).³ Example (2) effects a medial construal and describes the sound of cutting an onion directly in terms of **TOUCH** (*very sharp, jagged, rough*) and **MOTION** (*bouncy, flow*). Example (3) takes a proximal perspective on the sounds of someone eating chips by describing an auditory **EXPERIENCE** in terms of **EMOTION and PLEASANTNESS**. Unlike examples (1)–(3), which exemplify one perspective each, the description in (4) illustrates a mixed perspective that evokes **EVENT**, **SOUND**, as well as **EXPERIENCE** meanings. Example (4) accounts for relations between the elements of the soundscape, such as **CAUSE/EFFECT** (*EVENT traffic → SOUND a deep bass sound → EXPERIENCE feel heavy*) and **CONTRAST** (*sound of traffic vs. sounds of nature; makes you feel heavy vs. serves as a nice contrast*).

³ In all examples from the data, P stands for participant (individually numbered 1–214).
4 Findings and discussion

We set out to provide an overview of what speakers focus on when they describe everyday sounds of unknown sources, to account for the conceptual domains evoked by language in their descriptions and to discuss how these domains are cued linguistically. In this section, we present our findings with regard to the meaning components (*event*, *sound*, *experience*) and contingent deictic perspectives (distal, medial, proximal) that the participants make use of in their sound descriptions (Section 4.1). We then give an overview of the main conceptual domains (*space*, *time*, *touch* and *motion*) as well as the more specific conceptual structures within those domains that are evoked in the sound descriptions, while highlighting the meaning multitasking of wordings in the authentic production data (Section 4.2). Our analyses are accounted for with reference to the cognitive operations exhibited in language in use, in this case primarily comparison, perspective and salience construals. The discussion of the findings also highlights how wordings map onto conceptual structures in a dynamic fashion (Table 1), as each form–meaning pairing is both enabled and constrained by the context in which it is used, as described in Section 3.5.

4.1 Perspectives on soundscapes and their associated meanings

When we investigated the prevalence in the data of the three deictic perspectives on sound (distal, medial, proximal) and their associated meanings (*event*, *sound*, *experience*), we found that the vast majority of the responses (between 93 and 99% for the 20 sounds) describe an *event* (Figure 2). For *sound* meanings there is more variation across the sounds; between 14 and 64% of the responses to the 20 sounds include direct descriptions of *sound*. Experience meanings are the least frequent in the data; between 7 and 26% of the responses include direct description of an auditory *experience*.

We also found that if a response focuses on only one element of a soundscape, it is generally an *event*. Figure 3 shows the percentages of the responses to each sound that include only one type of meaning (*event*, *sound*, or *experience*), taking a single perspective on sound (distal, medial, or proximal), as well as the percentages of responses that are mixed and feature more than one type of meaning/perspective. Together, Figures 2 and 3 show the dominance of *event* descriptions in the data as compared to *sound* and *experience* descriptions. Figure 3 shows the relative prominence of four types of sound descriptions in the data, descriptions from an exclusively distal perspective (only *event* meanings), a fully medial perspective (only *sound*), a fully
proximal perspective (only EXPERIENCE), and mixed perspectives (two or three of the meaning types). As shown in Figure 3, responses that only describe EVENTS are frequent, as are mixed descriptions, while responses that focus only on SOUND or EXPERIENCE are quite infrequent. In other words, out of the three types of meanings coded for (EVENT, SOUND, EXPERIENCE, Table 4), only EVENT meanings are used with any frequency without support from the other two types of meanings. Direct descriptions of SOUND are less independent and are primarily used in combination with EVENT descriptions. Least independent are experiential descriptions, which very rarely occur as stand-alone descriptions. In other words, our results show that the vast majority of the responses include some description of an EVENT (Figure 2) and evoke sound meaning via activities, settings, and/or objects, surfaces and materials associated with the sounds.

The prominence of events in the sound descriptions is a novel finding that constitutes linguistic support for previous research, which has shown that events are central to people’s auditory experiences, as we discussed in Section 2.2. The finding is significant especially in light of the limited research available that considers how people actually communicate about everyday sounds. Our data show that when people are confronted with acousmatic everyday sounds and asked to describe them in detail to an imagined naïve person, the sounds are overwhelmingly presented via their causal events. However, while events are the most common of the soundscape elements (EVENT, SOUND, EXPERIENCE) in the responses, many of them also include descriptions of sound properties (quite loud snaps, a high-pitched scratching sound,

Figure 2: Percentages of responses that include descriptions of EVENT, SOUND, and/or EXPERIENCE for the 20 sounds, ordered by highest to lowest percent for SOUND descriptions.
occasional pops) and/or descriptions of perception and evaluation (I hear, hate this sound a lot, oddly satisfying), as Figure 2 shows. Moreover, while many responses account only for event, the second most common type of response features a combination of soundscape elements (Figure 3).

The three images in Figure 4a–c illustrate three different ways of affording selective prominence to the elements of a soundscape. The lighter ovals in the figures represent three different possibilities for such salience construal. Examples (5a), (5b), and (5c) from the data correspond to the three images in Figure 4.

(5)  

a. Someone digging gravel with a shovel. (Digging P143)

b. Someone digging with a metal shovel. There is crunching from the earth and sometimes resonance from the spade (Digging P59)

c. A person using a shovel to dig up earth or possibly snow after a snowfall. I can hear the shovel hitting the ground and lifting. (Digging P198)
Example (5a) illustrated in Figure 4a describes Digging exclusively as an event. Example (5b) as in Figure 4b describes an event, but also describes the sound in relation to this event (crunching from and resonance from). Example (5c) as in Figure 4c portrays Digging as an auditory experience (I can hear) of an event (the shovel hitting the ground …). Examples (5a)–(5c) all describe the sound Digging but they differ in terms of construed salience. In other words, they differ in the degree of prominence that they afford to the elements constitutive of the soundscape: an event that causes the sound, the sound itself, and a listener who experiences the sound, as well as the processes that connect these elements (generating, perceiving, simulating).

4.2 Conceptual domains and the meaning multitasking of words

In this section, we account for the conceptual domains evoked to describe sound and discuss how they are cued linguistically. The section focuses on space, time, touch, and motion, all prominent domains drawn upon to convey sound meanings. All conceptual structures discussed in this section instantiate both contentful and configurational pre-meanings (presented at the most general level in Table 1) that crystallize in language use through construal operations. In this section, we also follow up on the discussion of semantic theory and methodology in Sections 1 and 2 and use our production data to highlight the flexibility of meaning making. We emphasize that the idea of a one-to-one coupling between words and meanings is problematic. Our study shows that language forms in use multitask as they express meanings that flexibly draw upon multiple domains in response to contextual demands.

4.2.1 Space and time

Sound is dynamic. It moves through space and is extended and distributed over time. The language of sound reflects these spatiotemporal properties. For example, when people encounter environmental sounds, they break down the auditory input to selectively focus on aspects of the sounds vis-à-vis their goals, needs, or other circumstances, and may “segregate a sound of interest (the auditory figure) from other co-occurring sounds (the unattended auditory ground)” (Gregg and Samuel 2012: 998). Segregation into auditory figure and ground is reflected in the use of language in different ways and to varying degrees in our data. The more complex sounds that feature multiple and/or overlapping sounds (e.g., Forest and Grocery) are most frequently configured by means of figure/ground. Through salience construal, the figure/ground configuration prominently draws upon spatial relations, such as foreground/background, over/under, center/periphery to segregate and selectively emphasize
sound components, so that sounds are described as being in the background, above, over the top, underlying, beneath, interspersed, layered or as surrounding or sticking out from other sounds, as well as more dynamically involving change over time, as drowning out or breaking through other sounds.

Segregation of co-occurring sounds by means of spatial relations can be achieved through language in terms of event or sound. Physical space and auditory space can coincide, as in (6), so that a sound that is construed as spatially distant (far away) is also construed as being in the background (auditorily less prominent).

(6) **Distant murmuring sound in the background of many people talking and occasional giggle, lots of clamouring noises and constant beeping from a checkout till in a supermarket** (Grocery P191)

In (6) the *murmuring sound of many people talking* is conceptualized as both spatially distant and auditorily backgrounded. However, physical space and auditory space do not always correspond. Even when sound sources are at a roughly equal distance, as for *Dishes* described in (7), co-occurring and equidistant sounds can be segregated in terms of figure/ground.

(7) *The main sound is the hiss of running water. You can hear the clanging and banging of cutlery and pans in the background…* (Dishes P167)

Continuous sounds, such as the sound of running water, are often relegated to the auditory background (*hissing noise constant in the background*), while intermittent or random sounds are foregrounded (often implicitly), but (7) exemplifies the opposite pattern.

Salience construal by means of a figure/ground configuration of physical/auditory space exemplifies how multiple, co-occurring environmental sounds are segregated and afforded selective prominence through language. However, individual sounds can also be described by means of spatial construal. In descriptions of loudness, for instance, sounds can be construed as spatial objects with properties for size (small, large) or weight (heavy, light), and sounds can also be described as having spatial locations (low, high) or as moving through space (jumping, bouncing). In (8) a sound is described as growing as it gets louder.

(8) **A growing swish, like a crescendo…** (Apple P181)

A potential conceptual mapping between loudness and size is realized differently in examples (9) and (10), as is a mapping between duration and size. Both examples illustrate that words and meanings do not pair up in a one-to-one fashion and that the meanings of words have to be considered within their context of use. In (9) *small* evokes ‘low volume’ but also ‘short duration.’ In example (10) ‘low volume’ as a potential meaning of *little* is rejected (*but quite loud*) in favor of ‘short duration.’
Spatial construal, as in examples (8)–(10), is not the only means of expressing loudness. In our data, loudness is also described in terms of, for instance, **volume** (*quiet, loud*), **strength** (*strong, faint*), and **audition** (*inaudible, vague*). Additionally, words such as *hum*, *thump*, *clash*, and *tinkle* routinely convey loudness in combination with meanings for sound quality, pitch and duration. In use, these words can also express meanings configured according to, for instance, **force** (a *thump* involves a different amount and direction of force than a *clash*) or **deixis** (a *tinkle* indicates proximity of a listener to a sound source, Zwaan 2004: 44). **Force** and **deixis** are experientially linked to loudness because forceful interactions tend to be louder and proximity of a listener to a sound facilitates hearing. Language for loudness is just one example of the necessity for considering language in use in investigations of sensory meaning. Only communicatively situated language allows us to observe lexical multitasking that involves joint evocation of multiple domains.

Based on a cross-linguistic investigation of the coding of perception, Majid et al. write that their results suggest “that the most ‘natural’ mapping for sound contrasts [for pitch and loudness] may, in fact, reside in size rather than spatial location” (Majid et al. 2018: 11374). They draw this conclusion because, across a range of languages, the auditory stimuli used in their study primarily elicited “a big–small metaphor followed by pairs of nonantonymic contrasts: for example, *loud–soft, sharp–soft, strong–soft, strong–small*” (Majid et al. 2018: 11374). Their study was lexical both in aim (to study the codability of perception) and methods (using naming tasks). However, as we see for the English used in our data, one language can afford multiple ways of expressing loudness, many of which would not be elicited in response to *What sound is this?* (Majid et al. 2018: 11375). While it can be useful in a cross-linguistic comparison of vocabularies to establish a “most ‘natural’ mapping for sound contrasts” (Majid et al. 2018: 11374), it risks painting a somewhat skewed picture of people’s communication about sound because it potentially misses significant resources that languages offer.

In addition to being described as spatially configured and situated, sound is also temporally grounded through language. Previous research on audition has underscored the significance of temporal patterns to experiences of sound (Houix et al. 2012; VanDerveer 1979), and our data demonstrate that auditory temporal patterns
are also reflected in the language used to describe sound. Temporal properties of sound are evoked both through descriptions of causal events, as in (11), and in terms of sound attributes, as in (12).

(11) a continuous flow of water through small holes (Dishes P138)

(12) a continuous high-pitched hissing noise (Dishes P26)

Frequent descriptions of the temporal progression of sound in terms of continuity in our data is a linguistic finding in line with studies of auditory experiences, which have noted the significance of continuity to audition (Gygi et al. 2007; Houix et al. 2012).

In addition to continuity, the sound descriptions account for auditory temporal patterns of various kinds such as sequence (first, then, after), duration (brief), regularity (regular, random, chaotic), rhythm (rhythmic beeping), and tempo (moderate tempo). However, as example (13) illustrates, auditory temporal patterns are not communicated by single words such as intermittent or brief alone, but by full descriptions that establish and draw upon multiple domains. For instance, *The steady hum of air flowing through* … evokes temporal meaning (continuity, regularity, duration) in addition to expressing sound quality and force, direction and motion.

(13) The sound of a wood fire in a fireplace – contained, not racing on its own. The steady hum of air flowing through to the flue or a chimney or up from an open wood fire and the crack, crack, crack of dry wood as the fire licks up through that wood. (Fire P2)

The use of crack, crack, crack likewise conveys both sound quality and temporal features of the sound (discontinuity, rhythm, regularity), meanings that are realized jointly by the reduplicated crack and the source description (of dry wood …).

Temporal properties of individual sounds are frequently described in the data, as in a continuous high-pitched noise, but sounds are also segregated temporally. Sounds stemming from human activities such as eating, drinking or digging are often described as sequences of actions or sounds.

(14) Sounds like someone eating crisps, crunching them up. First the crunch and then jaws chewing. (Chip P213)

In addition to wordings such as first and then, as in (14), sequences of sounds are represented through the order in which the sounds, events, and experiences are described. However, for some sounds, there is no clear progression of the component sounds and there may be overlap between them. Such sounds can be afforded a temporal/spatial Gestalt in a description through which the sounds are construed as interacting. In (15) a constant low rumble is described as being punctuated by small slaps of water.
a constant low rumble of fluttering air punctuated by small slaps of water on a hard surface. Like how rain sounds when you are outside sheltering a picnic umbrella. (Fire P38)

In summary, sound is spatiotemporally grounded through language in use. In our data, individual sounds are described as having spatial properties (a small beeping sound, a growing swish), and sounds are also described as spatially situated and segregated (a low hum in the background), and temporally extended (a constant low rumble) and distributed (random beeps; crack, crack, crack). Such descriptions of spatiotemporal sound properties are often aligned with construals of sound as touchable and in motion, as the following two sections address.

4.2.2 TOUCH

There is a close affinity between audition and touch. In 1977 Schafer wrote that “Hearing and touch meet where the lower frequencies of audible sound pass over to tactile vibrations (at about 30 Hz). Hearing is a way of touching at a distance” (1977: 11). The close integration of these senses has since been explored in different studies that have found both “neural and cognitive ties between audition and the tactile modality” (Winter et al. 2017: 444). Our data indicate that people recruit tactile affordances and experiences of objects and surfaces in their descriptions of sound (a smooth/crunchy sound). When they are used to describe non-tactile perception, property words such as smooth in a smooth sound are sometimes regarded as instances of metaphor. It is argued that they effect a cross-domain mapping between TOUCH and SOUND. In this section, we will discuss some problems associated with considering these words in use as metaphorical (Paradis and Eeg-Olofsson 2013; Winter 2019b).

For there to be sound, something has to stir the air or another medium and give rise to pressure waves. In many instances, touch is a prerequisite for sound, especially if we consider touch in a broad sense as including not only tactile perception but also physical interaction between objects and surfaces, as in the tennis ball touched the net. As noted, wordings that evoke TOUCH are frequently used in the sound descriptions in our data, both in direct descriptions of SOUND, as in a scratchy sound (16), but also in descriptions of surfaces and materials (a smooth surface) that are part of sound-causing events.

It is a scratchy sound, like nails over a coarse fabric or a stuck zip (Textile P58)

In (16) the sound of ripping fabric is described as scratchy, and like introduces a comparative construal that evokes tactile, kinesthetic, and auditory features of two similar EVENTS (nails over a coarse fabric or a stuck zip). The actions of letting your
nails run over a coarse fabric or a stuck zip are associated with multisensory experiences, and in (16) these experiences are drawn upon to facilitate the description of the sound of ripping fabric (Hartman and Paradis 2018). This use of a scratchy sound is not metaphorical, but rather evokes sound by drawing metonymically upon the multisensory experiences introduced by like. Descriptions of objects/surfaces and sounds are closely interrelated in our data; a sharp knife can give rise to a sharp sound and a smooth surface can produce smooth sounds, as in (17) and (18).

(17) I can hear a sharp knife slicing into leaves, perhaps cabbage leaves or salad leaves. It sounds crisp and clear. I can also hear the sharp sound of a chopping board when the knife reaches the bottom of the leaves (Onion P163)

(18) Cars driving by on a road, a little distant. Making smooth “shh” sounds, as the road surface is very smooth. … (Traffic P45)

In addition to their perceptual qualities, objects and surfaces such as a sharp knife and a smooth road afford different types of actions that in turn evoke patterns for force, speed, and/or motion. Much like perceptual qualities, such action affordances can be drawn upon in descriptions of sound. For instance, a smooth surface affords unimpeded and temporally extended motion, which in turn can give rise to a uniform, continuous sound. A sharp knife affords effortless, swift motion, and the quality of the edge influences the force pattern and speed of the motion as well as any resulting sounds. In other words, the wordings a scratchy sound, a sharp sound, and a smooth sound in (16)–(18) rely not only on touch to describe sound, but on a constellation of domains that are selectively drawn upon to create meaning. If we remove these wordings (e.g., a sharp sound) from their communicative contexts, it may be tempting to describe them as metaphorical, but there are issues with their proposed metaphorical status that become clear when the words are considered in use. A metaphorical meaning presupposes an asymmetrical mapping across domains. For the meaning of a sharp sound to be metaphorical, there has to be a significant asymmetrical mapping between touch and sound, so that a sound is described in terms of a tactile experience. If, as we suggest, a sharp sound recruits multiple domains in its description of sound, the meaning construction is better explained as a metonymical mapping between sound and select parts of a matrix of experiential domains (Paradis and Eeg Olofsson 2013). Along the same lines, Winter argues “that many and perhaps all sensory adjectives have highly multisensory or supramodal meanings that encompass a much broader referential scope than is commonly admitted” (2019b: 107). This is certainly something we see evidence of in our data, and we would argue that wordings such as a sharp sound recruit meanings not only across multiple sensory domains, but even more widely across experiential domains. Example (19) describes the sound of fabric being ripped as quite broken not a smooth sound. This use of it’s
quite broken not a smooth sound does not primarily evoke TOUCH. Rather, this wording emphasizes the sound's temporal and force patterns so that they align with the manner of the causal action (pulling the tape from the roll bit by bit).

(19) Sounds like tape, begins ambiguous could be a zip or something, it's quite broken not a smooth sound as if they are pulling the tape from the roll bit by bit as it's very sticky (Textile P164)

While the use of not smooth arguably does evoke TOUCH in conjunction with very sticky, the sound description emphasizes force and temporal progression, so that the sound and its causal action are jointly construed as temporally distributed and uneven in force patterns.

4.2.3 MOTION

Sound is in motion and stems from activity, or as Howes and Classen put it: “There are no still sounds” (Howes and Classen 2014: 8). In our data, individual sounds are described as being in motion (jumpy, wriggly, swishy, fidgety), and MOTION also plays a part in the description of sound segregation. In (20) sounds are described as interacting through forceful motion.

(20) white noise of voices, metallic plinks and stabs individual sounds, crunches, knocks, wooden and crackling. A medley, a chaos of voice-mash with random voice sounds surfacing from the litter of the rest. Crashing, bumping, sense of space filled, crowding of sounds all crushed together. (Grocery P181)

The sounds are described as crashing, bumping, and being crushed together, foregrounding spatiotemporally situated MOTION along with FORCE and TOUCH. Random voice sounds are described as surfacing, presumably as they gain momentary prominence. Caballero and Paradis write that “sound events in language may be evoked through motion (e.g., descriptions of sound floating, lingering, or rising)” and that this “points to the dynamic nature of our perception and conceptualization of sound as propagated through space” (2020: 6, see also Huumo 2010; Strik Lievers and Winter 2018; Viberg 2015). The description in (21) likewise features meanings that draw upon MOTION and TOUCH.

(21) The sound varies a little and is quite soft but it sometimes bites and this bitey sound is very short and brief but the soft part of the sound is a continuous smooth flow. so this sound has two components, a soft component and a sharp bitey component. (Fire P177)

To allow sound segregation, the sound of a fire is described in (21) in terms of two components. These configurational components are construed comparatively so that
the sounds are contrasted according to time (short and brief vs. continuous), touch (soft vs. sharp), and motion (flow vs. bitey).

Examples (20) and (21) describe the sounds themselves as being in motion (jumpy, a flow), but sounds are also described via event-based motion. In (22) the action’s manner of motion is presented as a direct correlate to the sound.

(22) Teeth brushing, the brush moving around the mouth and the noise corresponding to the movement and vigour of the action (Teeth P58)

The noise is described as corresponding to the movement and vigour of the action, so that manner of motion is foregrounded and presented as observable via the sound.

In our data, wordings that describe the sound of movement often multitask by evoking sound and motion in addition to other domains. Walking in Heels is described by expressions such as clacking across and clomping up or down that conflate meanings for sound and motion. Prepositions are often used alongside such onomatopoeic wordings to indicate directional movement.

(23) shoes clip clopping across a hard floor and going up stairs, shoes with heels echoing across the floor (Heels P138)

Example (23) describes the sound of directional (horizontal and vertical) movement in an indoor space. To describe this sound, the language used draws upon a range of domains in addition to sound, such as action, motion, direction, rhythm, object, and surface. The description in (23) accounts for properties of the sound directly (clip clopping, echoing) while establishing its causal event through the same wordings (clip clopping across a hard floor, echoing across the floor). Clip clopping mimics qualities of the sound it represents, including its rhythm (Winter et al. 2017). Onomatopoeic wordings such as clip clop, bang, and tip tap are often used in our data as contextualized verbs (tip tapping on a window), which means that they establish sound through iconicity, but go beyond this “simplest kind of semiotic mapping, imagic iconicity” (Dingemanse 2012: 663) by construing the sound as event. Even onomatopoeic wordings that describe sound emission directly, such as goes dub dub dub dub in (24), rather than via a sound-causing action, as in clip clopping across a hard floor in (23), can gain meaning by association to an event.

(24) cars driving on a busy road maybe a highway. it sounds like you are on a bridge and the highway is below you. mostly normal car whooshing, but one loud motorcycle and one car that sounds like it has a flat tire that goes dub dub dub dub (Traffic P38)

As it mimics the sound, dub dub dub dub conveys sound meaning alongside a description of a car with a flat tire driving on a busy road. A lexical focus that disregards context (or establishes no context as the context) bypasses the discursive pooling of semantic
resources that takes place in actual language use. This potentially leads to an overvaluation of the contributions of individual words to meaning-making in communication, and at the same time potentially underestimates these same flexible and multitasking contributions.

5 Conclusion

Under the rather bold title *The language of sound*, we set out to explore: (i) what language users focus on when they describe everyday sounds of unknown sources, (ii) what conceptual domains are evoked in such sound descriptions and through what wordings, and (iii) some implications of using production data to get a more ecological picture of sound descriptions. In much of the literature on sound language, there is a tacit assumption that lexicalized meanings and designated domain specific words such *clank* and *swoosh* constitute the most natural way of communicating sensory experiences. In this article, we have raised methodological and theoretical concerns about the prevalence of such assumptions in sensory language research and suggested an alternative approach to such research, one that requires an ecological shift in the investigation and explanation of sensory language.

Firstly, we have shown that out of the three soundscape elements (*event*, *sound*, *experience*) and corresponding perspectives (distal, medial, proximal) that we considered in our analysis, language users most frequently adopt a distal perspective and focus on events in their descriptions of everyday sounds of unknown sources. Direct descriptions of sound attributes and the listeners’ auditory experiences are primarily expressed within such event descriptions. In other words, language users primarily focus on the events that may have caused the sounds.

Secondly, we have shown that there is no designated language of sound. When the aim is to describe an everyday sound to a person who cannot hear it, the sound can straightforwardly be described via features of its sound-causing event, including its setting (*a busy day at a shopping mall*) or an associated action (*chopping onions*). Words that come to mind as ‘sound words,’ such as *clank* and *swoosh* are primarily used within descriptions of sound-causing events, where such wordings can describe and causally link a sound to an action or object (*washing up, the clank of cutlery and rushing water pots and pans*) or evoke the sound through an action by conflating meanings for *sound* and *action* (*water tip tapping on a window*). Based on much of the available research on sensory language, it may be unexpected to find that sounds are not primarily described using designated sound words.

Our findings should not come as a surprise. After all, perception is not unimodal but multimodal. We do not experience the world around us using one sense at a time,
and linguistic meaning reflects the multimodality of human experience. The language we use to describe sensory experiences evokes meanings tied to multiple domains at the same time. Consequently, words and wordings multitask in their expression of meaning. A word for an object such as apple or a setting such as forest has the potential to evoke meaning for multiple senses, emotions, and actions. This is not an ‘extra’ but part of these words’ fundamental meaning potential. Using source descriptions in communication about sensory experiences is not less precise or less direct than using designated sensory lexical items such as clank and swoosh. In many ways, people’s experiences of objects and places may be more embodied (as people have physical experiences of these) than their experiences of lexicalized percepts. Also, conversely, because perception is multimodal, lexical items such as tip tap, boom, and splash convey meanings not only for sound, but also for actions, properties, or locations, meanings that can be configured according to, for instance, deixis, force, motion, direction. Language is flexible with respect to the route and direction through which these potential meanings are evoked. The methods used in studies of linguistic meaning and the models used to represent such meaning have to accommodate and reflect this inherent flexibility of language.

Thirdly, we set out to discuss some potential issues associated with the use of lexically oriented methods to study sensory language. We have questioned the extent to which it is possible to draw conclusions about how people communicate about sound based on data from tasks that elicit labels for sensory experiences or ratings of decontextualized words. While tasks of this type can yield valuable information about vocabularies, they tend to focus on wordings outside of a communicative context, making it difficult to assess the ecological validity of the findings vis-à-vis communication. Our data clearly indicate that communication about sound does not involve drawing upon collections of dedicated lexical items such as swish, clop, bang or smooth, loud, high-pitched, but rather accounts for what is heard and experienced, associations made, memories stirred, and reasoning about possible sources and settings, while also making the sounds accessible to someone who cannot hear them.

Our study highlights the need for a theoretical grounding of meaning in language that can underpin and guide methodological choices. In order to capture the language of sound, we need a theoretical and analytical framework that accommodates language use in a non-trivial way beyond stable couplings between individual words and meanings. We have proposed a framework for the modelling of combinations of form and meaning and their construal in situated communication that can handle meaning multitasking of words and that provides the requisite tools for description of meaning applications. Structured and flexible analyses that consider the complexities of language in use can facilitate comparisons between types of communication and across sensory modalities.
Data availability statement

The datasets generated and analysed during the current study are available in the OSF repository, osf.io/xv52g.

Acknowledgments: This research was funded by the Kamprad Family Foundation (grant number 20190045). We would also like to thank three anonymous reviewers for insightful and constructive feedback, Sara Farshchi and Victoria Johansson for input on the survey, and Eleni Tzimopoulou for work on the database.

References


Majid, Asifa & Niclas Burenhult. 2014. Odors are expressible in language, as long as you speak the right language. *Cognition* 130(2). 266–270.


