Article:

Ogden, Richard orcid.org/0000-0002-5315-720X and Keevallik, Leelo (2025) Issues of phonetics and social action in human-animal interaction. Language and Communication. pp. 113-120. ISSN: 0271-5309

https://doi.org/10.1016/j.langcom.2025.06.007

Issues of phonetics and social action in human-animal interaction.

Richard Ogden, Centre for Advanced Studies in Language & Communication,
Department of Language & Linguistic Science, University of York, UK
richard.ogden@york.ac.uk

Leelo Keevallik, Department of Culture and Society, Division of Language, Culture and Interaction, Linköping University, Sweden
leelo.keevallik@liu.se

Abstract

In encounters between humans and animals, both parties make use of sound, some of which are vocal. Since the anatomy of vocal tracts is different in different species, the production of sounds varies, while humans find ways to partially match the acoustics of animal sounds. Analytic challenges lie in the representation of the various sounds, as we need to move beyond the IPA, and in establishing when and how animals become participants in interspecies interaction.

Issues of phonetics and social action in human-animal interaction.

Introduction

This SI offers welcome studies of the mundane practices of people living close to animals, while some of these practices may even be disappearing as societies change. It has taken the multimodal turn in studies of social interaction (Nevile, 2015) for communication researchers to embrace the embodied nature of those everyday exchanges that have always been intuitive for people who live or work closely with animals, but have evaded scientific study. Research has furthermore tended to focus on vocal forms and their meaning, and on single agents, rather than networks of interacting participants. A focus on animal-directed vocal behaviour is an interesting twist to begin to question truths-taken-for-granted, such as that identical language competence is required from recipients of vocal messages. Furthermore, as Mondémé (2018, 2022) has repeatedly pointed out, interspecies interaction easily falls into the gap between the various analyses of human language and the species-specific analyses of communication systems that are either compared to that of human language or studied as vocal and embodied sign systems in their own right. The SI fills this gap, approaching the topic with different methodologies that range from sequential interaction analysis to ethnography and experimentation.

The SI demonstrates that vocalisations shape interactions between hearing humans and animals. Even though animals and humans inhabit very different social and sensorial worlds from each other (cf. Moore et al., 2016), these worlds meet in many different environments: from domestic settings for pet dogs and cats, who are likely to be treated more as family members, through to animals with whom we interact for pleasure, such as riding a horse, through to farmed animals. Overall, each of the many types of interaction brings with it different degrees and kinds of sentiment towards the animals; and vocalisations play many different roles.

Familiarity with the specific individuals, or at least the species, characterizes the analyzed episodes, and provides interesting parallels to humans speaking different languages or interacting in divergent activity contexts. Expert knowledge of the concerned species is also part of the methodological toolbox necessary to carry out the analyses. While the SI mainly targets human and animal sounds and their characteristics, the underlying assumption is that the sounds are potentially meaningful and functional in achieving intersubjectivity across the species, and the analyses try to capture how that plays out in different cases. In order to capture both the specificity of sounds and their function, we begin by commenting on the vocal features appearing in the collection and then devote the latter part of our commentary to action and participation.

What kinds of vocalisations do people direct to animals?

The papers in this SI show that human vocalisations directed at animals often exhibit features that deviate from the language spoken by the participants. For instance, several studies highlight the use of sounds absent from the local language's phonological system. Jääskeläinen's work documents numerous Finnish utterances containing a bilabial trill, a sound not found in Finnish. Szczepek Reed's analysis of instructions given to horses reveals the extensive use of clicks, employed in ways that differ from their function in spoken English or German. Beyond specific sounds, animal-directed vocalisations also differ in delivery. They might involve increased loudness, enabling the sounds to travel further than in normal human conversation, frequent repetition, and specific directional projection; or long-domain features such as palatalisation and/or lip protrusion or high pitch, which are audible in Harjunpää and Szczepek Reed’s data on talk directed at cats. These cues readily signal to an adult human bystander that the utterances are intended for animals, not for them.

What do these unique characteristics reveal? As Szczepek-Reed (this issue, p.42) notes, animal-directed vocalisations can employ sounds absent from a language's phonemic inventory (Bynon, 1976), ensuring they "[don't] sound like the rest of our words" (Corby, 2016, 0.47 s), which marks these utterances as special, aimed at animals, and intentionally designed to be distinct from communication within a human language community. The very practice of choosing features that are not part of the ambient language tells us that these utterances are intended for a different audience.

However, animal-directed utterances form systems of utterances which are used among and across human communities, and the sound-meaning association is conventionalised. This is rather reminiscent of the claim of Lexical Phonology, that different lexical strata in languages are differently structured (e.g. Mohanan, 1986; Itô and Mester, 1995 for Japanese): under this approach, animal-directed lexical items could form a separate system. Both the smaller number of contrasts and the distinctiveness of these sounds are human-centred in origin, though, and may not necessarily align with the actual auditory perception and responsiveness of animals like horses.

The papers also show that the animal-directed subsystems of human language contain adaptations to the type and age of animal. Jääskeläinen shows that in Finnish farming practices, the calls differ according to species; and the calls used to young animals are often different from those used to older ones, as Szczepek Reed & Harjunpää discuss in the case of adult cats vs. kittens. Szczepek-Reed shows that clicks and trills, along with a form of the syllable ‘ho’, are common in certain Western European traditions of horse-riding, even though the same sound may have a different use in different cultures. These practices seem to suggest that certain types of sound are felt to be appropriate, and perhaps even appealing, to specific kinds of animals. To the extent this holds true, we can consider these animal-directed vocalisations to be recipient-designed. It would be interesting to see how such practices work in other parts of the world. Where humans’ vocalisations are produced as responses to animals’, and match some feature(s) of the animal’s vocalisation, this reflects an understanding on the human’s part that the animal’s vocalisation were communicative, even if they do not treat their meaning as always apparent.

How to deal with animal sounds?

Different species have different physical resources for making sounds. Human speech uses the potential of the vocal tract: the lungs, oral and nasal cavities as well as articulators such as the larynx, tongue and lips: speech requires the mastery and coordination of very precise articulations. Animal vocal tracts are differently configured than humans', in particular with respect to the shape and position of the larynx, and the ways in which it can be used as an articulator.

Linguistic phonetic description is built on human anatomy and the affordances of the human vocal tract as they are deployed in spoken languages. These facts present us with three main problems when it comes to interaction between species:

1. How to represent in writing the sounds that animals make
2. How to depict, mimic, or reproduce in the human vocal tract sounds that animals make
3. How to transcribe the humans' depiction of the animal's vocalisation.

Representing in writing the sounds that animals make

Decolonising linguistics in the context of interspecies communication requires us to treat animals on their own terms. Following this principle, if we were to develop a scheme for annotating animal sounds, we would need to develop species-specific 'phonetic alphabets'. This would mean having an understanding of a species' anatomy and how this is deployed to make sounds; then working out the composition of animals' sounds, on the assumption that at least some of them, like the sounds of human speech, are composites of respiration, phonation and articulation. For example, a cat-centred account of feline sounds would need to account for a whole range of vocalisations, including meow, purr, trill, hiss, snarl and chirp (Schötz et al., 2017) with reference to feline anatomy. Similarly for other animals; though if one accounted for these sounds using terms endogenous to the animals' noise-making capacity, the resulting 'phonetic alphabet' might look different for each species; and this is before we think of features classed as 'non-segmental' on the IPA chart (which are not especially detailed or precise for human speech), like duration, pitch, loudness, and so on.

While it is tempting to say, as Cornips (this issue) does, that a cow's moo has a resonance like /u/, this is not strictly accurate for a cow, since a cow's moo is not accompanied by distinctive lip-rounding as a human [u] is. On the other hand, a voiced sound produced with closure at the lips and with nasal airflow is also what a human [m] is; but there are differences in the two anatomies which make the dynamics of the production of the sounds different, and [m] in a cow's system of vocalisations does not and presumably cannot stand in contrast with other sounds, such as [n], which require a great deal of articulatory control.

For the purposes of reproducing sounds that animals make, however, humans necessarily use their own vocal tract. If we read Cornips' claim that cows' moos are close to the sound [u], but are made with closed lips, then we can easily see that [m], perhaps with a back resonance (i.e. raising of the back of the tongue), [mˠ], is the best reproduction of the cow’s moo that we can produce with the human vocal tract. Why a back resonance? One reason is that it has a lower second formant in humans, and lower formant values are associated with bigger resonating cavities.

Another resource for representation is orthography. Norrthon and Nilsson depict a horse’s snort as ‘phrrrh’; here we have to understand the <rrr> as capturing the trilling of the horse’s lips (where <r> would normally refer to trilling of the tongue tip or uvula in humans), and <h> as marking breath accompaniment – the equivalent of the IPA [ʙ̥], or the symbol [ψ] in the traditional Finnish system that Jääskeläinen uses. Peltola, Wu and Grandgeorge represent a dog’s growl as ‘wrouh: wrou::h’ using French orthography, where the implication of <r> is a uvular sound and <w> implies labiodental friction. These are orthographic expressions which humans can understand as representing animal sounds, without any claim to ‘accuracy’. They certainly cannot be read as words of human language, since they do not conform to orthographic norms of words in the languages of the ambient human cultures in the papers.

On the other hand, Norrthon and Nilsson use IPA vowel symbols to represent the sounds of a cat, including diacritics that indicate where on the cardinal vowel chart the sound fits, and nasalisation. This has to be read as instructions to a human competent at reading the IPA on how to produce cat sounds, since cats cannot produce vowels in any meaningful linguistic sense: the vowel symbol is used as a proxy for a feline sound made with an open vocal tract and periodic vocal fold vibration, and representing a quality that reminds a human listener of a particular human articulatory configuration (cf. Ekström et al. 2024). Thus the vowel symbol does not make an explicit claim about the position of the cat’s tongue, or the height of its jaw opening, but rather gives humans a sense of what sound the cat made.

So, while there are some similarities between human and animal vocal tract configurations, and the noises that they can produce, there are significant differences too; and it is not trivial to devise a scheme that is endogenously valid for each animal. Our transcriptions are necessarily human-centred, because any kind of transcription is a human artifact, and comes with human cultural background and human conventions for interpretation.

Humans using their vocal tract to depict the sounds that animals make

Harjunpää and Szczepek Reed illustrate humans mimicking purring, using a trilled [r] sound (‘kurr’ Their Example 3 also contains a case where a cat produces a trill, and the human produces a stretch of creaky voice in the response (line 5). These two human vocalisations orient to the low-frequency pulsing or trilling sounds that cats can make. To purr, a cat does not produce an alveolar trill (such as we see in human reproductions such as [purr] or [kurr]), but produces a low frequency vibration at the glottis: a feline equivalent of creak. However, the rate of trilling of the human tongue against the alveolar ridge is a better match in frequency than creaky voice. Furthermore, it can be difficult for most humans to willingly sustain creaky voice for long, whereas a cat can purr while both exhaling (as humans do) or inhaling (as humans generally don’t do; and if they do, the sound of the inhale is very different from that of the exhale).

Peltola et al. illustrate a child imitating a dog’s growl, which is repesented as ‘wrouh: #wrou::h’, and is described as having two parts. The child’s imitation of the dog’s growl is represented by ‘r(h)o::: r(h)a:::’, and is described as a ‘two-part combination of a growl and exhalation’ – meaning that the transcript picks up the trilling and origin of the growl with ‘r’, which in French is uvular and the turbulent airflow of the exhalation with ‘(h)’, but it seems that the temporal organisation of the two parts of the child’s production is not identical to the dog’s. Nonetheless, the child reproduces, in his own vocal tract, the essential elements of the dog’s production.

In these examples, humans pick up on features of animals’ vocalisations and produce something that bears acoustic similarity, but which is not – and cannot be – a strict copy of the animal’s vocalisations.

Transcribing the sounds humans make to depict animals

The third issue should now be apparent: using the IPA to depict animals’ vocalisations does not represent the animals’ own sounds, but constitutes a set of instructions to a human about how the animal’s vocalisations could be imitated with human anatomy. In cases where humans replicate some aspect of an animal’s vocalisation, this constitutes an analysis, in human terms, of what matters about the sound that is produced. After all, we have our own phonetic systems, and phonetic systems are both rooted in the body, and in the linguistic systems of the speaker’s language(s). We are used to pitch being deployed for tone and intonation and e.g. affective displays; we are used to attending to certain kinds of acoustic features (such as pitch movements; low-frequency trilling; harsh vs. soft onsets; loud vs. quiet sounds; voiced vs. voiceless sounds), and we probably impose these basic categories onto animal vocalisations. Considering these vocalisations from the perspective of animal articulation and perception, this may or may not be correct; and it is also likely that we are missing features that are important to animals. This would not be surprising, given that different species have different auditory systems, vocalise to their conspecifics, and inhabit very different social and sensorial worlds. So for example Cornips’ careful multimodal study of greetings, looking at cows’ vocalisations alongside ear and tail posture, head posture (height), smelling, touching seems to be heading in the right direction, because it leaves open the possibility that our human perspective on meaningful interactional contributions is not appropriate for cows.

Acoustics as a mediator

In summary, transcribing the sounds animals make in a non human-centric way would require us to develop a phonetic alphabet for each species whose vocalisations we wish to write down; and to do so in a way that respects their anatomy, and the set of vocalisations they make. Phonetic history is full of attempts to build a transparent, physiologically motivated writing system for human speech:

Bell’s Visible Speech is one example, Pitman’s shorthand is another. These systems were invented in the Victorian era, but they never took off. The symbols they used were unfamiliar, and symbols often resembled each other, and so were visually confusing in a way that e.g. the distinction between the voiceless-voiced pairs of plosives [p b], [t d] and [k g] are not. So developing numerous writing systems for animals based on their physiology and with consistent correspondences between symbol and articulation is probably only practicable or useful for a highly specialised audience.

The common world that humans and animals inhabit is sound, even if our auditory and perceptual systems are not the same; and human spoken languages trade on relations between articulation and acoustics, even though our phonetic categorisation of sounds relies heavily on articulation. Humans and animals may not perceive sound identically; and we may not know how animals attend to elements of sound. But clearly sound is at the heart of our calls to animals, and at least an important element of their calls to us, and to each other. So one solution to this conundrum is to afford a more central place to acoustics, as several of the papers in the SI do. Cornips shows pitch traces of cows on a semitone scale, which we know for humans captures the salience of pitch movements more clearly, especially at lower frequencies, than for example a linear scale. Harjunpää and Szczepek Reed provide examples of f0 traces of cats, and of humans matching the pitch of cats, albeit in a lower octave – itself a human (and culturally rooted) analysis of what counts as ‘same’ to us. So humans treat the pitch as important, and worth repeating; and although humans cannot match the cat’s absolute pitch, we can produce something that at least to human ears is a match, at one octave lower. Indeed, recent work in speech technology uses acoustics as the foundation for transcribing animal sounds (Hagiwara et al., 2024).

So the auditory/acoustic/perceptual domain is where our sensory worlds overlap, much more so than articulation; and it probably provides a better basis for describing and presenting our vocalisations to one another, and understanding the basis on which we interact across species.

Do animals perform social actions and do they use vocalizations to do that?

Spoken human languages are used to accomplish actions such as asking and telling, complaining and agreeing. Any human vocalisation can be interpreted and acted upon by other humans, since vocalisations display many things which are hard to verbalise (Keevallik and Ogden, 2020). As shown in this SI, humans act on animal sounds as well. However, while the ascription of action is difficult enough (Levinson, 2013), it may be even more difficult to claim that a particular behavior by an animal counts as an action in human terms, such as a "request for help" or a "greeting" (see the discussion in Mondémé, 2023). A human greeting, for example, refers to discrete embodied and audible action that is used to publicly ratify another person's social copresence (Pillet-Shore, 2012), including participants “moving together and bodily addressing one another” (Goffman, 1981, p.130). This kind of bodily reorientation between an animal and a human is described in several papers, such as a cow running towards the human (Cornips) and a cat orienting to a human who enters the stable (Norrthon and Nilsson). Animals are thus shown to make themselves bodily available for an encounter, whether a sound occurs or not. Importantly, animal within-species greetings among primates (baboons) can be organized through entirely embodied actions with no sounds involved (Mondada and Meguerditchian, 2022). All of this points to the relevance of embodied behavior for sequential engagement for humans and animals alike.

The audible greeting between humans, at the same time, can be varied. A ‘hi’ can be produced softly, without prosodic prominence to convey a neutral stance, or loudly, with high pitch and a lengthened form to display a special status of the relationship (Pillet-Shore, 2012). Humans have been argued to regularly use the latter format with animals (Nilsson and Norrthon, 2024). Animal greetings, on the other hand, can involve a variety of embodied features, including ear movement and sniffing. One of the main methodological questions is, which segment or aspect of this behavior can be seen as the counterpart of the human audible greeting. How can we distinguish it from an introduction (through sniffing?) or a recognition (by licking?). Can we take it for granted that animals fit into human rituals, or should these encounters also be described from the animals’ own perspective, with careful regard to their perceptual capacities? A complementary approach would build on the strength of the ethnomethodological conversation analysis by taking a members’ perspective, but this necessitates expert knowledge of the species.

In any case, it is evident that several vocalisations and embodied behaviors by animals are treated by knowledgeable humans as meaningful contributions to ongoing activities, including meeting each other and engaging in play. In human interaction, speakers and hearers build actions collaboratively, and with the help of culturally meaningful objects (Goodwin, 2007). In this SI, we can witness how a cat is treated as asking for food through head turn (gaze) and spatial positioning next to the empty bowl (Norrthon and Nilsson), and how a horse repeatedly and rhythmically displays compliance with instruction (Szczepek Reed). As an example of social action sequences within non-human species, orangutans have been argued to accomplish embodied requests and offers in a similar way to human infants (Rossano and Liebal, 2014). Mondémé (2023) discusses first and second actions by cats, dogs, and horses (see also Cornips et al. 2023). These results suggest that future research on action sequences in interspecies interaction may be fruitful, among other things in order to discover whether sustained turn-taking is upheld, or whether the intersubjective episodes only emerge sporadically in co-presence.

During the past decades, it has become increasingly clear that human social actions are multimodal rather than conducted unimodally through language (Goodwin, 1995; Streeck, 1996; Mondada, 2006). In this SI, we can witness that cattle calls are multimodal in their embodied performance, including holding up hands close to mouth and involving objects, such as twigs and bread (Jääskeläinen). We can see that besides calling out a name, a dog’s attention is attracted with embodied signals such as whistles and claps (Mitchell et al.), and while engaging with horses, cows, cats and dogs, humans touch them (Cornips, Harjunpää and Szczepek Reed, Norrthon and Nilsson, Peltola et al.). In fact, social actions can entirely be carried out by the human body (Mondada, 2016; Keevallik, 2018), which opens up possibilities for similar analyses on animal behavior. Even though primarily targeting vocalisations, this SI also documents ear movements by cows and horses, body shapes by dogs and cats, head turns by cows and cats – all of which can potentially be meaningful contributions to social action, at least to humans who are experienced with the particular species. It seems to be virtually impossible to single out sound, if we are to understand the meaning conveyed by animals and humans alike.

Furthermore, social actions are the dialogical achievements of participants who display their understanding incrementally, locally, and in real time. Human interaction evolves sequentially, as people take turns at producing actions, recognising and displaying their recognition of another's actions, and producing a next action (Schegloff, 2007). As this SI witnesses, humans treat animals as both initiating and responding to actions. For example, they treat horses', dogs' and cats' vocalisations as meaningful first actions, and ascribe emotional states to them on the basis of these vocalisations (Norrthon and Nilsson). Humans can affiliate through matching their next contributions to those of animals, such as growling back to dogs, or repeating cats' pitch movements (Peltola et al, Harjunpää and Szczepek Reed; Mondémé (2018) calls these morphisms). When a cow produces *mmmm* while gazing towards a recipient, and positioning their ears to the side, some humans treat it as a first greeting (Cornips), children respond to dogs' barks, even contesting them in next turns, thereby treating them as meaningful first actions (Peltola et al.). At the same time, animals are trained to respond to human initiating actions, such as cattle calls that make the cows' approach a relevant response (Jääskeläinen). Horses are socialised to respond to specific human vocalisations by adjusting their movement, sometimes reinforced by the rider through other embodied means (Szczepek Reed). Intriguingly, however, there is quantitative evidence that dogs tend to not comply with attention-getting devices by humans (Mitchell et al.). No doubt, humans keep trying to achieve responsivity, as it constitutes the foundation of mutual engagement. Humans also respond to animal initiations and treat animals as understanding their actions, regardless of whether, e.g., a cow recognizes their own name. They often talk to animals in complex phrases and full sentences (Cornips, Norrthon and Nilsson, Jääskeläinen). This suggests it is rather the voice itself and its various qualities that are relevant for animals, perhaps conveying a non-threatening approach, such as a lot of the cat-directed speech is high pitch and palatalised (Harjunpää and Szczepek Reed), thereby giving the impression of a small body producing it (Ohala 1994; Gussenhoven 2002). All the while, talk may rather be produced for overhearing humans.

In summary, animals are often treated as producing both initiating and responsive actions, which results in at least short sequences of mutual engagement. Animal as well as human actions, however, are both vocal and embodied, and vocalisations inevitably provide only part of the meaning. A further issue is how to analytically establish that animals are part of focused interactions at all.

When are animals participants?

The studies of this SI treat animals as participants. Yet, participation is a complex matter even for humans: it is organized in relation to the material setting, through the placement of the individual bodies, as well as through how individuals choose to contribute to the evolving interaction. Already Goffman talked about the difference between co-presence and focused interactions, where co-presence refers to a circumstance where individuals can merely perceive what each other is doing, and sense to be perceived (Goffman, 1966, p.17). Focused interactions, on the other hand, involve participation in an encounter that comprises joint attention to some interactional event (Goffman, 1966, p.83). These two concepts immediately provide a challenge regarding human-animal interaction. When are animals and humans merely co-present? When and how do animals become participants in an event, such as an equestrian training session (Sczcepek Reed) or playing a game (Mitchell et al., Peltola et al.)?

Among other things, focused interactions feature embodied orientations by the individuals, including particular body positions and gaze directions. In human-human interactions, a particular spatial organisation is endemic to focused interactions. F-formation (or facing formation) refers to an organisation where participants have turned their lower bodies towards a common center (Kendon, 1977, p.179). In the middle of this formation of bodies there is an area of visual focus, involvement, and engagement. How does such a formation get established between a human and an animal? Literature on human initiations highlights the importance of the interactants' mutual approach, which leads to an F-formation and also enables physical contact (Goffman, 1966; Duranti, 1992; DeStefani and Mondada, 2018). In the current SI, Cornips discusses approach as a centerpiece of an encounter between a human and a cow, at the same time also pointing out the obstacles in a barn that restrain the animals from approach and, thus, prohibit a focused interaction. Jääskeläinen provides third-person accounts of cattle calls that aim to reduce the distance between the caller and the called (but not to accomplish a F-formation), while Peltola et al. similarly point out that differences in proximity lead the children to adjust their vocal delivery to dogs. In other studies, we can see a sustained organisation of a focused interaction, such as between the coach, the rider, and the horse (Szczepek Reed), or notice at least a momentary mutual gaze orientation when the cat turns their head towards the person filming (Harjunpää and Szczepek Reed). While we have good methodological tools for analysing how humans treat "others" as participants, including infants and robots, it needs a different competence to document whether and how each non-human species treats, e.g., the camera person as a participant. It is also analytically important to distinguish between instances where humans merely ascribe a participatory role and agency to an animal (such as when interpreting their distress), vs. when animals act as participants.

In the current collection of studies it is mostly humans who orient themselves with regard to animals, including to get them to play (Mitchell et al., Peltola et al.) or follow them with a camera. Animal initiation is captured in a few cases, through approach and vocalisation (Cornips; Harjunpää and Szczepek Reed). Responsivity is likewise a central way of establishing participation, such as responding vocally, with embodied action, or even allowing someone to touch oneself. Responsivity can include either individuals or collectivities – a greeting can single out one cow (Cornips) while cattle calls treat cows as a group (Jääskeläinen). In some papers in this SI the responsiveness is presumed, or itself foundational for human action, such as in the case of cattle calls (Jääskeläinen) or when instructing a horse (Szczepek Reed). In other cases, responsivity could be detailed more in future multimodal analysis, such as when establishing whether a horse is a passive recipient of a kiss or an active participant in it (Norrthon and Nilsson), especially since we know that kissing between humans is a mutual accomplishment (Katila and Cekaite, 2023) and it also has a vocal component (Katila et al., 2023).

Overall, the studies in this SI provide convincing proof that animals (pets) and humans can indeed engage in focused interactions, forming "clusters of individuals who extend one another a special communication license and sustain a special type of mutual activity" (Goffman, 1966, p.83). However, these engagements, just as with humans, are in a constant flow of transition, and allow for a variety of qualitatively different participation frameworks. Individuals may emerge as addressed or unaddressed hearers, but also bystanders, overhearers, or eavesdroppers (Goffman, 1981). When are animals addressed hearers and when are they simply bystanders? When are humans bystanders or overhearers in animal-animal interaction (Harjunpää and Szczepek Reed, Norrthon and Nilsson)? As has been shown previously, pets can be specifically granted the role of a principal who is responsible for the words, when humans use their co-presence as a tool of conveying messages to other humans (Tannen, 2004). In the current SI, Szczepek Reed shows how horses are addressed alongside the rider, particularly through specific sounds that instruct both of them. Thus, participation takes a variety of forms and it can also evolve dynamically and reflexively due to various local concerns. Among other things, there are humans who are clearly more qualified than others to treat animals as true participants rather than mere bystanders, resulting from their extensive expertise with animals in general, or the particular individuals concerned. In parallel with transformations in human interaction that rely on interactional histories (Deppermann, 2018), within-species developments have been documented between bonobo infants and mothers (Rossano, 2023), suggesting that similar adjustments in human-animal long-term involvements are highly likely. On the other hand, expertise may also include knowledge about the limits of animals' capacity to act and participate. In this SI we see abundant evidence of how animals are treated as addressed hearers.

Future directions

One of the main challenges emerging in the SI is the need to theorise how to represent animal sounds so that we do not rely on the system that is configured on the human vocal tract. For that reason, acoustics seems to be a firmer ground for departure. These challenges are shared by researchers trying to transcribe machine sound (Pelikan, 2024). A productive way forward could be to study extended engagements, i.e., activities beyond paired actions (greetings, requests for food, directives and compliances). We could, for example, examine how animals treat human vocalisations that imitate their own sounds, as is done briefly in Harjunpää (2022: 81-82). Do the parties continue to build some element of the other's vocalisations into their own? In that case, which parameters of human vocalisations do animals pick up on?

A related challenge is how to analytically establish animal participation, i.e., when animals are not merely co-present with humans but accomplish sequential actions. How do we determine initiation and responsiveness in different species and how are they made sense of across the species? In linguistically oriented studies we mostly focus on linguistically significant sounds. Also in this SI, phonetic analysis has been hugely insightful in revealing systematicity in the functioning of different sounds. Nonetheless, there are other aspects of conduct that offer opportunities for further research. For example, Jääskeläinen's interesting historical account, which is based on written records of cattle calls, can be complemented with the analysis of embodiment and mobility. Indeed, we may gloss the function of these sounds as generically "calling" animals but judging from the documentary she refers to, *Karjan kutsumahuudot* 'Cattle calls' (https://yle.fi/a/20-108884), these complex calling routines are used in a variety of situations. That includes calling from a distance while the caller is actively moving (presumably towards the animals), calling while moving alongside the animals, occasionally while someone else is chasing the animals from behind (with different sounds), calling single individuals rather than the entire herd, etc. These different embodied organisations show a huge potential for research on the mutual adjustments and reflexive use of the vocal practices, to truly understand not only the fine detail of human-animal interaction and the role of sound in it, but also the status of participation among the individuals at every moment.

There is no doubt that human-animal interactions are a mixture of successes, blatant ignoring (Mitchell et al.), and misunderstandings (Peltola et al.), some of which humans learn to deal with by adapting their own vocal behavior (Sczcepek Reed). Notably, this SI features aligning interactions, targeting moments of mutual engagement rather than disalignments, reprimands, refusals to play, or even leave-taking. Broadly speaking, humans can certainly behave in a way that is threatening for the animals, sometimes inadvertently (which perhaps the camera person was to the hiding kitten; Harjunpää and Szczepek Reed) but sometimes out of the need to guide animals to safety or, alternatively, to safeguard humans. A collection of papers on those events would be a nice complement to the current SI. Perhaps one setting to look for those would be in encounters with wild animals.

The overall pathos of the SI is to highlight the interactive capacities of non-human animals. The papers treat single animals as named interactants rather than merely using the name of the species (Cornips, Norrthon and Nilsson, Mitchell et al., Peltola et al.), thereby subscribing to the overall quest of ecolinguistics not to lump together non-human animals under abstract category labels. We also see respect to animal contributions, be it in the form of responsiveness in play engagement (Peltola et al.) or when humans repeat the pitch contours produced by animals, respectfully treating these as intentional (Harjunpää and Szczepek Reed).Ecolinguistics has hitherto mostly focused on the ways humans use language, either to critically address environmental discourse though the toolbox of systemic-functional linguistics, or to promote ways of linguistic expression that treat animals as sentient beings with their own desires and inner lives (Stibbe, 2015; see the epistemological and historical overview in Zhou, 2022). This SI offers a new domain of research along the same vein, providing analytical tools to engage with human-animal interaction from a non-hierarchical perspective. By showing a new path for analyzing interspecies interaction, the SI promotes inclusive linguistics, as suggested by Cornips (2019), and launches a new empirical branch of ecolinguistics.

Acknowledgements

This paper is partly supported by Riksbankens Jubileumsfonds grant P21-0447 Sounding for Others: Distributed Agency in Action.

References

Bynon, J., 1976. Domestic animal calling in a Berber tribe. In: McCormack, W.C., Wurm, S.A. (Eds.), Language and Man: Anthropological Issues. Mouton, pp. 39–66.

Corby, D., 2016. How to Teach Your Horse to Stop//Voice Commands//Horse Training Tips with Deanna Corby Dressage. Uploaded 12/05/2016. https://www.youtube.com/watch?v¼df-C3QzzG3Q. (Accessed 29 November 2024).

Cornips, L., van Koppen, M., Leufkens, S., Melum Eide, K., & van Zijverden, R. (2023). A linguistic-pragmatic analysis of cat-induced deixis in cat-human interactions. *Journal of Pragmatics, 217,* 52–68. https://doi.org/10.1016/j.pragma.2023.09.002

De Stefani, E., Mondada, L., 2018. Encounters in Public Space: How Acquainted Versus Unacquainted Persons Establish Social and Spatial Arrangements. Research on Language and Social Interaction. 51 (3), 248–270. https://doi.org/10.1080/08351813.2018.1485230

Deppermann, A., 2018. Changes in turn-design over interactional histories: the case of instructions in driving school lessons. In: Deppermann, A., Streeck (Eds.), Time in Embodied Interaction: Synchronicity and Sequentiality of Multimodal Resources John Benjamins, Amsterdam, pp. 293–324.

Duranti, A., 1992. Language and Bodies in Social Space: Samoan Ceremonial Greetings. American Anthropologist. 94 (3), 657–691. https://doi.org/https://doi.org/10.1525/aa.1992.94.3.02a00070 Ekström, A. G., Vila, L. C., Schötz, S., & Edlund, J. (2024, July). A single formant explicates the ubiquity of “meow”. In *Proceedings of the 4th International Workshop on Vocal Interactivity in-and-between Humans, Animals and Robots (VIHAR 2024)* (pp. 74-78).

Cornips, L., 2019. The final frontier: Non-human animals on the linguistic reserach agenda. Linguistics in the Netherlands. 36 (1), 13–19. https://doi.org/https://doi.org/10.1075/avt.00015.cor

Goffman, E., 1966. Behavior in public places. Free Press, New York, NY.

Goffman, E., 1981. Forms of talk. University of Pennsylvania Press, Pennsylvania.

Goodwin, C., 1995. Co-Constructing Meaning in Conversations With an Aphasic Man. Research on Language and Social Interaction. 28 (3), 233–260. https://doi.org/10.1207/s15327973rlsi2803_4

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

Gussenhoven, C. (2002, April). Intonation and interpretation: phonetics and phonology. In Proceedings of speech prosody (Vol. 2, pp. 47-57). Université de Provence Aix-en-Provence.

Hagiwara, M., Miron, M., Liu, J.-Y., 2024. ISPA: Inter-Species Phonetic Alphabet for transcribing animal sounds. In arXiv [cs.SD]. arXiv. http://arxiv.org/abs/2402.03269

Harjunpää, K.(2022). Repetition and Prosodic Matching in Responding to Pets’ Vocalizations. *Langage et société,* 176(2), 69-102. https://doi.org/10.3917/ls.176.0071.

Itô, H., Mester, R. A. (1995). Japanese Phonology. In J. Goldsmith (Ed.), The Handbook of Phonological Theory (pp. 817–838). Blackwell Publishing.

Katila, J., Cekaite, A., 2023. Trajectories of love: Affective reciprocity during kissing in the everyday life of romantic couples. Journal of Pragmatics. 214, 38–53. https://doi.org/https://doi.org/10.1016/j.pragma.2023.06.010

Katila, J., Hofstetter, E., Keevallik, L., 2023. Cries of Pleasure and Pain: Vocalisations Communicating How Touch Feels in Romantic Relationships. Research on Language and Social Interaction. 56 (4), 330–349. https://doi.org/10.1080/08351813.2023.2272529

Keevallik, L., 2018. What Does Embodied Interaction Tell Us About Grammar? Research on Language and Social Interaction. 51 (1), 1–21. https://doi.org/10.1080/08351813.2018.1413887

Keevallik, L., Ogden, R., 2020. Sounds on the Margins of Language at the Heart of Interaction. Research on Language and Social Interaction. 53 (1), 1–18. https://doi.org/10.1080/08351813.2020.1712961

Kendon, A., 1977. Studies in the Behavior of Social Interaction. Indiana University.

Levinson, S.C., 2013. Action Formation and Ascription. In: Sidnell, J., Stivers, T. (Eds.), The Handbook of Conversation Analysis. Wiley-Blackwell, Oxford, UK, pp. 101–130. https://doi.org/https://doi.org/10.1002/9781118325001.ch6

Mohanan, K.P., 1986. The theory of lexical phonology. D. Reidel Publishing Company.

Mondada, L., 2006. Participants’ online analysis and multimodal practices: projecting the end of the turn and the closing of the sequence. Discourse Studies. 8 (1), 117–129. https://doi.org/10.1177/1461445606059561

Mondada, L., 2016. Multimodal resources and the organisation of social interaction. In: Andrea, R., Louis de, S. (Eds.), Verbal Communication. De Gruyter Mouton, Berlin, pp. 329–350. https://doi.org/doi:10.1515/9783110255478-018

Mondada, L., Meguerditchian, A., 2022. Sequence organisation and embodied mutual orientations: openings of social interactions between baboons [Article]. Philosophical Transactions of the Royal Society B: Biological Sciences. 377 (1859), Article 20210101. https://doi.org/10.1098/rstb.2021.0101

Mondémé, C., 2018. How do we talk to animals? Modes and pragmatic effects of communication with pets. Langage et société. No 163 (1), 77–99. https://shs.cairn.info/journal-langage-et-societe-2018-1-page-77?lang=en

Mondémé, C., 2022. Why study turn-taking sequences in interspecies interactions? Journal for the Theory of Social Behaviour. 52 (1), 67–85. https://doi.org/https://doi.org/10.1111/jtsb.12295

Mondémé, C., 2023. Sequence organisation in human–animal interaction. An exploration of two canonical sequences. Journal of Pragmatics. 214, 73–88. https://doi.org/https://doi.org/10.1016/j.pragma.2023.06.006

Moore, R. K., Marxer, R., Thill, S., 2016. Vocal interactivity in-and-between humans, animals, and robots. Frontiers in Robotics and AI. 3, 191823. https://doi.org/10.3389/frobt.2016.00061

Nevile, M., 2015. The Embodied Turn in Research on Language and Social Interaction. Research on Language and Social Interaction. 48 (2), 121–151. https://doi.org/10.1080/08351813.2015.1025499

Nilsson, J., Norrthon, S., 2024. Opening interspecies encounters – Greetings between humans and nonhuman animals. Journal of Pragmatics. 229, 40–55. https://doi.org/https://doi.org/10.1016/j.pragma.2024.05.007

Ohala, J. J. (1994). The frequency codes underlies the sound symbolic use of voice pitch. In L. Hinton, J. Nichols, & J. J. Ohala (Eds.), Sound Symbolism (pp. 325–347). Cambridge University Press.

Pelikan, H., 2024. Transcribing human–robot interaction: methodological implications of participating machines. In: Haddington, P., Eilittä, T., Kamunen, A., Kohonen-Aho, L., Oittinen, T., Rautiainen, I., Vatanen, A. (Eds.), Ethnomethodological conversation analysis in motion: emerging methods and new technologies. Routledge, London, pp. 42–62.

Pillet-Shore, D., 2012. Greeting: Displaying Stance Through Prosodic Recipient Design. Research on Language and Social Interaction. 45 (4), 375–398. https://doi.org/10.1080/08351813.2012.724994

Rossano, F., 2023. How to study interactional history in non-human animals? Challenges and opportunities. In: Haddington, P., Eilittä, T., Kamunen, A., Kohonen-Aho, L., Oittinen, T., Rautiainen, I., Vatanen, A. (Eds.), Ethnomethodological conversation analysis in motion: emerging methods and new technologies. Routledge, London, pp. 21–41.

Rossano, F., Liebal, K., 2014. “Requests” and “offers” in orangutans and human infants. In: Drew, P., Couper-Kuhlen, E. (Eds.), Requesting in Social Interaction. John Benjamins, Amsterdam, pp. 335–364. https://doi.org/https://doi.org/10.1075/slsi.26.13ros

Schegloff, E.A., 2007. Sequence organisation in interaction: A primer in conversation analysis. Cambridge University Press. https://doi.org/10.1017/CBO9780511791208

Schötz, S., van de Weijer, J., & Eklund, R. (2017). Phonetic characteristics of domestic cat vocalisations. In The 1st International Workshop on Vocal Interactivity in-and-between Humans, Animals and Robots, VIHAR 2017, 25–26 August 2017, University of Skövde, Sweden (pp. 5-6).

Stibbe, A., 2015. Ecolinguistics: Language, Ecology and the Stories We Live By. Routledge, London. https://doi.org/10.4324/9781315718071

Streeck, J., 1996. How to do things with things. Human Studies. 19 (4), 365–384.

Tannen, D., 2004. Talking the Dog: Framing Pets as Interactional Resources in Family Discourse. Research on Language and Social Interaction. 37 (4), 399–420. https://doi.org/10.1207/s15327973rlsi3704_1

Zhou, W., 2021. Ecolinguistics: A half-century overview. Journal of World Languages. 7 (3), 461-486. https://doi.org/doi:10.1515/jwl-2021-0022