NON-ARBITRARINESS ACROSS MODALITIES: BOOTSTRAPPING THE EMERGENCE AND DEVELOPMENT OF QUESTIONS

CHRISTINE CUSKLEY[1], JOHANNES M. HEIM[*1,2]

*Corresponding Author: johannes.heim@abdn.ac.uk
[1]Language Evolution, Acquisition, and Development Group, Newcastle University, Newcastle upon Tyne, UK
[2]Language and Linguistics, University of Aberdeen, Aberdeen, UK

Over the last decade, language evolution research has been the driving force behind a renewed interest in non-arbitrary mappings (Cuskley & Kirby, 2013), as well as multimodal theories of language development (Vigliocco et al., 2014). These theories have focused largely on lexical iconicity in language systems. Here, we begin to extend this line of research by considering how non-arbitrariness plays a role in language development and emergence beyond the lexicon, moving from lexical to pragmatic meaning across different modalities. In other words, we focus on non-arbitrariness in situated language (Murgiano et al., 2021; Özyürek, 2021). Specifically, we consider the role of indices of dominance in scaffolding the development of questions. By surveying the state of the art in research on prosody, gesture, and signing, we outline evidence for cross-linguistically attested, iconic gestural scaffolding of questions across signed and spoken languages. This supports development of indexical use of pitch in question prosody, which has analogies in non-arbitrary dominance signalling. We conclude by synthesizing this evidence, presenting a novel non-arbitrary bootstrapping theory for the emergence of questions.

1. Introduction

Making our desires known through linguistic directives, such as questions, lies at the very heart of human communication. Questions play a key role in driving conversations by engaging in a collaborative effort in building common ground (Roberts, 2012) and have a prominent place in accounts of first language acquisition (Thornton, 2016). While morphosyntactic cues to questions are arbitrarily aligned and exhibit considerable cross-linguistic variation, different modalities – gesture and intonation - are aligned across many languages. For example, while both English and Mandarin use rising intonation to encode questions, English combines rising intonation with a sentence-initial auxiliary 'do' and Mandarin with a question particle. Moreover, questions in both languages will be accompanied by raised eyebrows, hands, and shoulders.

Intonation and by extension, gesture (Bolinger, 1983) can mirror features associated with dominance asymmetries (Ohala, 1983). We argue that questions inherently involve a temporary dominance asymmetry within a conversational dyad, where the speaker uses upward movement to encode that they need something from the addressee (materially or informationally).

We propose that a rich inventory of developmentally robust and cross-culturally attested multimodal, that is: non-morphosyntactic cues scaffolds the emergence of complex linguistic questions, both in development and language evolution. This paper will draw on previous literature for synthesising three distinct lines of evidence for this novel theory: i) the cross-linguistic ubiquity of non-morphosyntactic cues to questionhood ii) the importance of such cues in early pragmatic development, iii) the parallels of grammaticalized facial cues in spoken and signed languages. On this basis, we show that non-arbitrary cues are likely to have played a key role in the emergence of questions given them a role as a catalyst in language development, the key features occurring consistently across languages and dominant signalling modalities.

2. Non-arbitrariness across modalities and languages

Questions of whatever morphosyntactic form engage the addressee to respond to the speaker's epistemic stance (Heim, 2019). Questions are widely considered the marked form of human conversation as opposed to assertions, which occur more frequently (Huddleston, 1994). Correspondingly, an auxiliary-initial question word order, which make up the majority of questions in the caregiver input of an English-speaking child (Estigarribia 2009), occurs less often than an auxiliary-medial word order (Newport et al., 1977). Morphosyntactic marking of questions is present in 80% of the world's languages (Dryer, 2008), and comes in the shape of word order changes, interrogative clitics and particles, and/or verbal inflection (Siemund, 2001). Content questions have an additional cue form of a variable corresponding to the element unknown (e.g., *wh*-pronouns in English) and sometimes omit other cues present in yes-no questions. While morphosyntactic marking varies significantly across languages, other modalities, such as intonation and gesture, show considerable cross-linguistic consistency.

By far the most common cue to mark questions is rising intonation (Ultan, 1978), although prosodic details vary alongside how pitch is employed at the lexical level. Pitch plays a key role in other kinds of non-arbitrariness, particularly linked to size sound symbolism via the frequency code hypothesis (Ohala, 1994). The mechanism of motivatedness behind the frequency code is that a larger resonant chamber is an index for a larger producer, but motivatedness can have multiple mechanisms (Ahlner & Zlatev, 2010). For example, while high vowels are generally considered "smaller" than low vowels, Diffloth (1994) describes Bahnar, a language where high vowels are large and low vowels are small, mediated by the size of the tongue within the mouth (which occupies a larger volume when producing high vowels like /i/ relative to low vowels like /a/).

Despite some cross-linguistic variation in how question intonation is deployed, it is nonetheless a reliable indexical cue, motivated by dominance asymmetries in pragmatic context across languages. For example, Gussenhoven & Chen (2000) show that peak alignment, peak height, and end pitch are associated with questions by speakers of Dutch, Chinese, and Hungarian; although these languages vary in their specific encoding of questions. In general, animals and humans exhibit a strong association between pitch and dominance (Puts et al., 2007), with lower pitch signalling larger size and greater dominance.

The manual-gestural modality shows similar non-arbitrariness in marking questions, arguably tied more straightforwardly to dominance: hands and eyebrows are raised (Krahmer & Swerts, 2005). This contrasts markedly with negation, where facial articulators show the opposite direction: eyebrows are lowered, and lips are pressed against each other (Benitez-Quiroz et al., 2016). Hence, just like with intonation, a downward movement expresses dominance, an upward movement a lack thereof. One other prevalent gesture associated with the absence of knowledge is the palm-up (epistemic) gesture, which is also associated with submissiveness (Cooperrider et al., 2018) and echoes the upward pitch contours often associated with questions in speech (Ultan, 1978). In contrast to co-speech manual gestures, House (2002) argues that facial cues are less reliable than manual ones for encoding questions, at least in spoken languages. Sign languages, however, are more reliable in this context: They generally mark questions with raised eyebrows (Cecchetto, 2012) including ASL, Sign Language of the Netherlands, and Kazakh-Russian Sign Language (Kimmelman et al., 2020). This facial gesture is also a widespread marker of submissiveness in primates (Keating, 1985), lending further support to the idea that dominance indices played a catalyst role in language emergence.

Areas that require further research are cues such as breathiness, duration effects**,** and speech rate (Dehé et al., 2022)**,** which have all been found to encode questions in several languages. Moreover, in Yélî Dnye (Levinson, 2010), questions are marked by mutual gaze making rather than intonation. This nonetheless arises from systematic dominance asymmetries within a conversation (with the speaker asserting dominance, rather than telegraphing submission).

3. Non-arbitrariness in pragmatic development

Non-arbitrary cues to questionhood are not only attested widely across languages and modalities, but also in language development. Here, too, non-arbitrary cues tend to be non-morphosyntactic. Intonation and gesture provide an avenue into pragmatic development, and questions in child directed speech which use non-morphosyntactic cues are ubiquitous (Kania, 2016). Morphosyntactic encoding of questions, on the other hand, has a long developmental trajectory, with the first questions marked by word order occurring from around 20 months in Italian, German and Swedish, but only from 30 months in English. It takes years for children to master the morphosyntactic variation in questions, especially when

they include negation (Guasti, 2017). With non-morphosyntactic cues appearing early, the acquisition of morphosyntactic cues can benefit from multimodal scaffolding in development, building on motivated manual and facial gestures, which then bootstrap indexical use of pitch.

We assume manual and facial gestures to be the starting point of this developmental process, because they are the first means available for communication (Iverson & Goldin-Meadow 2005). Gestures frequently serve as precursors for what is later expressed using more complex linguistic structure (Capirci et al., 2005). Children then combine vocalizations and gestures for relating non-arbitrary signs to previous experience with a rapid increase of gestures during the first year (Burkhardt-Reed et al., 2021). Pointing, too, serves as a cue to questions when the addressee is knowledgeable (Begus & Southgate, 2012). From 14 months, infants use prelinguistic vocalizations to differentiate conversational strategies (Grünloh & Liszkowski, 2015). Combining these findings with the those of intonational cues, we find that non-arbitrary forms are well-suited for bootstrapping linguistic development (Bohn et al., 2019).

Intonational cues follow early on. Sensitivity to different sentence melodies is evident form as early as 5 months (Frota et al., 2014). Intonational distinctions of questions from other speech acts occur as early as 7 months, and adult-like intonation is produced by English-speaking children in their second year of life (Prieto Vives & Esteve Gibert, 2013). Yet, infants show notable variation as to when this command equals adult performance. Correspondingly, children use different strategies for asking questions up to age 11 (Patel & Grigos, 2006). German children aged 2;8-2;10 (age; month) show striking variability in producing question intonation and appear not to reach prosodic targets before age 3;0 (Lleó & Rakow, 2011). Hübscher and colleagues (2019) show that facial and prosodic cues for expressing knowledge states are employed at the age of three, while lexical items are only employed later. This strongly suggests that non-arbitrary, non-morphosyntactic cues play a central role in advancing pragmatic development and can act as a platform for later integrating morphosyntactic cues.

4. Non-arbitrariness in signed and spoken languages

Parallels in the non-arbitrariness of encoding questions across spoken and signed languages add a third piece of evidence supporting our hypothesis of the catalytic function of these cues. Broadly speaking, non-arbitrariness has been shown to play an important role across both infant-directed speech and signing (Brand et al., 2002) across different modalities. Perniss and colleagues (2015) demonstrate that non-arbitrariness plays a greater role in referring to non-visible referents in child-directed signing compared to adult-directed speech. Laing and colleagues (2017) show that prosody in onomatopoetic words (e.g., *quack*) is more marked in infant-directed speech than in equivalent conventional words (e.g., *duck*). In both studies, however, the focus was very much on the non-arbitrary signalling of lexical meaning, where the benefits for language

development are well-established (Ota et al., 2018; although see Nielsen & Dingemanse, 2021). This suggests that children acquiring language – regardless of the dominant modality (speech or sign) – exploit non-arbitrariness. We argue this extends to cues which aid in acquiring pragmatic concepts, such as questions.

As in section 2, we focus here on the directionality of these cues, which shows consistent trends across spoken and signed languages. Zeshan (2004) reports that most signed languages use a combination of eyebrow raise, eyes wide open, eye contact, a head forward position or forward body posture to ask questions. These non-manual interrogative signs accompany manual/lexical question markers, such as content question words or dedicated morphemes. The non-manual signs are key components of marking questions across many sign languages (Kimmelman et al., 2020). Their frequent presence in questions remind us that these are the marked speech acts compared to assertions. Remarkably, the directionality of these upward and forward oriented non-manual signs in signed languages mirrors that of the intonational patterns observed in many spoken languages (Dixon, 2012). The same applies for the upward movement of co-speech gestures in spoken languages.

Even more remarkable is the fact that directionality changes exhibit the same patterns across modalities. Content and polarity questions in English, for instance, have a fall and rise as their standard intonational contours, respectively (Bartels, 1997). Reversing the directionality for these types is meaningful, too, as their function changes from and open to a closed question (Hedberg et al., 2017), where 'closedness' evokes a set expectation about the answer. We find the analogous pattern for non-manual signs in sign-languages. Italian Sign Language has raised eyebrows in polar questions and lowered eyebrows in content questions (Zeshan, 2004) just as spoken Italian has a falling-rising contour for polar questions and a falling contour for content questions (D'Imperio, 2002). The paradigm is complete with lowered eyebrows reported for closed or loaded questions, as reported for Israeli Sign Language (Meir, 2004). The similarity between spoken and signed questions also has some equivalence in morphosyntax where both signed and spoken languages use the words for indefinite pronouns and question markers (Dixon, 2012). Studies that address the role of non-arbitrariness across infant-directed signing and speaking with respect to interactional strategies are still in their infancy and researchers have just begun the explore the richness of multimodal cues in the learner's input. (Murgiano et al., 2021)

5. Conclusions

By comparing non-morphosyntactic cues across languages, ages, and modalities, we have argued that their non-arbitrariness presents an opportunity for bootstrapping morphosyntactic development of questions. By drawing on a wide range of findings from production and perception data, we establish that non-morphosyntactic cues are likely to be non-arbitrary, transcend cultural and linguistic borders, and precede morphosyntactic competence in infancy. These

properties make non-arbitrary cues to questionhood a strong candidate for playing a central role in pragmatic development, and by extension in language emergence due to their prevalence and early availability. Moreover, this theory opens specific hypothesis spaces which can be probed with more specific experimental and typological work. For example, using iterated artificial language learning, we might expect that readily available indexical cues might outcompete complex morphosyntactic cues for questions in a dyadic face to face context. We outline a possible experimental approach wherein participants are trained on an artificial language with complex morphosyntactic cues for questionhood used irregularly (i.e., in 60% of questions) in the first generation. In a face-to-face condition, we predict that these cues will either stagnate or fade out of the system over iterated "generations", in favor of indexical cues. In a virtual condition without the availability of indexical cues, we predict the cues will regularize and come to be more reliable cues to questionhood. In addition, typologically, we might expect that the 20% of languages without complex morphosyntactic cues (Dryer, 2008) are likely to have richer non-morphosyntactic cues. Assuming non-arbitrary, non-morphosyntactic cues are to some extent present even in the many languages with complex morphosyntactic marking, this points to the former type of cues as an evolutionarily older strategy. In summary, we have established (i) the consistent cross-linguistic availability of non-arbitrary question cues, (ii) the use of these cues in bootstrapping the acquisition of question, and (iii) evidence that similar non-arbitrary cues play an important role in both acquisition and evolution. Together, these lines of evidence point to a novel theory for the emergence of questions, where non-arbitrary cues from intonation, gesture, and signing play a catalyst role in the language emergence and development.

References

Ahlner, F., & Zlatev, J. (2010). Cross-modal iconicity: A cognitive semiotic approach to sound symbolism. *Sign Systems Studies*, *38*(1/4), 298–348.

Bartels, C. (1997). *Towards a compositional interpretation of English statement and question intonation*. University of Massachusetts Amherst.

Begus, K., & Southgate, V. (2012). Infant pointing serves an interrogative function. *Developmental Science*, *15*(5), 611–617.

Benitez-Quiroz, C. F., Wilbur, R. B., & Martinez, A. M. (2016). The not face: A grammaticalization of facial expressions of emotion. *Cognition*, *150*, 77–84.

Bohn, M., Call, J., & Tomasello, M. (2019). Natural reference: A phylo- and ontogenetic perspective on the comprehension of iconic gestures and vocalizations. *Developmental Science*, *22*(2), e12757.

Bolinger, D. (1983). Intonation and gesture. *American Speech*, *58*(2), 156–174.

Brand, R. J., Baldwin, D. A., & Ashburn, L. A. (2002). Evidence for 'motionese': Modifications in mothers' infant-directed action. *Developmental Science*, *5*(1),

Burkhardt-Reed, M. M., Long, H. L., Bowman, D. D., Bene, E. R., & Oller, D. K. (2021). The origin of language and relative roles of voice and gesture in early communication development. *Infant Behavior and Development*, *65*, 101648.

Capirci, O., Contaldo, A., Caselli, M. C., & Volterra, V. (2005). From action to language through gesture: A longitudinal perspective. *Gesture*, *5*(1–2), 155–177.

Cecchetto, C. (2012). Sentence Types. In *Sign Language: An international handbook* (pp. 292–315). De Gruyter Mouton.

Cooperrider, K., Abner, N., & Goldin-Meadow, S. (2018). The Palm-Up Puzzle: Meanings and Origins of a Widespread Form in Gesture and Sign. *Frontiers in Communication*, *3*.

Cuskley, C., & Kirby, S. (2013). Synesthesia, cross-modality, and language evolution. *Oxford Handbook of Synaesthesia*, *20*, 869–907.

Dehé, N., Braun, B., Einfeldt, M., Wochner, D., & Zahner, K. (2022). The prosody of rhetorical questions: A cross-linguistic view. *Linguistische Berichte*, *269*, 3–42.

Diffloth, G. (1994). I: Big, a: Small. In L. Hinton, J. Nichols, & J. Ohala (Eds.), *Sound symbolism* (pp. 107–130). CUP.

D'Imperio, M. (2002). Italian intonation: An overview and questions. *Probus*, *14*(1).

Dixon, R. M. (2012). *Basic linguistic theory, vol. 3: Further grammatical topics*. Oxford: OUP.

Dryer, M. S. (2008). Polar Questions. In *World Atlas of Language Structures*. Max Planck Digital Library. http://wals.info/feature/116

Frota, S., Butler, J., & Vigário, M. (2014). Infants' Perception of Intonation: Is It a Statement or a Question? *Infancy*, *19*(2), 194–213.

Grünloh, T., & Liszkowski, U. (2015). Prelinguistic vocalizations distinguish pointing acts*. *Journal of Child Language*, *42*(6), 1312–1336.

Guasti, M. T. (2017). *Language acquisition: The growth of grammar*. MIT press.

Gussenhoven, C., & Chen, A. (2000). Universal and language-specific effects in the perception of question intonation. *6th International Conference on Spoken Language Processing (ICSLP 2000)*, 91–94.

Hedberg, N., Sosa, J. M., & Görgülü, E. (2017). The meaning of intonation in yes-no questions in American English: A corpus study. *Corpus Linguistics and Linguistic Theory*, *13*(2), 321–368.

Heim, J. M. (2019). *Commitment and engagement: The role of intonation in deriving speech acts* [PhD Thesis]. University of British Columbia.

House, D. (2002). Perception of question intonation and facial gestures. *TMH-QPSR Fonetik*, *44*, 41–44.

Hübscher, I., Garufi, M., & Prieto, P. (2019). The development of polite stance in preschoolers: How prosody, gesture, and body cues pave the way. *Journal of Child Language*, *46*(5), 825–862.

Huddleston, R. (1994). The contrast between interrogatives and questions. *Journal of Linguistics*, *30*(2), 411–439.

Kania, U. (2016). *The Acquisition and Use of Yes-no Questions in English: A Corpus-Study from a usage-based perspective* (Vol. 36). Narr Francke Attempto.

Keating, C. F. (1985). Human Dominance Signals: The Primate in Us. In S. L. Ellyson & J. F. Dovidio (Eds.), *Power, Dominance, and Nonverbal Behavior* (pp. 89–108). Springer.

Kimmelman, V., Imashev, A., Mukushev, M., & Sandygulova, A. (2020). Eyebrow position in grammatical and emotional expressions in Kazakh-Russian Sign Language: A quantitative study. *PLOS ONE*, *15*(6), e0233731.

Krahmer, E., & Swerts, M. (2005). How children and adults produce and perceive uncertainty in audiovisual speech. *Language and Speech*, *48*(Pt 1), 29–53.

Laing, C. E., Vihman, M., & Keren-Portnoy, T. (2017). How salient are onomatopoeia in the early input? A prosodic analysis of infant-directed speech. *Journal of Child Language*, *44*(5), 1117–1139.

Levinson, S. C. (2010). Questions and responses in Yélî Dnye, the Papuan language of Rossel Island. *Journal of Pragmatics*, *42*(10), 2741–2755.

Lleó, C., & Rakow, M. (2011). Intonation targets of yes/no questions by Spanish and German monolingual and bilingual children. In E. Rinke & T. Kupisch (Eds.), *Hamburg Studies on Multilingualism* (Vol. 11, pp. 263–286). John Benjamins.

Meir, I. (2004). Question and Negation in Israeli Sign Language. *Sign Language & Linguistics*, *7*(2), 97–124.

Murgiano, M., Motamedi, Y., & Vigliocco, G. (2021). Situating Language in the Real-World: The Role of Multimodal Iconicity and Indexicality. *Journal of Cognition*, *4*(1), 38.

Newport, E. L., Gleitman, H. & Gleitman, L. R. (1977). Mother I'd rather do it myself: some effects and non-effects of maternal speech style. In Snow, C. E. & Ferguson, C. A... (eds), Talking to children: language input and acquisition. Cambridge: C.U.P.

Nielsen, A. K., & Dingemanse, M. (2021). Iconicity in Word Learning and Beyond: A Critical Review. *Language and Speech*, *64*(1), 52–72.

Ohala, J. (1983). The Origin of Sound Patterns in Vocal Tract Constraints. In P. F. MacNeilage (Ed.), *The Production of Speech* (pp. 189–216). Springer.

Ohala, J. (1994). The frequency code underlies the symbolic use of voice pitch. In L. Hinton, J. Nichols, & J. Ohala (Eds.), *Sound Symbolism* (pp. 325–347). CUP.

Ota, M., Davies-Jenkins, N., & Skarabela, B. (2018). Why Choo-Choo Is Better Than Train: The Role of Register-Specific Words in Early Vocabulary Growth. *Cognitive Science*, *42*(6), 1974–1999.

Özyürek, A. (2021). Considering the Nature of Multimodal Language from a Crosslinguistic Perspective. *Journal of Cognition*, *4*(1), 42.

Patel, R., & Grigos, M. I. (2006). Acoustic characterization of the question–statement contrast in 4, 7 and 11 year-old children. *Speech Communication*, *48*(10), 1308–

Perniss, P., Zwitserlood, I., & Özyürek, A. (2015). Does space structure spatial language? *Language*, *91*(3), 611–641.

Prieto Vives, P., & Esteve Gibert, N. (2013). Prosodic structure shapes the temporal realization of intonation and manual gesture movements. *Journal of Speech, Language, and Hearing Research. 2013 Jun; 56: 850-64.*

Puts, D. A., Hodges, C. R., Cárdenas, R. A., & Gaulin, S. J. (2007). Men's voices as dominance signals: Vocal fundamental and formant frequencies influence dominance attributions among men. *Evolution & Human Behavior*, *28*, 340–344.

Roberts, C. (2012). Information structure: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics*, *5*, 6–1.

Siemund, P. (2001). Interrogative constructions. *Language Typology and Language Universals*, *2*, 1010–1028.

Thornton, R. (2016). *Acquisition of Questions* (J. L. Lidz, W. Snyder, & J. Pater, Eds.; Vol. 1). Oxford University Press.

Ultan, R. (1978). Some general characteristics of interrogative systems. In J. Greenberg (Ed.), *Universals of Human Language* (pp. 83–124). Stanford UP.

Vigliocco, G., Perniss, P., & Vinson, D. (2014). Language as a multimodal phenomenon. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *369*(1651), 20130292.

Zeshan, U. (2004). Interrogative Constructions in Signed Languages: Crosslinguistic Perspectives. *Language*, *80*(1), 7–39.