Title

Communicative factors in the emergence of phonological dispersion

Permalink

https://escholarship.org/uc/item/7332b5w1

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

Authors

Roberts, Gareth
Kuang, Jianjing
Clark, Robin

Publication Date

2024

Communicative factors in the emergence of phonological dispersion

Gareth Roberts (gareth.roberts@ling.upenn.edu)[*]

Jianjing Kuang(kuangj@sas.upenn.edu)[*]

Robin Clark (rclark@sas.upenn.edu)[*]

[*]Department of Linguistics, University of Pennsylvania
3401-C Walnut Street, Philadelphia, PA 19104 USA

Abstract

We investigated the emergence of dispersion in phonological systems using an established experimental paradigm in which pairs of participants play a non-linguistic communication game, taking turns to select discrete colors from a continuous underlying space and send them to each other to communicate animal silhouettes. Over time participants established sets of signals made up of combinatorial color units, analogous to the phonemes of natural language. This allowed us to investigate the role of interactive pressures on the emergence of organizational structure in phonological inventories, principally dispersion. We manipulated minimum signal length (as a means of investigating the role of coarticulation) and the presence of probabilistic noise. We also manipulated the nature of the underlying color space. There was an effect of colorspace but not of noise or minimum signal-length. However, dispersion occurred at above-chance levels in all conditions. Our results provide evidence for the role of communicative interaction in the emergence and cultural evolution of phonological structure.

Keywords: cultural evolution; phonology; combinatoriality; emergence of structure; language; communication; experiment

Introduction

Phonological systems consist of sets of combinatorial units drawn from a continuous space. This is true of both vocal languages (Ladefoged & Disner, 2012) and sign languages (Brentari, 2011). In vocal languages, vowels are for instance typically considered to be drawn from a trapezoidal space in which one dimension corresponds in acoustic terms to the first formant and in articulatory terms to tongue height while the other corresponds to the second formant and to tongue backness. Notably, it has long been observed that vowels—and other phonological units—are not simply selected randomly from the underlying phonetic space but exhibit structural organization that can be characterized in terms of dispersion and symmetry (de Boer, 2000).

Certain classes of account explain such organization in terms of such notions as descriptive complexity and *markedness* (Chomsky & Halle, 1968; Jakobson & Halle, 1956). That is, systems that are less "marked" and can be described using fewer features are more compressible, and thus more easily learnable, and thus impose lower cognitive costs on their users. (See Blevins, 2004, and de Boer, 2001, on the potential for circularity in such approaches.) Other accounts have attempted to ground distinctive features and markedness in terms of the physical realities of the articulatory system and their constraining influence on individual phonemes (e.g., Stevens & Keyser, 2010; Carré, Divenyi, & Mrayati, 2017; Flemming, 2001).

Other accounts focus instead on the functional advantages of dispersion for the system as a whole (e.g., Lindblom, 2003). This includes dispersion theory, which emphasizes perceptual distinctiveness (Liljencrants & Lindblom, 1972). That is, greater perceptual distance between signals provides protection against noise (in its information-theoretic sense; Wiley, 2017). This is potentially somewhat in tension with pressures acting on the producer to reduce effort, both in terms of reducing metabolic production effort but also in terms of selecting areas of the phonetic space that are easier to locate reliably over repeated productions (Harris, 2005; Stevens, 1989; Stevens & Keyser, 2010). Given this, later versions of dispersion theory focus on achieving *sufficient* dispersion rather than on *maximizing* dispersion (Lindblom & Maddieson, 1988). The demands of the producer and perceiver are not necessarily at odds, however. A perceptually well-dispersed vowel system may have vowels located in the center and in the corners, which also satisfy production demands for effort reduction.

Roberts and Clark (2020) conducted a non-linguistic referential communication game experiment (for an overview of such approaches, see Nölle & Galantucci, 2022) to tease these pressures apart and to investigate the role of communicative interaction more broadly in the emergence of phonological dispersion. Participant dyads took turns as *Sender* and *Receiver* to communicate a set of animal silhouettes to each other. To communicate, the Sender would move their finger around on a rectangular trackpad; any given finger position corresponded to a color (displayed dynamically on their screen) drawn reliably from an underlying colorspace. By holding their finger in place for 1 s the Sender could send that color to the Receiver. Each round lasted 20 s in total and the Sender could send series of as many colors as they wished during this time. Dyads thus established sets of signals, composed of color-based "phonemes". Roberts and Clark (2020) manipulated the nature of the underlying color space such that, in the *Outer-edge* condition, the color corresponding to the very center of the trackpad was black and colors became brighter the further away from the center the Sender's finger was. This meant that the corners and edge of the trackpad, which are the easiest parts of the pad to find reliably for the Sender, also produced the most distinct col-

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

ors. In the *Inner-edge* condition, the colors became lighter as the Sender's finger moved away from the center but only up until they reached an imaginary line 30% of the way in from the edge of the pad; after that they abruptly became darker again, with the very corners of the pad corresponding to black. This had the consequence that the pressures acting on production and perception were poorly aligned so that participants could satisfy one but not the other. Roberts and Clark (2020) found—in line with dispersion theory—that participants took account of perceptual pressures and created systems that were significantly less dispersed from the point of view of production in the Inner-edge condition than the Outer-edge condition. This helped maintain perceptual distinctiveness but came at a cost, with dyads also being significantly less successful at establishing reliable signals in this condition. In the Outer-edge condition, dyads produced systems qualitatively resembling natural-language vowel systems, which exhibited greater levels of dispersion than would be expected by chance. Roberts and Clark (2023) conducted a post hoc exploration of the original data, shedding light on the ways in which such dispersion emerged over time through interactive processes.

Our experimental study

We conducted a preregistered replication of Roberts and Clark (2020) with three changes, designed to extend the paradigm and investigate the role of further theoretically important variables.

The first change involves the inclusion of a minimum signal-length as a means of investigating coarticulation. In our experiment, the Sender had to select a minimum number of colors without lifting their finger from the pad before the colors would send. The minimum number was manipulated as an independent variable with two levels: 1 vs. 2. We predicted that dispersion would be reduced when minimum signal length was greater, as a pressure to maintain distinction between adjacent units in a signal would lead participants to fill more of the space with signal units

The second change involves an explicit probabilistic noise variable, so that (in conditions where noise was present) the same location on the trackpad would not necessarily always produce the same color every time. We predicted that noise would motivate greater dispersion by encouraging participants to make more uses of the corners and edges of the space, which are generally easier to locate for the Sender and which reduce the degrees of freedom on which noise can operate.

The third change simply involves changing the underlying colorspace. Roberts and Clark's (2020) Inner-edge condition provided a pressure *against* dispersion. However, their Outer-edge condition was not neutral with regard to dispersion, and in fact encouraged it. We therefore therefore included a colorspace that was not dark in the center or corners, for which there was not a clear perceptual pressure for dispersion (although we would expect the production-based pressure to locate reliable locations to remain).

Method

We conducted a preregistered experiment. The preregistration document can be read at
https://aspredicted.org/QY1_MPZ.

Participants

164 university students,[1] none of whom were colorblind, participated in dyads for course credit or $15.

Outliers We excluded as outliers dyads who did not follow task instructions, such as attempting to use verbal communication with each other (of which there were none), as well as dyads who failed to send signals in a high number of turns (defined as 2 or more SDs from the mean). Four such dyads were excluded. The pattern of results does not change if they are included.

Materials

Participants sat in separate cubicles, each with a computer (a mid-2014 Apple iMac with a 21.5" screen), running custom-designed software written in Python (Python Software Foundation,www.python.org) and Kivy (Virbel, Hansen, & Lobunets, 2011; www.kivy.org), and a wireless multitouch trackpad (a 2009 Apple Magic Trackpad, measuring 13.01cm by 13.13cm). Participants could not see each other from their cubicles or hear each other easily.

Procedure

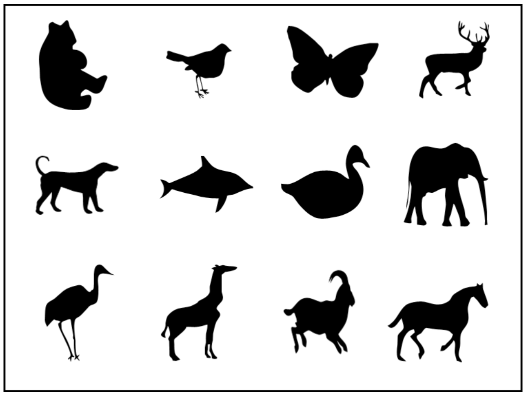

Figure 1: Referents used in experiment

Pairs of participants played a cooperative communication game, taking turns to be *Sender* and *Receiver*. Each participant (henceforth also *player*) in a dyad sat in a separate cubicle and saw a screen divided vertically into two halves. (For the most part, the screen looked much the same whether the player was Sender or Receiver; Figures 2 and 3). In the left half of the screen—the *referent panel*—a set of *referents* were displayed (black animal silhouettes, the same as those

---

[1] We planned to collect data from 80 dyads (i.e., 160 participants) but accidentally gathered data from two extra pairs. The pattern of results remains the same if they are included or excluded. Here we include them.

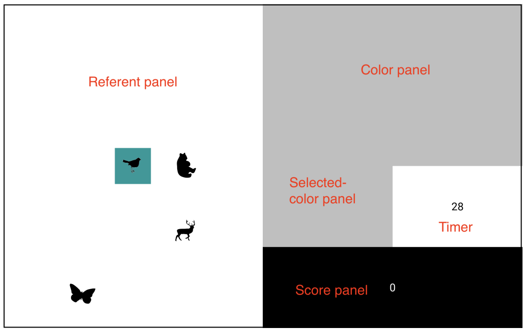

Figure 2: Screenshot of sender screen. Labels in red are for clarity and were not shown present in the experiment.

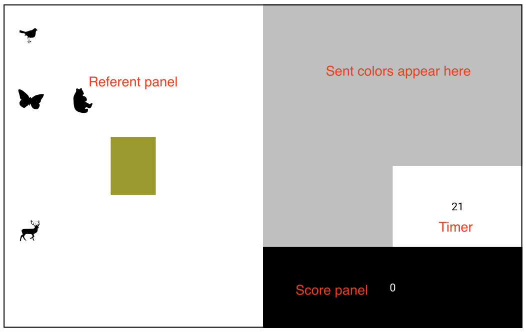

Figure 3: Screenshot of receiver screen. Labels in red are for clarity and were not shown present in the experiment.

used by Roberts & Clark, 2020; Fig. 1). The top right quarter of the screen, the *color panel*, appeared gray by default, but would change color depending on the behavior of the Sender. The same was true of a smaller section immediately below it—the *selected-color panel*—which was also gray by default and took up a quarter of the width of the screen as a whole and a quarter of the height. (See below for a description of how the color panel and the selected-color panel worked.) To the right of the selected-color panel, a timer was displayed on a white background. Below this, taking up half the width of the screen, was a *score panel* displaying the dyad's joint score against a black background.

The referent panel differed slightly for the Sender and the Receiver. First, the referents were not in the same places (i.e., were redistributed at random) from round to round. Second, no referent was ever in the center of the Receiver's referent panel; the Sender, on the other hand, always had one referent in the center, against a teal background (Fig. 2). This varied from round to round and was selected at random, by the server, from the set of available referents. Third, the Receiver had a yellow cursor that could be moved around the referent panel by using the arrow keys on the computer keyboard (Fig. 3); the Sender had no such movable cursor.

The Sender's task was to convey to the Receiver which referent was highlighted in the center of their referent panel by sending series of colors to the Receiver (see below), and the Receiver's task was to move their cursor to the correct referent and press enter. Both players would then receive feedback: The correct referent would be highlighted in the referent space for the Receiver and the chosen referent would be highlighted for the Sender. This happened whether or not the Receiver chose correctly. If the Receiver did choose correctly, the dyad would score one point; their total point score was displayed throughout the game in the score panel at the bottom of the screen. After players started to do well at signaling the referents, more were added, in groups of four, up to a total of twelve. This would occur if, for all referents in the referent panel, the Receiver had selected them correctly at least 75% of the time over the previous four rounds in which they had occurred (cf. Roberts, Lewandowski, & Galantucci, 2015). Once referents were added, they were never removed and would continue to occur as targets.

The feedback stage started immediately after the Receiver selected a referent. If they selected no referent, this would happen after 30 s (in which case the dyad scored no point for that round). The feedback stage always lasted for two seconds, after which a new round would begin. Whatever the outcome of the round, the players would swap roles for the following round. The game lasted for 80 min in total, and would finish at the end of the current round when the 80 min mark had been passed. At the start of the experiment, players played four practice rounds that differed from the ordinary rounds in three ways: First, they lasted 90 s rather than 30 s; second, the players' score from these rounds did not carry over into the normal rounds; third, players were reminded at the start of each round whether they were Sender or Receiver. Beyond being told to move a finger around the pad and observe the screen, and to hold a finger down for 1s to send a color, players were not instructed how to use the signaling medium, but rather had to explore it on their own.

Signaling medium To convey to the Receiver which referent to select, the Sender needed to send a series of colors. This could be achieved by placing a single finger on the trackpad, which would produce a color in the color panel on the top right of the Sender's (though not Receiver's) screen. This color was dependent on the coordinates of the Sender's finger and would change in real time as the Sender moved their finger. To select a color to send, the Sender had to hold their finger in place on the trackpad for one second or longer. As confirmation that this color had been selected, it would then appear for two seconds in the selected-color panel that occupied a space directly underneath the color panel, to the left of the timer. Having selected a color, the Sender could move their finger to other parts of the pad to select more. Only when they lifted their finger from the pad would the series of selected colors be sent to the Receiver, for whom each color would appear for two seconds in the Receiver's color panel in the order they had been selected. (This two-second period was fixed and was not influenced by how long the Sender held their finger down; in other words, duration was not a variable property that could be employed for communication.) We will refer in what follows to these colors as *units* or *color units*, each of which should be understood as being specified by a pair of xy coordinates in the articulatory space and a set of CIELAB coordinates in the perceptual space. A set of units sent simultaneously in the game will be referred to as a *signal*. After sending a signal, the Sender would see the units comprising it displayed in the selected-color panel as a set of simultaneously displayed colored rectangles. The first color would be represented by a rectangle the same size as the selected-color panel; the second color would be represented by a smaller rectangle inside the first; and so on.

The Sender had to select a minimum number of colors before raising their finger from the pad; if they raised their finger having selected fewer than this number, a red bar would appear in the selected-color panel and no colors would be sent to the Receiver. For half the participants the minimum number was one; for the remaining half of participants, the minimum number was two (see Section *Conditions*). The Sender could send as many series of colors as they liked—including none at all—within the time available (20 s).

The relationship between the Sender's finger position and the color produced was based on a CIELAB colorspace. This was done by having the a* axis vary according to either the x- or y-coordinate and the b* axis vary according to the other. Both these axes ranged from -128 to 127, such that the edge of the trackpad corresponded to one end of this range while the other edge corresponded to the other end. Which edge corresponded to which end, as well as which xy axis corresponded to which CIELAB axis, was counterbalanced between trials. The lightness value L* was varied between dyads. Half the dyads experienced the *light-center* colorspace, in which L* was set to 100 for all positions on the trackpad. The remaining dyads experienced the *dark-center* colorspace, in which L* varied from 0 to 100 depending on the proximity of the Sender's finger to the center of the pad, with the very center of the pad corresponding to a value of 0 and the corners of the pad corresponding to a value of 100. (See Figures 4 and 5 for examples.) The colorspaces were counterbalanced with the noise and signal-length conditions.

Conditions Aside from the underlying colorspace we manipulated two main independent variables, each of which had two between-subjects levels. The first of these concerned minimum signal-length, that is, the minimum number of colors that the Sender had to select before their colors would be sent. This could be one or two. In neither condition was the maximum number of colors specified (although the time limit provided a practical limit on how many could be selected). The goal of this was to investigate the effect of coarticulation on the systems produced.

The second independent variable concerned the presence or absence of noise. In the *Noise condition*, the colors that the Sender selected would be affected randomly by probabilistic noise. That is, 25% of the time when the Sender selected a color, the x- and y-coordinates of the Sender's finger would be shifted in a random direction by between 10 and 20% of

Figure 4: Example colorspace from Light-center condition

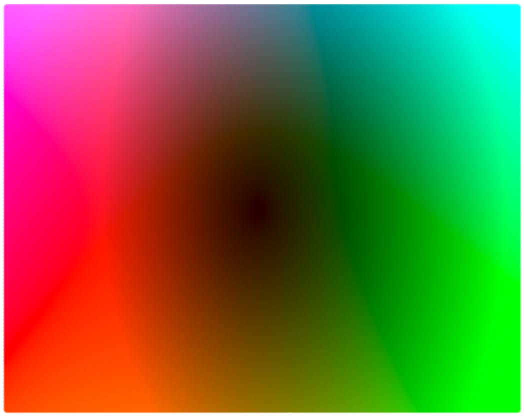

Figure 5: Example colorspace from Dark-center condition

the overall size of the space. The direction would be decided by an independent random 50/50 choice of direction for each of the two coordinate axes. The amount of movement along each axis was also generated independently as a uniform random value between 0.1 and 0.2. The color was then generated (and shown to the Sender) based on the new coordinates. For instance, if the Sender's finger were located in the exact upper-right corner of the space, there was a 25% chance that the color shown might be drawn from a point up to 20% of the way down the right edge of the space, up to 20% to the left of the corner along the top edge, or somewhere in-between those two points. In the *No-noise condition* there was no such noise imposed, and the same finger position would always produce the same color for a given participant.

These two independent variables were crossed with each other in a 2×2 design; as described above, they were also fully counterbalanced with the two colorspaces.

Results

Game length and success

Mean game length was 264.95 turns ($SD = 57.68$). Variability was due to the fact that a round lasted 30 s by default but would end immediately when the Receiver selected a referent. Some dyads tended to use close to the entire time each round while others used as little as 10 s. The total number of turns was related to success, $r(76) = 0.66, p < 0.001$, most likely because speed in selecting referents is a proxy for Receiver confidence. 39.7% of pairs successfully established signals for at least eight referents, while four pairs (5%) successfully established signals for all 12, and 32% did so for at least some of the final four referents. Following Roberts and Clark (2020) we calculated a success index as $(\sum_1^{n_r} s)/12n_r$, where n_r is the number of rounds and the numerator is thus a cumulative count of s, the number of successfully established words in a given round (with 12 being the maximum possible given the number of referents).[2] The mean success index was 0.35 ($sd = 0.15$), but scores ranged from 0.005 to 0.69. A linear model with success index as the dependent variable and colorspace, noise, and minimum signal length as predictors, along with their interactions, found an effect of minimum signal length, $\beta = -0.19, SE = 0.06, t = -3.17, p = 0.002$, with participants finding the game harder when minimal signal length was 2. There was no effect for either of the other predictors, nor any interactions.

Phoneme sets

For each player we took the final successful signal for each referent and took the set of units from all these signals as representing the "raw" combinatorial units of the player's system. However, units are likely to be repeated between signals, so we established a final set of signals using a k-means clustering analysis (with 25 initial configurations) based on the units' xy coordinates (Steinley, 2006). We established a value for k via silhouette analysis, using the factoextra library in R (Kassambara & Mundt, 2017).

Dispersion

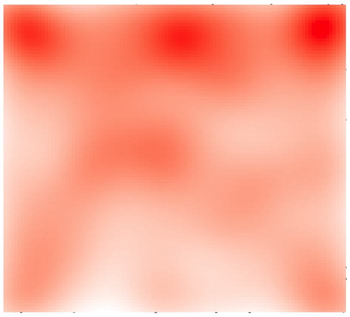

Figure 6: Heatmap of color units from final successful signals (all conditions)

Figure 6 shows a heatmap of all color units from final successful signals across conditions. As can be seen, there is clear evidence of dispersion. Following Roberts and Clark (2020) we measured dispersion in terms of mean pairwise distance between final units in a set, normalized by dividing it by the highest possible score for that number of phonemes, resulting in a value between 0 and 1. The overall mean pairwise distance was 0.62 ($sd = 0.11$). We calculated chance level by generating random sets of points of the same sizes as the real sets (before clustering), applying k-means clustering, and calculating dispersion as for the real data. We did this 100 times and took the overall mean as the level to be expected by chance. This was 0.56. We compared the real data with this value by generating a normally distributed random set of data of the same length as the original data for each measure, using the standard deviation of the real data but the mean of the random data. Based on this, participants' sets were significantly more dispersed than chance: $\beta = 0.074, SE = 0.02, t = 4.52, p < 0.001$.

As our primary preregistered analysis, we ran a linear model with pairwise distance as dependent variable and noise and minimum signal length as predictors along with their interaction terms. There were no significant effects ($p > 0.6$). Contrary to our expectations, in other words, neither noise nor coarticulation seems to have affected dispersion.

As a secondary preregistered analysis, we also ran a model with mean pairwise distance as dependent variable and colorspace as predictor, with random intercepts for noise and minimum signal length. We found a significant effect of colorspace, with greater dispersion in the light-center condition: $\beta(74) = 0.06, SE = 0.02, t = 2.79, p < 0.01$. Figure 7 shows a heatmap of signals in this condition.

There was a relationship between mean pairwise distance and success (Figure 8). We conducted a mixed model

---

[2] No player could score 1, as that would require them to have successfully communicated all twelve referents several times before the start of the game. Correcting for this would have needlessly complicated an index intended as a relative measure.

with success index as dependent variable and pairwise distance as the predictor, with random intercepts for colorspace, minimum-signal length, and noise conditions: $\beta(71.6) = 0.82, SE = 0.11, t = 7.79, p < 0.001$.

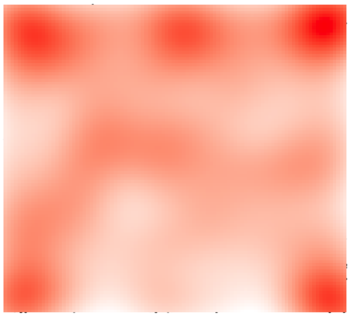

Figure 7: Heatmap of all color units from final successful signals for Light-center colorspace

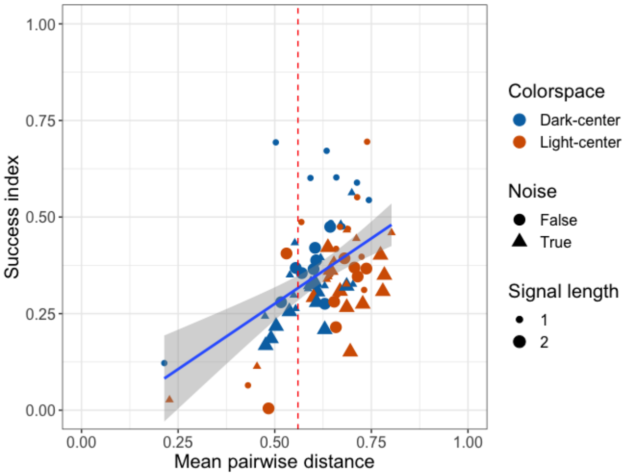

Figure 8: Relationship between pairwise distance and success. Red dotted line indicates chance level pairwise distance.

Discussion

We replicated Roberts and Clark's (2020) non-linguistic referential communication game experiment to investigate the role of communicative factors including noise and coarticulation in the emergence of structural organization in phonological inventories. To our surprise, we did not find an effect of either noise or coarticulation. However, consistent with Roberts and Clark (2020), we did find an effect of underlying colorspace. In spite of the unexpected lack of a result for noise and coarticulation, we consider these results to be interesting and encouraging in several ways for the paradigm.

First, the effect of colorspace lends further evidence for the role of the topology of the signaling space in phonological structure (cf. Stevens & Keyser, 2010). Second, the lack of an effect of coarticulation is encouraging for the robustness of Roberts and Clark's (2020) result, since it suggests that their results were not an artifact of participants selecting units in isolation. (It's worth adding that we did find within-signal effects whereby signal-initial units were at more extreme locations of the space than later ones; but this did not impact the system as a whole.) The lack of a noise effect seems likely due to the fact that communication in this paradigm is inherently non-trivial and, as a result, somewhat noisy. In other words, noise-driven dispersion might well have been at ceiling. In future work a better approach to investigating the role of noise would be to reduce it rather than increase it.

Overall, we hope that this work will provide an impetus to make further progress within this and similar paradigms towards investigating the structure of phonological inventories.

Acknowledgements

We thank many members of the Cultural Evolution of Language Lab for running trials; we thank Colin Twomey for advice on analysis; and we gratefully acknowledge funding from the National Science Foundation (award number 1946882)

References

Blevins, J. (2004). *Evolutionary Phonology: The Emergence of Sound Patterns*. Cambridge: Cambridge University Press.

Brentari, D. (2011). Sign language phonology. In J. Goldsmith, J. Riggle, & A. C. L. Yu (Eds.), *The Handbook of Phonological Theory* (pp. 691–721). Malden, MA: Wiley-Blackwell.

Carré, R., Divenyi, P., & Mrayati, M. (2017). *Speech: A Dynamic Process*. Berlin/Boston: de Gruyter.

Chomsky, N., & Halle, M. (1968). *The Sound Patterns of English*. Cambridge, MA: MIT Press.

de Boer, B. (2000). Self-organization in vowel systems. *Journal of Phonetics*, *28*(4), 441–465. doi: 10.1006/jpho.2000.0125

de Boer, B. (2001). *The Origins of Vowel Systems*. Oxford: Oxford University Press.

Flemming, E. (2001). Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology*, *18*, 7–44. doi: 10.1017/S0952675701004006

Harris, J. (2005). *Vowel reduction as information loss* (C. J. Ewen, J. Durand, & P. Carr, Eds.). John Benjamins.

Jakobson, R., & Halle, M. (1956). *Fundamentals of Language*. The Hague: Mouton.

Kassambara, A., & Mundt, F. (2017). *Package 'factoextra'*.

Ladefoged, P., & Disner, S. F. (2012). *Vowels and consonants*. John Wiley & Sons.

Liljencrants, J., & Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language*, *48*(4), 839–62. doi: 10.2307/411991

Lindblom, B. (2003). Patterns of phonetic contrast: Towards a unified explanatory framework. In *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS)* (pp. 39–42).

Lindblom, B., & Maddieson, I. (1988). Phonetic universals in consonant systems. *Language, Speech, and Mind*, 62–78.

Nölle, J., & Galantucci, B. (2022). Experimental semiotics. In A. M. García & A. I. nez (Eds.), *The routledge handbook of semiosis and the brain* (p. 4). Taylor & Francis.

Roberts, G., & Clark, R. (2020). Dispersion, communication, and alignment: An experimental study of the emergence of structure in combinatorial phonology. *Journal of Language Evolution*, *5*(2), 121–139.

Roberts, G., & Clark, R. (2023). The emergence of phonological dispersion through interaction: An exploratory secondary analysis of a communicative game. *Frontiers in Psychology*, *14*, 1130837.

Roberts, G., Lewandowski, J., & Galantucci, B. (2015). How communication changes when we cannot mime the world: Experimental evidence for the effect of iconicity on combinatoriality. *Cognition*, *141*, 52–66. doi: 10.1016/j.cognition.2015.04.001

Steinley, D. (2006). K-means clustering: A half-century synthesis. *British Journal of Mathematical and Statistical Psychology*, *59*(1), 1–34.

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics*, *17*, 3–45.

Stevens, K. N., & Keyser, S. J. (2010). Quantal theory, enhancement and overlap. *Journal of Phonetics*, *38*(1), 10–19. doi: 10.1016/j.wocn.2008.10.004

Virbel, M., Hansen, T., & Lobunets, O. (2011). Kivy – A framework for rapid creation of innovative user interfaces. In *Workshop-Proceedings der Tagung Mensch & Computer 2011. überMEDIEN|ÜBERmorgen.*

Wiley, R. H. (2017). How noise determines the evolution of communication. *Animal Behaviour*, *124*, 307–313.