



## Iconic bias in Italian spatial demonstratives

Ian Joo, Yu-Yin Hsu & Emmanuele Chersoni

**To cite this article:** Ian Joo, Yu-Yin Hsu & Emmanuele Chersoni (2022) Iconic bias in Italian spatial demonstratives, *Australian Journal of Linguistics*, 42:1, 57-74, DOI: [10.1080/07268602.2022.2056410](https://doi.org/10.1080/07268602.2022.2056410)

**To link to this article:** <https://doi.org/10.1080/07268602.2022.2056410>



© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 13 May 2022.



Submit your article to this journal [↗](#)



Article views: 726



View related articles [↗](#)



View Crossmark data [↗](#)

RESEARCH ARTICLE



# Iconic bias in Italian spatial demonstratives

Ian Joo , Yu-Yin Hsu and Emmanuele Chersoni

The Hong Kong Polytechnic University

## ABSTRACT

An iconic pattern across spoken languages is that words for ‘this’ and ‘here’ tend to have high front vowels, whereas words for ‘that’ and ‘there’ tend to have low and/or back vowels. In Italian, there are two synonymous Italian words for ‘here’, namely *qui* and *qua*, and two synonymous words for ‘there’, *lì* and *là*. *Qui* ‘here’ and *là* ‘there’ are iconic because *qui* has the high front vowel /i/ and *là* has the low vowel /a/, whereas *qua* ‘here’ and *lì* ‘there’ are counter-iconic, since their vowels are the opposite. Based on corpus, survey and computational data, we demonstrate that (i) *qui* ‘here’ and *là* ‘there’ have been consistently used more frequently throughout history compared to *qua* ‘here’ and *lì* ‘there’, respectively; and (ii) in present-day Italian, *qui* ‘here’ tends to refer to a location that is closer to the speaker than *qua* ‘here’ does, whereas *là* ‘there’ tends to refer to a location that is further away from the speaker than *lì* ‘there’ does. In summary, the iconic demonstrative pronouns (*qui* and *là*) are used more frequently and are closer to the prototypical meanings of ‘here’ and ‘there’. We argue that their frequency and prototypicality are motivated by their iconic power. This case study shows how iconicity may work as pressure on language use and language change.

## ARTICLE HISTORY

Accepted 26 February 2022

## KEYWORDS

Iconicity; Italian; spatial deixis; language change; sound symbolism

## 1. Introduction

Spoken languages tend to express the proximal deixis with front high vowels (such as ‘this’ and ‘here’ in English), while the distal deixis is often expressed with low and/or back vowels (such as ‘that’ and ‘there’ in English). Previous studies have repeatedly demonstrated this pattern based on sizeable samples of diverse languages (Johansson & Zlatev, 2013; Tanz, 1971; Ultan, 1978; Woodworth, 1991).

Such tendency of vowel-to-distance mapping has been observed in experimental settings as well. Lockwood and Dingemanse (2015) provide an overview of experimental studies on sound symbolism, including the association between sound and shape, size, colour, taste and speed. Although their review does not mention research on the sound symbolism of distance, Rabaglia et al.’s (2016) experiments demonstrate that vowel-to-distance mapping can influence the human perception of physical distance.

**CONTACT** Ian Joo [ian.joo@connect.polyu.hk](mailto:ian.joo@connect.polyu.hk)

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

**Table 1.** Proximal and distal demonstratives with vowel quality as their minimal contrast.

Language	Proximal	Distal
Hungarian	<i>itt</i> /it:/ 'here'	<i>ott</i> /ot:/ 'there'
Indonesian	<i>sini</i> /sini/ 'here'	<i>sana</i> /sana/ 'there'
Khmer	នេះ /nih/ 'this; here'	នោះ /nuh/ 'that; there'
Tamil	இது /itu/ 'this'	அது /atu/ 'that'

In one of their experiments, the English-speaking participants were asked to throw a beanbag towards a dog toy after being informed of the toy's name. The participants who were instructed that the toy was called [dob], [nup] or [kuv] (all with a back vowel) threw the beanbag further away from themselves than did those who were informed that the toy was named [dib], [nɛp] or [kiv] (all with a front vowel). This shows that the vowel-to-distance mapping may have psychological effects on human behaviour. Thus, it is not surprising that human languages may reflect this cognitive bias.

Table 1 shows some examples in which proximal and distal demonstrative pronouns show a minimal contrast to vowels in different languages. This type of vowel-to-distance mapping, as will be explained in §2.1, is motivated by **iconicity** – the resemblance between form and meaning.

Spatial deixis in Italian consists of two pairs of proximal and distal demonstrative pronouns: *qui* /kwi/ and *qua* /kwa/, both meaning 'here', and *lì* /li/ and *là* /la/, both meaning 'there'. Strik Lievers and Miola (2018, p. 76) suggest that *lì* is semantically changing into the medial demonstrative in contemporary Italian, whereas *là* remains as the distal demonstrative. But as we will show in §5, not all speakers make such a distinction, and many speakers perceive *lì* and *là* as synonyms.

Interestingly, each pair of these Italian synonyms differs only in the vowel (/i/ v. /a/). Particularly, one word in each pair conforms to the iconic mapping between vowel and distance: *qui* with a front high vowel expressing proximal deixis, and *là* with a low vowel expressing distal deixis, while the other does not (*qua* and *lì*).

This study adopts three different methodologies (corpus, survey and computational) to investigate whether this iconic discrepancy in Italian spatial demonstrative pairs has motivated uneven historical distribution and subtle semantic distinctions. Based on three methodologies, we make the following observations:

- (i) iconic pronouns (*qui* 'here' and *là* 'there') have been used more frequently than their non-iconic counterparts (*qua* 'here' and *lì* 'there') throughout history (corpus study in §4); and
- (ii) the pronouns with /i/ (*qui* 'here' and *lì* 'there') tend to express places that are closer to the speaker compared to the pronouns with /a/ (*qua* 'here' and *là* 'there') (survey in §5 and computational study in §6).

We then argue that iconicity may have motivated (i) and (ii). In other words, *qui* 'here' and *là* 'there' have been favoured over *qua* 'here' and *lì* 'there' throughout history due to their iconic power, and the iconic nature of the vowel /i/ may have motivated *qui* 'here' and *lì* 'there' to refer to places that are closer to the speaker in preference to *qua* 'here' and *là* 'there'. (We do not, however, rule out the possibility that (i) may have motivated (ii) or vice versa, as we discuss briefly in §7.)

## 2. Background

### 2.1 *The iconic association between vowel and distance*

Why is /i/ associated with proximity and /u/ and /a/ with distance? Vainio (2021) argues that vowel-to-distance mapping is part of the larger cognitive association between vowel and magnitude, referred to as **magnitude symbolism**. Many experimental studies have demonstrated that we tend to associate a small size with high, front and unrounded vowels, and a large size with low, back and/or rounded vowels (Knoeferle et al., 2017; Sapir, 1929; Shinohara & Kawahara, 2010). According to Vainio's (2021) view, both vowel-to-distance mapping and vowel-to-size mapping are different realizations of the same magnitude symbolism. This theory is intuitively persuasive, since distance and size are simply different dimensions of physical quantity.

Some studies have suggested that /u/, /a/, and other low and/or back vowels may be associated with large size because their articulation involves widening the (front) oral cavity (Masuda, 2007; Shinohara & Kawahara, 2010). Based on this theory, we suggest that magnitude symbolism is an **iconic** association, since it involves the meaning (the magnitude) resembling the form (the size of the oral cavity when articulating a vowel).

Eberhardt's (1940) experiment with deaf and hearing child participants strongly supports the articulation theory. In her experiment, the deaf group and the hearing group were asked to judge whether made-up words with different vowels referred to something big or to something small. Both groups tended to judge words with /ɔ/ and /a/ as referring to something big and those with /i/ as referring to something small. Since the deaf children were prelingually deaf, the only reasonable explanation for their vowel-to-size mapping was articulatory. However, the deaf and hearing children did not rate the size of the vowels identically. For example, the deaf children judged /u/ as being neither big nor small, whereas the hearing children judged it to be big-sounding. Thus, as Eberhardt concluded, magnitude symbolism is not uniquely motivated by acoustics or articulation, but by a combination of both.

Based on these studies, we conclude that vowel-to-distance mapping is an iconic association between physical distance and the size of the oral cavity when articulating a vowel (although this may not be the only motivating factor). Thus, the Italian demonstratives *qui* and *là* iconically represent the meanings 'here' and 'there', whereas *qua* and *lì* are counter-iconic.

### 2.2 *Iconicity-driven language change*

Jespersen (2010) pointed out that, in many different languages, words that are semantically related to smallness contain the vowel /i/; accordingly, the fact that these words contain this iconic character is an important factor in their survival throughout language change:

... [T]he fact that a word meaning little or little thing contains the sound [i], has in many, or in most, cases been strongly influential in gaining popular favour for it; the sound has been an inducement to choose and to prefer that particular word, and to drop out of use other words for the same notion, which were not so favoured. In other words, sound-symbolism makes some words more fit to survive and gives them a considerable strength in their struggle for existence. (pp. 288–289)

If this hypothesis is true, we would expect the Italian iconic demonstrative words to have increased in popularity in comparison to their non-iconic counterparts due to their phonosemantic match. In other words, *qui* and *là* must have been used more frequently throughout history than were *qua* and *lì*.

Johansson and Carling (2015) investigated how various Indo-European languages have constantly changed throughout history to fit their demonstrative pronouns into the vowel-to-distance mapping schema. They observed a general diachronic pattern in which demonstrative pronouns tend to be constantly reconstructed towards an iconic vowel-to-distance mapping. For example, the English pronouns *this* and *that* are derived from Old English neuter proximal pronoun *þis* and neuter distal pronoun *þæt*, respectively. The motivation for sound symbolism can explain why only the neuter forms *þis*/*þæt*, whose vowels are iconically mapped onto distance, survived into modern English, while the feminine forms *þes*/*sē* and the masculine forms *þeos*/*sēo*, whose vowels are not iconically mapped onto distance, dropped out. This example and many others imply that there is constant pressure for language change towards iconicity, which explains how many contemporary forms of demonstrative pronouns in different languages show a similar vowel-to-distance mapping pattern. We can thus predict that similar pressure must have pushed Italian speakers to use *qui* 'here' and *là* 'there' more frequently than their counter-iconic counterparts, *qua* 'here' and *lì* 'there'.

Moreover, Benczes (2020) hypothesized that iconicity could motivate semantic change. As an example, she mentioned the case of *buxom*, which had changed in meaning from the Old English *bugan* 'to bend' to its contemporary meaning of 'big-breasted'. This semantic change must have been critically motivated by English words that begin with *b* and share similar meanings, such as *breast* and *bosom*. This may be related to the iconic association between labial sounds and softness, since human lips are soft. Sakamoto and Watanabe (2018) experimentally demonstrated the perceptual association between softness and bilabial sounds among Japanese speakers. Moreover, cross-linguistically, lexemes meaning 'breast' and 'ash', both of which are soft substances, tend to have [+labial] phonemes, such as /u/ or /m/ (Blasi et al., 2016; Johansson et al., 2020; Joo, 2020).

In view of this hypothesis regarding iconic pressure on semantic change, it is noteworthy that the two pairs of Italian spatial demonstrative pronouns, while consisting of fully interchangeable synonyms, do exhibit some subtle semantic differences. For example, some speakers of Italian report that *qui* denotes a more concrete, specific space that is close to the speaker, whereas *qua* refers to a broader space that is not as close to the speaker.

Maiden and Robustelli's reference grammar of Italian (2007) notes:

There is a further difference between the forms with the vowel *a*, and those with *i*. Those with *i* are essentially 'punctual': they refer to clearly defined, focused, 'points' in space or time; those in *a* are 'areal', and have a vaguer, more diffuse reference, which helps to explain why only the *a* forms are encountered in certain expressions denoting general but not specific position, or general motion in a particular direction: *qua e là* 'here and there' (as in *girava qua e là* 'he wandered hither and thither'), *di là* 'over there', 'over that way', 'beyond', *di qua* 'over here', 'this way', 'on this side', *più in là* 'further away', *più in qua* 'closer in', *quaggiù* 'down here', *laggiù* 'down there', *quassù* 'up here', *lassù* 'up there'. There is thus a distinction between *Non passare di là* 'Don't go that way', and (the rather unusual) *Non passare di lì* 'Don't go through that specific spot'. (p. 89)

Nobile (2011, pp. 123–124) suggests that the vowels of *qui/qua* and *lì/là* are iconically related to the specificity of the place these pronouns designate: /i/ (with smaller oral cavity) is similar to the ‘punctual’ sense of *qui* and *lì*, whereas /a/ (with larger oral cavity) is similar to the ‘areal’ sense of *qua* and *là*.

Maglio et al. (2014) experimentally demonstrated that front vowels are perceptually associated with precision compared to back vowels: in one of their experiments, participants were asked to visually divide an image of a city into as many regions as possible. When the city’s name contained a front vowel (*Fleeg*, *Theek* or *Cheetle*), participants divided the city into finer divisions than when the city’s name contained a back vowel (*Floog*, *Thook* or *Chootle*), showing that “people apply greater precision to targets labeled with front (vs. back) vowels” (p. 1086). Thus, there is experimental evidence to believe that the /i/ vowel of *qui* and *lì* is iconically related to specificity, compared to the /a/ vowel of *qua* and *là*.

Whether the majority of the Italian speakers agree with such differences in spatial precision is unclear, and there may be inter-speaker variations regarding what semantic differences the two pairs convey. This begs the following questions: if the iconic demonstrative pronouns and their non-iconic counterparts have gone through a subtle semantic divergence, have the iconic pronouns changed towards the more prototypical senses of ‘here’ and ‘there’ compared to their non-iconic synonyms?

### 3. General research question

Previous research has shown that: (i) certain words of spoken languages show an iconic association between sound and meaning, such as between vowel quality and spatial distance; and that (ii) such cross-linguistic patterns may be motivated by diachronic pressure on language change. Previous studies for the second argument have either provided cross-linguistic evidence based on quantitative data (Johansson & Carling, 2015) or a theoretical framework based on qualitative observation (Benczes, 2020). Another methodology that can be employed is to focus on how iconic pressure has occurred within a single language and then observe this diachronic mechanism at a finer level.

Italian provides a perfect environment for such a micro-level study, because it has two minimal pairs of spatial demonstrative pronouns, differing only in terms of vowel-to-distance mapping, whose usage can be observed throughout a well-recorded written history.

To observe how these minimal pairs of iconicity have behaved differently in the Italian language in the past and present, the current study attempts to survey and investigate the association between iconicity and the semantic differences in demonstrative pronoun pairs in Italian. In this paper, we aim to seek answers to a synchronic and a diachronic question.

- Diachronic question: have Italian speakers generally preferred the use of iconic demonstrative terms to non-iconic demonstrative terms throughout history?
- Synchronic question: is there a (subtle) difference between the meanings of iconic demonstrative terms and the meanings of non-iconic demonstrative terms? If so, are the meanings of iconic demonstrative terms a better fit for the prototypical meanings of proximal and distal deixis?

Section 4 presents a corpus study addressing the diachronic question, while §5 and §6 present two experiments to investigate the synchronic question.

## 4. Corpus study

### 4.1 Research question

Does the cross-linguistic preference for mapping front vowels with proximal deixis and mapping back vowels with distal deixis influence which words Italian speakers choose to express spatial deixis? In other words, do Italian speakers prefer to use *qui* for ‘here’ and *là* for ‘there’, since these words iconically express their meaning, unlike their non-iconic counterparts (*qua* and *lì*)?

### 4.2 Methodology

We used the MIDIA corpus (*Morfologia dell’Italiano in DIAcronia* [Morphology of Italian in Diachrony])<sup>1</sup> to investigate whether the iconic demonstrative terms in Italian were used more frequently than were non-iconic ones in different periods and in different genres. The MIDIA corpus is a diachronic corpus of written Italian that contains approximately 7.5 million tokens from 800 texts dating from the thirteenth to the early twentieth centuries; these are divided into five time periods and seven genres. We searched for the occurrence of *qui*, *qua*, *lì* and *là* in different periods and genres throughout the entire corpus.

### 4.3 Results

Table 2 shows that *qui* and *là* were preferred during five time periods and in seven genres in comparison to *qua* and *lì*, respectively. Two-way ANOVAs investigating the correlation between the number of occurrences of *qui* v. *qua* (or *là* v. *lì*) and different periods revealed that *qui* and *là* have generally been preferred to *qua* and *lì* ( $p < 0.001$ ). However, we did not find a pattern indicating an increasing preference for *qui* and *là* over time: the ANOVAs did not show a significant correlation between the number of occurrences and the pronoun–period interaction.

We conducted 35 two-sided binomial tests based on the parameters shown in Table 3. When corrected for multiple comparisons (Benjamini & Hochberg, 1995), the results show that, in 30 out of 35 period–genre combinations (five periods  $\times$  seven genres), *qui* was significantly preferred to *qua*; in 31 out of 35 period–genre combinations, *là* was significantly preferred to *lì* (FDR = 10%).

### 4.4 Discussion

This section shows that the synonymous pairs *qui/qua* and *lì/là* have both had a biased distribution favouring the iconic pronouns (*qui* and *là*) over their non-iconic counterparts (*qua* and *lì*). This confirms Jespersen’s (2010) hypothesis that iconicity leads to a preference for certain words.

---

<sup>1</sup>[www.corpusmidia.unito.it](http://www.corpusmidia.unito.it)

**Table 2.** The occurrence of each demonstrative pronoun in each period and genre and whether the occurrences of *qui* and *là* were more frequent than were the occurrences of *qua* and *lì*, respectively (\*, FDR = 10%).

Period	Genre	<i>qui</i>	<i>qua</i>	<i>lì</i>	<i>là</i>	<i>qui</i> > <i>qua</i>	<i>lì</i> < <i>là</i>
1200–1375	Exposition	48	10	0	50	*	*
1200–1375	Legal	42	68	4	34		*
1200–1375	Literary prose	134	38	3	118	*	*
1200–1375	Personal	3	0	0	0		
1200–1375	Poetry	190	87	32	219	*	*
1200–1375	Scientific	78	21	0	28	*	*
1200–1375	Theatre and mimesis	16	5	0	6	*	
1376–1532	Exposition	182	83	29	114	*	*
1376–1532	Legal	36	19	11	24	*	*
1376–1532	Literary prose	167	41	7	58	*	*
1376–1532	Personal	315	207	35	73	*	*
1376–1532	Poetry	227	100	17	122	*	*
1376–1532	Scientific	170	37	17	53	*	*
1376–1532	Theatre and mimesis	400	140	7	129	*	*
1533–1691	Exposition	127	19	6	24	*	*
1533–1691	Legal	53	22	8	34	*	*
1533–1691	Literary prose	78	61	2	59		*
1533–1691	Personal	66	41	5	54	*	*
1533–1691	Poetry	249	76	4	159	*	*
1533–1691	Scientific	237	45	8	20	*	*
1533–1691	Theatre and mimesis	317	166	4	161	*	*
1692–1840	Exposition	111	26	9	52	*	*
1692–1840	Legal	55	0	3	18	*	*
1692–1840	Literary prose	176	48	20	98	*	*
1692–1840	Personal	202	28	13	50	*	*
1692–1840	Poetry	197	47	2	120	*	*
1692–1840	Scientific	125	28	3	35	*	*
1692–1840	Theatre and mimesis	400	92	19	80	*	*
1841–1947	Exposition	188	25	19	78	*	*
1841–1947	Legal	5	2	0	5		
1841–1947	Literary prose	151	64	111	174	*	*
1841–1947	Personal	195	33	65	84	*	
1841–1947	Poetry	210	72	40	244	*	*
1841–1947	Scientific	146	21	2	56	*	*
1841–1947	Theatre and mimesis	478	452	78	252		*

**Table 3.** Parameters of binomial tests.

Number of successes	Occurrence of <i>qui</i> (or <i>là</i> )
Number of trials	Occurrence of <i>qui</i> and <i>qua</i> (or <i>lì</i> and <i>là</i> ) + 1
Probability of success	0.5

We also wanted to see if this uneven frequency distribution persists in current-day Italian. As a follow-up survey, we checked the frequency of these four pronouns in the contemporary corpus of *Italian Web 2016* (accessed via [www.sketchengine.eu](http://www.sketchengine.eu)). As expected, the frequency of *qui* (2,603,931) was much higher than that of *qua* (218,016). The frequency of *là* (457,860), however, was roughly equal to that of *lì* (540,358).

Although we expected the preference for the iconic pronouns to increase over time, the results from the MIDIA corpus did not reveal this trend. Thus, the gap between the iconic pronouns and their non-iconic counterparts has not increased over time. Moreover, in contemporary Italian, *lì* (non-iconic) does not seem to be any less frequent than *là* (iconic). This suggests that the non-iconic pronouns may have served some purpose



**Table 4.** Participant information.

Gender	19 females, 12 males, 1 other
Age	Median = 28, sd $\approx$ 3.44
Native language(s)	Italian (except one whose sole native language is Tuscan, an Italian dialect). Ten participants reported one or two Italian dialects as additional native languages.
Occupation	Eighteen participants reported having an academic occupation (such as student or teacher)

that prevented them from becoming extinct and maintained them at a certain level of frequency, even though they were generally used less often than the iconic pronouns throughout history.

Such a purpose may be found in their subtly different semantic connotations. As previously mentioned, some speakers of Italian consider that *qui/là* are not entirely semantically identical to *qua/li*, however subtle the difference may be. If the iconic pronouns and the non-iconic pronouns have (slightly) different semantic meanings, this would explain how the non-iconic pronouns have managed to survive despite the presence of their iconic rivals.

In the following sections, we investigate whether iconicity has influenced (or is influencing) the semantic change in the Italian words for ‘here’ and ‘there’. Even though *qui* is generally interchangeable with *qua* and *là* is generally interchangeable with *li*, given that *qui* and *là* are phonologically closer to the typical words for ‘here’ and ‘there’, are they also semantically closer to the prototypical meanings of ‘here’ and ‘there’?

## 5. Experimental study

### 5.1 Research question

Do Italian native speakers consider *qui* to be semantically different from *qua*, and *li* from *là*? Are the semantic distinctions described by native speakers consistent, and if so, in what way? Do these distinctions support the hypothesis that the iconic demonstrative pronouns (*qui* and *là*) semantically resemble prototypical proximal and distal terms more closely than do their non-iconic counterparts?

### 5.2 Methodology

Thirty-two Italian speakers were recruited to participate in an online survey. Their demographic information is shown in Table 4.<sup>2</sup>

The participants were asked to answer the following open-ended questions, which were presented in Italian.

- *Secondo Lei, il senso/significato delle parole ‘qui’ e ‘qua’ è lo stesso? Se no, qual’è la differenza?* [According to you, are the sense/meanings of the word *qui* and that of *qua* the same? If not, what is the difference?]

<sup>2</sup>Although it is unlikely that a modern Italian only speaks an Italian dialect as a native language and not also standard Italian, we state the participant’s report as it is.

- *Secondo Lei, il senso/significato delle parole 'lì' e 'là' è lo stesso? Se no, qual'è la differenza?*  
[According to you, are the sense/meanings of the word *lì* and that of *là* the same? If not, what is the difference?]

In the survey, we informed the participants that it was not a test, and that there were no correct or incorrect answers; we also asked the participants to answer based on their intuition. The survey was approved for the human subjects ethics review by the Hong Kong Polytechnic University Institutional Review Board (Reference number: HSEARS20210104001).

### 5.3 Survey results

Table 5 summarizes the responses to the survey. Of the 32 participants, 18 responded that *qui* and *qua* were different and 14 that they were the same, while 19 responded that *lì* and *là* were different and 13 responded that they were the same. Regarding the participants who felt that the meanings were different for each pair, even if subtly so, the most significant two opinions were that *qui* and *lì* expressed a **specific location** or a **location close to the speaker**, in contrast to *qua* and *là*.

Eight participants answered that *qui* referred to a more specific place than did *qua*, while only one participant felt that the opposite was true, with *qua* being more specific. Nine participants expressed that *lì* was more specific than was *là*, whereas only one participant expressed the opposite opinion that *là* was more specific.

Four participants noted that *qui* referred to a space that was closer to the speaker, while one considered *qua* to be closer. Eight participants answered that *lì* was closer to the speaker than was *là*, and no one reported the opposite.

Less popular opinions were that *lì* was closer to the listener (2), *là* was closer to the listener (1), *qua* was closer to the listener (1), *qua* had a more concrete sense (1) and that *qua* was more colloquial (1).

### 5.4 Survey discussion

The answers to the open-ended survey illustrated that approximately half of the participants judged *qui* and *lì* to be semantically different from *qua* and *là* (with percentages over 50%). The participants who thought that they were different tended to be of the opinion that *qui* and/or *lì* denote a location that was more specific or was closer to the speaker.

The fact that *qui* and *lì*, both of which contain /i/, may express a location that is closer to the speaker when compared to *qua* and *là*, respectively, confirmed our initial hypothesis that *qui* (closer to the speaker) and *là* (further away from the speaker) were closer to the prototypical meanings of proximal and distal deixis. The relative vicinity of *lì* compared to *là* confirms the abovementioned observation from Strik Lievers and Miola (2018, p. 76) that *lì* is changing into the medial demonstrative in contemporary Italian, at least for some speakers. There is no doubt that closeness to the speaker is associated symbolically with the vowel /i/, as several cross-linguistic studies have confirmed that high front vowels are common in proximal distal pronouns in languages across the world.

**Table 5** Survey responses

No.	%	Response
18	56.3%	<i>Qui</i> is different from <i>qua</i>
19	59.4%	<i>Li</i> is different from <i>là</i>
9	28.1%	<i>Li</i> is more specific than <i>là</i>
8	25%	<i>Qui</i> is more specific than <i>qua</i>
8	25%	<i>Li</i> is closer to the speaker than <i>là</i>
4	12.5%	<i>Qui</i> is closer to the speaker than <i>qua</i>
2	6.3%	<i>Li</i> is closer to the listener than <i>là</i>
1	3.1%	<i>Là</i> is closer to the listener than <i>li</i>
1	3.1%	<i>Qua</i> is closer to the listener than <i>qui</i>
1	3.1%	<i>Qua</i> has a more concrete sense than <i>qui</i>
1	3.1%	<i>Qua</i> is more colloquial than <i>qui</i>

What about the specificity of *qui* and *li* when compared to *qua* and *là*? The specificity of a location, or **spatial boundedness**, is lexically coded in the spatial demonstratives of some languages (Imai, 2003). In Malagasy, *eo* refers to a bounded, specific location and *eny* to an unbounded, non-specific location:

- (1) Malagasy
- |    |  |              |           |
|----|--|--------------|-----------|
| a. | apetraho   | <b>eo</b>    | ity/ito   |
|    | put  | <b>there</b> | this/this |
|    | 'Put this <b>there</b> .' (Imai, 2003, p. 111, slightly modified)                    |              |           |
| b. | apetraho   | <b>eny</b>   |           |
|    | put  | <b>there</b> |           |
|    | 'Put it ( <b>somewhere</b> ) <b>there</b> .' (Imai, 2003, p. 111, slightly modified) |              |           |

As we have mentioned in §2.2, experimental findings support that the vowels of *qui/qua/li/là* are iconically related to their (un)specificity. However, this distinction of boundedness may not be due to iconicity alone, but also to the etymological origin of the demonstratives instead. *Qui* comes from the Latin *ecce hīc* 'see here!' whereas *qua* is from the Latin *ecce hāc* 'see this way!'. Similarly, *li* comes from the Latin *illic* 'there' and *là* from the Latin *illāc* 'by that way'. Since the notion of HERE/THERE is more bounded than is the notion of THIS/THAT WAY, it is possible that the distinction of boundedness is a direct result of the Latin heritage rather than being an Italian innovation.

In summary, the proximities indicated by *qui* and *li* may be due to their vowel /i/, although their boundedness may not be only due to the iconic value of /i/ but may result from their historical origin. While both iconicity and historical origin may have played a role, it is difficult to confirm which one of them was the main cause.

## 6. Computational study

The experiment in §5 is limited in several ways: it is an open-ended survey based on a small group of participants and biased towards an academic background (more than half having an academic occupation).

In order to confirm the personal observations retrieved from the experiment in §5, we adopted a computational method to test whether the two main points observed in the previous experiment – the difference in distance and specificity – can be observed in a computational setting. To do this, we used Word2Vec (Mikolov et al.,

2013), a model that is commonly used in the field of Natural Language Processing (NLP).

## 6.1 Method

Word2Vec is a neural network model that has been trained on a large amount of textual data to produce vector representations of words. The primary assumption underlying Word2Vec training is that words occurring in similar contexts have similar meanings, and should therefore be assigned similar vector representations. The similarity between the vectors, which are also referred to as *word embeddings*, is assessed as the cosine of their angle.

Models such as Word2Vec have been extremely successful in modern NLP, since they allow researchers to easily build data-driven word representations from textual corpora. More importantly, studies adopting cognitively motivated evaluation datasets have shown that word embeddings exhibit significant correlations with human performance on psycholinguistic and neurolinguistic tasks (Hollenstein et al., 2019; Mandera et al., 2017).

In this section, we present the experiments conducted via Word2Vec using the vector representations of Italian demonstratives. In particular, we were interested in measuring the semantic similarity between Italian demonstratives and pairs of words related to concepts of ‘distance’ and ‘specificity’. For our analysis, we used a recent Italian version of Word2Vec (Di Gennaro et al., 2021) that has been trained on a combination of three corpora: an April 2019 dump of the Italian Wikipedia, a series of articles extracted from the main categories of Italian Google News (World, Nation, Business, Technology, Entertainment, Sports, Science and Health), and a collection of anonymized chats between the Laila customer care chatbot and the application users,<sup>3</sup> yielding a total of 2.6 GB of raw text.

## 6.2 Experimental pairs

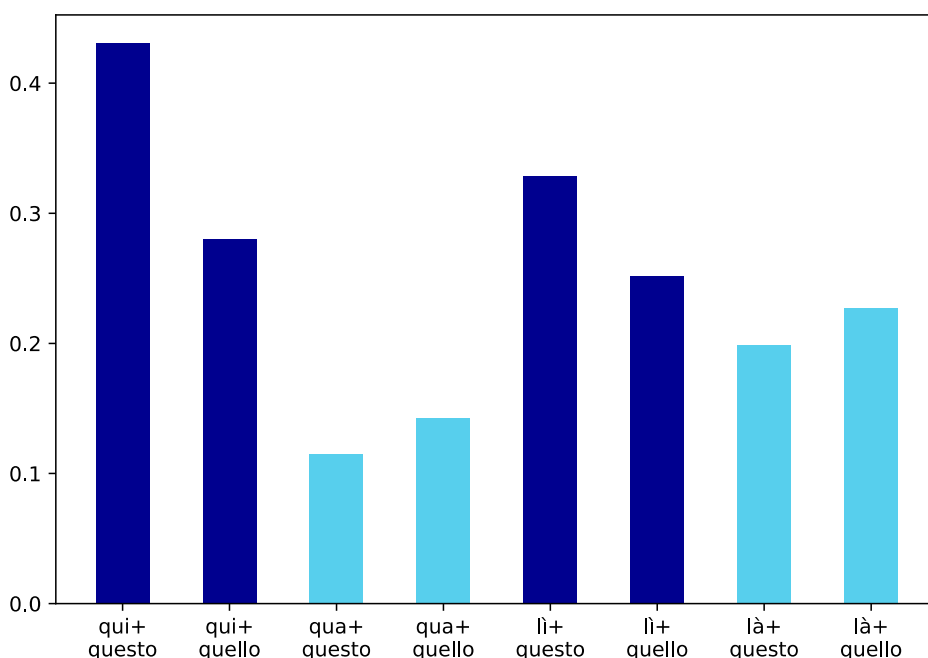
In the first part of the experiment, we selected the following three word pairs, which all relate to the concept of distance, to investigate the meaning of Italian spatial demonstratives:

- *questo* ‘this’ and *quello* ‘that’
- *vicino* ‘near’ and *lontano* ‘far’
- *io* ‘I’ and *tu* ‘you’

These pairs consist of basic words used frequently in everyday conversations. The concepts ‘this’, ‘far’, ‘I’ and ‘you’ are found in the Leipzig–Jakarta List (Tadmor, 2009), a list of basic concepts used to compile the fundamental vocabulary of different languages. Intuitively, these words can be conceived of as the reference points for distance-related meanings in the vector space: if a demonstrative refers to a location that is close to the speaker, we would expect it to be more similar to the first terms in the pairs.

---

<sup>3</sup><https://www.laila.tech/>



**Figure 1.** Word2Vec similarity between the Italian demonstratives and *questo/quello* 'this/that'.

For the last pair, it should be noted that the second person singular is also related to proximity, as the hearer is often close to the speaker (Joo, 2020, pp. 9–10). The hearer is, however, still further away from the speaker compared to the speaker themselves. Thus, *io* 'I' implies a greater degree of proximity than *tu* 'you'.

In the second part of the experiment, we tested Italian demonstratives and the following three word pairs related to specificity:

- *proprio* 'exactly' and *intorno* 'around/near'
- *preciso* 'precise' and *impreciso* 'imprecise'
- *stretto* 'narrow' and *largo* 'wide'

These three antonym pairs consist of one word designating specificity, either in an abstract sense ('exactly' and 'precise') or a physical sense ('narrow'), and another word meaning the opposite of these senses of specificity. The concept 'wide' is also found in the Leipzig–Jakarta List. While the concepts 'exactly' and '(im)precise' are not quite basic, they are fairly straightforwardly related to the concept of specificity.

### 6.3 Results and discussion

The results for the similarity between the demonstratives and the targeted word pairs were as follows. In the experiment concerning distance, Figure 1 shows that *questo* 'this' is more similar to *qui* and *lì* than *quello* 'that' is, while the reverse is true for *qua* and *là*. Figure 2 shows a similar pattern, as *vicino* 'near' is more similar to *qui*. As expected,

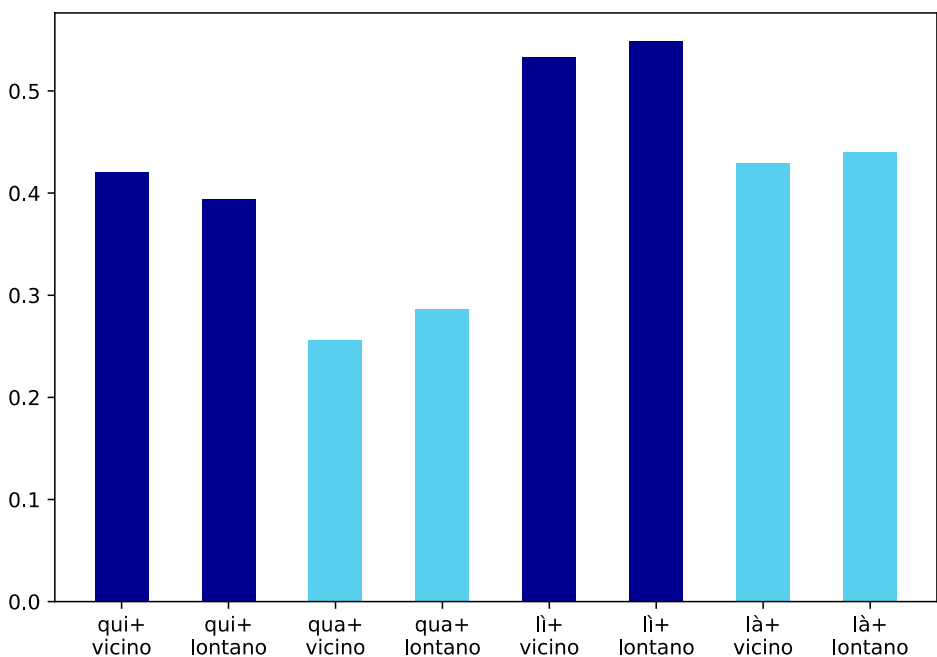


Figure 2. Word2Vec similarity between the Italian demonstratives and *vicino/lontano* ‘near/far’.

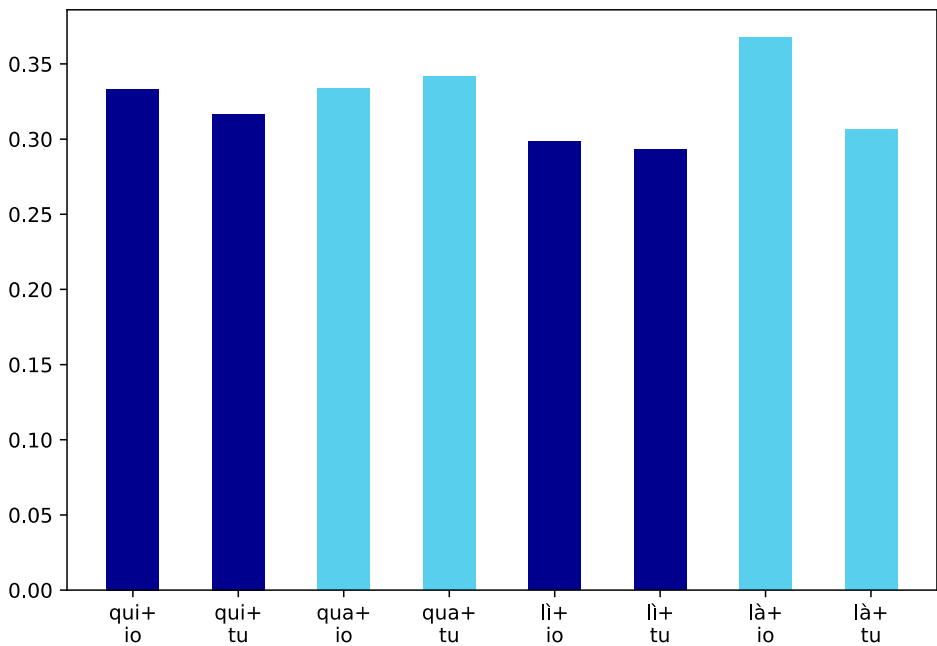
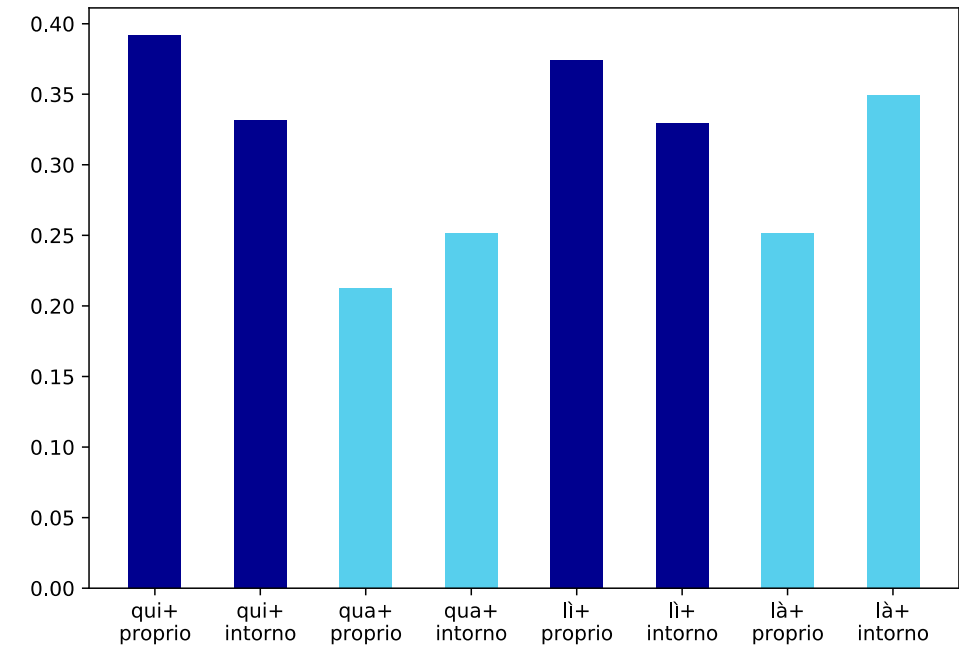
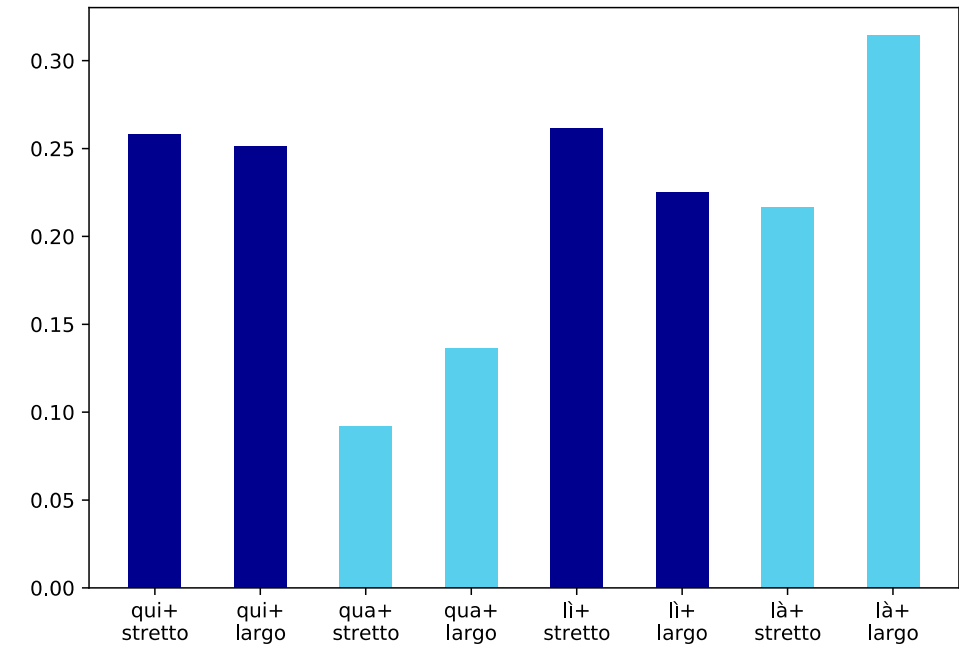


Figure 3. Word2Vec similarity between the Italian demonstratives and *io/tu* ‘I/you’.

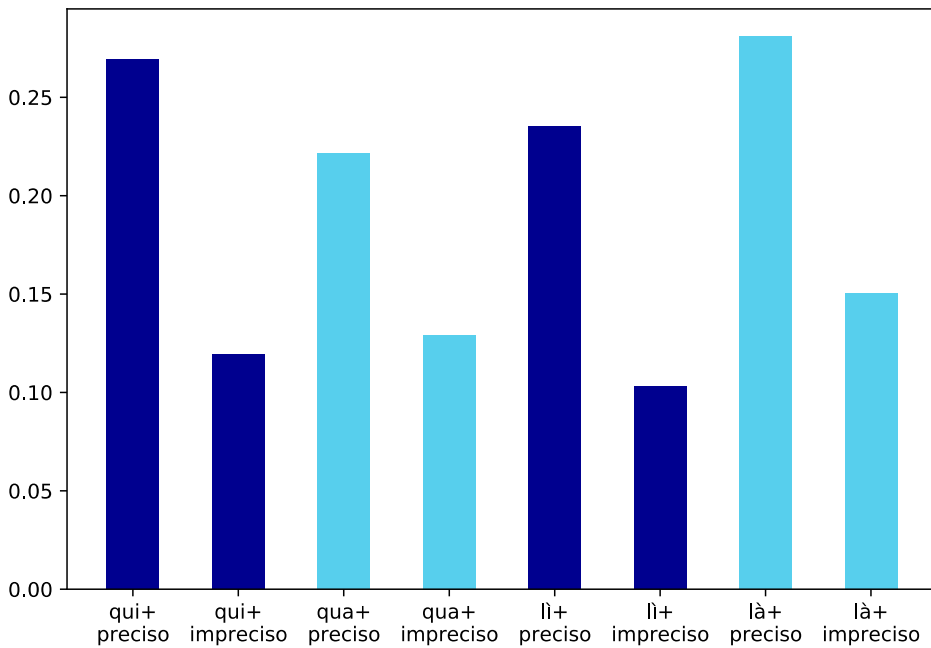
*lontano* ‘far’ is closer to *qua* and *là* than *vicino* is. However, for *lì*, *lontano* is more similar than *vicino* is, which was contrary to our expectation. Figure 3 shows that *io* ‘I’ is more similar to *qui* and *lì* than *tu* ‘you’ is, whereas the opposite is true for *qua*. *Là* shows an



**Figure 4.** Word2Vec similarity between the Italian demonstratives and *proprio/intorno* ‘exactly/around’.



**Figure 5.** Word2Vec similarity between the Italian demonstratives and *stretto/largo* ‘narrow/wide’.



**Figure 6.** Word2Vec similarity between the Italian demonstratives and *preciso/impreciso* 'precise/imprecise'.

unexpected pattern in which *io* is closer to *là* than *tu* is. Overall, although there were some exceptions, words implying proximity were generally more similar to demonstratives with *-i* than were their distal antonyms, whereas the opposite was true for demonstratives with *-a*. Thus, the Word2Vec similarity data confirmed the speakers' intuitions discussed in our survey results, namely that demonstrative pronouns with *-i* are more proximal than are those with *-a*.

In the specificity experiment concerning the similarity between Italian demonstratives and the target word pairs, Figures 4 and 5 show the expected patterns, as *qui* and *lì* are closer to *proprio* 'exactly' and *stretto* 'narrow' than to *intorno* 'around/near' and *largo* 'wide', while the reverse is true for *qua* and *là*. However, Figure 6 shows that all four demonstrative pronouns are closer to *preciso* 'precise' than to *impreciso* 'imprecise'. Despite there being two exceptions (*qua* and *là* are closer to *preciso* than to *impreciso*), the results largely point towards a pattern in which *qui* and *lì* are semantically closer to specificity, while *qua* and *là* are closer to non-specificity.

Thus, although not without exceptions, the results obtained via the Italian Word2Vec model were generally consistent with the intuitive responses retrieved from the questionnaire: *qui* and *lì* are closer to the speaker and are more specific than *qua* and *là*.

## 7. Conclusion

In this paper, we shed light on the possibility that iconicity has motivated the difference in the frequency and the semantic nuance of Italian demonstrative pronouns. By using a corpus, a survey and a vector-based experiment, our results revealed that *qui* and *là* have



been used more frequently throughout history and that they are currently being used in the more prototypical sense of 'here' (close to the speaker) and 'there' (far from speaker). We argued that iconicity could be the driving force for the two results we obtained. Although we cannot completely rule out the possibility that iconicity is only a coincidental correlation with the frequency differences and the semantic distinctions, and not the cause thereof, the results were at least in line with our hypothesis that a word that is more iconic may tend to be used more frequently and may represent a prototypical meaning of its kind.

We also cannot rule out the possibility that prototypicality caused frequency, or vice versa, instead of iconicity directly causing the two. For example, it could be the case that *qui* and *là* were used more frequently because they were close to the prototypical meanings of 'here' and 'there', since more prototypical meanings are often referred to more frequently. Conversely, it could be the case that *qui* and *là* were used more frequently, and that their frequent usage caused their meaning to change into the more prototypical meanings of 'here' and 'there'. Neither case, however, would change the fact that iconicity is the ultimate cause of both frequency and prototypicality, directly or indirectly.

The contrast in iconicity can affect domains other than frequency and prototypicality, such as grammaticalization. Strik Lievers and Miola (2018), in their study on Italian phrasal verbs including *lì* or *là* (such as *lasciare lì/là* 'to give up, lit. to leave there'), observe that phrasal verbs with *lì* are more likely to be grammaticalized and lose their deictic sense than those with *là*. This suggests that *là* is more likely to retain its original deictic meaning and unlikely to be grammaticalized into a non-deictic sense due to its iconic value.

As a single case study, this paper alone cannot confirm the claim that iconicity causes a lexeme to be used more frequently and become close to the prototype in its semantic field. However, it is a step towards this idea which, if proven to be correct, would give us a better understanding of what drives lexical survival and semantic change. We invite future studies to investigate our hypothesis in greater depth.

## Acknowledgements

We would like to thank the editor and the anonymous reviewer for their constructive comments. Our gratitude also goes to Lavinia Salicchi for distributing the survey to the participants.

## Data availability statement

The data that support the findings of this study are openly available in GitHub at: <https://github.com/ianjoo/IconicBiasInItalianSpatialDemonstratives>.

## Disclosure statement

No potential conflict of interest was reported by the authors.




## Notes on contributors

**Ian Joo** is a PhD student at the Hong Kong Polytechnic University.

**Yu-Yin Hsu** is an Assistant Professor at the Hong Kong Polytechnic University.

**Emmanuele Chersoni** is a Research Assistant Professor at the Hong Kong Polytechnic University.

## ORCID

Ian Joo  <http://orcid.org/0000-0001-7678-1227>  
 Yu-Yin Hsu  <http://orcid.org/0000-0003-4087-4995>  
 Emmanuele Chersoni  <http://orcid.org/0000-0001-8742-0451>

## References

- Benczes, R. (2020). Sound symbolism and semantic change. In K. Allan (Ed.), *Dynamics of language changes: Looking within and across languages* (pp. 253–264). Springer Singapore. [https://doi.org/10.1007/978-981-15-6430-7\\_16](https://doi.org/10.1007/978-981-15-6430-7_16).
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Blasi, D. E., Wichmann, S., Hammarström, H., Stadler, P. F., & Christiansen, M. H. (2016). Sound-meaning association biases evidenced across thousands of languages. *Proceedings of the National Academy of Sciences*, 113(39), 10818–10823. <https://doi.org/10.1073/pnas.1605782113>
- Di Gennaro, G., Buonanno, A., Di Girolamo, A., Ospedale, A., Palmieri, F. A. N., & Fedele, G. (2021). An analysis of Word2Vec for the Italian language. In A. Esposito, M. Faundez-Zanuy, F. C. Morabito, & E. Pasero (Eds.), *Progresses in artificial intelligence and neural systems* (pp. 137–146). Springer. [https://doi.org/10.1007/978-981-15-5093-5\\_13](https://doi.org/10.1007/978-981-15-5093-5_13).
- Eberhardt, M. (1940). I. A summary of some preliminary investigations of the deaf. *Psychological Monographs*, 52(1), 1–5. <https://doi.org/10.1037/h0093460>
- Hollenstein, N., de la Torre, A., Langer, N., & Zhang, C. (2019). Cognival: A framework for cognitive word embedding evaluation. *Proceedings of the 23rd Conference on Computational Natural Language Learning*, 538–549. <https://doi.org/10.18653/v1/K19-1050>.
- Imai, S. (2003). *Spatial deixis* [PhD thesis]. State University of New York at Buffalo.
- Jespersen, O. (2010). Symbolic value of the vowel i. In O. Jespersen (Ed.), *Selected writings of Otto Jespersen* (pp. 515–535). Routledge.
- Johansson, N., Anikin, A., Carling, G., & Holmer, A. (2020). The typology of sound symbolism: Defining macro-concepts via their semantic and phonetic features. *Linguistic Typology*, 24(2), 253–310. <https://doi.org/10.1515/lingty-2020-2034>
- Johansson, N., & Carling, G. (2015). The de-iconization and rebuilding of iconicity in spatial deixis: An Indo-European case study. *Acta Linguistica Hafniensia*, 47(1), 4–32. <https://doi.org/10.1080/03740463.2015.1006830>
- Johansson, N., & Zlatev, J. (2013). Motivations for sound symbolism in spatial deixis: A typological study of 101 languages. *Public Journal of Semiotics*, 5(1), 3–20. <https://doi.org/10.37693/pjos.2013.5.9668>
- Joo, I. (2020). Phonosemantic biases found in Leipzig–Jakarta lists of 66 languages. *Linguistic Typology*, 24(1), 1–12. <https://doi.org/10.1515/lingty-2019-0030>
- Knoeferle, K., Li, J., Maggioni, E., & Spence, C. (2017). What drives sound symbolism? Different acoustic cues underlie sound-size and sound-shape mappings. *Scientific Reports*, 7(1), 1–11. <https://doi.org/10.1038/s41598-017-05965-y>
- Lockwood, G., & Dingemanse, M. (2015). Iconicity in the lab: A review of behavioral, developmental, and neuroimaging research into sound-symbolism. *Frontiers in Psychology*, 6, 1246. <https://doi.org/10.3389/fpsyg.2015.01246>
- Maglio, S. J., Rabaglia, C. D., Feder, M. A., Krehm, M., & Trope, Y. (2014). Vowel sounds in words affect mental construal and shift preferences for targets. *Journal of Experimental Psychology: General*, 143(3), 1082–1096. <https://doi.org/10.1037/a0035543>
- Maiden, M., & Robustelli, C. (2007). *A reference grammar of modern Italian* (2nd ed.). Routledge.
- Mandera, P., Keuleers, E., & Brysbaert, M. (2017). Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *Journal of Memory and Language*, 92, 57–78. <https://doi.org/10.1016/j.jml.2016.04.001>

- Masuda, K. (2007). The physical basis for phonological iconicity. In E. Tabakowska, C. Ljungberg, & O. Fischer (Eds.), *Insistent images* (Vol. 5, pp. 57–71). John Benjamins Publishing Company.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv Preprint arXiv:1301.3781*.
- Nobile, L. (2011). Words in the mirror: Analysing the sensorimotor interface between phonetics and semantics in Italian. In P. Michelucci, O. Fischer, & C. Ljungberg (Eds.), *Semblance and signification* (Vol. 10, pp. 101–131). John Benjamins Publishing.
- Rabaglia, C. D., Maglio, S. J., Krehm, M., Seok, J. H., & Trope, Y. (2016). The sound of distance. *Cognition*, 152, 141–149. <https://doi.org/10.1016/j.cognition.2016.04.001>
- Sakamoto, M., & Watanabe, J. (2018). Bouba/kiki in touch: Associations between tactile perceptual qualities and Japanese phonemes. *Frontiers in Psychology*, 9, 295. <https://doi.org/10.3389/fpsyg.2018.00295>
- Sapir, B. Y. E. (1929). A study in phonetic symbolism. *Journal of Experimental Psychology*, 12(132), 225–239. <https://doi.org/10.1037/h0070931>
- Shinohara, K., & Kawahara, S. (2010). A cross-linguistic study of sound symbolism: The images of size. *Annual Meeting of the Berkeley Linguistics Society*, 36(1), 396–410. <https://doi.org/10.3765/bls.v36i1.3926>
- Strik Lievers, F., & Miola, E. (2018). Lì ('there') and là ('over there') in Italian phrasal verbs. *Archivio Glottologico Italiano, CIII*, (1), 75–97. <https://doi.org/10.1400/273483>
- Tadmor, U. (2009). Loanwords in the world's languages: Findings and results. In M. Haspelmath, & U. Tadmor (Eds.), *Loanwords in the world's languages: A comparative handbook* (pp. 55–75). De Gruyter Mouton.
- Tanz, C. (1971). Sound symbolism in words relating to proximity and distance. *Language and Speech*, 14(3), 266–276. <https://doi.org/10.1177/002383097101400307>
- Ullan, R. (1978). Size-sound symbolism. In J. Greenberg (Ed.), *Universals of human language* (Vol. 2, pp. 527–568). Stanford University Press.
- Vainio, L. (2021). Magnitude sound symbolism influences vowel production. *Journal of Memory and Language*, 118, 104213. <https://doi.org/10.1016/j.jml.2020.104213>
- Woodworth, N. L. (1991). Sound symbolism in proximal and distal forms. *Linguistics*, 29(2), 273–300. <https://doi.org/10.1515/ling.1991.29.2.273>