UiO : University of Oslo

Writing systems, distributional properties of writing, and sound symbolic associations.

Candidate: Tomasz Bauc

Master's Thesis in Linguistics: LING4190

Programme: Linguistics

Credits: 60

Semester: Summer 2023

Department of Linguistics and Scandinavian Studies (ILN)

Abstract

This study investigated the relationship between writing systems and sound symbolic associations. This was achieved by evaluating the results of an experiment, which measured distributional properties of participants' writing and the congruence of their responses with the previously established sound symbolic association patterns. The participants were grouped by the primary writing system of their first language (L1) to allow for the cross-script comparison. Three groups of N = 20 participants were included, one for users of each: Latin scripts, Simplified Chinese characters and Thai script.

The experiment conducted involved two tasks. The participants first completed a task involving writing down sounds of non-word audio stimuli using their respective writing systems. The second task followed an established model of matching images to audio stimuli. The individual results obtained were analysed to derive two distributional metrics: the average token count measuring the number of written representations per speech sound (for Latin and Thai scripts) or speech sound sequence (for Simplified Chinese characters) and the entropy of the distribution of those tokens. These two metrics were then tested for correlation with scores in the sound symbolic association task.

The statistical analysis of the results confirmed negative correlations of both average token count and calculated entropy with scores obtained in the sound symbolic association task, both with the strength of approximately $r = -0.52$ and the significance $p = 0.02$, in the Latin script group *only*. Additionally, the negative trend of increase in distributional properties' scores accompanying decrease in sound symbolic associations test scores was observed in cros-group comparison.

Taken together, these results confirm the previously observed association between the writing system itself and sound symbolism, and provide a new insight into the possible (conditionally) correlated factor of distributional properties. The study concludes that the chosen metrics, along with other L1- and usage-based factors constitute good candidates for further examination in terms of relationship with sound symbolism and broader iconicity.

Foreword

This paper is a final product of years of study and months of independent work. The project reported herein has been an incredible, if at times tough, learning experience as well as a first taste of what the author hopes to be a future career. The subject matter investigated here is a union of author's academic and personal interests, combining elements of experimental research into cognition, phonetics & phonology, quantitative methods and mathematical models with personal interests in writing systems and asian languages. The project has been a long, hard and at times lonely endeavour, and the author wishes to express deep gratitude to all who have supported him on this journey. Special thanks go out to my supervisor, Timo B. Roettger for invaluable advice and candid feedback along the entire way, mr. Ed for technical advice, Sara for extensive academic advice and help with review, แต๊ว for huge help with participant recruitment, and LG for help with testing, recruitment and support throughout the process.

Table of Contents

1. Introduction

The phenomenon of sound symbolism is well established within linguistic, and psycholinguistic in particular, research (Hinton et al., 1995). The core idea is that of a direct link between acoustic properties of speech sounds and non-acoustic properties of meanings they are associated with. The most iconic example of this are the “maluma-takete” (Köhler, 1947) and “bouba-kiki” (Ramachandran, 2001) experiments demonstrating higher than chance probability of round visual shapes being associated with pseudo-words made up of rounded and labial consonants and vowels (/b/,/m/,/u/) and spiky shapes with pseudo-words containing voiceless, post-alveolar and velar, consonant and front vowel sounds (/t/,/k/,/e/,/i/). An important question on this matter has been that of the influence of vision (Bottini et al., 2019) and visual appearance of graphemes representing these sounds on this observed association (Cuskley et al., 2017). While in the latin-based scripts this similarity is readily observable, it is much less obvious in other writing systems.

This project attempts to evaluate two examples of such writing systems, namely Thai script and Simplified Chinese characters and compare them to Latin script. This was achieved through evaluation of the *bijectivity* (the 1 to 1 correspondence of sound and grapheme) in their users’ writing. This was measured by distributional properties of token count and entropy, and the correlation of these factors to scores in sound symbolism associations tests attained by their users, as well as through direct comparison of these scores achieved by groups of users of different scripts.

1.1. Iconicity

The central concept for this project, iconicity, can be most broadly described as any non-arbitrary relationship, or resemblance (Dingemanse et al., 2020), between linguistic form and meaning. Iconicity can be direct, as when a logographic symbol like a hieroglyph resembles the meaning it expresses by direct visual similarity, or conceptual like when an image of an arrow is intuitively associated with the concept of direction. Iconicity is present wherever linguistic features are associated not by a rule of a language, but rather by *natural* similarity.

Iconicity has traditionally been perceived in opposition to arbitrariness or *convention* in language, however more recent research suggests its role is instead better understood as complementary, i.e. instead of being two competing forces, iconicity and arbitrariness appear to have different, non-conflicting functions with arbitrariness allowing for conceptual complexity and iconicity providing a bridge between the abstract linguistic structures and human sensory experience. (Perniss & Vigliocco, 2014).

Iconicity has been noted already by classical scholars. Plato's Cratylus (~400 BC) features Socrates discussing it at length and expressing fondness for *natural* (iconic) language forms. Beyond some early interest, however, iconicity has attracted little mainstream scientific inquiry throughout the post-classical and early modern times. Some of the strongest opposition to iconicity's importance comes in the form of one of the most fundamental assertions of the influential Ferdinand de Saussure, who proposed language's arbitrariness as his very first axiom (Saussure, 1966). Consequently, much more effort has been dedicated to the study of the *symbolic* (arbitrary) phenomena and features in language, with iconicity recognised to exist in some readily observable instances (e.g. onomatopoeic expressions), yet prompting little rigorous investigation into its role or nature.

In recent times, with new interest in multimodal forms of communication (particularly sign languages, but also, for example, emoji) within linguistic research and new theoretical approaches (e.g. functional linguistics) emerging, the role of iconicity, as both latent and active element of language, has been revisited by multiple authors. Croft (2001) discusses the matter at length in one of the chapters of his *Radical Construction Grammar*, concluding its role is to motivate the emergence of fundamental (symbolic) pairings of form and meaning and selection between competing constructions. From the perspective of general psychology, a study by Perniss et al. (2010) suggests iconicity helps in "reducing the gap between linguistic form and conceptual representation to allow the language system to "hook up" to motor, perceptual, and affective experience". Thus the relationship between iconicity and symbolism increasingly appears to be cooperative rather than competitive.

Direct relations between form and meaning can be found across all levels of language. Some easy examples of iconicity can be observed in pragmatic and discourse devices. Speakers commonly adjust pitch and speech patterns to imitate participants, actions or events reported, and even copy their gestures and behaviour. Sign languages rely heavily on iconicity, with entire vocabularies built upon similarity between the gestures and the meaning they refer to. In semantics, a range of sound-imitating words can be identified, with different amounts of iconicity employed; from half-symbolic like "yikes", "pow" or "kazaam" through semi-universal (of clear similarity in form) across languages like animal sounds "meow" or "moo" to near-all language encompassing words like "mama" and "papa" (Nöth, 2001), which have been suggested to derive from the earliest articulation attempts of human babies conventionalised into semantic meaning - a powerful reminder of just how central iconicity can be to language.

Beyond onomatopoeic expressions and other directly imitative devices, some much less obvious examples of iconicity have been demonstrated. At discourse level, iconicity of sequence, i.e. the correlation between the order of constituents and the order of events they describe, has been observed (Diessel, 2008). Syntactic constructions can be shown to follow similar pattern (c.f. tense iconicity in Croft, 2001, p. 336). They have also been shown to involve other flavours of iconicity, e.g. of proximity, where the distance between the constituents appears to correspond to the distance (spatial and/or conceptual) between the objects etc, which they refer to. On the morphology level, iconicity of complexity has been suggested, where more complex meaning tends to be expressed by more complex morphological structures. For example, Langendonck (in Geeraerts & Cuyckens, 2007) reports observations of unmarked verb (present tense) and noun (indefinite) forms showing a tendency for representations consisting of fewer morphemes than the marked ones.

1.1.1. Sound symbolism

Sound symbolism refers to an *iconic* relationship between phonemes themselves, or their phonetic or phonological qualities and features, on one hand and aspects of semantic meaning on the other. In the broadest sense, sound symbolism refers to any non-arbitrary relationship between sound and meaning.

The subject has been discussed already in classical times. One of the earlier examples comes from Plato's *Cratylus*, which, beyond a broader discussion of iconicity notes how certain letters represent sounds (the distinction between letter and sound is less explicit and they sometimes are used interchangeably) suited for specific meaning (Klensovsak et al., 2011). Even before Plato, the association of phonemes with meaning has been discussed in the ancient Indian Upanishads (Kawahara et al., 2020) and some ancient Japanese texts.

Ultimately, sound symbolism has gained mainstream recognition in the XX-th century with a series of experimental investigations demonstrating interesting outcomes. A prominent example of this is Edward Sapir's *A study in phonetic symbolism* (1929). In the study, the author reports experiments showing "a word with the vowel a is likely to symbolize something larger than a similar word with the vowel i, or e, or e, or ä" (the mil/mal effect), suggesting a correlation between a phonetic vowel quality and the concept of size. Around similar time, another author, the German psychologist Wolfgang Köhler, reported in his *Gestalt psychology* (1947), observations of a relationship between specific consonant qualities and the mental associations of smooth and spiky shapes. In his experiments, the author found participants consistently match a shape with round edges to the dummy word "maluma" and a spiky one to the label "takete". Tanz (1971) collated examples showing "universal sound symbolism in a constellation of words in natural languages relating to proximity and distance" both physical and temporal. Ramachandran & Hubbard (2001) used similar shapes with the labels "bouba" and "kiki", and this convention has been used in the follow up studies examining the different aspects of this effect.

Further research has produced interesting insights into different aspects of the now famous "bouba-kiki effect". Blasi et al. (2016) ran a large scale cross-linguistic analysis and report an analysis of "word lists covering nearly two-thirds of the world's languages" finding "considerable proportion of 100 basic vocabulary items carry strong associations with specific kinds of human speech sounds, occurring persistently across continents and linguistic lineages". Shinohara & Kawahara (2010) ran an experiment testing how "images of size (small or large) are affected by three phonetic factors: the height of vowels, the backness of vowels, and voicing in obstruents" with speakers of Chinese, English, Japanese, and Korean, finding that "these three factors contribute to the images of size, with only a few exceptions". Sidhu & Pexman (2015) tested "whether or not the Bouba/Kiki effect extends to existing lexical stimuli; in particular, real first names". They found a range of associations linking "femaleness and round shapes, and maleness and sharp shapes" and documented that "adjectives previously judged to be either descriptive of a figuratively 'round' or a 'sharp' personality were associated with names containing either round- or sharp-sounding phonemes, respectively". The authors conclude that "sound symbolic associations extend to existing lexical stimuli", a new example of sound symbolic mapping. A large scale study by Wichman et al. (2010) examined phonetic properties of a "selection of basic vocabulary in nearly one half of the world's languages" finding "commonalities among sound shapes for words referring to same concepts" demonstrating the prevalence of sound symbolic associations in world's languages. Another large scale study by Haynie et al. (2014) involved a "statistical investigation of the evidence for several common patterns of sound symbolism, using data from a sample of 120 [Australian] languages" finding "evidence for the expected associations" noting however the sound symbolic patterns being "more complicated than expected".

Another line of inquiry aimed to deconstruct the physical basis of sound symbolism. Westbury (2005) demonstrated, by measuring reaction times in an interference task, that the bouba-kiki effect appears to be "pre-semantic". Fort & Schwartz (2022) attempted to shed light on the origins of the phenomenon "by combining mathematical findings largely unknown in the field, with computational models and novel experimental evidence". They conclude that "this effect relies on two acoustic cues: spectral balance and temporal continuity" and that "it is not speech-specific but rather rooted in physical properties of objects, creating audiovisual regularities in the environment", and that the source of the effect lies in the physical (acoustic) properties of the speech sounds. Similarly, Knoeferle et al. (2017) ran a series of experiments attempting to detect "whether different sets of acoustic cues predict size and shape symbolism, respectively". Their analysis is focused on vowel formants (characteristic frequencies), and found that "Visual size judgments were predicted by vowel formant F1 in combination with F2, and by vowel duration. Visual shape judgments were, however, predicted by formants F2 and F3." They also note that sound symbolism is a "process that is not based merely on broad categorical contrasts, such as round/unround and front/back vowels. Rather, individuals seem to base their sound-symbolic judgments on specific sets of acoustic cues, extracted from speech sounds, which vary across judgement dimensions". Thus different acoustic components of speech sounds appear to correlate with meaning of different physical properties. In a broad review, Sidhu & Pexman (2018) point to five potential types of factors involved in the occurrence of sound symbolism: "statistical co-occurrence between phonetic features and associated stimuli in the environment, a shared property among phonetic features and stimuli; neural factors; species-general, evolved associations; and patterns extracted from language".

Other studies focused on the relevance of sound symbolism for language acquisition. Monaghan et al. (2012) found that "sound symbolism resulted in an advantage for learning categories of sound-shape mappings but did not assist in learning individual word meanings". Imai & Kita (2014) suggest a more extensive role of sound symbolism in early language learning by infants. They propose a "bootstrapping hypothesis" positing a cohesion-producing function for sound symbolism, easing the process of language acquisition by providing cross-sensory cohesion, and suggest a similar role in early human language emergence.

Yet another direction of research has been to determine the universality of the phenomenon across speakers of different backgrounds and cultures. A recent one, by Cwiek et al. (2022) has demonstrated that despite differences, the outcome of the bouba-kiki effect experiments remains robust for speakers of a wide variety of languages using multiple different writing systems.

1.2. Writing and orthography

As linguistic studies shifted from the traditional prescriptive approach to the more recently common descriptive attitude, written language has become a less popular object of interest. Despite, what Linell (2005) terms the Written Language Bias in Linguistics (a bias certainly recognisable in the design of the present study), that is the systematic bias in language analysis resulting from the structure and mechanics of writing, the *primacy of speech* (Lyons, 1981), i.e. the view of speech as the primary mode of language, and writing as a technical mode of delivery, has resulted in a sidelining of writing-related phenomena within the mainstream linguistic research.

Following a spike in interest in sound symbolism sparked by Ramachandran & Hubbard's work, a question has emerged of the influence of writing on the observed effect.

Cuskley et al. (2019) report that "letter curvature is strong enough to significantly influence word–shape associations even in auditory tasks, where written word forms are never presented to participants" and conclude that "many previous investigations of the bouba–kiki effect may not have given appropriate consideration or weight to the influence of orthography among literate subjects". Drijvers et al. (2015) documented a difference in sound symbolic associations strength in individuals affected by dyslexia, thus demonstrating sound symbolic associations' vulnerability to processes that disrupt cross-modal abstraction. Bottini et al. (2019) have demonstrated significant sound-shape correspondence effect in early blind individuals (matching objects round or spiky in shape to the touch), while also showing that, in sighted individuals, the shape of letters themselves is associated with the shape of objects (words written with spiky letters tend to "sound spiky" etc).

A study by Westbury et al. (2018) examined participants' judgements on the suitability of randomly generated non-words as representants of 18 semantic categories, interestingly they found that "some previously unsuspected dimensions show strong symbolic effects; and that features, phonemes, and letters may all contribute to sound symbolism", suggesting a link to writing.

Another interesting experiment was conducted by Cwiek et al. (2022) as part of their cross-linguistic study. To demonstrate the universality of the bouba-kiki effect, the authors ran a large scale study testing sound symbolic associations of speakers of a multitude languages, including users of multiple different scripts, finding that majority of participant groups (18 out of 25) matched the two sounds to the round and spiky images in a manner congruent with the sound symbolic assumption (bouba - round, kiki - spikey) with a higher-than-chance probability. In an attempt to control their results for the influence of orthography, they ran a survey in which participants were asked to match the written forms of "bouba" and "kiki" in multiple scripts to the spiky and rounded images. Despite finding higher-than-average scores of congruence between the written forms (of nine different scripts) and the images used, they conclude that their analysis "does not indicate a reliable effect of the orthography bias measure". The current study attempts to replicate these results along the investigation of its narrowly specified object of interest.

2. Motivation

Inspired in part by the orthography bias sub-study in Cwiek et al. (2022), this project aims to put the influence of writing systems on sound symbolic associations front and centre. One functional property, which appears to vary between different scripts, is the degree of *bijectivity*, or one-to-one-ness, in their sound-to-grapheme mapping. Furthermore, this property can also vary between different languages using the same script. For example, the sound /k/ can be represented by only the letter "k" in Norwegian or Polish, by letters and letter sequences "c", "q" and "ch" in Italian, and by all four, and more (e.g. "ck"), in English. These differences can be traces of historic sound shifts, splits and mergers, which languages' orthographies have not yet corrected for, or a result of mismatch in phonetic inventories between the language the writing system is adopted for and the one it was designed for, etc.

Reading and writing are amongst the most common everyday tasks involving conversion between sound and shape. Thus a question arises whether properties of an individual's writing correlate to how strongly they associate specific sounds with specific graphic symbols and their shape, and whether the strength of this association is representative of, or even mediates, the strength of their sound-to-shape associations in general. This study aims to explore the interaction between these two factors through measurement of two distributional properties: average count of different graphic representations (tokens) for the same sound and the entropy of their distribution, and correlating them with sound symbolic association test scores.

The underlying question, which this study tests a specific instance of, is whether such associations are transferable or context specific, i.e. whether the association between speech sound and written symbol, or between an acoustic property and a shape, is generalised onto broader cognitive classifications and to what degree. The level, or degree, of such generalisation cannot be quantified based on the results (nor is it aimed for) but they can provide a hint towards understanding how integrated our linguistic knowledge is into the broader framework of cognitive processes.

Furthermore, the study attempts to investigate in more detail the specific cases of two non-Latin based scripts of different structure and functionality in the context of sound symbolic associations, which has been dominated by Western perspective of the Latin alphabet as the "default" writing system.

Finally, an important goal of this project is to create a technical framework for future investigation of writing-related phenomena with modular design allowing for ease of replication, adjustment and expansion, with all tools used published on OSF (http://osf.io/znc5u/) in the spirit of open science.

3. Methodology

The study employed online survey as its primary data collection tool. The data collected was analysed to produce individual scores of two types: one for the degree of bijectivity in participant's writing (measured by average token count and entropy) and another for sound symbolic association. These scores were then tested for correlation using statistical tools. A group-wise analysis of average scores obtained followed for a broader overview of the results.

3.1. Hypothesis

The narrowly defined hypothesis of this study has been that of existence of a correlation between:

(a) writing system used, and;
(b) distributional properties of an individual's writing measured by:
  (1) average token count and;
  (2) entropy

and:
(c) strength of sound symbolic associations.

Of these, the correlation between (a) and (c) was assessed through direct comparison of results scored by different participant groups. The correlation between (b) and (c) was assessed using statistical tests of correlation (Pearson's and Spearman's depending on the properties of data).

3.2. Global design features

Certain design decisions affected the entirety of the project. The main motivation for these has been an attempt to balance resource limitations on one hand and achieving two main aims on the other. The first aim has been to design a methodology that is easy to replicate and/or adjust to include a broader range of input data, participant groups, etc, while the other has been to expand and improve on previous models of experimental inquiry into the subject of sound symbolism.

3.2.1. Scope

The focus of this study's design is similar to that of the classical experiments in the area. However, the design departs from the two-label, two-image approach commonly employed, in favour of a set of total eight speech sounds as main focus, and total four images illustrating the two visual features. The experiments were performed by two groups of participants of non-latin script L1-s, i.e. Thai (using Thai writing) and Chinese (using simplified Chinese characters).

The choice of scope has been motivated by generalisability considerations. A commonly used model, where participants match one of two audio stimuli to one of two images, e.g. the sounds /buba/ and /kiki/ to one image of round shape and one image of spiky shape, provides results highly confident for its selected stimuli, but it's less convincingly indicative of potential results beyond them. While high-confidence results are always desirable, the main idea of sound symbolism is that of cognitive associations between certain types of sounds, and/or their phonetic and phonological features on one hand, and mental representations of certain types of shapes on the other, rather than between the specific non-words and specific images. In other words, the sound symbolism line of inquiry seeks to describe a relationship between internalised, potentially abstract, acoustic and visual features, not only their select specific instances.

In principle, an "ideal" study would involve a broader range of the most common and universal speech sounds and visual shape types, however the scale of such study was beyond the resource limitations of this one-man project. An analogous logic guided the choice of writing systems included in the study. The decision to investigate non-latin script users has been an attempt to account for a previously sidelined factor, non-inclusion of more has been due to resource considerations, i.e. the workload involved was beyond the capacity of the author alone.

3.2.2. Writing systems

To achieve a degree of generalisability, three writing systems have been selected for analysis in this study: Latin-based scripts, Thai script and Simplified Chinese characters. Each comes with its own structural and conceptual characteristics. In particular each has a different level of nominal bijectivity, i.e. different constraints on the number of possible representations of the same sound.

An important immediate observation here is that the latin characters used to encode the sounds selected for the study show alignment with the visual features investigated. Not necessarily representative of the entire Latin alphabet, the “round” sounds /b, m, o, u/ are represented by visually round graphs, while the “spiky” sounds /k, t, i/, and to some extent /e/ (although it also has round features), show a tendency for pointiness. This can be speculated to introduce a bias for mental associations not generalisable to the entirety of Latin alphabet. Additionally, similar visual alignment does not hold for Thai or Chinese writing systems.

3.2.2.1. Latin scripts

Likely the most thoroughly studied writing system, and in particular in the context of orthography, the Latin alphabet has been demonstrated to feature certain traces of sound symbolism. This can be readily seen in the case of the bouba-kiki pair. Every single grapheme in the (latin) written form associated with the round shape (“bouba”) has round elements, and every single grapheme in the form associated with the spiky shape (“kiki”) has some spiky features. For this project, another relevant feature of this writing system is its widespread usage. Consequently, it could influence the mental associations of all participants of the experiments (as all were at least communicative in English and hence familiar with the writing).

By design, the Latin alphabet was meant to be entirely bijective, i.e. each grapheme was originally meant to represent only one sound, and each sound is meant to be encoded by only one grapheme. Because of historical phonemic shifts, splits and mergers this is not entirely the case in practice. Additionally, as the script was adopted to new languages, with phonologies different from that of Latin, the bijectivity has been diluted further and further. Today, many languages employing latin writing will feature multiple articulatory realisations of the same grapheme and/or multiple ways to encode the same sounds. In English for example, the sound /u/ can be encoded in a variety of ways (soup, suit, root, put, etc.) while the grapheme “u” can be used to encode different sounds (put, cut, mute).

3.2.2.2. Thai script

Across the majority of the approximately 70 million population of Thailand (Eberhard et al., 2023), and minor populations abroad, the Thai script is the primary writing system used, and the official written form in the country. Thai script is visually similar to, yet not mutually intelligible with, Lao and Khmer scripts, having evolved from an older form of the latter. The script encodes all consonant sounds of modern Thai phonology with one or more symbols, and all vowels, usually with multiple symbols and in certain contexts none (implicit). Notably, of the sounds used in this study, /m/ is the only one with but a single written form.

Of the writing systems featured in this project, Thai is the “middle-of-the-road” one. Deriving ultimately from the same roots as the latin script, it can be said to be in similar distance from ideographic representations. While Thai writing is designed to encode *sound*, like latin script, the range of characters available to encode it is at times much broader. It does not, however, reach the sheer multitude of representations of the same sound present in Chinese characters.

Because of phonological rules in Thai affecting the phonetic qualities of word-initial and -final consonants, and a set of orthographic rules giving finals different written forms, this study was developed to focus on consonant sounds in medial position. Taking this constraint into account, the written representations of the eight sounds used in this study, and sounds close to them present in the Thai phonetic inventory, are summarised in Table 3.1:

sound	grapheme(s)	notes
/k/ $/k^h/$	ก ข ค ฆ	
/d/ /t/ $/t^h/$	ฎ ด ฏ ต ฐ ฑ ฒ ถ ท ธ	
/b/ /p/	บ ป	
/m/	ม	
/e/ /e:/ /ɛ/ /ɛ:/	เcะ เc แcะ แc	“c” denotes a consonant grapheme for the sound preceding the vowel /o/ can also be implicit (not realised in writing)
/i/ /i:/	cิ cี	
/o/ /o:/ /ɔ/ /ɔ:/	โcะ โc เcาะ cอ	
/u/ /u:/	cุ cู	

Table 3.1 Thai script representations of the sounds used in the study.

While not all of the Thai graphemes above will, in practice, be used an equal amount in modern writing, all are present in the writing system, while a couple more are only used in literary, historical or religious texts. Consequently, the table includes all graphemes that can plausibly be expected to be used in transcriptions of the sounds of interest to the study. It also illustrates the ambiguity of the Thai writing system in terms of sound-to-grapheme correspondence. Notably the written forms for voiced sounds /b/ and /d/ are also included, as these can be used to denote the /p/ and /t/ sounds, respectively, albeit normally only in word-final position. This variety of written representations increases even further in strictly-final word positions. As this study did not include stimuli with word-final consonant sounds, those are, however, of less relevance for the project.

3.2.2.3. Simplified Chinese characters

Across the 1.35 billion (Eberhard et al., 2023) speakers of Sinitic (Chinese) languages, two varieties of a single writing system are used. The Traditional Chinese characters have evolved from the ancient oracle bone script used more than 3 thousand years ago. As part of efforts to promote literacy in China, the Simplified Chinese characters have been developed in the XX-th century (1950-60s) and have become the dominant written form across a majority of the Sinitic speakers. The simplification had the general aim of reducing the number of *strokes* (individual lines each character is written by hand with) and thus the detail level of the characters. An example of differences is illustrated in table 3.2:

Traditional Chinese	媽媽現在頭疼
Simplified Chinese	妈妈现在头疼
phonetic transcription (Standard Mandarin)	māmā xiànzài tóuténg
meaning transcription	mother now head.pain
meaning	mother is having a headache

Table 3.2 Comparison of simplified and traditional Chinese characters.

Notably 在 and 疼 do not vary, 媽 is straightforwardly reduced to 妈, while 頭 is transformed significantly into 头. Whereas "simple" characters (those written with few strokes) are commonly identical between the two systems, other "complex" ones are transformed to the point of no longer being legible for users of Traditional Chinese writing without prior training. Conversely, users of Simplified Chinese characters can recognise the Traditional characters, which have undergone minimal or no simplification with ease, but generally require training to read and write others. The simplified form, along with its users, is the object of interest in this study.

An important feature of Chinese writing is that the symbols encode the meaning of monosyllabic words. Where a Chinese word consists of two or more syllables, each of those can also be analysed as an independent word with either semantic or grammatical meaning. A common occurrence is two-syllable words made up of the same word repeated twice (for example: 妈妈 /māmā/ "mom" or 爸爸 /bàba/ "dad") the double-syllable word is then usually of the exact same meaning as the component word (妈 and 爸 in previous example). Multi-syllable words can also be made up of different characters, often with seemingly unrelated meaning, for example, in: 马铃薯 "potato", 马 translates to "horse", 铃 means "bell" and 薯 means "yam". This type of words can in effect be analysed as compounds with meaning ranging from direct composition of constituents, through semi-metaphorical to fully conventionalised and unrelated to that of each of its "building blocks". Notably, while the meaning of some characters can be grammatical (for example 的 /de/ identifies the preceding word as an owner/possessor, 了 /le/ is a past tense marker, 吗 /ma/ marks interrogative forms, etc.) or emotive (啦 /la/, 咯 /lo/, 啊 /a/, and others, change the mood/tone of the sentence to imperative, pleading, surprised, etc.) or in other ways abstract, every character *does* carry meaning. This is in stark contrast to, for example, the Latin alphabet, which allows its users to denote any combination of sounds, regardless of presence or absence of semantic or other meaning in its segments or entirety.

A related unique (within this study) feature of Chinese writing is that it does not aim to encode sound, but meaning. This produces two significant effects relevant for this study. First, there is a broad range of homophones, each with their own character, encoded without any trace of apparent similarity in their written forms. For example the sound /lì/ can be encoded by: 力 ("force"), 利 ("profit"), 立 ("to stand"), 丽 ("pretty"), 例 ("example") and multiple others. The characters can additionally have multiple meanings. Consequently the inventory size of Chinese characters ranges in tens of thousands. At the same time, the characters can be read differently by speakers of different Chinese languages, and even Japanese and Korean, where Chinese characters have been adopted to local languages.

Of the three scripts included in this project, Chinese writing is arguably the most ideographic. The characters are derived from symbols imitating the meanings they convey, and while both of the other writing systems in the study historically derive from ideographic scripts, Chinese is the only one retaining today clearly identifiable similarities between the visual shapes and meaning. This may not be of direct consequence for the subject matter of this project but there is at least one possible indirect effect of relevance. A question arises whether there is a qualitative difference between the processing of, at least partially, iconic writing system and an apparently fully abstract one.

The cumulative effect of these features of Chinese writing could be a greater cognitive separation between the sound of the language and the visual shape of its writing. The results of the present study could be interpreted as a clue to this expectation's accuracy.

3.2.3. Synchronic approach

Another important design decision, particularly in the context of sound-to-graph investigations, has been to approach the subject synchronically. While taking into consideration the changing nature of the relationship between speech and writing within each language community would obviously be ideal (and at least theoretically possible), such extension would call for a significant increase in the scope of the study. On the other hand, the majority of inquiries into the subject have historically been synchronic, and have produced meaningful results.

3.2.4. Data not collected

A range of information potentially available has been intentionally left uncollected. The majority of such information has simply fallen beyond the scope of the study, and had been deemed excessive to collect or store.

3.2.4.1. Participant details

The only information collected about the participants themselves was that acquired by pre-experiment survey of their first language (L1) and knowledge of other languages. This is notable as a range of socio-economic factors could theoretically have an influence. For example, due to such factors, individuals may have more or less time, ability, habit, opportunity, motivation etc. to devote to reading, which could lead to different strengths in writing-related mental associations.

The decision to not collect, or analyse, this type of data has been motivated mostly by resource limitations, but also by ethical considerations. As this study has more phonetic-cognitive focus, the socio-economic data has been evaluated to be of less priority, and not worth the risks inherent in its collection and storage, or the cost of mitigation of those risks.

3.2.4.2. Metadata

While some information about the circumstances of participants' performance of the experimental tasks has been *observed*, as some participated in author's side-by-side presence, none of the data about time, place, device and similar circumstances has been recorded or analysed, and (encrypted) timestamps etc. collected by the pcibex platform have been removed in the process of data extraction and preparation for analysis. Notably, the participants did *not* perform the experimental tasks in a standardised or uniform environment beyond the mode of delivery (online survey).

3.2.4.3. Tone

Both Thai and Chinese languages are predominantly tonal. This aspect of phonology was, however, disregarded in the process of stimuli selection and construction. It is therefore an open question how participants perceived and internalised the tone in audio recordings used as stimuli, or what effect it may have had on their answers. Whereas in Thai script, tone is usually encoded by its own diacritics, or an additional grapheme preceding the syllable, making it possible to isolate, in Chinese languages a change in tone can produce a new word, with new meaning and, crucially, different character to encode it. This issue can, of course, be accounted for with use of a stimuli set including specific examples of different tones. As this would mean a five or more times larger set of stimuli components, this was beyond the scope of this project.

3.2.5. Potential bias sources

The specific design decisions, both on conceptual and practical level, inevitably introduce sources of bias, which need to be taken into account when evaluating the study's results and their reliability.

3.2.5.1. Author's background (positionality statement)

Relevant for this study, the author is of Polish origin and L1, fluent in English and Norwegian. These three languages, despite rudimentary knowledge of Thai and Chinese, certainly shape the author's outlook on language and the way it operates. Likewise, the author's education in Poland and Norway has followed Western scientific traditions, resulting in a point of view built upon the methods used and conclusions drawn by Western thinkers.

The author is also a functionalist, with some background in mathematics, contributing to a mechanistic view of both human language and cognition. While acknowledging the existence of phenomena not fitting this approach, the author is most concerned with those that do, and tends to analyse new information using mathematical models and logical constructions, easily accepting conclusions that fit mechanistic explanations. This shines clearly through in the design of the present study, with focus on quantitative analysis of functionally defined factors, as well as in the straightforward, at times naive, style of analysis and discussion.

3.2.5.2. Participant selection and background

Two important sources of potential bias stem directly from the way participants for the study have been recruited. Because the recruitment was conducted from the author's own social network (directly, and indirectly from its extension, i.e. friends of friends etc), the vast majority of participants were L2 English speakers. Consequently, they were all familiar with the Latin alphabet and this ability can be expected to have had an effect on their internalised sound and writing knowledge structures. This effect can be ameliorated by including more, or exclusively, single-writing system users, that was however beyond the ambitions of this project.

The second consequence of this form of participant recruitment is that they were not fully representative of wider demographics, as the majority were close to the author's own age, as well as of similar life circumstances, e.g. university students were heavily overrepresented. It is not immediately obvious how this would influence the results of the study on this particular subject, or what kind of effect it would have.

Both of these sources of potential bias can be remedied by replicating the experiments with more balanced participant groups without any necessary changes to the study's design otherwise.

3.3. Stimuli construction

An important design feature of the present study is the set of speech sounds selected for examination. On one hand, the intention has been to expand on the traditionally two-label sound symbolism experimental design (Bouba-Kiki, maluma-takete, etc), while including sounds shown by most prominent experiments to be involved. On the other hand, the scope of the study has been limited to maintain a degree of thoroughness in the investigation of the sounds selected. Consequently four consonant: /b/, /m/, /t/ , /k/ and four vowel: /o/, /u/, /e/, /i/ sounds have been selected for the study. Notably, /a/ is not included, as it has previously been used in stimuli for both “spiky” (“takete”) and “smooth” (“bouba”, “maluma”) shape labels. The sounds have further been divided into two groups according to the previous research designs: the “rounded” group /b/, /m/, /o/ and /u/ and the “spiky”: /k/, /t/, /e/ and /i/.

While the first group can be easily defined as “rounded and labial” sounds, the other group escapes such elegant description. The vowels included are both front, the consonants are articulated quite far from each other (postalveolar and velar), and while they are both plosives (*and* stops) so is one (*and* both) of the “rounded” consonants used. Defining them simply as voiceless consonants risks gross overgeneralisation, they are however amongst the few most universal *unvoiced plosives* of human languages (Stefanuto & Vallée, 1999). Notably, both sounds participate in the phenomenon of fronting, for example in child speech (Lowe et al., 1985).

The sounds selected were combined into syllables, consisting of one consonant and one vowel from the same grouping, yielding the primary sound sequences used throughout the study. Thus all audio stimuli have been made up of two or more of these primary components. Table 3.3 shows the selected syllables:

<table>
<tr><th colspan="2" rowspan="2"></th><th colspan="4">Consonants</th></tr>
<tr><th>/b/</th><th>/m/</th><th>/k/</th><th>/t/</th></tr>
<tr><th rowspan="4">Vowels</th><th>/o/</th><td>/bo/</td><td>/mo/</td><td colspan="2" rowspan="2"></td></tr>
<tr><th>/u/</th><td>/bu/</td><td>/mu/</td></tr>
<tr><th>/e/</th><td colspan="2" rowspan="2"></td><td>/ke/</td><td>/te/</td></tr>
<tr><th>/i/</th><td>/ki/</td><td>/ti/</td></tr>
</table>

Table 3.3 Sound sequences used in the study.

The choice of this syllable-wise approach has its source in the particular feature of Chinese characters to encode single-syllable words rather than their constituent sounds, making it impossible to disassemble them into smaller phonetic components. As this stimuli structure does not cause complications with other writing systems, the approach was adopted for all participant groups to support comparability.

The audio stimuli used in the experiments were then generated using an online text-to-speech service (TTSMP3) using multiple of its “voices”. Two main principles guided the choice of voices for the audios: diversity and clarity. The diversity consideration informed the choice of multiple voices of different backgrounds, in terms of gender and languages. The reasoning for the choice of diverse voices has been to simulate natural human language and its variety of speakers. At the same time, the voices were evaluated for their pronunciation of specific sound combinations with the aim of producing audio stimuli of high clarity.

3.4. Set up

Participants took the two experiments consecutively, in the order of their description (bijectivity test first, sound symbolic associations test afterwards). This sequencing has been chosen deliberately reflecting the experiments' progressing overtness. The experiments were accessed through the pcibex platform, participants were encouraged to find a comfortable, quiet place and use earphones if possible. A demonstration version of the survey used for the Latin scripts group can be viewed under the following link: https://farm.pcibex.net/r/SQUpzZ/ . Full scripts for all surveys used (one for each participant group, with minor adjustments in expected completion time and explicit references to groups' scripts) are available in the OSF repository: http://osf.io/znc5u/

3.4.1. Experiment 1 (bijectivity test)

The degree of bijectivity was tested in an experiment where participants were asked to write down the non-word stimuli containing the evaluated sounds using their L1-associated writing systems. The question this experiment sought to answer is how close to a bijective mapping the relation between the spoken sounds and written representations is, both on individual and group levels.

3.4.1.1. Procedure

Participants heard a total of 12 audio recordings each of (machine generated speech) non-word sounds and were asked to note them down in their L1-associated writing system. The experiment instructions were framed as a game, where the aim is to pass on "secret codes" to an unseen teammate in another room by means of written text. Participants were presented with the audios and instructed to write down their answers on paper first, before being asked to type them in a form at a later stage. This step was introduced to prevent bias arising from Chinese writing input methods, which usually rely on pinyin (latin alphabet based) input and frequency based typing suggestions, and, in the case of Thai, possible bias arising from the arrangement of symbols on the Thai writing keyboard, and its pages (usually two pages, first one displaying by default and another one available upon additional key press).

The stimuli consisted of 5 open syllables each. The initial ones were always /na/, /la/ and /fa/, in rotating order, while the remaining 4 were non-repeating, random sequences of the 8 syllables permitted in the study, pre-generated by a simple python script. See Appendix A for code used. Table 3.4 lists the sound sequences used in the experiment 1:

trial	sound sequence	trial	sound sequence	trial	sound sequence
1	/namomutibu/	**5**	/nabotetimo/	**9**	/nabokitite/
2	/labokikete/	**6**	/labukikemu/	**10**	/labukemumo/
3	/samubutiki/	**7**	/satemokebu/	**11**	/samokibubo/
4	/fabomoteke/	**8**	/famubokiti/	**12**	/famuteketi/

Table 3.4 Sound sequences used in experiment 1.

In total, each individual set of answers provided six responses for each of the eight syllables in the stimuli component set.

3.4.1.2. Rationale & vulnerabilities

This task aimed to produce a measure of bijectivity (1-to-1-ness) between speech sounds of the auditory stimuli, and written representations, for individual participants, as well as each writing system users group. Within the framework of this project's design, this was intended to produce a measure for each individual, and each writing system, which then could be used as a variable to be tested for correlation with their sound symbolic associations test results at a later stage of analysis.

The most apparent vulnerability of this task's design can be observed to arise from the limited nature of stimuli set. For the purposes of this project, considering the number of participants available, it was not practical to test entire phonetic inventories of participants' L1's, and therefore the measure produced will not necessarily be extendable to a confident assessment of the writing systems as whole. On the other hand, the experimental model can be easily extended to include more complete stimuli sets without major changes to the overall design. This ease of extension is, as stated, an intentional aspect of the project's design.

A second vulnerability that can be identified is the study's limited assessment of speakers' language varieties beyond their primary writing system. While this can be expected to produce more uncontrolled variety in participants' answers, it can be viewed as a feature of the study's particular design, rather than a flaw. With additional metadata collected from participants, a more detailed study can be conducted to investigate more narrowly defined groups, including speakers of more narrowly defined language varieties. Such study would, however, provide answers to slightly differently formulated questions, including potentially questions of dynamics within each writing system user group treated in this project as units. Another aspect of the particular choice made for the present study is further discussed in the "Ethics" section.

3.4.2. Experiment 2 (sound symbolic associations test)

The second experiment followed the usual procedures used in the sound symbolic line of research. It aimed to measure the "accuracy" of participants in matching audio stimuli made up solely of sounds from one of the two pre-defined groups to smooth or spiky visual shapes. The aim of this experiment was to produce sound symbolism scores required for the correlation with bijectivity analysis.

3.4.2.1. Procedure

Participants were presented with an audio recording and a pair of 2 images. The audio stimuli were /tiki/, /momu/, /bubo/ and /kete/. The images used were the "bouba" and "kiki" images (sourced from Cwiek, 2020) in test 1 and 3 (/tiki/ and /bubo/), and the "maluma" and "takete" images from Sapir's 1929 study in test 2 and 4 (/momu/ and /kete/). Figure 3.5 illustrates the four images used:

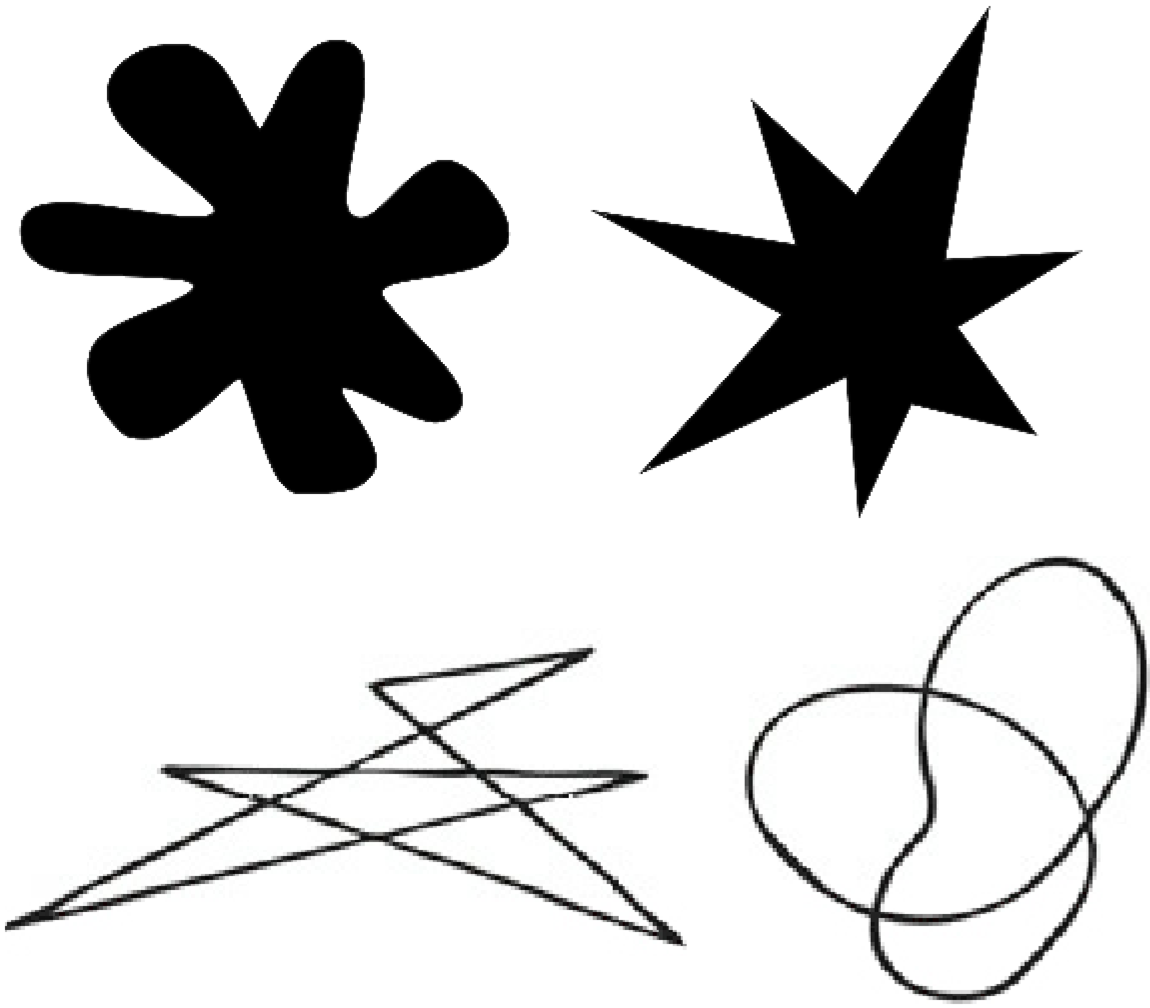


Figure 3.5 Images used for visual stimuli in experiment 2. Clockwise from top left: “bouba”, “kiki” (from Cwiek, 2020), “maluma” and “takete” (from Sapir 1929)

The image order was reversed in each of the two repeated uses. Table 3.6 illustrates the sequence of audio and visual stimuli in the procedure.

trial	audio stimuli	image 1	image 2	congruent answer
1	/tiki/	“bouba”	“kiki”	2
2	/momu/	“maluma”	“takete”	1
3	/bubo/	“kiki”	“bouba”	2
4	/kete/	“takete”	“maluma”	1

Table 3.6 The procedure in experiment 2.

Thus the answers considered congruent with sound symbolic association premise were, in order: 2, 1, 2, 1. For each such answer, the participant was given 1 score, with a total calculated in the data analysis stage.

3.4.2.2. Rationale & vulnerabilities

The main purpose of this test was to produce a measure of sound symbolism scores for participating individuals and writing system groups.

This was a slightly modified version of the classic sound symbolic association experiments. Instead of the traditional non-words, this experiment used similar ones containing only sounds from the subset selected for the project. Additionally, the stimuli were modified to not feature repeated syllables. The original motivation for this was to avoid the impression of meaning for the Chinese participants, as many Chinese words are formed by repetition of a single open syllable, notably both /keke/ and /titi/ would bear close resemblance to the realisations of Chinese words (哥哥 and 弟弟) meaning older and younger brother respectively, in some dialects.

Visual stimuli used were sourced from two of most prominent sound symbolism experiments, with audio stimuli following the same patterns, the overarching idea of this experiment was to test the associations between the sound two groups (/t/, /k/, /e/ and /i/ vs /b/, /m/, /o/ and /u/) and smooth/curved vs “spiky” visual shapes.

3.5. Participants

The participants for the experiments have been recruited by means of social networking (convenience sampling). Approximately 20% of the participants were recruited by the author directly, the rest were recruited by these primary participants from their own networks. The target group size of $N_G = 20$ was set for a total of $N_T = 60$ for the entire study. The inclusion criteria were defined as “any adult whose first language (L1) uses the designated group’s writing system as its primary”.

Within the Simplified Chinese characters and Thai script groups, this resulted in inclusion of speakers of Chinese or Thai languages respectively. In the Latin scripts group the participants L1’s included primarily Norwegian and Polish, with multiple ($n > 1$) additional representatives of Vietnamese, Indonesian and English, as well as single instances of Italian, Luganda, and Tagalog.

3.6. Data collection

During data collection, new entries were checked for completeness of answers in order to monitor the progress of the process. Entries, which did not follow the experiment instructions, e.g. entries in simplified Chinese characters or Thai script groups submitted in latin alphabet, entries lacking answers to one or more questions in the experimental tasks, etc, were discarded. These constituted approximately 20% of submitted answer sets in the Chinese group and approximately 10% in the Thai group. The data was collected using a separate pcibex form (identical except for time approximations and specific references to the group’s writing systems in instructions) for each defined participant group until the point of $N = 20$ eligible answer sets. After that the pcibex experiments were un-published and no further data collected.

3.7. Analysis

All data collected from participants was cleansed, transformed to the required format and analysed using python 3.0 scripts in a jupyterhub notebook provided by jupyterhub.uio.no . The functions used were imported from packages:

- numpy (for numerical operations and data structures)
- scipy (for statistical analysis functions)
- pandas (for descriptive statistics)
- matplotlib (for plots and visualisations)

All scripts used were published in the OSF repository: osf.io/znc5u

3.7.1. Data preparation

Upon completed collection the data downloaded from the pcibex platform was prepared for further analysis. All metadata (timestamps, IP hash, etc.) was stripped and permanently deleted. The responses were controlled for stated L1 and categorised accordingly. Incorrectly filled responses were discarded and the remaining data was collated into script-specific data sets. This dataset, including some of the calculated metrics, is accessible through the OSF repository used for this study: osf.io/znc5u/

3.7.2. Metrics

From each individual's responses three values were derived: two for distributional properties and one for sound symbolism scores. To measure distributional properties the token count was calculated, first for each stimuli segment by counting the number of different written symbols used by the participant to encode it, then an average between the eight stimuli segments. Next the formula for entropy used in information theory was employed:

$$H = -\sum_i p_i \log p_i$$

In the formula, i stands for the number of different graphemes used by the participant to encode a given measured sound, and p_i stands for the probability of each (i-th) grapheme being used (calculated by dividing the number of its instances by the number of occurrences of the sound). This produced a measure of entropy for each sound with the value of 0 (when participant used one grapheme per sound) or more, increasing as participant uses more graphemes for every instance of the sound. The measures thus calculated were then averaged to produce a single value for each participant.

Notably, while entropy is related to token count, it presents a more nuanced picture of how the tokens are distributed. Thus, where token count can be considered a superficial property of writing, entropy serves to illustrate a specific aspect of writing with its own implications.

For example, for a sound segment with 1 instance of written representation X and 11 instances of representation Y will result with a score of:

- token count = 2
- **entropy = 0.287**

While if the sound is represented by 2 tokens, with 6 instances recorded each, the score becomes:

- token count = 2
- **entropy = 0.693**

This reflects higher uncertainty as in the first example, the representation with 11 instances is highly likely to be chosen, while in the second example the outcome is more uncertain due to equal probability of the two different representations occuring. Thus a very important difference in usage is captured by entropy whereas a simple token count would leave it unnoticed.

The value measured for sound symbolic associations strength was a simple tally of participant's answers congruent with the sound symbolic paradigm in the range from 0 (for no congruent answers) to 4 (all congruent).

3.7.3. Statistics

The three metrics obtained were tested for correlation, for the total pool of participants as a whole, and for each group of participants separately. The idea guiding this choice was to collect analyse intra-group trends to detect irregularity resulting from individual factors other than L1 and writing system (largely common inside the Chinese characters and Thai script groups, and partially common in the Latin scripts group), and to analyse the scores obtained from the totality of participants in search of trends introduced by allowing such factors and using a larger dataset, more likely to return valid results (as groups themselves, at N = 20, were close to the minimum required for statistical methods robustness).

The average scores of all three values for each group have also been calculated for the purpose of qualitative comparison.

3.7.4. Rationale & vulnerabilities

In mathematics (information theory) entropy is used to measure the level of uncertainty within the value set of a function. Low entropy indicates low uncertainty, with the minimum of 0 indicating no variation in the value set at all (entirely predictable outcome), and positive values indicating the degree of uncertainty. In the context of the data collected in experiment 1, entropy provides a way to measure the level of variation in participant's answers thus making it a very suitable tool for the purpose of measuring how close to bijective the individual's mapping between speech sounds and written symbols is.

It is important to note that different writing systems allow for different levels of flexibility in written representations. This is especially true of the three writing systems selected for this study. The latin scripts allow for very little options in encoding the sounds present in the stimuli set, Thai script provides significantly more, while Chinese characters set is the only to allow its users (in principle) to write the exact same sound in six (or more) ways. This asymmetry in systemic constraints will inevitably skew the entropy results. Moreover, as the study is limited to only a subset of non-randomly selected speech sounds, and hence also a subset of written symbols, the results should only be taken as an indication of the trend. A more complete picture of the phenomenon can be achieved in a study with broader scope. Also, one conceivable way to account for these constraints set by the writing systems themselves could be to evaluate entropy in terms of percentage of maximum allowed entropy permitted by the writing system, i.e. to calculate, for each participating writing system, the maximum entropy (most evenly distributed) case permitted and measure each participant's entropy score as a percentage of entropy allowed by the system. This, however, would only be practical for some scripts and would not account for entropy achievable with different number of observations recorded.

3.8. Ethics

The subject matter of this project has been of little public interest or debate and participation has therefore been judged as unlikely to bring negative outcomes to the participants. Additionally, no identifying data has been collected about the participants and any timestamps, IP hash information and similar technical and metadata automatically recorded by the pcibex platform has been deleted in the process of data cleansing. The original databases of collected data on the pcibex servers have been purged as part of the same process as a precaution.

Because the subject is not of general interest, the results are unlikely to produce significant benefits to the participants or the wider public, beyond the specialists in the field of linguistics and cognitive sciences, they are however hoped to induce interest in the respective writing systems included.

3.9. Publication

All audio materials, images, scripts and (cleansed) data collected have been published in an OSF repository at osf.io/znc5u/ . Additionally, the final report will be added to the repository, and submitted to the National Research Council of Thailand in compliance with the terms and conditions of the research permission granted to the author, once such publication is permitted by the author's institute.

4. Results

The data collected was cleansed, transformed where necessary and prepared for the analysis.

4.1. Data cleansing and preparation

The data was downloaded from pcibex and deleted from its servers. Subsequently, fields containing TextInput labels and final values were stored in a new database and the csv files deleted. Where required, the cells containing answers to first and second experimental tasks were sorted to reflect the sequencing of stimuli in part 1, and the order of tasks in part 2. Answers to part 2 were standardised to numeric form, i.e. answers of the form “image 1”, “left”, “image 2”, “right”, etc. were replaced by corresponding numeric values 1 or 2. This new dataset was then used for the rest of the analysis process.

4.2. Data collected

The cleansed data set contained the participants’ submitted L1 and L2s information, written representations for the eight audio stimuli in experiment 1 and four answers for sound symbolic associations in experiment 2, as well as answers to the feedback request on the final page of the survey. Within each set of answers further metrics were derived as follows:

- For each stimuli component (syllable for the Chinese group, single sound for other groups) the number of different written representations (tokens) was counted along with each representation’s number of occurrences (frequency).
- For each participant the average number of tokens per stimuli component was calculated by dividing the sum of token counts for all stimuli components by eight.
- For each stimuli component the value of entropy was calculated based on the number of different tokens and their frequencies, following the formula presented in the “Method” section.

- For each participant the average entropy was calculated by dividing the sum of entropies by eight.
- For each participant the number of answers in experiment 2 congruent with the sound symbolic assumption was counted.

The last three of the values above, i.e: the average number of tokens, the average entropy and the number of congruent experiment 2 answers were collated into a new dataset and this dataset was used as the basis of further analysis. The dataset, including metrics derived, is available for inspection in the osf repository (osf.io/znc5u).

4.2.1. Data quality

As data submitted was monitored during the collection process, the dataset used for analysis was already quality-checked. In rare instances, where answers submitted were shorter than the corresponding stimuli, the author judged which part of the stimuli was missing. This was done based on the author's ability to read the corresponding scripts in consultation with native speakers of the language in question when necessary. As this was not a common occurrence, the influence of these judgement calls on the outcomes of the study was judged to be low.

4.3. Data evaluation

The data collated was first evaluated using descriptive statistics and tested for normality of distribution as to assess the statistical tools viable for its analysis. This was done first groupwise then with a dataset combining answers from all groups.

4.3.1. Groupwise

In the first phase of the analysis process each participant group was evaluated separately.

4.3.1.1. Simplified Chinese characters

The three values in the set were first analysed using descriptive statistics. The summary is presented in Table 4.1:

	value range	Mean	Standard Deviation
average token count	2.375 - 5.875	4.0125	0.96
average entropy	0.657 - 1.763	1.227	0.31
s-s associations	0 - 4	2.5	1.1

Table 4.1 Descriptive statistics (values rounded) for the results from the Chinese characters group.

Some interesting observations could be made already at this stage:

The average token count (the number of different symbols used to encode the same syllable) in this group was approximately 4, with typical scores between 3 and 5. Additionally, while this has not been quantified in this analysis, multiple instances of participants using the same symbol for different (if usually similar, e.g. /bo/ and /mo/) sounds are evident in the dataset. This can be interpreted to mean that the mapping between the sound and symbol is, for the users of simplified Chinese characters, far from bijective. On the other hand, this can also suggest that the sounds in the stimuli audios were perceived differently by different participants. A quite broad range of values suggests different strategies used by the participants, as well as variance depending on the specific sound concerned.

The mean value of 2.5 for S/S associations in the group suggests higher than chance (62.5%) probability of providing congruent responses in the sound symbolic association test of experiment 2.

Next, the values obtained were tested for normality of distribution. First, they were visualised using histograms:

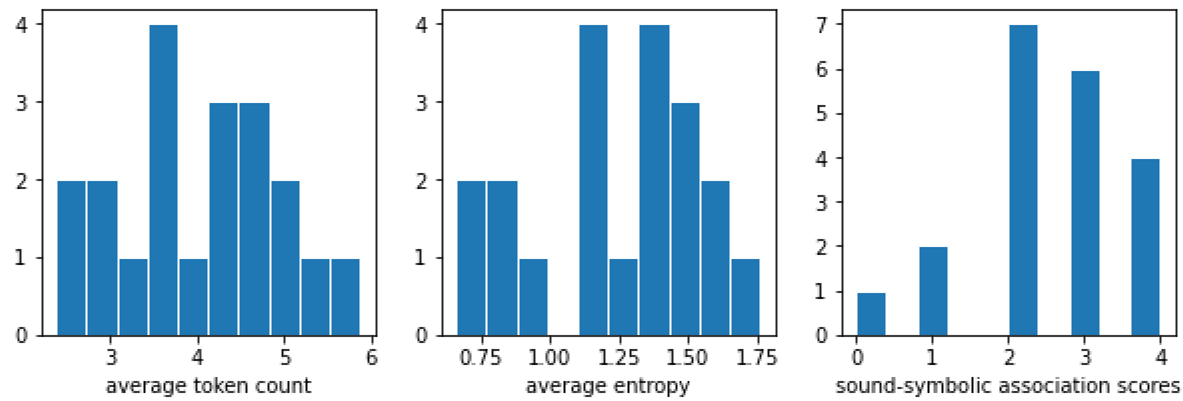


Figure 4.2 Frequency of values in the data collected from the Chinese characters group.

With the distributions approximately bell-curved, probability distribution (QQ) plots were taken:

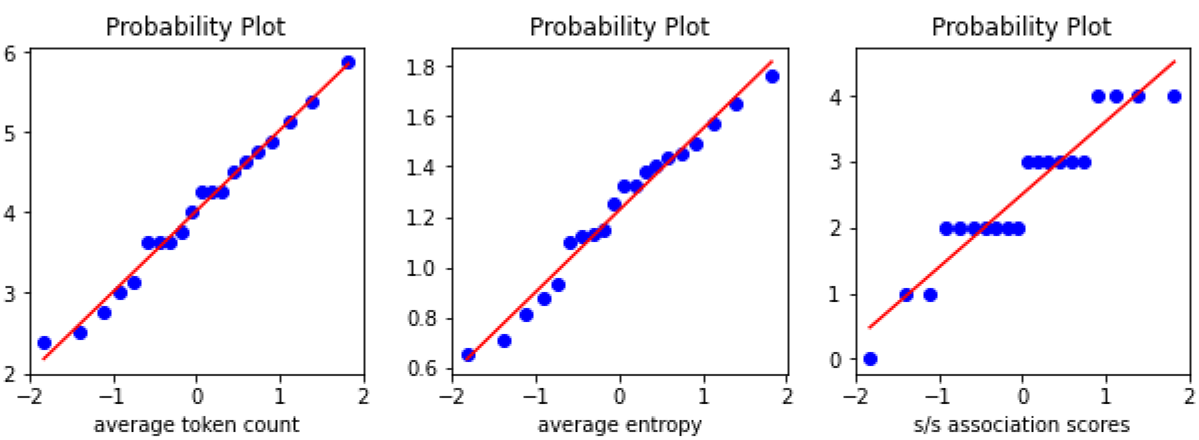


Figure 4.3: Probability distribution (QQ plots) in the data collected from the Chinese characters group.

Finally the Shapiro-Wilk test was applied to confirm:

	average token count	**average entropy**	**s-s association scores**
test statistic (W)	0.98	0.97	0.91
significance (p)	0.93	0.75	0.06

Table 4.5 Normality statistics for the data collected from the Chinese characters group.

Thus, as for all three variables $p > 0.05$, the data appears to be normally distributed. While with the small sample sizes the formal normality tests may have high error rates, the distributions do appear to be normally distributed judging by the histogram and probability distribution (QQ) plot images.

4.3.1.2. Thai script

The process was repeated for this group, the first step was descriptive statistics:

	value range	**Mean**	**Standard Deviation**
average token count	2.125 - 5.125	3.15	0.782
average entropy	0.428 - 1.397	0.789	0.226
s-s associations	0 - 4	2.6	1.142

Table 4.6: Descriptive statistics (rounded) for the results from Thai script group.

Again, some initial observations can be made at this stage:

The average token count in this group was just above 3, with most values falling in the range of approximately 2.4 - 3.9, this is slightly lower (somewhat predictably) than the Chinese group. Again, similar to the first group, instances of the same symbols being used to encode different sounds abound in the participants' responses, suggesting both a non-bijective mapping system, but also differences in the way sounds in audios were perceived. A relatively broad range of values, as with the Chinese group, suggests individual differences in strategies adopted for the performance of the experimental task.

The mean value of 2.6 in sound symbolic associations task suggests higher than chance (65%) probability of congruent answers, 2.5% higher than in the first group.

Next, the values were visualised in histograms to assess normality of distribution:

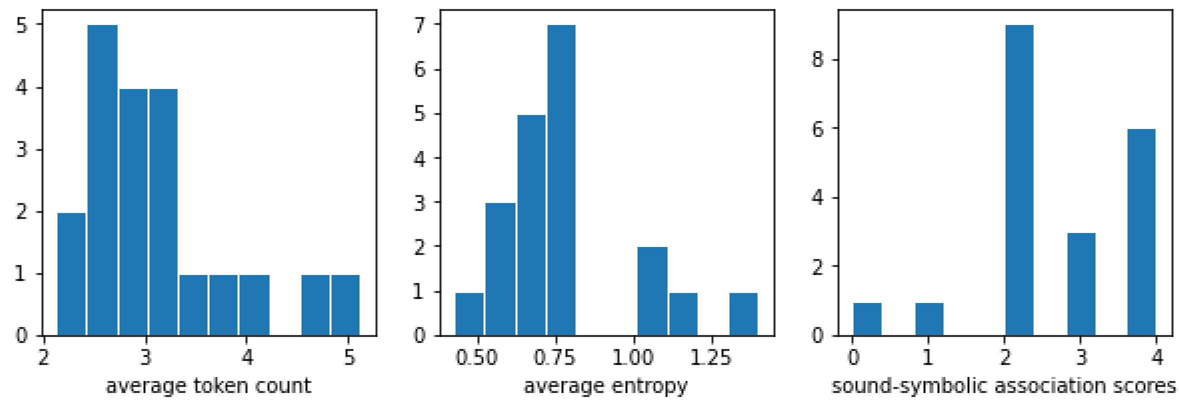


Figure 4.7 Frequency of values in the data collected from the Thai script group.

Next probability distribution (QQ) plots were produced:

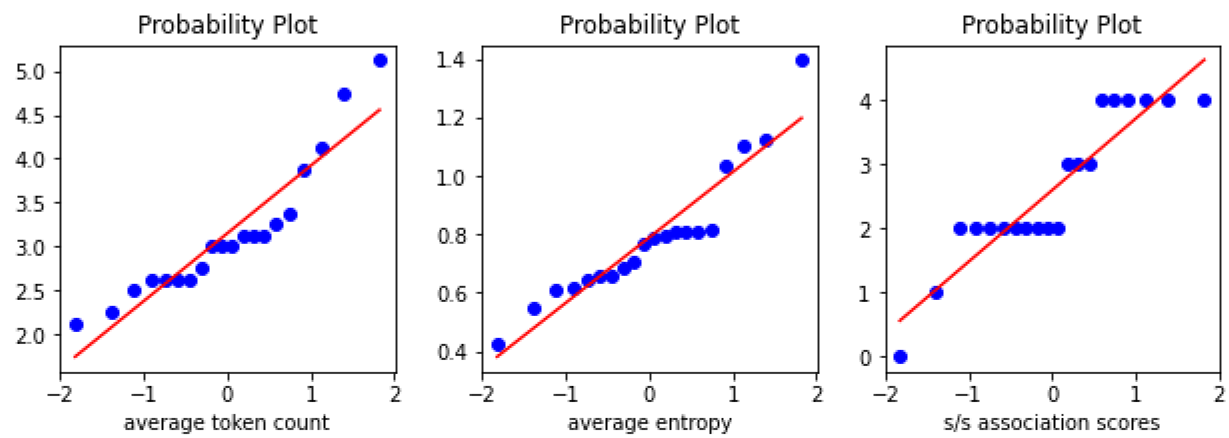


Figure 4.8 Probability distribution (QQ plots) in the data collected from the Thai script group.

Again, following the histograms and probability plots being inconclusive, Shapiro-Wilk test was performed:

	average token count	**average entropy**	**s-s association scores**
test statistic (W)	0.88	0.90	0.86
significance (p)	0.02	0.04	0.01

Table 4.9: Normality test results, Thai script group.

Here, unlike the Chinese group, with all significance values of $p < 0.05$, the conclusion of the test is that the data is not distributed normally. This is, due to small sample size, less reliable, but the result was taken to be valid.

4.3.1.3. Latin scripts

Following the same process, data collected from the latin scripts group was analysed, first using descriptive statistics:

	value range	Mean	Standard Deviation
average token count	1.375 - 3.125	1.931	0.469
average entropy	0.126 - 0.712	0.345	0.168
s-s associations	1 - 4	2.85	1.137

Table 4.10 Descriptive statistics (rounded) for the results from the Latin scripts group.

Of note in this group is the much lower average token count, with the average of approximately 2 distinct written representations per sound and majority of scores in the range of approximately 1.5 - 2.5. Somewhat unsurprisingly, the scores were lower than both Chinese and Thai groups, as is the range of values, suggesting even less flexibility in written representations than the other groups.

The average score of 2.85 on the sound symbolic association test translates into 71.25% probability of congruent answers, highest of all groups in the study.

Next, the values were examined for normality of distribution, first using histograms:

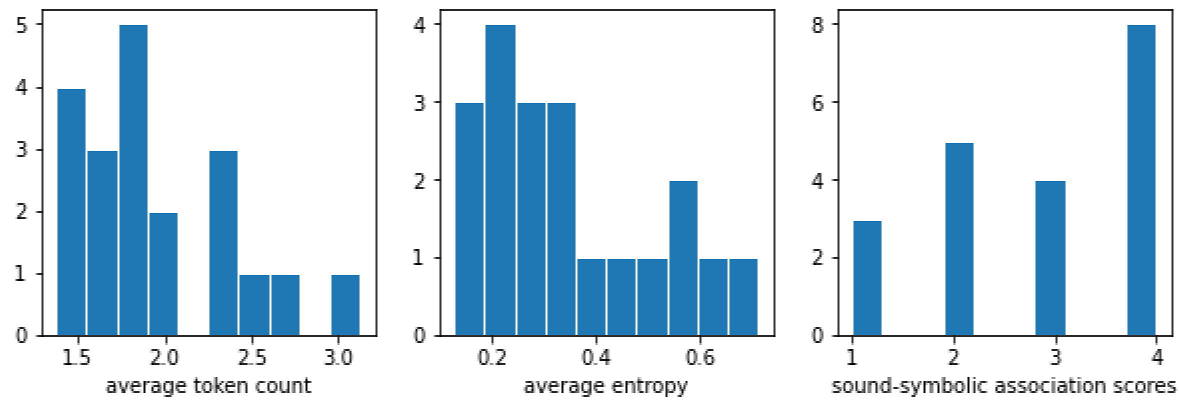


Figure 4.11 Frequency of values in the data collected from the Latin scripts group.

Again, data looking symmetrical, probability plots were produced:

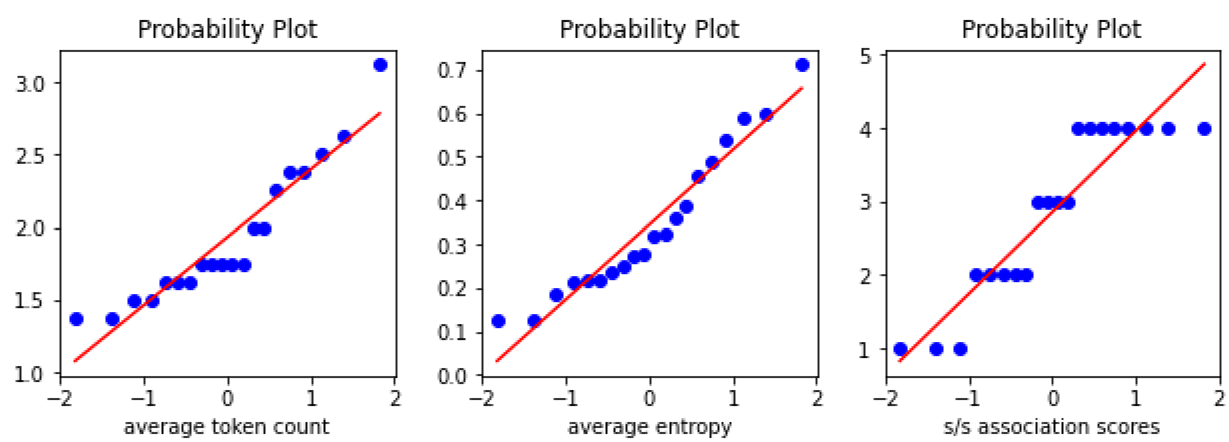


Figure 4.12: Probability distribution (QQ plots) in the data collected from the Latin scripts group.

Finally, Shapiro-Wilk test was taken:

	average token count	**average entropy**	**s-s association scores**
test statistic (W)	0.9	0.93	0.83
significance (p)	0.04	0.13	0.002

Table 4.13: Normality test results, Latin scripts group.

In this group, the values of token count and sound symbolic association scores were not distributed normally (both scoring $p < 0.05$), while, interestingly, entropy was ($p = 0.13 > 0.05$). Again, while histograms and probability plots may suggest otherwise, this result was accepted as valid.

4.3.2. Combined

The procedure was then repeated using a database consisting of the values calculated for *all* participants. Starting with descriptive statistics:

	value range	**Mean**	**Standard Deviation**
average token count	1.375 - 5.875	3.03125	1.143
average entropy	0.126 - 1.763	0.787	0.434
s-s associations	0 - 4	2.65	1.117

Table 4.14 Descriptive statistics (rounded) for the results from all participants.

These paint an interesting picture of the results obtained by the totality of 60 participants from the three groups in the study when treated as one whole:

A mean of approximately 3 written symbols used per sound, with the majority in the range of approximately 2 - 4 suggests (subject to limitation discussed earlier) the relationship between sound and written form is far from bijective for the vast majority of the participants, and by extension, the population at large, notably regardless of the writing system.

The mean value of 2.65 for sound symbolic associations signifies a 66.25% probability of congruent answers, another endorsement of the phenomenon’s universality.

Next, the database was analysed for normality of distribution. Starting with histograms:

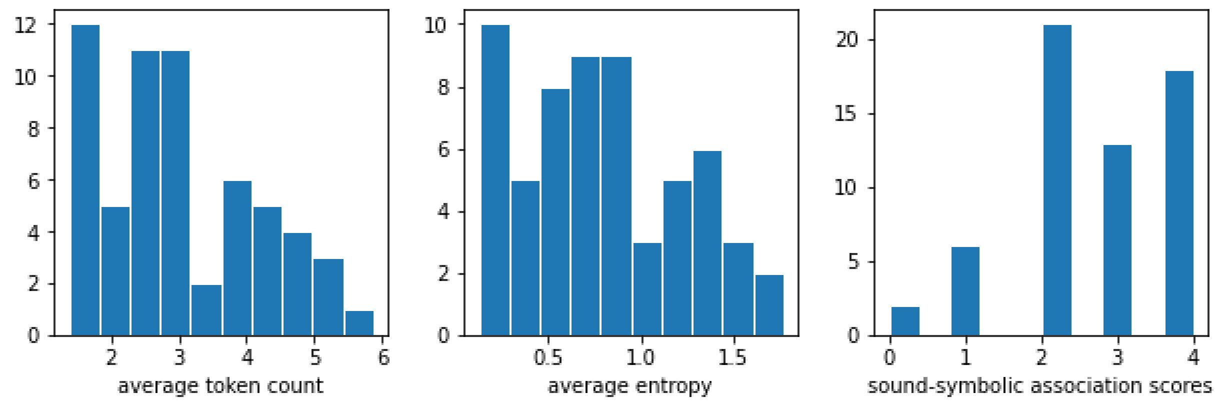


Figure 4.15: Frequency of values in the data collected from all participants.

Following previously established procedure, probability plots were produced:

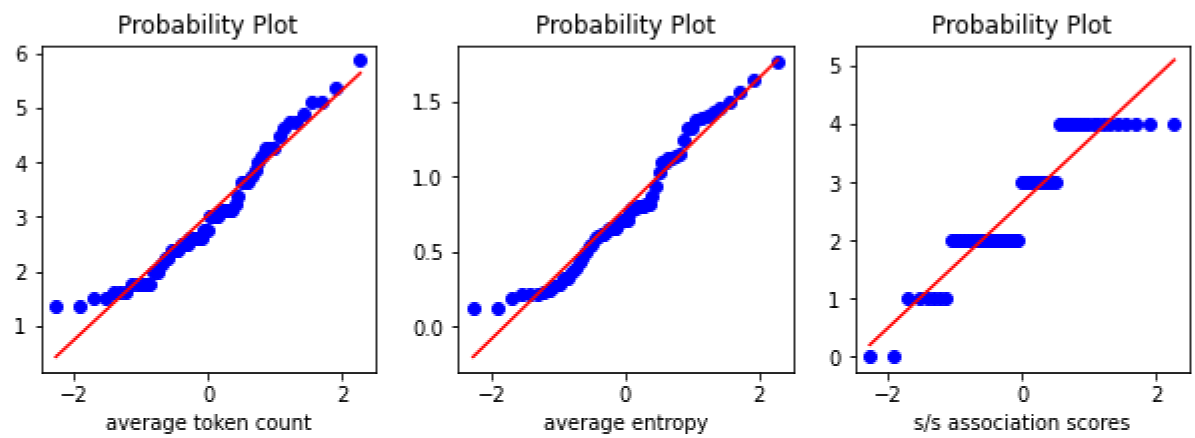


Figure 4.16: Probability distribution (QQ plots) in the data collected from the entire sample.

And then a confirmatory Shapiro-Wilk test:

	average token count	**average entropy**	**s-s association scores**
test statistic (W)	0.95	0.95	0.86
significance (p)	0.02	0.03	0.00*

Table 4.17 Normality test results, entire sample. * - the value obtained was approximately 0.0002

As evident from the table above, taken as a whole, none of the values collected and calculated from the experimental results appears to distribute normally.

4.3.3. Comparison

Following the data evaluation, the scores obtained and calculated were compared between the groups and the entirety of the sample in an attempt to compare the performances of each group with each other. For this purpose, the means and standard deviations for each of the three analysed variables were compared using bar plots:

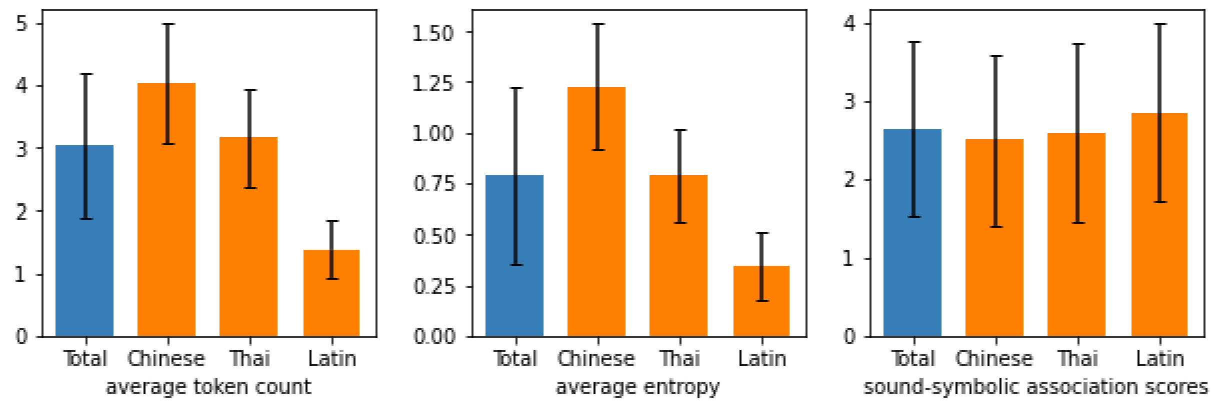


Figure 4.18: Comparison of mean scores (error bars represent standard deviation) in the three variables across the groups in the study.

As can be seen, the scores of the Thai script group across the three variables reflected most closely those of the totality of participants, with simplified Chinese characters group scoring distinctly higher and Latin scripts group lower on the counts of token count and entropy, with the sound symbolic association scores following an opposite trend, i.e. Chinese group scoring lowest and Latin group highest, this can be interpreted as an indication of a negative relationship, i.e. it may appear that the distributional properties (token count and entropy) follow a trend opposite to that of sound symbolic associations.

4.4. Analytics

Having established the properties of the distributions of all variables, Pearson and Spearman tests were used to test the values calculated for correlation. This was, again, performed first within each group and then for the entire participant sample as a whole.

4.4.1. Groupwise

To test the project's underlying hypothesis of a relationship between the degree of bijectivity in writing and sound-symbolic associations strength, the values of average token count and average entropy were tested for correlation with achieved sound-symbolic association scores.

4.4.1.1. Simplified Chinese characters

Before selecting the suitable tool for the analysis of correlation between average token count and entropy on one hand, and the sound symbolic associations scores on the other, the data was visualised using scatterplots to evaluate the nature of potential relationships:

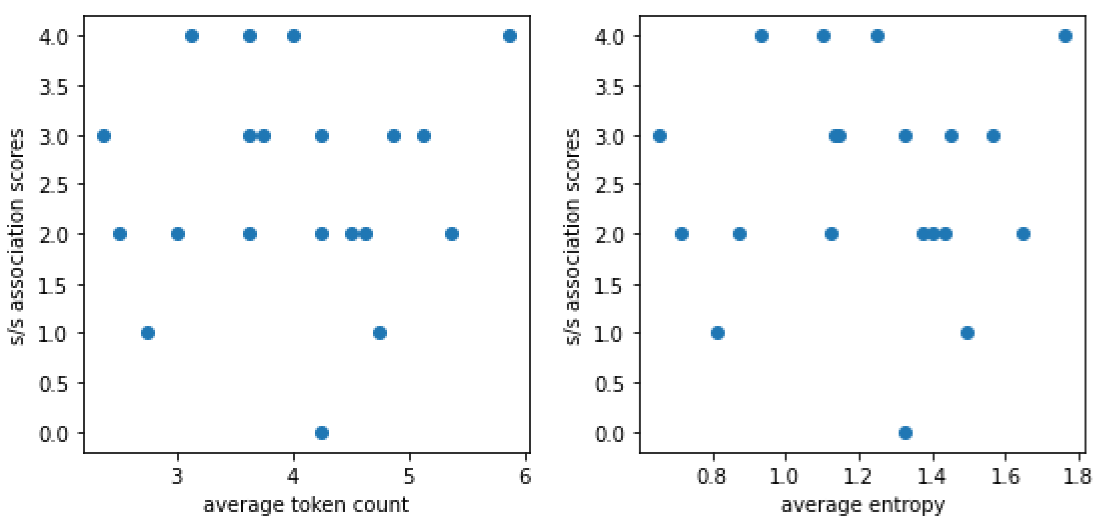


Figure 4.19: Distributional properties against sound symbolic association values, Chinese group.

Little could be concluded from the inspection of scatterplots. This could be partially due to the relatively low number of observations, but no hint of a relationship was immediately visible. Because all three variables in the simplified Chinese characters users group were distributed normally, Pearson's test of correlation was conducted first to test for linear relationships, followed by Spearman’s test to detect potential non-linear correlations. The tests were conducted for correlation between sound symbolic association scores and average token count first, and between the association scores and average entropy after that. The results of these tests are presented in Table 4.20:

Correlation with sound symbolic association scores:				
	Pearson’s test		**Spearman’s rank correlation**	
	statistic (r)	**significance (p)**	**coefficient (r)**	**significance (p)**
average token count	0.056	0.814	0.010	0.966
average entropy	0.023	0.921	0.025	0.916

Table 4.20 Correlation test results, Chinese characters group.

As can be seen, both Pearson’s and Spearman’s test results provide a sound refutal of the hypothesised relationship between both average token account or average entropy with sound symbolic association scores, with significance values all $p >> 0.05$ suggesting absence of relationship evident in the data collected, linear or otherwise.

4.4.1.2. Thai script

Following the procedure established in the analysis of simplified Chinese characters group, the results collected from the Thai script group were first visualised using scatterplots:

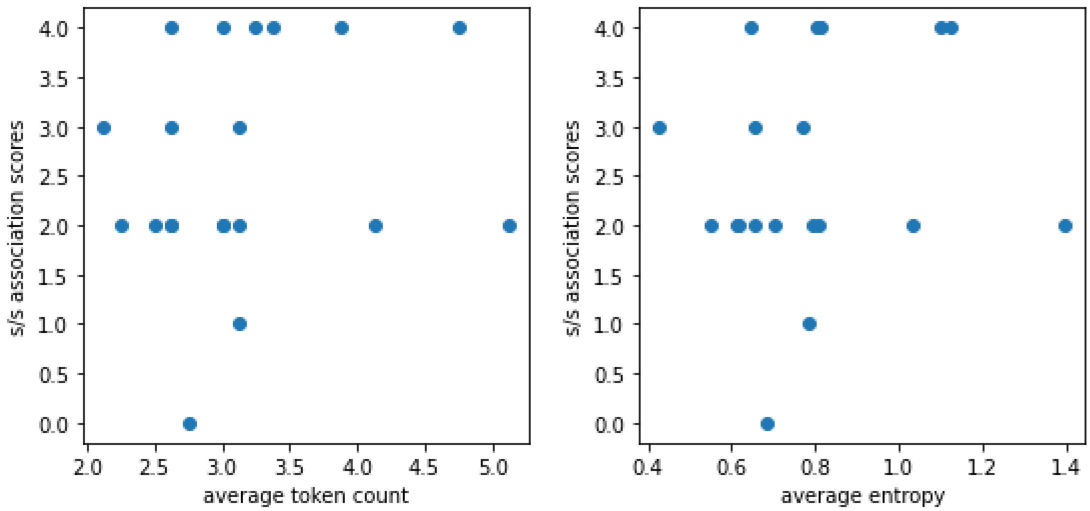


Figure 4.21 Distributional properties against sound symbolic association values, Thai script group.

Again, little could be read from the scatterplot. Additionally, as the data has previously been established not to be distributed normally, the Person's test was skipped and only Spearman's rank correlation test was performed to detect any potential monotonic relationship. The results are presented in Table 4.22:

Correlation with sound symbolic association scores:		
	Spearman's rank correlation	
	coefficient (r)	**significance (p)**
average token count	0.218	0.355
average entropy	0.268	0.254

Table 4.22 Correlation test results, Thai script group.

As both tests produce results with significance of $p > 0.05$, the results suggest an absence of relationship between average token count or average entropy with sound symbolic association scores obtained.

4.4.1.3. Latin scripts

As with the other groups, the data collected was first analysed using histograms:

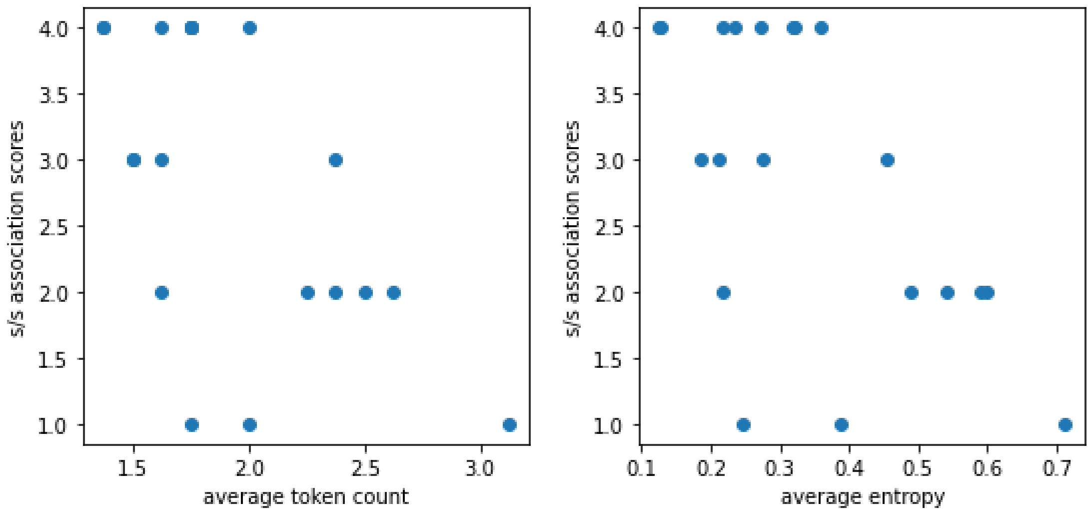


Figure 4.23 Distributional properties against sound symbolic association values, Latin scripts group.

Here, some indication of a negative correlation could be hinted by the top left - to - bottom right distribution of data points. Because the sound symbolic association scores collected in this group were not distributed normally, Pearson's test was skipped and only Spearman's correlation coefficient calculated. Table 4.24 presents the results:

Correlation with sound symbolic association scores:		
	Spearman's rank correlation	
	coefficient (r)	**significance (p)**
average token count	-0.525	0.017
average entropy	-0.517	0.019

Table 4.24 Correlation test results, Latin scripts group.

Interestingly, the results collected from this group show a negative, medium (~0.5) strength correlation for both average token count and entropy with sound symbolic association scores. This can be interpreted as any increase in sound symbolic association scores can be attributed to the *decrease* of token count or entropy approximately half of the time.

4.4.2. Combined

Lastly, the familiar routine was followed once more for the dataset consisting of all data points put together. Starting with a scatterplot to evaluate data visually:

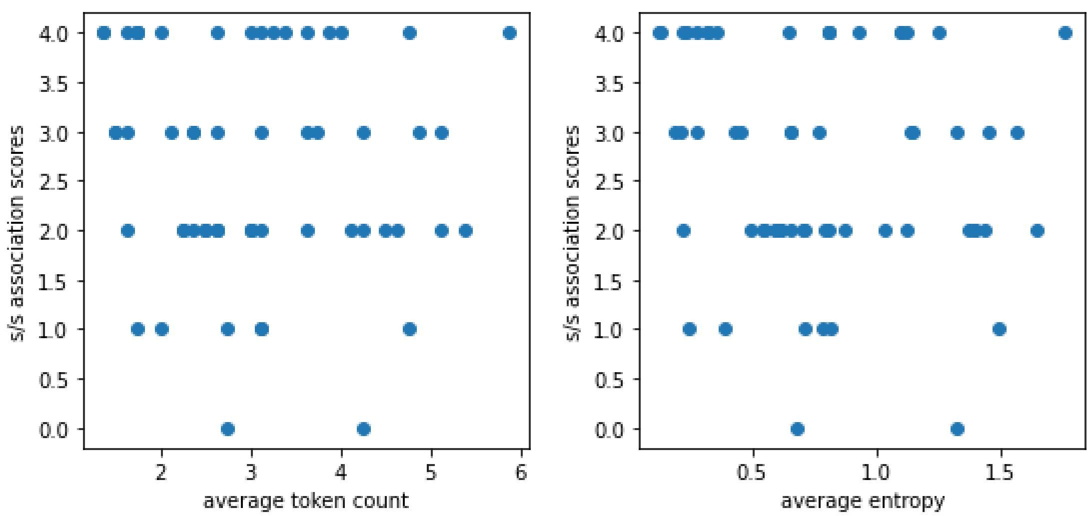


Figure 4.25 Distributional properties against sound symbolic association values, whole sample.

Once again, no obvious relationship was visible, and since the data had earlier been established not to be normally distributed, the Spearman's rank correlation test was performed to detect any monotonic relationship not apparent in the visualisations. The results are presented in table 4.26.

Correlation with sound symbolic association scores:		
	Spearman's rank correlation	
	coefficient (r)	significance (p)
average token count	-0.146	0.267
average entropy	-0.156	0.234

Table 4.26 Correlation test results, entire sample.

Once again, the tests produced results with $p > 0.05$, suggesting the hypothesised correlation was not found to be statistically significant.

5. Discussion

The study set out to investigate the relationship between the distributional properties of individuals' writing and the strength of their sound symbolic associations. This relationship was evaluated using measures of participants' token count (the number of different written representations used for each sound or sound sequence) and the entropy (a measure of uncertainty, or disorder) of the tokens' distribution, and the scores achieved in a sound symbolic association task. The participants were recruited from three groups, each representing users of a different writing system, with the intention of evaluating similarities and differences between their users. The measures of token count and entropy were tested for correlation with sound symbolic association scores within each group and for the totality of participants as a whole.

5.1. Findings

For each group, and for the totality of all participants, correlation tests were performed. First for average token count and sound symbolic association scores, then for entropy and s/s association. No correlations were detected in the data obtained from the simplified Chinese characters and Thai script groups, nor in the dataset containing combined data from the participants of all three groups.

In the Latin scripts group, however, negative, medium strength correlations were detected for both average token count ($r = -0.525$, $p = 0.017$) and entropy ($r = -0.517$, $p = 0.019$) with sound symbolic association scores.

Additionally, in a cross-group comparison of mean values for the three variables a negative trend could be observed, with mean scores of sound symbolic association following a tendency opposite to that of both average token count and entropy, i.e. the groups scoring highest (simplified Chinese characters) and lowest (Latin alphabet) in both token count and entropy were also found scoring lowest and highest, respectively, in sound symbolic associations.

Finally, across the study, participants have responded in sound symbolic associations task in a manner consistent with the previously documented "bouba-kiki effect" with higher than chance probability, with 62.5% answers in the simplified Chinese characters group, 65% of the answers in the Thai script group and 71.25% answers in the Latin scripts group, for an average of 66.25% across the study, congruent with the effect.

5.2. Interpretations

The study set out to investigate if and how the distributional properties of an individual's writing influence the strength of their sound symbolic associations. Specifically, the aim was to evaluate whether individuals whose writing is closer to bijective (one-to-one) would score higher on sound symbolic associations. The results provide limited support for this hypothesis. While the direct correlations were only confirmed in one of the three participant groups, it is notable that the relationship thus indicated has the expected direction (low scores in distributional properties correlate with high sound symbolic associations scores), considerable strength ($r \approx -0.5$) and high significance ($p \approx 0.02$) for both distributional metrics. Had these correlations been weaker, it could be tempting to dismiss this finding as due to imperfections of the study. The high significance, however, together with the general trend of lower scores on distributional properties accompanying higher scores on sound symbolic associations in the cross-group comparison, point to a real possibility of interaction.

These general findings are in line with some of the past research. The idea of correlation between writing system and sound symbolic associations was previously demonstrated by Cwiek et al. (2022). They found that "Languages that predominantly use the Roman script had numerically higher bouba/kiki matches (descriptive average: 75%) than languages that use other scripts (63%)." Notably, these scores are close to the findings of this study, and the group scores follow the same order, with Chinese group scoring lowest, Thai higher than Chinese, and Roman script (Latin alphabet) higher than both (Figure 4.18).

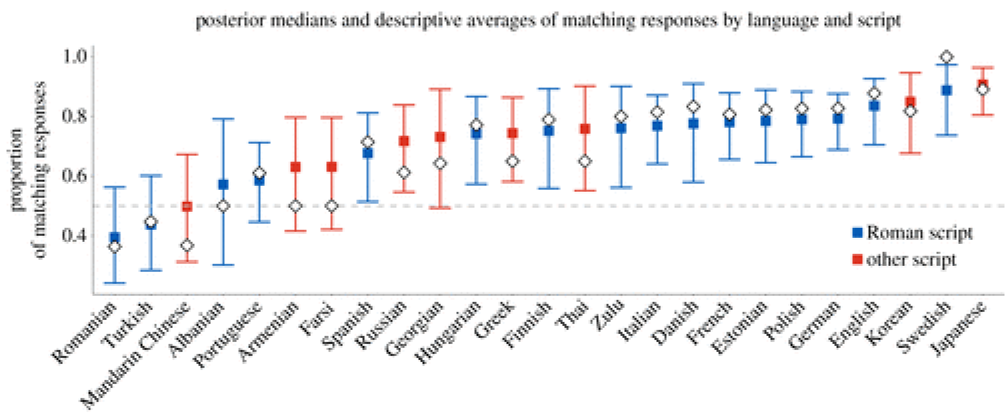


Figure 5.1 Bouba/kiki effect strength by L1. Cwiek et al. (2022)

While their participants were grouped by L1, not writing, this would very likely map directly onto respective writing systems as they were the only representatives of their language families within the participant group, and the only language groups using these writing systems as primary. The sound symbolic associations have also been previously demonstrated to be mediated by vision. Bottini et al. (2019) conducted experiments with both sighted and early-blind individuals, testing both strength and patterns of sound symbolic associations. They conclude that "early blindness does not prevent the emergence of SSC [sound-shape correspondences], and differences between sighted and visually impaired people may be due the indirect influence (or lack thereof) of orthographic letter shape". Considering the nature of writing systems examined in the present study (discussed more in the Background section), and the notable roundness of Latin alphabet's graphemes like "b", "o", or "m", and the spikiness of "t", "k" or "i", the results obtained can be seen to align with both of these studies.

Some additional observations can be made. The most immediate of those might be the fact of the correlations being detected only within the results collected from the Latin script group. This is interesting because the group was the only significantly linguistically diverse one in the study. Notably, virtually all participants in the simplified Chinese characters and Thai script groups declared largely mutually intelligible L1s, whereas the Latin alphabet group included representatives of several distinct linguistic families. One possible interpretation is that the variance in the results of the participants in Chinese and Thai groups is due to individual strategies for dealing with the constraints of their language and writing system, i.e. *corrected* for the effect of L1 and writing system. The Latin group, on the other hand, can be seen to demonstrate the effect of the distributional properties of individuals' writing resulting from a broader range of factors, like personal habits and experience, socio-economic factors, or linguistic background. This can be further supported by the results of group-wise comparison showing significant differences in average scores of distributional properties scores between the participant groups of the study. Taken as a whole, this can suggest that the distributional properties are modulated by the constraints of the writing system itself first, and then by more general and individual factors. The first of these steps can be seen at play in the results showing different value ranges recorded for each group (Figure 4.18) while the second in the variance in answers within the groups. The findings of this study can thus be said to confirm the direct influence of the writing system on the sound symbolic associations, while failing to conclusively show a direct effect of individual strategies and other individual characteristics beyond the users of Latin-based scripts. A more definitive answer could be provided by replicating the current study with participant groups of more uniform L1 within one or more of the writing systems examined. Such study could involve multiple groups of participants narrowly defined by specific language varieties using a common writing system.

Another observation that can be made is that, should the distributional properties of individuals' writing turn out to *only* correlate with sound symbolic associations strength for users of one specific writing system, or only for some systems, this could signify a qualitative difference in the operation of the sound symbolic mechanisms in their users.

Finally, accepting the general existence of a correlation between the distributional properties in writing and sound symbolic associations, a competing interpretation could be that it is the latter driving the former, i.e that the strength of sound symbolic associations in an individual shapes the way they use their writing systems within those systems' constraints. This is an interesting and logically consistent notion, which could explain the previously observed "influence" of writing as having been misinterpreted in terms of direction. Lastly, neither might be true, with the correlation originating in a common cause, i.e. both distributional properties of an individual's writing and the strength of their sound symbolic associations shaped by another common source factor.

5.2.1. Statistical error

When interpreting the results of the study it is necessary to bear in mind the consequences of potential type 1 and type 2 errors in the statistical results.

5.2.1.1. Type 1

Should the results obtained in the correlations tests in the Latin script group turn out to be a false positive, the conclusion becomes rather simple, and fairly consistent with the results obtained in the other groups. In such an event the interpretation becomes that the hypothesis of distributional properties correlating to sound symbolic association scores is false, and the two processes operate independently of each other. However, given the significance levels of both Pearson's and Spearman's test, both at approximately $p = 0.02$, the probability of a false positive is relatively low (~2%).

5.2.1.2. Type 2

Compared to false positive, the probability of false negative is of more serious concern. This is due to the sample sizes for the individual groups in the study, $N = 20$, are only barely large enough for the statistical tests to return reliable results. Such a scenario is conceptually plausible as it would signify the distributional properties of *all* participants' writing correlate with the strength of their sound symbolic associations in a consistent fashion. The interpretation then becomes that the hypothesised relationship is indeed present across different groups. This, together with the group-wise comparison results, could be understood as demonstrating an important and universal relationship.

5.3. Implications

The results obtained in this study affirm the relationship between the writing systems and the strength of sound symbolic associations. They also lend limited support to the idea of this strength of associations being correlated with the distributional properties of individuals' writing, although apparently conditionally. Depending on the interpretation, these findings could signify a range of consequences.

Accepting a correlation between distributional properties of writing and sound symbolic associations strength implies a role of language usage in shaping broader cognitive processes. If the way we use our writing system correlates with the way we associate concepts cross-modally, this could suggest that either our behaviour (writing) itself shapes the way we conceptualise information, or that these associations shape our (linguistic) behaviour. Alternatively, both could be true, with a circular feedback model where writing influences mental associations, which in turn influence the way we write. This last interpretation would likely be the safest, although the exact mechanism of the influence, in either direction, remains uncertain.

Beyond these general considerations, one of the most interesting findings of this study has been the detection of the correlation within a single participant group only. Should this be representative of qualitative differences in language (or other cognitive) processing, questions arise as to the origin and the extent of such differences. While reminiscent of notions of linguistic relativism, such phenomenon could also have functional explanations. Humans do learn through repetition and writing is a repetitive task. Interestingly, the level of repetitiveness is directly related to the size of the a writing system's graphemic inventory, and comparing the inventory sizes for the systems included in this study we can find that these sizes typically include around thirty consonant and vowel symbols in Latin-based alphabets, around sixty in the Thai script (some of them used rarely or in narrowly defined circumstances, i.e. literary, religious etc. contexts), and *thousands* in the simplified Chinese characters system. It is likely that users of writing systems based on fewer symbols use each symbol more frequently (on average). This could result in the associations between the sound represented and the grapheme becoming stronger, which, in turn, could enhance the strength of sound symbolic associations. Such explanation would notably be consistent with the finding of this study (Figure 4.18).

Alternatively, accepting the statistical results at face value, another conclusion that can be drawn is that the hypothesised relationship is *only* present in the users of Latin scripts specifically. Following this line of reasoning, the question arises as to the cause of such anomaly. One possible explanation could be that the source of this phenomenon is in the notion that inspired this line of inquiry in the first place. For no writing system included in this study is the similarity between the shape of the graphemes and the shapes involved in the sound symbolic associations test as vivid and direct as for the Latin script. This would imply that different writing systems come with different levels of visual iconicity, i.e. different degree of sound symbolic congruence between their graphemes and the visual shapes associated with the sounds they represent. This, in turn, would imply that sound symbolism either (a) plays a role in the design of writing systems, (b) influences their evolution, or, conversely, (c) is mediated or modulated by the inherent sound symbolic properties of the writing. Assessing the first (a) of these propositions, it is important to consider that many writing systems have been demonstrated to have ideographic origins. Analysing the early forms of writing systems for traces of sound symbolic congruence could shed the light on the importance of this factor in the emergence of written forms. Similarly, the second (b) proposition would imply that as written forms evolve, sound symbolism influences the direction of their change. This notion would be in line with the assertion of Croft (2001) about iconicity in general, placing it in the role of a motivator of choice between competing forms. Similar conclusion is also drawn by Pernis et al. (2010), who propose that iconicity provides cross-modal cohesion to abstract linguistic pairings. Lastly, the third (c) proposition would suggest that writing systems have different degrees of retained sound symbolic congruence, which conditions their users towards stronger or weaker sound symbolic associations. An alternative interpretation to this would be that the sound symbolic-compatible regularities prominent in Latin scripts are the main *cause* of their users' susceptibility to those associations; this would, however, fail to align with the observation of distributional properties as a correlating factor.

Finally, it is notable that the significant correlations were detected in the users of the single writing system familiar to *all* participants of the study, yet this familiarity seems to have failed to translate into similar outcomes when performing the experimental task in other scripts. This could suggest that the effects of the writing system on sound symbolic associations are not generalised beyond the context of its actual usage, i.e. the strength of those associations may be contextually dependent on the task being performed, and by extension other circumstances like the language used, mode of communication, etc. This would present sound symbolism as a multivariate process activating with a varying strength depending on a complex set of parameters.

5.4. Limitations

The results obtained in the study, along with their interpretations, must be considered in the context of the limitations imposed by the design choices and the practical realities around the way the project was conducted.

5.4.1. Study design

A range of limitations trace back to the study's selected premise and design, ranging from the choice of writing systems, stimuli, and experimental setup to the participant groups' make up and the circumstances in which these participants performed the experimental tasks.

5.4.1.1. Scope

The structural differences between the three writing systems included in this study must be noted. From a functional perspective, Thai and Latin scripts operate by the exact same mechanism of linking speech sounds with graphic symbols, the major difference (further mediated by the specific language) being the availability of different representations for the same sounds. simplified Chinese characters, on the other hand, follow an entirely different model, one where the link is between *meaning* and graphic symbols. This major qualitative difference could mean that different experimental design is required to capture the specific processes responsible for the variance in the results collected from this group.

Likewise, the choice to investigate users of only three writing systems, two of them of similar origin and functionality, out of some few hundred in use globally, limits the generalisability of the findings. This is especially so considering that two of these writing systems (Latin script and Simplified Chinese characters) are amongst the most widely used in the world, which, on one hand, may be representative of some dominant properties, but may fail to accurately reflect the diversity of features and mechanisms present in human writing systems.

5.4.1.2. Stimuli choice

Another significant factor is that of the effect of study design choice of stimuli components. The sounds selected for the stimuli components may be perceived in different degrees of contrast to each other depending on the participants' L1 phonology. This can have a direct effect on both token count and entropy as measured within the design of this study. An example of this effect can be seen in the frequency of overlap between the representations of sounds /m/ and /b/ in the responses submitted by the participants of different groups. While some examples of this can be found in the answers obtained from the Latin scripts group, the prevalence of this overlap was clearly higher in the other two groups. Similar trends could be observed for other sound pairs like /t/ and /k/, or /e/ and /i/. Such differences in the perception of the speech sounds and the contrasts between them would have a direct result on the results of this study, with the most probable effect being to artificially inflate the values of distributional properties in an asymmetrical fashion. One way to account for this could be to introduce an additional step in the experiment, where the participants are asked to record themselves repeating the audio stimuli. This would allow for the analysis of the properties of the written form submitted based on the sound as perceived by the participants rather than the study author.

5.4.1.3. Experimental setup

One feature of the experimental set up in particular can be identified as having the potential to affect the outcomes. The order of stimuli presented to the participants was not randomised. This could result in the answers to subsequent questions being affected by the experience of dealing with the earlier ones. In particular, in experiment 2, the participants have performed the task with the images from question 1 and 2, used again in questions 3 and 4 (although in switched order, i.e. the images on the left in question 1 and 2 were displayed on the right in questions 3 and 4, respectively, and vice versa), with new audio stimuli for each try. As a result, the first two questions were answered by the participants with entirely novel images, while in the last two questions those images were already familiar to the participants, with one of the shapes already selected to match another audio stimulus. While this novel vs familiar difference can not be remedied beyond limiting the number of tries to match the number of contrasting image sets, randomising the order in which they are displayed could help to ameliorate the potential bias. Analogous assessment is also valid for experiment 1, with the participants being exposed to specific stimuli in a specific order, with a similar effect and solution possible.

Another design choice with the potential to affect the outcomes has been the form of the experimental task in experiment 1 itself. While the participants in the Latin and Thai script groups appeared comfortable with the procedure, a fairly common feedback from the participants in the Simplified Chinese characters group has been that of the task being very difficult. This was reported to be due to more than one reason. Firstly, some participants struggled with the concept of writing sound in Chinese characters as such, this is understandable considering that while such a task is possible to perform it goes against the way the characters are normally used, i.e. recording of meaning. The second reported issue has been the confusion in writing own dialect in Chinese characters, which were perceived by users as "Mandarin". This is understandable when considering the way the Chinese characters are used today, i.e. on electronic devices, where they are selected through a process involving typing the corresponding pinyin (latinised) form in Mandarin Chinese and then selecting the desired symbol from a list of suggestions. What speakers of other dialects of Chinese had to do in order to perform the experimental task was then to first conceptualise the sounds of audio stimuli as meaningful, similar-sounding words in their own dialects, and then use written forms, which usually bore an association with the sound of another dialect (Mandarin) to record them in writing. While, as intended, the effect of this process has been to some degree contained by the intermediate step of writing on paper before typing, it is clear that the nature and order of operations performed by the participants of this group has been different from the other two. The direct result of this issue would likely be weakened comparability of the results obtained from the Chinese group vs the others.

5.4.1.4. Sample

From a technical perspective the design choice of three groups of N = 20 participants has been the bare minimum requirement for the statistical analysis methods employed in the study. While the results obtained can be viewed as indicative of the relationship recorded, those conclusions would require confirmation from a more extensive study to be considered convincing in terms of statistical significance, with type 2 error, discussed earlier, the major factor of concern.

5.4.2. Other

Beyond the design-specific limitations, it is worth noting that assuming the reliability of the results obtained, those results conclude with a presence of a correlation between writing system and writing system, and the distributional properties of individual's writing in users of Latin scripts. What the results do not show is the direction of any causal interaction, i.e. accepting the existence of a relationship, its nature and mechanisms cannot be elucidated from the results obtained in this study.

5.5. Future research

Following the findings of the present study, three main directions of future research can be identified

5.5.1. Secondary data analysis

The data collected from the participants of the study allows for further analysis. A study by Fort et al. (2015) concludes that "consonants have a greater influence than vowels in the bouba-kiki effect". Likewise, Cwiek et al. (2022) observe that "bouba was reliably matched with the round shape in 22 out of 25 languages, whereas kiki was reliably matched to the spiky shape in only 11 out of 25 languages". These observations can be tested in the data collected, along with a detailed analysis for each of the participating eight sounds. Furthermore, an analysis of distributional properties for different sounds can be conducted to deepen our understanding as to the nature of sound symbolic associations. Similarly, an analysis similar to presented in this study (token count and entropy vs sound symbolic associations), but focused on one specific sound at a time can be conducted to compare their inherent properties. Such findings are not presented here as they go beyond the established premise of the study.

5.5.2. Confirmatory

While interesting, the findings of this study come with a significant level of uncertainty. To strengthen their reliability these results would need to be replicated by similar inquiries into the subject. Such future inquiries could include replication of the experiments performed in this study with larger and/or differently selected (e.g. monolingual individuals) or categorised (e.g. by specific L1 variety) population samples, different sets of writing systems or audio/visual stimuli, or differently controlled experiment performance circumstances (standardised environment, e.g. in-lab).

The present study's design provides a framework for further investigation of the same or similar questions, with all scripts and materials used available for scrutiny, reuse and adaptation to experimental examination of the same or related research problems.

5.5.3. Exploratory

Following the directions set out by the different implications of the current study, further questions could be investigated.

Accepting the general premise of writing systems influencing sound symbolic association, it would be interesting to examine how this influence manifests in users of multiple writing systems by testing their sound symbolic association scores following tasks requiring the use of different scripts. This could further our understanding of the nature of such influence, in particular to establish whether it is a constant or dynamic factor.

Following the more specific notion of distributional properties of writing correlating with sound symbolic associations strength, questions as to other aspects of writing could be pursued. An example of this could be an investigation of the influence of writing systems' apparent iconicity, e.g. through measurement of performance in a task of matching sounds to written forms in unfamiliar writing systems. This would help us to understand whether and, if so, to what extent such iconicity inherent in a writing system influences sound symbolic associations. Similar investigations could be conducted into aspects of writing like segment type, comparing users of alphabets to those of syllabic scripts, graphemic inventory size or the degree of overtness in writing (examining scripts that do and don't overtly encode all sounds, e.g. arabic script). All such features of functional variance could be potential candidates for interacting factors.

Conversely, following the idea of a relationship with opposite direction, i.e. of sound symbolism influencing writing, the nature of the constraint set by the writing systems could be investigated in order to determine the specific properties responsible. Likewise the extent of this influence on writing systems themselves could be investigated, diachronically comparing different stages of writing systems in terms of sound-shape correspondence *change*.

5.5.4. Peripheral

Pursuing the notion of interaction between writing and sound symbolic associations, the interplay between the properties of writing and other manifestations of sound symbolism could be explored. As a growing body of research appears to suggest other relationships of similar cross-modal nature, for example between trilled /r/ and roughness (Winter et al., 2022), it could be interesting to see whether such related phenomena can be demonstrated to interact in a similar way.

Another related direction to explore would be to pursue the link between properties of languages themselves and sound symbolic associations, e.g. by comparing phonetic inventory sizes, the presence or absence of phonological features like vowel length or tonality. Assuming an influence of L1 on sound symbolic associations warrants a wide range of investigations into the many distinctive features varying between languages as potential sources of influence.

6. Conclusions

The present research project attempted to provide new insights on the potential phenomena of relevance for sound symbolic associations. The broad aim of directly confirming the involvement of writing systems as a factor influencing the strength of such associations has been achieved in demonstrating clear differences in performance of users of quantitatively and qualitatively different scripts in sound symbolic associations tasks. The specific aim of mapping the relationship between distributional properties of individuals' writing and the strength of sound symbolic associations has been achieved to the extent of detecting such a relationship in users of Latin scripts *only*. These results indicate that distributional properties of writing are correlated to the strength of sound symbolic associations within the constraints set by the writing systems themselves. Given the sample sizes used in the study, and the asymmetry of results (the correlation detected in one participant group only), it is recommended to treat these findings as indicative but not conclusive.

Given the novel (to the best of author's knowledge) approach in choice of metrics compared, the results provide an interesting insight into the nature of sound symbolism and its interplay with writing systems and habits. If replicated, such results could suggest a new, interesting direction for further inquiry into the role of writing, and other modes of communication, in shaping sound symbolic associations, or other iconic phenomena. Similarly, other unique and differentiating factors tied to writing systems and language varieties could be considered for analysis of correlation following the findings of this study. Furthermore, an experimental framework used in this study is presented as potentially useful in such future investigations.

To explore the relationships suggested by these findings, further research is required to confirm the validity of the results and identify other factors involved as well as the direction of interaction along with its potential sources. This can be achieved by both direct and conceptual replication, with the present study's tools published and available for reuse.

While differences in performance in sound symbolic associations tasks based on both L1 and writing systems have been demonstrated before, the source of such differences has not yet been fully explained. The results of this study are humbly hoped to be a little step in the direction of such explanation.

7. References

Blasi, Damián & Wichmann, Søren & Hammarström, Harald & Stadler, Peter & Christiansen, Morten. (2016). *Sound-meaning association biases evidenced across thousands of languages.* Proceedings of the National Academy of Sciences of the United States of America. 113. http://doi.org/10.1073/pnas.1605782113

Bottini, Roberto & Barilari, Marco & Collignon, Olivier. (2019). *Sound symbolism in sighted and blind. The role of vision and orthography in sound-shape correspondences*. Cognition. 185. 62-70. http://doi.org/10.1016/j.cognition.2019.01.006

Croft, W. (2001). *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.

Cuskley, Christine & Simner, Julia & Kirby, Simon. (2017). *Phonological and orthographic influences in the bouba-kiki effect*. Psychological research. 81. http://doi.org/10.1007/s00426-015-0709-2

Ćwiek Aleksandra, Fuchs Susanne, Draxler Christoph, Asu Eva Liina, Dediu Dan, Hiovain Katri, Kawahara Shigeto, Koutalidis Sofia, Krifka Manfred, Lippus Pärtel, Lupyan Gary, Oh Grace E., Paul Jing, Petrone Caterina, Ridouane Rachid, Reiter Sabine, Schümchen Nathalie, Szalontai Ádám, Ünal-Logacev Özlem, Zeller Jochen, Perlman Marcus and Winter Bodo (2022). *The bouba/kiki effect is robust across cultures and writing systems* Phil. Trans. R. Soc. B3772020039020200390 http://doi.org/10.1098/rstb.2020.0390

Diessel, Holger. (2008). *Iconicity of sequence: A corpus-based analysis of the positioning of temporal adverbial clauses in English*. Cognitive Linguistics. 193. 465-490. http://doi.org/10.1515/COGL.2008.018

Dingemanse, M., Perlman, M., & Perniss, P. (2020). *Construals of iconicity: Experimental approaches to form–meaning resemblances in language*. Language and Cognition, 12(1), 1-14. doi:10.1017/langcog.2019.48

Drijvers, L., Zaadnoordijk, L., & Dingemanse, M. (2015). *Sound-symbolism is disrupted in dyslexia: Implications for the role of cross-modal abstraction processes*. In D. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, et al. (Eds.), Proceedings of the 37th Annual Meeting of the Cognitive Science Society (CogSci 2015) (pp. 602-607). Austin, Tx: Cognitive Science Society.

Eberhard, David M., Gary F. Simons, and Charles D. Fennig (eds.). 2023. *Ethnologue: Languages of the World. Twenty-sixth edition*. Dallas, Texas: SIL International. Online version: http://www.ethnologue.com

Fort, M., Martin, A., & Peperkamp, S. (2015). *Consonants are More Important than Vowels in the Bouba-kiki Effect*. Language and Speech, 58(2), 247–266. https://doi.org/10.1177/0023830914534951

Geeraerts, D.,Cuyckens, H. (2007). *The Oxford Handbook of Cognitive Linguistics*, Oxford University Press.

Haynie, H. J., Bowern, C., & LaPalombara, H. (2014). *Sound Symbolism in the Languages of Australia*. PLOS ONE, 9(4), e92852. https://doi.org/10.1371/journal.pone.0092852

Hinton, L., Nichols, J., & Ohala, J. (Eds.). (1995). *Sound Symbolism*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511751806

Imai, M., & Kita, S. (2014). *The sound symbolism bootstrapping hypothesis for language acquisition and language evolution*. Philosophical Transactions of the Royal Society B, 369(1651), 20130298. https://doi.org/10.1098/rstb.2013.0298

Kawahara, S., Godoy, M. C., & Kumagai, G. (2020). *Do Sibilants Fly? Evidence from a Sound Symbolic Pattern in Pokémon Names*. Open Linguistics, 6(1), 386–400. https://doi.org/10.1515/opli-2020-0027

Klensovsak, D., & Zaefferer, D. (2011). *Testing Cratylus' Hypothesis: Aspectual Iconicity*. Proceedings of the Annual Meeting of the Cognitive Science Society, 33. Retrieved from https://escholarship.org/uc/item/4740c8gm

Knoeferle, Klemens & Li, Jixing & Maggioni, Emanuela & Spence, Charles. (2017). *What drives sound symbolism? Different acoustic cues underlie sound-size and sound-shape mappings*. Scientific Reports. 7. 5562. http://doi.org/10.1038/s41598-017-05965-y

Köhler, W. (1947). *Gestalt Psychology: An Introduction to New Concepts in Modern Psychology*. Liveright.

Koriat, A., Levy, I. *The symbolic implications of vowels and of their orthographic representations in two natural languages*. J Psycholinguist Res 6, 93–103 (1977). https://doi.org/10.1007/BF01074374

Linell, P. (2005). *The Written Language Bias in Linguistics: Its Nature, Origins and Transformations* (1st ed.). Routledge. https://doi.org/10.4324/9780203342763

Lowe, R., Knutson, P., & Monson, M. H. (1985). *Incidence of Fronting in Preschool Children*. Language Speech and Hearing Services in Schools, 16(2), 119–123. https://doi.org/10.1044/0161-1461.1602.119

Lyons, J. (1981). *Language and Linguistics*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511809859

Meir, Irit , & Tkachman, Oksana (2014). *Iconicity*. obo in Linguistics. doi: 10.1093/obo/9780199772810-0182

Monaghan, P., Mattock, K., & Walker, P. (2012). *The role of sound symbolism in language learning*. Journal of Experimental Psychology: Learning, Memory, and Cognition, 38(5), 1152–1164. https://doi.org/10.1037/a0027747

Nöth, W. (2001). *Semiotic foundations of iconicity in language and literature*. The motivated sign: Iconicity in language and literature, 2, 17-28.

Perniss P., Thompson, R. L. & Vigliocco, G. (2010) *Iconicity as a general property of language: evidence from spoken and signed languages*. Front. Psychology 1:227. http://doi.org/10.3389/fpsyg.2010.00227

Perniss Pamela and Vigliocco Gabriella (2014) *The bridge of iconicity: from a world of experience to the experience of language* Phil. Trans. R. Soc. B3692013030020130300

Ramachandran, Vilayanur & Hubbard, Edward. (2001). *Synaesthesia - A Window Into Perception, Thought and Language*. Journal of Consciousness Studies. 8 (12). 3-34.

Sapir, E. (1929) *A study in phonetic symbolism*. Journal of Experimental Psychology, 12(3), 225–239. http://doi.org/10.1037/h0070931

Saussure, F., In Bally, C., & Sechehaye, A. (1966). *Course in general linguistics*. New York: McGraw-Hill.

Shinohara, K., & Kawahara, S. (2010). *A Cross-linguistic Study of Sound Symbolism: The Images of Size*. Proceedings of the Annual Meeting of the Berkeley Linguistics Society, 36(1), 396. https://doi.org/10.3765/bls.v36i1.3926

Sidhu, D.M., Pexman, P.M. (2018). *Five mechanisms of sound symbolic association*. Psychon Bull Rev 25, 1619–1643. https://doi.org/10.3758/s13423-017-1361-1

Sidhu, D. M., & Pexman, P. M. (2015). *What's in a Name? Sound Symbolism and Gender in First Names*. PLOS ONE, 10(5), e0126809. https://doi.org/10.1371/journal.pone.0126809

Stefanuto, M, Vallée, N (1999): *Consonant systems: from universal trends to ontogenesis*, In ICPhS-14, 1973-1976

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

Westbury, C. (2005). *Implicit sound symbolism in lexical access: Evidence from an interference task*. Brain and Language, 93(1), 10–19. https://doi.org/10.1016/j.bandl.2004.07.006

Westbury, C., Hollis, G., Sidhu, D. M., & Pexman, P. M. (2017). *Weighing up the evidence for sound symbolism: Distributional properties predict cue strength*. Journal of Memory and Language, 99, 122–150. https://doi.org/10.1016/j.jml.2017.09.006

Wichmann, S., Holman, E. W., & Brown, C. H. (2010). *Sound Symbolism in Basic Vocabulary*. Entropy, 12(4), 844–858. MDPI AG. Retrieved from http://dx.doi.org/10.3390/e12040844

Winter, B., Sóskuthy, M., Perlman, M., & Dingemanse, M. (2022). *Trilled /r/ is associated with roughness, linking sound and touch across spoken languages*. Scientific Reports, 12(1). https://doi.org/10.1038/s41598-021-04311-7

Appendix A: Stimuli generation script with example strings

In [2]:

```
import numpy as np
y=np.array([])
x=np.array(["ti","ki","te","ke","mo","bo","mu","bu"])

for n in range(6):
    np.random.shuffle(x)
    y=np.concatenate((y,x))

for n in range(6):
    print(y[8*n]+y[8*n+1]+y[8*n+2]+y[8*n+3])
    print(y[8*n+4]+y[8*n+5]+y[8*n+6]+y[8*n+7])
```

```
timobote
bumukike
tebumomu
tikekibo
tetibubo
kekimomu
tekekibo
butimomu
moteboki
kemutibu
timoboke
kitemubu
```

Appendix B: Entropy calculation script with example results

In [2]:

```
import numpy as np

a1 = [11,1]
a2 = [6,6]
a3 = [11,1]
a4 = [11,1]
a5 = [12]
a6 = [10,2]
a7 = [11,1]
a8 = [12]

n = 1.*np.sum(a1)

e1 = 0
for i in range(len(a1)):
    e1 = e1 - a1[i]*np.log(a1[i]/n)/n
e2 = 0
for i in range(len(a2)):
    e2 = e2 - a2[i]*np.log(a2[i]/n)/n
e3 = 0
for i in range(len(a3)):
    e3 = e3 - a3[i]*np.log(a3[i]/n)/n
e4 = 0
for i in range(len(a4)):
    e4 = e4 - a4[i]*np.log(a4[i]/n)/n
e5 = 0
for i in range(len(a5)):
    e5 = e5 - a5[i]*np.log(a5[i]/n)/n
e6 = 0
for i in range(len(a6)):
    e6 = e6 - a6[i]*np.log(a6[i]/n)/n
e7 = 0
for i in range(len(a7)):
    e7 = e7 - a7[i]*np.log(a7[i]/n)/n
e8 = 0
for i in range(len(a8)):
    e8 = e8 - a8[i]*np.log(a8[i]/n)/n

print(e1)
print(e2)
print(e3)
print(e4)
print(e5)
print(e6)
print(e7)
print(e8)
print((e1+e2+e3+e4+e5+e6+e7+e8)/8.)
print((len(a1)+len(a2)+len(a3)+len(a4)+len(a5)+len(a6)+len(a7)+len(a8))/8.)
```

```
0.2868359830561607
0.6931471805599453
0.2868359830561607
0.2868359830561607
0.0
0.45056120886630463
0.2868359830561607
0.0
0.2863815402063616
1.75
```

Appendix C: Statistics scripts with results

In [1]:

```
#Data input

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from scipy import stats

Cc = [3.625,3.625,3.125,3,4,4.25,4.75,4.875,3.625,2.75,5.875,4.25,2.375,2.5,4.625,5.37
CcS = pd.DataFrame(Cc)
Ce = [1.122229431,1.136387684,0.9327776423,0.8750153772,1.250097078,1.324397358,1.4926
CeS = pd.DataFrame(Ce)
Cb = [2,3,4,2,4,3,1,3,4,1,4,2,3,2,2,2,3,3,2,0]
CbS = pd.DataFrame(Cb)

Tc = [2.25,3,4.125,3.375,2.625,3.125,3.125,3.25,2.125,3,2.5,3.875,3,2.625,4.75,2.625,2
TcS = pd.DataFrame(Tc)
Te = [0.5485201425,0.7046586955,1.035021407,0.807623517,0.6113749152,0.785267153,0.769
TeS = pd.DataFrame(Te)
Tb = [2,2,2,4,2,1,3,4,3,4,2,4,2,3,4,2,0,4,2,2]
TbS = pd.DataFrame(Tb)

Lc = [1.5,1.625,2.5,1.5,2,2.375,2.625,1.75,2.25,2,1.625,1.625,1.75,1.375,1.375,1.75,1.
LcS = pd.DataFrame(Lc)
Le = [0.184349298,0.2758936228,0.5891040849,0.2129283051,0.3887920037,0.5406640905,0.5
LeS = pd.DataFrame(Le)
Lb = [3,3,2,3,1,2,2,4,2,4,2,4,4,4,4,4,1,3,1,4]
LbS = pd.DataFrame(Lb)

Ac = np.concatenate((Cc,Tc,Lc))
AcS = pd.DataFrame(Ac)
Ae = np.concatenate((Ce,Te,Le))
AeS = pd.DataFrame(Ae)
Ab = np.concatenate((Cb,Tb,Lb))
AbS = pd.DataFrame(Ab)
```

In [2]:

```
#Descriptive statistics for the Chinese group

print(CcS.describe())
print(CeS.describe())
print(CbS.describe())
```

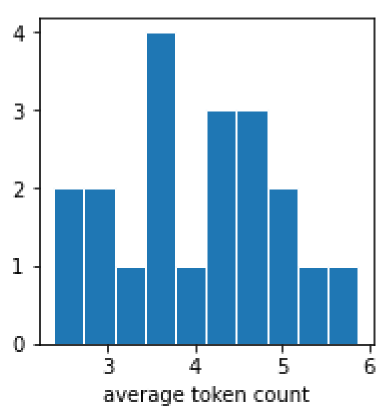
```
                0
count  20.000000
mean    4.012500
std     0.960486
min     2.375000
25%     3.500000
50%     4.125000
75%     4.656250
max     5.875000
                0
count  20.000000
mean    1.226559
std     0.310229
min     0.656682
25%     1.058515
50%     1.287247
75%     1.438780
max     1.762878
                0
count  20.000000
mean    2.500000
std     1.100239
min     0.000000
25%     2.000000
50%     2.500000
75%     3.000000
max     4.000000
```

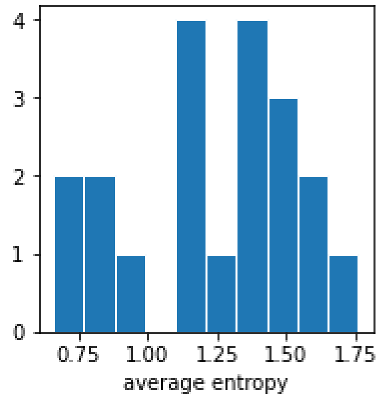
In [3]:

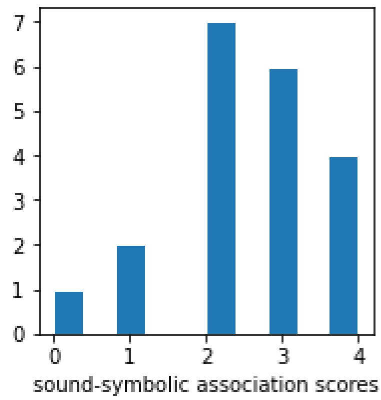
```
#histograms for the Chinese group

plt.figure(figsize=(3,3))
plt.hist(CcS, edgecolor='white')
plt.xlabel("average token count")

plt.figure(figsize=(3,3))
plt.hist(CeS, edgecolor='white')
plt.xlabel("average entropy")

plt.figure(figsize=(3,3))
plt.hist(CbS, edgecolor='white')
plt.xlabel("sound-symbolic association scores")
```

Out[3]:

```
Text(0.5, 0, 'sound-symbolic association scores')
```

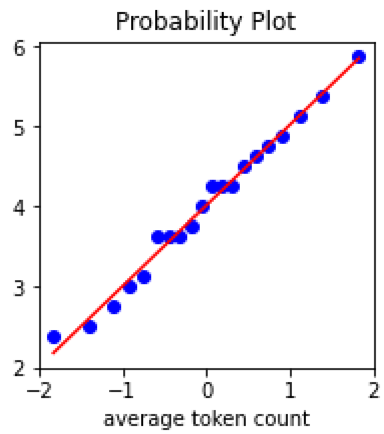


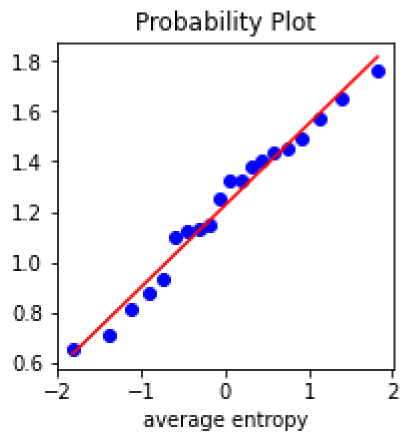


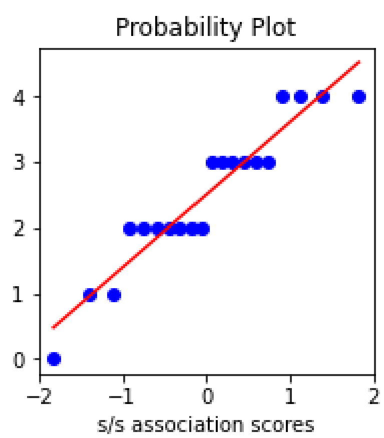


In [4]:

```
# QQ plots for the Chinese group:

fig = plt.figure(figsize=(3,3))
plot1=fig.add_subplot()
stats.probplot(Cc,plot=plot1)
plt.xlabel("average token count")
plt.ylabel("")
plt.show()

fig = plt.figure(figsize=(3,3))
plot2=fig.add_subplot()
stats.probplot(Ce,plot=plot2)
plt.xlabel("average entropy")
plt.ylabel("")
plt.show()

fig = plt.figure(figsize=(3,3))
plot3=fig.add_subplot()
stats.probplot(Cb,plot=plot3)
plt.xlabel("s/s association scores")
plt.ylabel("")
plt.show()
```







In [5]:

```
#Shapiro-Wilk for the Chinese group:

print(stats.shapiro(Cc))
print(stats.shapiro(Ce))
print(stats.shapiro(Cb))
```

```
ShapiroResult(statistic=0.9802921414375305, pvalue=0.9378172755241394)
ShapiroResult(statistic=0.9698916673660278, pvalue=0.7525834441184998)
ShapiroResult(statistic=0.9094840288162231, pvalue=0.06230909377336502)
```

In [6]:

```
#Descriptive statistics for the Thai group

print(TcS.describe())
print(TeS.describe())
print(TbS.describe())
```

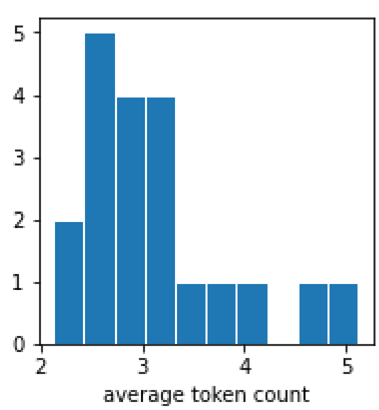
```
               0
count  20.000000
mean    3.150000
std     0.781783
min     2.125000
25%     2.625000
50%     3.000000
75%     3.281250
max     5.125000
               0
count  20.000000
mean    0.789254
std     0.225606
min     0.427878
25%     0.653451
50%     0.777260
75%     0.808798
max     1.397225
               0
count  20.000000
mean    2.600000
std     1.142481
min     0.000000
25%     2.000000
50%     2.000000
75%     4.000000
max     4.000000
```

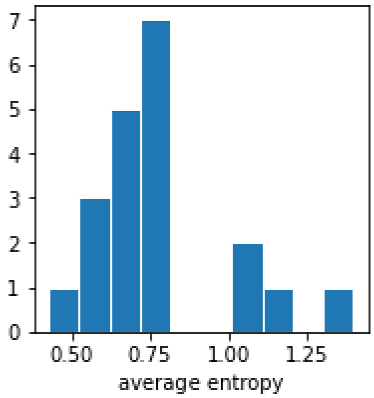
In [7]:

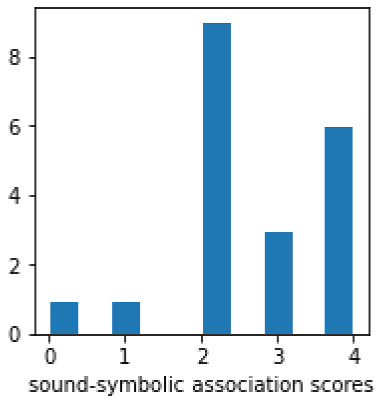
```
#histograms for the Thai group

plt.figure(figsize=(3,3))
plt.hist(TcS, edgecolor='white')
plt.xlabel("average token count")

plt.figure(figsize=(3,3))
plt.hist(TeS, edgecolor='white')
plt.xlabel("average entropy")

plt.figure(figsize=(3,3))
plt.hist(TbS, edgecolor='white')
plt.xlabel("sound-symbolic association scores")
```

Out[7]:

```
Text(0.5, 0, 'sound-symbolic association scores')
```

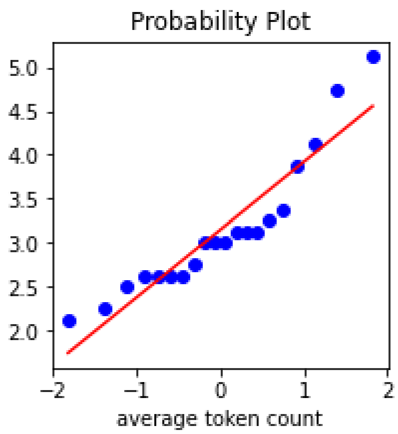


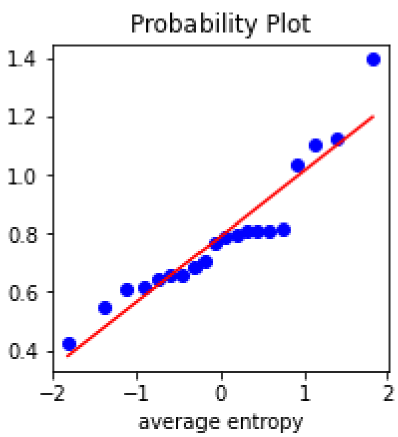


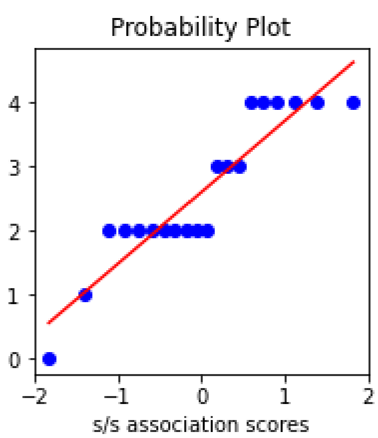


In [8]:

```
# QQ plots for the Thai group:

fig = plt.figure(figsize=(3,3))
plot1=fig.add_subplot()
stats.probplot(Tc,plot=plot1)
plt.xlabel("average token count")
plt.ylabel("")
plt.show()

fig = plt.figure(figsize=(3,3))
plot2=fig.add_subplot()
stats.probplot(Te,plot=plot2)
plt.xlabel("average entropy")
plt.ylabel("")
plt.show()

fig = plt.figure(figsize=(3,3))
plot3=fig.add_subplot()
stats.probplot(Tb,plot=plot3)
plt.xlabel("s/s association scores")
plt.ylabel("")
plt.show()
```







In [9]:
```
#Shapiro test of normality for the Thai group

print(stats.shapiro(Tc))
print(stats.shapiro(Te))
print(stats.shapiro(Tb))
```

```
ShapiroResult(statistic=0.8779652714729309, pvalue=0.01626601628959179)
ShapiroResult(statistic=0.900734543800354, pvalue=0.04256824031472206)
ShapiroResult(statistic=0.8555128574371338, pvalue=0.006602911744266748)
```

In [10]:
```
#Descriptive statistics for the Latin group

print(LcS.describe())
print(LeS.describe())
print(LbS.describe())
```

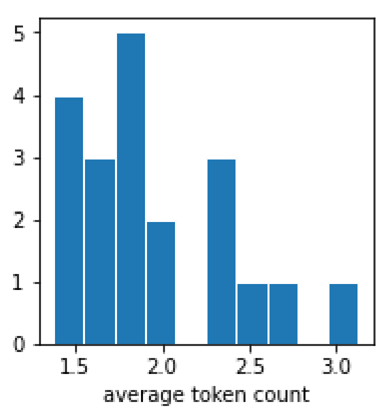
```
               0
count  20.000000
mean    1.931250
std     0.468542
min     1.375000
25%     1.625000
50%     1.750000
75%     2.281250
max     3.125000
               0
count  20.000000
mean    0.344618
std     0.167841
min     0.126037
25%     0.217975
50%     0.297224
75%     0.463138
max     0.712015
               0
count  20.000000
mean    2.850000
std     1.136708
min     1.000000
25%     2.000000
50%     3.000000
75%     4.000000
max     4.000000
```

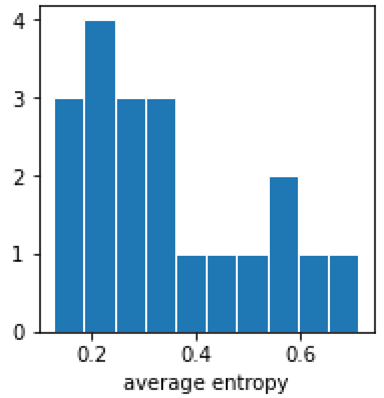
In [11]:

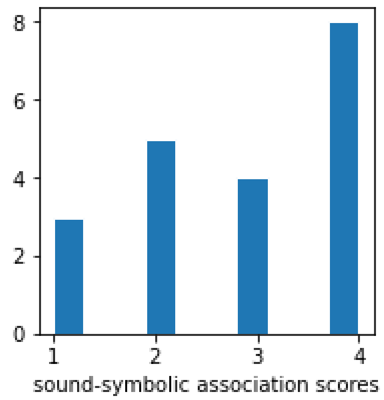
```
#histograms for the Latin group

plt.figure(figsize=(3,3))
plt.hist(LcS, edgecolor='white')
plt.xlabel("average token count")

plt.figure(figsize=(3,3))
plt.hist(LeS, edgecolor='white')
plt.xlabel("average entropy")

plt.figure(figsize=(3,3))
plt.hist(LbS, edgecolor='white')
plt.xlabel("sound-symbolic association scores")
```

Out[11]:

```
Text(0.5, 0, 'sound-symbolic association scores')
```

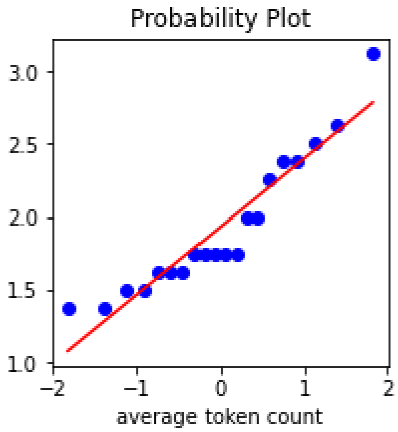


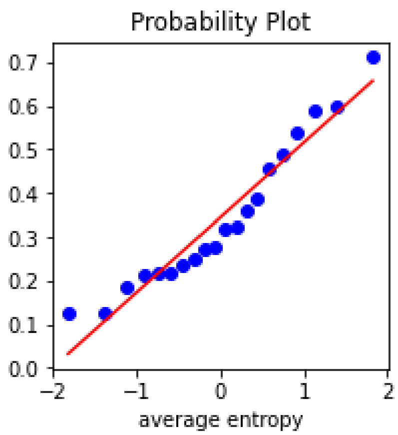


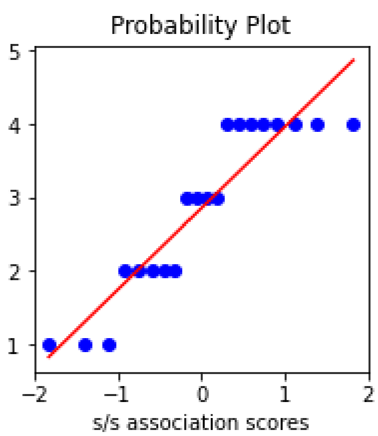


In [12]:

```
# QQ plots for the Latin group:

fig = plt.figure(figsize=(3,3))
plot1=fig.add_subplot()
stats.probplot(Lc,plot=plot1)
plt.xlabel("average token count")
plt.ylabel("")
plt.show()

fig = plt.figure(figsize=(3,3))
plot2=fig.add_subplot()
stats.probplot(Le,plot=plot2)
plt.xlabel("average entropy")
plt.ylabel("")
plt.show()

fig = plt.figure(figsize=(3,3))
plot3=fig.add_subplot()
stats.probplot(Lb,plot=plot3)
plt.xlabel("s/s association scores")
plt.ylabel("")
plt.show()
```







In [13]:

```
#Shapiro test of normality for the Latin group

print(stats.shapiro(Lc))
print(stats.shapiro(Le))
print(stats.shapiro(Lb))
```

```
ShapiroResult(statistic=0.8970510363578796, pvalue=0.03632199019193649)
ShapiroResult(statistic=0.9265977740287781, pvalue=0.13278581202030182)
ShapiroResult(statistic=0.8317881226539612, pvalue=0.0026799554470926523)
```

In [14]:

```
#Descriptive statistics for the entire sample:

print(AcS.describe())
print(AeS.describe())
print(AbS.describe())
```

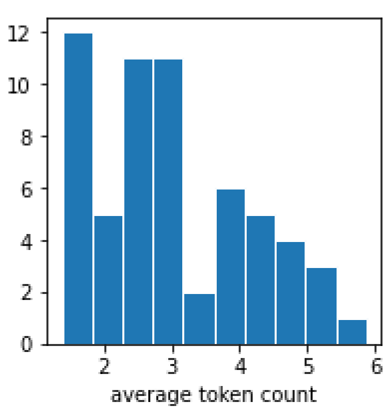
```
              0
count  60.00000
mean    3.03125
std     1.14278
min     1.37500
25%     2.21875
50%     2.87500
75%     3.78125
max     5.87500
               0
count  60.000000
mean    0.786810
std     0.433927
min     0.126037
25%     0.447523
50%     0.713043
75%     1.122793
max     1.762878
               0
count  60.000000
mean    2.650000
std     1.117276
min     0.000000
25%     2.000000
50%     3.000000
75%     4.000000
max     4.000000
```

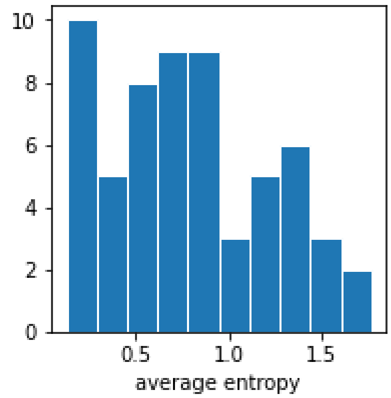
In [15]:

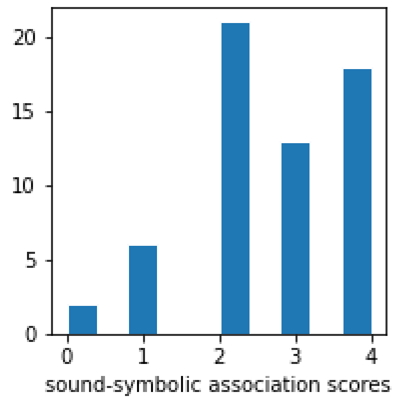
```python
#histograms for the entire sample:

plt.figure(figsize=(3,3))
plt.hist(AcS, edgecolor='white')
plt.xlabel("average token count")

plt.figure(figsize=(3,3))
plt.hist(AeS, edgecolor='white')
plt.xlabel("average entropy")

plt.figure(figsize=(3,3))
plt.hist(AbS, edgecolor='white')
plt.xlabel("sound-symbolic association scores")
```

Out[15]:

```
Text(0.5, 0, 'sound-symbolic association scores')
```

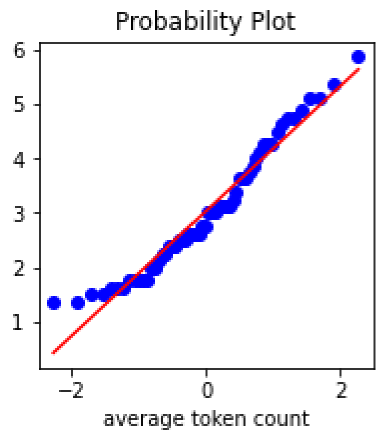


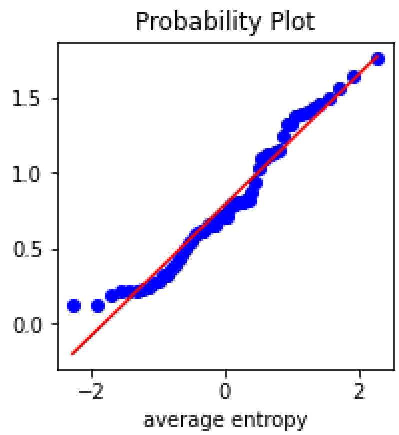


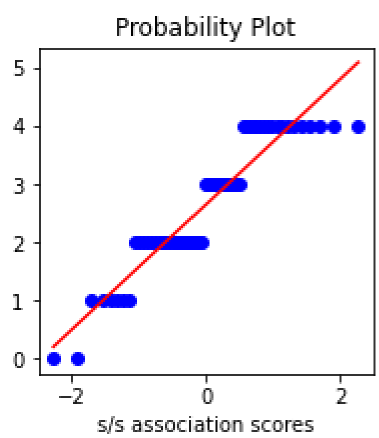


In [16]:

```
# QQ plots for the entire sample:

fig = plt.figure(figsize=(3,3))
plot1=fig.add_subplot()
stats.probplot(Ac,plot=plot1)
plt.xlabel("average token count")
plt.ylabel("")
plt.show()

fig = plt.figure(figsize=(3,3))
plot2=fig.add_subplot()
stats.probplot(Ae,plot=plot2)
plt.xlabel("average entropy")
plt.ylabel("")
plt.show()

fig = plt.figure(figsize=(3,3))
plot3=fig.add_subplot()
stats.probplot(Ab,plot=plot3)
plt.xlabel("s/s association scores")
plt.ylabel("")
plt.show()
```







In [17]:

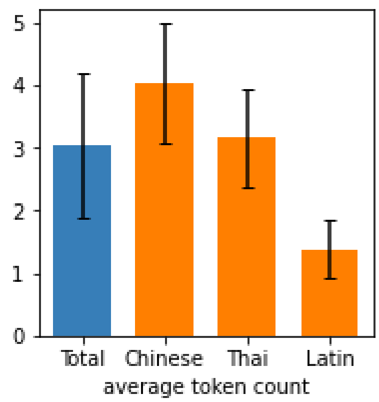
```
#Shapiro test of normality for the entire sample

print(stats.shapiro(Ac))
print(stats.shapiro(Ae))
print(stats.shapiro(Ab))
```

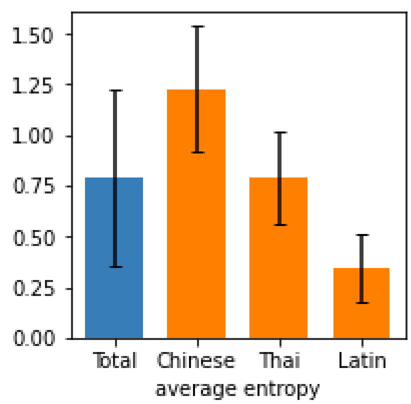
```
ShapiroResult(statistic=0.9511832594871521, pvalue=0.017710423097014427)
ShapiroResult(statistic=0.954643726348877, pvalue=0.025885462760925293)
ShapiroResult(statistic=0.8758575320243835, pvalue=1.933111525431741e-05)
```

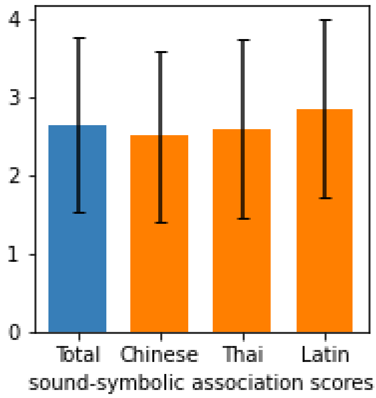
In [18]:

```
#Cross-group mean + std comparison:

labels = ["Total", "Chinese", "Thai", "Latin"]
cS = [3.03125, 4.012500, 3.150000, 1.375000]
errcS = [1.14278, 0.960486, 0.781783, 0.468542]
eS = [0.786810, 1.226559, 0.789254, 0.344618]
erreS = [0.433927, 0.310229, 0.225606, 0.167841]
```

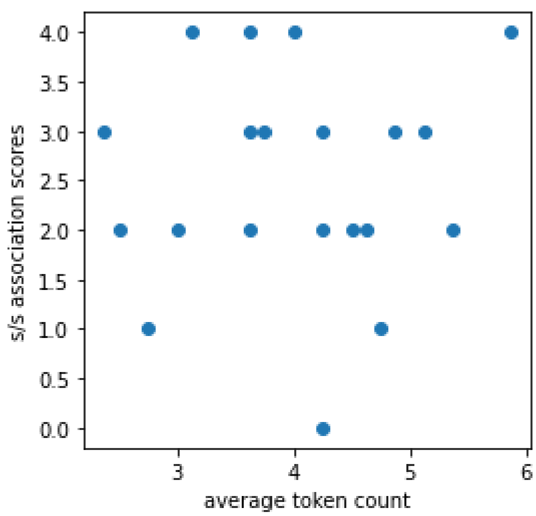
```
bS = [2.650000, 2.500000, 2.600000, 2.850000]
errbS = [1.117276, 1.100239, 1.142481, 1.136708]

colors = ["#377eb8","#ff7f00","#ff7f00","#ff7f00"]

plt.figure(figsize=(3,3))
plt.bar(labels, cS, color=colors, width=.7)
plt.errorbar(labels, cS, yerr=errcS, fmt="none", color="black",capsize=3)
plt.xlabel("average token count")

plt.figure(figsize=(3,3))
plt.bar(labels, eS, color=colors, width=.7)
plt.errorbar(labels, eS, yerr=erreS, fmt="none", color="black",capsize=3)
plt.xlabel("average entropy")

plt.figure(figsize=(3,3))
plt.bar(labels, bS, color=colors, width=.7)
plt.errorbar(labels, bS, yerr=errbS, fmt="none", color="black",capsize=3)
plt.xlabel("sound-symbolic association scores")

plt.show()
```

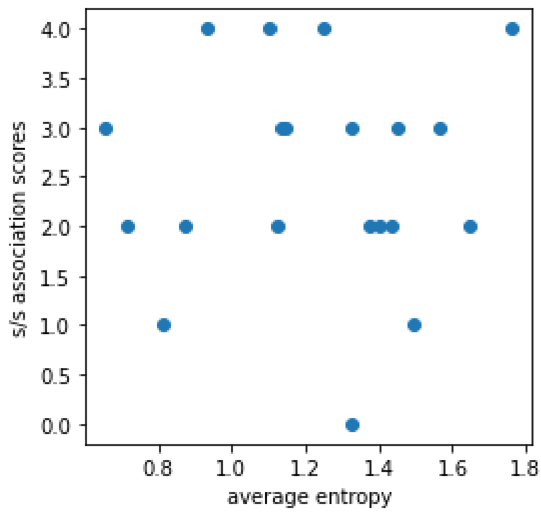






In [19]:
```
#scatterplots for the Chinese group:

plt.figure(figsize=(4,4))
plt.scatter(Cc,Cb)
plt.xlabel("average token count")
plt.ylabel("s/s association scores")
plt.show()

plt.figure(figsize=(4,4))
plt.scatter(Ce,Cb)
plt.xlabel("average entropy")
plt.ylabel("s/s association scores")
plt.show()
```

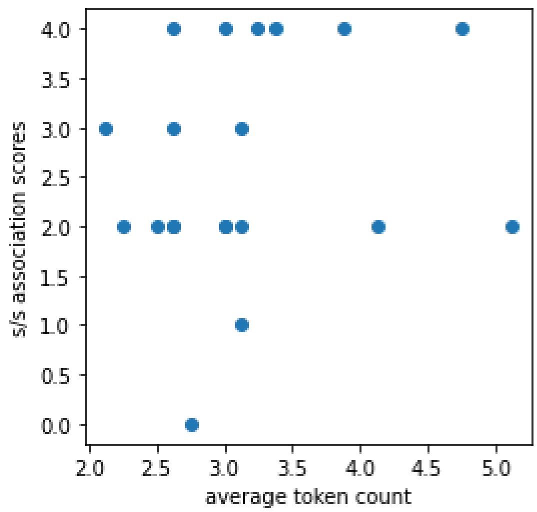


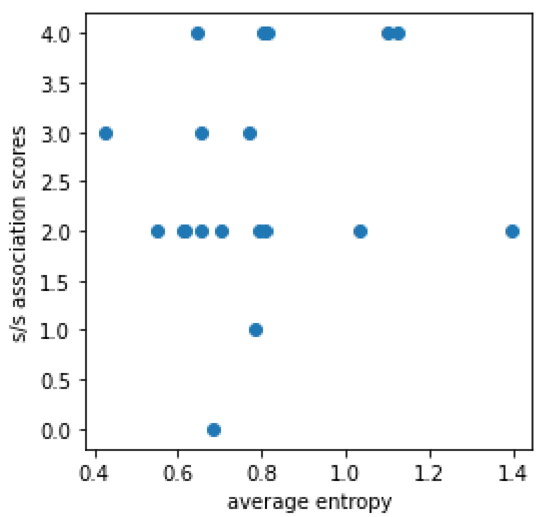


In [20]:
```
#correlations tests for Chinese group:

print(stats.pearsonr(Cc,Cb))
print(stats.spearmanr(Cc,Cb))


print(stats.pearsonr(Ce,Cb))
print(stats.spearmanr(Ce,Cb))
```

```
(0.05603003239876689, 0.8145010679759168)
SpearmanrResult(correlation=0.010203013031189584, pvalue=0.9659469962080711)
(0.02358938777934503, 0.9213642494099029)
SpearmanrResult(correlation=-0.02503946090933977, pvalue=0.9165460782633877)
```

In [21]:
```
#scatterplots for the Thai group:

plt.figure(figsize=(4,4))
plt.scatter(Tc,Tb)
plt.xlabel("average token count")
plt.ylabel("s/s association scores")
plt.show()

plt.figure(figsize=(4,4))
plt.scatter(Te,Tb)
plt.xlabel("average entropy")
plt.ylabel("s/s association scores")
plt.show()
```

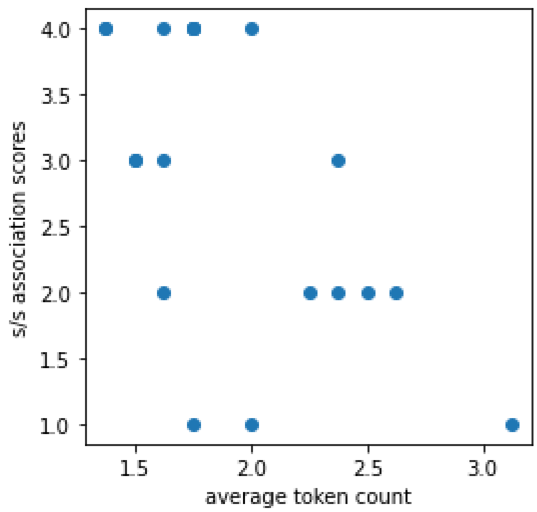


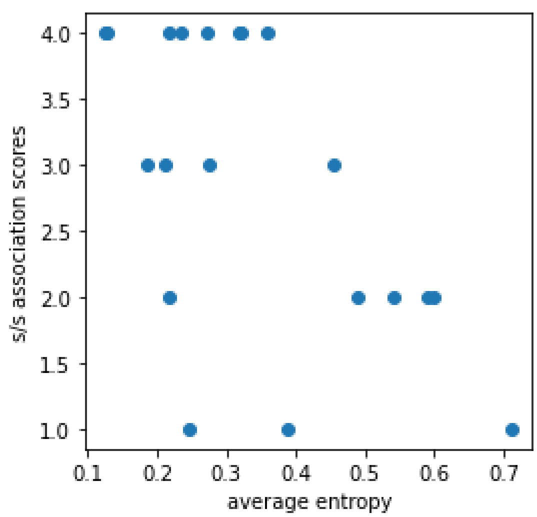


In [22]:
```
#correlations tests for Thai group:

print(stats.spearmanr(Tc,Tb))
print(stats.spearmanr(Te,Tb))
```

```
SpearmanrResult(correlation=0.21823374447197813, pvalue=0.3553100098070423)
SpearmanrResult(correlation=0.2676345631192341, pvalue=0.25395529372002634)
```

In [23]:
```
#scatterplots for the Latin group:

plt.figure(figsize=(4,4))
plt.scatter(Lc,Lb)
plt.xlabel("average token count")
plt.ylabel("s/s association scores")
plt.show()

plt.figure(figsize=(4,4))
plt.scatter(Le,Lb)
plt.xlabel("average entropy")
plt.ylabel("s/s association scores")
plt.show()
```

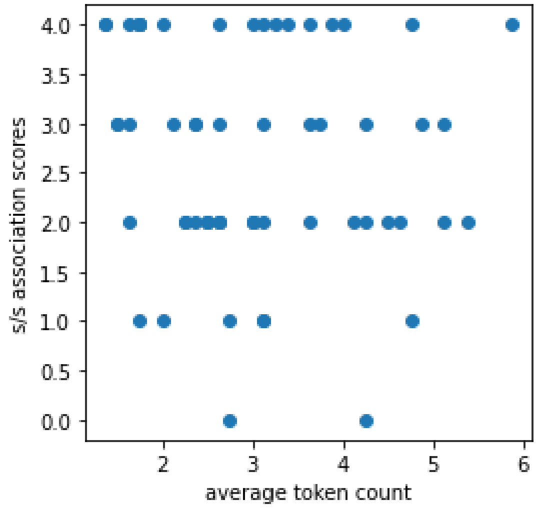


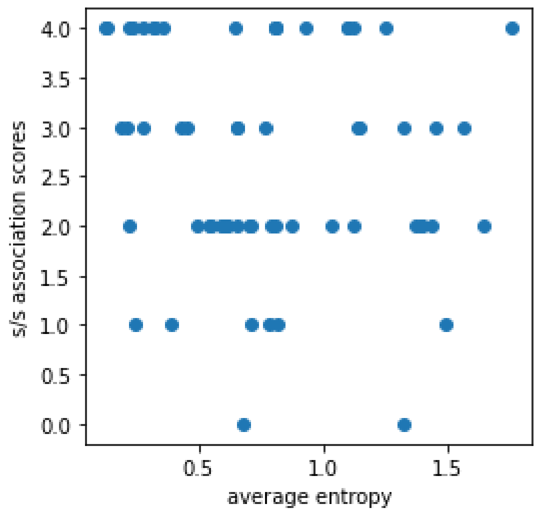


In [24]:
```
#correlations tests for Latin group:

print(stats.spearmanr(Lc,Lb))
print(stats.spearmanr(Le,Lb))
```

```
SpearmanrResult(correlation=-0.5253964493862663, pvalue=0.017361016232284168)
SpearmanrResult(correlation=-0.5174736487453586, pvalue=0.019450722432289)
```

In [25]:
```
#scatterplots for the entire sample:

plt.figure(figsize=(4,4))
plt.scatter(Ac,Ab)
plt.xlabel("average token count")
plt.ylabel("s/s association scores")
plt.show()

plt.figure(figsize=(4,4))
plt.scatter(Ae,Ab)
plt.xlabel("average entropy")
plt.ylabel("s/s association scores")
plt.show()
```





In [26]:

```
#correlations tests for entire group:

print(stats.spearmanr(Ac,Ab))
print(stats.spearmanr(Ae,Ab))
```

```
SpearmanrResult(correlation=-0.14563968614898104, pvalue=0.26686045598674035)
SpearmanrResult(correlation=-0.15586211310896758, pvalue=0.23436492669937137)
```