1 Introduction

We come to experience the world via our senses, traditionally divided into sight, sound, touch, taste, and smell. The senses do not live in isolation but are deeply connected. Perceptual psychology provides ample evidence for interactions between the senses, such as illusions, where one sense can change what the other one perceives. In the double flash illusion, for example, a single light flash (sight) accompanied by two beeps (sound) makes people see two light flashes (Shams et al., 2000). There also is evidence for a variety of crossmodal correspondences (Spence, 2011), such as people matching higher-pitched sounds to brighter images (e.g, Marks, 1974) and lower-pitched sounds to rounder shapes (e.g., O'Boyle & Tarte, 1980). Finally, there is a phenomenon called synesthesia, rare in the general population (Deroy & Spence, 2013; Simner et al., 2006, 2009), where certain individuals have vivid atypical sensory experiences often involving connections between the senses, such as automatically seeing colors when hearing particular sounds.

Just as the senses are richly interwoven in perception, they are also deeply connected in language (e.g., Alvarado et al., 2024; Classen, 1993, Ch. 3; Winter, 2019a). The rare phenomenon of synesthesia lends its name to linguistic synesthesia, otherwise known as synesthetic metaphor[1] (key references in the annotated bibliography Strik-Lievers, 2023). Linguistic synesthesias combine linguistic expressions that are strongly associated with different modalities, thus generating a conflict between separate sensory concepts (Strik-Lievers, 2017). For example, in the English adjective-noun pair *smooth melody*, the adjective is associated with touch, but it modifies a noun that describes an auditory concept. Linguistic synesthesias are generally seen as involving a metaphorical transfer *from* one sense (source) *to* the sense the expression is about (target) (but see Paradis & Eeg-Olofsson, 2013; Rakova, 2003; Winter, 2019a, 2019b). In adjective-noun pairs such as *smooth melody*, the noun is the head of a noun phrase, and ultimately what the phrase is about. Under a metaphorical analysis, this noun is seen as the target of the intersensory transfer, whereas the modifying adjective is seen as the source. The phrase *smooth melody* would then be characterized as featuring a touch→sound (source→target) transfer. Syntactic configurations other than adjective-noun pairs can usually be analyzed in the

---

[1] This team of authors has different views on the nature of linguistic synesthesias, with one author seeing them as metaphors (Strik-Lievers, 2017), the other one seeing them as literal expressions (Winter, 2019b). For the topic of this study, this theoretical issue is irrelevant. The asymmetries that are the focus of this Element can be described irrespective of how the expressions are classified; both metaphorical projection and contextual modification of literal expressions can model asymmetries (cf. discussion in Winter, 2019a, pp. 102–103). We chose to adopt the label "linguistic synesthesia" here as it is commonly employed in the literature and does not commit ourselves to a particular interpretation, unlike the equally frequent "synesthetic metaphor."

same way, in which case the target always relates to the sensory modality the expression is about, whereas the modality conflicting with it is the source. For instance, the copular clause *This melody is smooth* would be treated as also featuring a touch→sound transfer.

Starting with Stephen Ullmann (1937, 1945, 1946, 1947, 1959[1951]), many researchers have noted that there are striking patterns in linguistic synesthesias within individual languages as well as across languages, specifically with respect to which senses are more likely to be combined with which others, and whether a sense tends to be a source or a target (e.g., Day, 1996; Kumcu, 2021; Ronga et al., 2012; Shen, 1997; Strik-Lievers, 2015; Winter, 2019a; Zhao et al., 2019). For example, in adjective-noun pairs, researchers have noted that touch (source) adjectives are commonly combined with auditory (target) nouns (e.g., *smooth melody*, *rough/abrasive/hard/blunt sound*, etc.), but the reverse happens very rarely, that is, auditory adjectives such as *loud*, *quiet*, or *squealing* rarely modify tactile nouns such as *touch* or *feeling*. This asymmetry between sound and touch has been found in a number of studies on different languages (e.g., Day, 1996; Kumcu, 2021; Ronga et al., 2012; Shen, 1997; Strik-Lievers, 2015; Ullmann, 1959; Winter, 2019a; Zhao et al., 2019).

Ullmann paved the way for future investigations of linguistic synesthesia. Based on the analysis of literary texts in French, Hungarian, and English, he established the following generalizations about linguistic synesthesia (1959, pp. 280–283):

(i)   Intersensory transfers tend to go from "lower" to "higher" modalities
(ii) Touch is the most common source modality
(iii) Sound is the most common target modality

Later work reinterpreted Ullmann's generalizations in the context of implicational hierarchies in linguistic typology, with Viberg (1983) being the first to formulate the following linear representation (p. 159):

(iv) touch > heat > taste > scent > sound > sight

As a Western-European scholar, it is perhaps no surprise that already in his earliest paper, Ullmann (1937) relied on the Western cultural idea that hearing and especially sight are more advanced than touch, taste, or smell (Classen, 1993). The hierarchy above could then be characterized in terms of the senses being sorted from "lower" to "higher," with the former mapping onto the latter more likely than the reverse. Or, in other words, presumed-to-be lower senses are common sources in linguistic synesthesia; presumed-to-be higher senses are common targets. Perhaps aided through Viberg's representation, Ullmann's generalizations have later often been interpreted as a hierarchy that is "linear" (Jo, 2022, p. 284; Kumcu, 2021, p. 241; Zhong et al., 2023, p. 3). Here, we refer

to any proposal that assumes either a linear ordering or a binary divide between the lower and higher senses as "the hierarchy of the senses," notwithstanding important differences between theoretical proposals (for discussion, see Shinohara & Nakayama, 2011; Winter, 2019a, Ch. 8). We will deal with different versions of the hierarchy later in this study.

Linguistic synesthesias have been studied both experimentally and observationally. In observational studies, such as those by Ullmann and the others mentioned above, the evidence consists in the frequency with which specific combinations of the senses appear in texts or dictionaries, and the prevalence of each sense as a source versus as a target. On the other hand, experimental studies (Fishman, 2022; Nakamura et al., 2010; Sakamoto & Utsumi, 2014; Shen & Gil, 2008; Shinohara & Nakayama, 2011; Werning et al., 2006; Winter & Strik-Lievers, 2023; Zhong et al., 2023) use linguistic synesthesias as stimuli.[2] For example, Shen and Aisenman (2008) show that Hebrew linguistic synesthesias aligning with the hierarchy of the senses, such as *sweet fragrance* (taste→smell), are judged to be more natural and are more easily retained in memory than those conflicting with the hierarchy, such as *fragrant sweetness* (smell→taste). Similar results come from Shinohara and Nakayama (2011), who show that synesthetic adjective-noun phrases in Japanese that conform to the hierarchy are judged to be more understandable than those phrases that violate the hierarchy. Similarly, German speakers tested by Werning et al. (2006) rate hierarchy-congruent synesthesias as more accessible than incongruent ones. With both sources of evidence – experiments and observational studies – pointing toward a similar set of asymmetries between the senses, such as between sound and touch, the question naturally arises: What explains these asymmetries? *Why* do so many patterns look, at first sight, like they follow a hierarchy of the senses?

Explanations for the hierarchy abound (for reviews, see Fishman, 2022; Winter, 2019a, Ch. 9), with Williams (1976) being the first prominent paper to speculate that asymmetries in linguistic synesthesia may stem from extralinguistic asymmetries between the senses, in his case, biological asymmetries. He observed that the way the senses relate in his data loosely parallels the

---

[2] It is worth pointing out that observational and experimental evidence for the hierarchy of the senses cannot be treated as independent from each other, as experiments in this space are inherently language-based. Even in studies where stimuli are novel instances of linguistic synesthesia, not attested in corpora, the frequency of use of the individual lexemes, together with other linguistic features, can influence outcomes. Werning et al. (2006), for instance, show that accessibility ratings depend not only on the sensory modality of the lexemes but also on their corpus frequency and morphological features, with more frequent and nonderived adjectives enhancing the accessibility of the synesthetic expressions in which they appear (see also Winter & Strik-Lievers, 2023).

development of the senses in human phylogenetic evolution as well as in the ontogenetic maturation of the senses in infants, with touch being an evolutionarily ancient sensory modality that is already active at birth and the other senses sequentially emerging thereafter. Thus, Williams (1976) seeks to explain observations about asymmetries between sensory terms *in language* with something that is external to language, such as biological facts about perception.

Seeking a direct link between language and language-external facts also characterizes some strands of modern work on linguistic synesthesia, such as proposals which see Ullmann's generalizations from the perspective of embodiment (e.g., Zhao et al., 2019, 2022). The embodied cognition framework – not without its fair share of critics (e.g., Goldinger et al., 2016; Hickok, 2014; Leshinskaya & Caramazza, 2016; Mahon & Caramazza, 2008; Mahon & Hickok, 2016; Morey et al., 2021) – emphasizes the role of low-level processes involving perception and action in higher-level processes such as language and cognition. In the context of linguistic synesthesia, researchers have attempted to distinguish the senses in terms of the "degree of embodiment" (Zhong et al., 2023, p. 2), or variously in terms of how the senses differ with reference to notions such as "accessibility" or "concreteness" (Shen, 1997; Shen & Aisenman, 2008; Shen & Cohen, 1998; Shen & Gadir, 2009; Shen & Gil, 2008). Shen (1997, pp. 54–55), for example, suggests that sensory modalities involving direct contact with the source of the stimulus, such as touch, are more accessible than those not involving contact, like sight. Shen and colleagues also characterize touch and, to a lesser extent, taste and smell as more "concrete" senses than hearing and sight (Shen & Cohen, 1998; Shen & Gil, 2008), and this difference in "concreteness" is supposed to lie behind the observed asymmetries between the senses in language.

These proposals are situated within cognitive linguistics and explicitly link the hierarchy of linguistic synesthesia to conceptual metaphor theory, the general proposal that metaphors are rooted in facts about human cognition, specifically the tendency to think and talk about the abstract in terms of the concrete (Gibbs, 1994; Kövecses, 2002; Lakoff & Johnson, 1980, 1999). With other biological or perceptually grounded explanations, these proposals share the view that factors external to language govern what senses form likely sources or targets. This has parallels with the typological literature on how sensory meanings are connected in the semantics of perception verbs across languages, where Viberg (1983) has observed comparable hierarchical tendencies that are also often understood to be at least partially grounded in facts about human perception (Evans & Wilkins, 2000; Viberg, 2001).

Alongside approaches that seek language-external explanations, such as "embodiment," there are studies that have provided evidence for language-internal

factors determining some of the observed asymmetries. For example, Strik-Lievers (2015) focused on the internal composition of the sensory vocabulary with respect to part-of-speech categories (cf. Strik-Lievers & Winter, 2018), showing that in her data, those senses that are more lexically differentiated in the adjectival domain also feature more commonly as sources. This helps explain why sight→sound expressions (e.g., *bright sound*) are more common than sound→sight expressions: the word list used by Strik-Lievers (2015) features more adjectives for sight than sound, and given this distribution, sight has a greater opportunity to act as a source. Winter (2019a) provided additional evidence that the composition of the sensory vocabulary can help explain asymmetries observed in corpus data. He also provided empirical evidence for three other lexical factors: word frequency, iconicity, and emotional valence. For example, adjectives that are relatively more emotional are more likely to be used as sources in linguistic synesthesia, which helps explain why taste commonly features as source (e.g., *sweet melody*), given that taste vocabulary has strong emotional connotations (Bagli, 2017; Winter, 2016). Uncovering another lexical factor that may play a role in linguistic synesthesia, Petersen et al. (2008) provide some evidence consistent with the idea that scalar adjectives are more likely used as sources than nonscalar ones. As the sight vocabulary features many nonscalar color terms, this helps explains why other than when combined with sound, sight does not commonly feature as a source in linguistic synesthesias. These studies clearly show that language-internal explanations need to be considered alongside language-external ones (cf. Fishman, 2022).

In this study, we argue that before even beginning to discuss what explains the hierarchy of the senses, we need to revisit the available evidence base and ascertain whether the observational data actually lines up with common formulations of the hierarchy; that is, do the available data patterns follow what the hierarchy predicts? Most studies have essentially reapplied the conceptual mold of the hierarchy that was already present in Ullmann to more and more datasets. The predominant perspective has been one of seeking to confirm the hierarchy of the senses. With this comes a particular vantage point from which the data is seen: the existence of a hierarchy of the senses (be it cultural, cognitive, or linguistic in nature, depending on the explanatory approach) is usually *assumed*, and empirical studies are aimed at assessing the degree to which observational and experimental data conform to it. Thanks to a history that spans nearly a century of data collection efforts, we can use the existing data to revisit this vantage point and explore the extent to which different analytical techniques yield different insights.

Starting with Ullmann, scholars working on linguistic synesthesia have published their results in such a way that the summary data can be easily extracted from publications (see Section 3). Moreover, since data presentation

generally follows Ullmann's original work quite closely, the data structures that can be extracted from published studies are highly comparable. As a result of this, we are now in the opportune position to conduct the first ever meta-analysis in this field of study, the statistical combination of results from different studies. The time is ripe to take stock of the available evidence and combine data sources for more general inferences on the nature of linguistic synesthesia. This has potentially profound influences for theory development in this space, as the accumulation of evidence allows us to characterize the empirical foundation that motivates theories of linguistic synesthesia in a more precise manner.

Our analyses probe the extant data using a range of analytical techniques that differ with respect to how much they take the hierarchy of the senses as given. In our analyses, we slowly move from a "top-down" perspective of the data towards more bottom-up approaches that make no assumptions about whether there is a hierarchy of the senses prior to seeing the data. By looking at the same data with different analytical approaches, we can, for the first time, assess how the hierarchy of the senses might or might not emerge under different methodological choices, thus also demonstrating how these choices directly impact theory building. Our study is thus concerned with the explanandum, the thing to be explained, and not the explanans, the explanation itself. As we first need to capture a phenomenon before seeking to explain it, the question we address here is arguably more fundamental than the explanatory question and directly impacts it. We ask: What is actually in the data that we seek to explain?

2 Why a Meta-Analysis of Linguistic Synesthesia Is Needed

There are many benefits of combining datasets into a meta-analysis (for general discussion, see Gurevitch et al., 2018; Schmidt, 1992). First, the hierarchy is often presumed to be cross-linguistically universal, or near-universal (e.g., Shen & Gil, 2008; Ullmann, 1959; Williams, 1976); however, individual studies generally focus on only one language at a time, or at most a handful of languages. By combining datasets from different languages, we move research on linguistic synesthesia a small step closer to linguistic typology, where claims about universals are generally based on samples of dozens, often even hundreds of languages (see, e.g., Dryer, 1992; Johansson et al., 2019; Koptjevskaja-Tamm et al., 2024; Urban, 2011; Winter et al., 2022; Youn et al., 2016). In stark contrast to the principles of linguistic typology, the literature on linguistic synesthesia has often been quick to claim universality on the basis of a couple of languages. For example, Williams (1976) spoke of universals against the backdrop of an analysis focused on just English and Japanese; Popova (2005)

says that Shen's (1997) extension of linguistic synesthesia research to Hebrew has confirmed the "universal character" of the directionality of mappings (p. 398); Shen and Gil (2008) consider the case for the universal nature of the hierarchy of the senses to be "substantially boosted" if Indonesian would also show the same tendencies (p. 6). Throughout all this research on linguistic synesthesia, we never find that different languages are combined into the same statistical model to assess cross-linguistic generalizability in a formal fashion.

In quantitative linguistic typology, on the other hand, it is common practice to create statistical models aimed at making inferences that go beyond individual languages. This is done by combining data from many different languages into a single dataset (see, e.g., Bickel, 2011; Dryer, 1992), and more specifically, to statistically factor out variation that is due to dependencies between languages emerging from genealogical relatedness and language contact (Bickel, 2015; Cysouw, 2010; Jaeger et al., 2011; Winter & Grice, 2021). If, for example, many different languages from the same language family show a particular asymmetry in linguistic synesthesia, such as touch→sound or taste→smell, this asymmetry could have emerged only once in the ancestral language, which would paint a substantially less universal picture than if the same asymmetry is found again and again across languages from different families. Thus, without formally addressing these dependencies, such as whether languages are genealogically related, claims of universality stand on uncertain grounds (Dunn et al., 2011; Roberts et al., 2015; Roberts & Winters, 2013). The meta-analytic dataset we discuss below, although clearly falling short of the balanced samples that commonly feature in typological studies, is the so-far largest cross-linguistic analysis of linguistic synesthesia, involving 38 datasets from 14 languages, including English, Hungarian, Romanian, French, Italian, German, Korean, Japanese, Spanish, Ancient Greek, Chinese, Latin, Turkish, and Tzotzil. This allows us to put claims of universality to a stronger test than was possible in any one of the individual studies this data comes from.

With respect to amplifying cross-linguistic generalizability, a meta-analysis also provides the opportunity to resolve between "seemingly contradictory research outcomes" (Gurevitch et al., 2018, p. 175), in this case, with respect to studies focused on different languages. For example, Yu (2003, p. 22) observed that his Chinese data "by chance or not, shows no evidence of sight being a less frequent destination than sound," thus apparently contradicting Ullmann's third generalization cited above. It is, however, a common misconception that if a result is found in study A and not in study B, we can logically conclude that the two studies produced different results (cf. "misconception 3" in Vasishth & Nicenboim, 2016, pp. 354–355). Instead, the results from both studies need to be combined to formally test for a difference. The meta-analytic

literature is replete with examples where conclusions change when evidence is accumulated across studies; for example, a meta-analysis can reveal effects that are absent when looking at any one study in isolation (e.g., Garg et al., 2008). So, to truly know whether Yu's remarks on Chinese being exceptional are valid, data from Chinese would have to be combined with data from other languages to statistically assess the difference between languages.[3]

A second advantage of performing a meta-analysis is increased statistical power, that is, one's ability to detect statistically reliable patterns, or in the context of a Bayesian analysis, increased precision with which an effect is captured. Both statistical power and precision are greatly affected by how many data points are available for statistical inference, and all else being equal, it is desirable to have more power/precision, which means that it is generally desirable to have more rather than fewer data points. In some specific cases, theoretical claims about the nature of linguistic synesthesia hinge directly on dataset size, as we discuss in Section 3.4 below.

A third advantage of performing a meta-analysis is that analytical approaches vary across studies on linguistic synesthesia, with different studies reporting different measures that are not always comparable (Winter, 2019a, pp. 214–216). For example, several studies have relied on a measure that reports the average percentage of cases that are deemed congruent with the hierarchy of the senses; for example, Shen (1997) reported 95% for Hebrew, Winter (2019a) reported 86% for English, Jo (2019) reported 85% for Korean, Kumcu (2021) reported 95% for Turkish, and Strik-Lievers (2015) reported 62% for English and 74% for Italian. As discussed in Winter (2019a, pp. 214–215), these figures, however, are hard to compare because different studies treat different intersense transfers as congruent with the hierarchy, depending on what specific theoretical model is taken as a baseline. Kumcu (2021) explicitly demonstrates this for his Turkish data, where the number of hierarchy-congruent cases varies from 68% to 95% depending on different interpretations of the hierarchy. A meta-analysis allows us to streamline such analyses to arrive at more consistent results by applying the same analytical methods to all datasets, thus facilitating comparisons across studies, and across languages.

Finally, a meta-analysis offers an opportunity to look at old data with fresh eyes (for an example, see Winter, 2022). Our methodological approach detailed below

---

[3] In fairness to Yu (2003), he did not actually analyze his Chinese data statistically. This, however, makes it even harder to assess whether the data patterns he highlights are truly exceptional. While qualitative research clearly has an important role to play in research on linguistic synesthesia, claims about the hierarchy of the senses can only be addressed statistically because hierarchical tendencies are inherently statistical generalizations. Ullmann himself stressed the statistical nature of his generalizations (Ullmann, 1959, p. 276).