

# A corpus-based study on synesthesia in Korean ordinary language

Charmhun Jo

Faculty of Humanities,  
The Hong Kong Polytechnic  
University, Hong Kong  
jch337@hotmail.com

## Abstract

Synesthesia means an involuntary neurological phenomenon where “sensory events in one modality take on qualities usually considered appropriate to another” (Marks, 1982, p. 15). More generally, it indicates an experiential mapping of one sense domain with another, such as “sweet sound”. The study reported in this paper is to test Ullmann’s (1963) theoretical framework of “hierarchical distribution” through the synesthetic data coming out of Korean National Corpus (KNC), focusing on modern daily Korean. The research questions here are (a) what are the routes for Korean synesthetic transfers like?, (b) what are the predominant source and target domain for the transfers?, and (c) what are the universal and/or culture-specific aspects in the association? Based on Strik Lievers et al.’s (2013) methodology, the study extracts synesthetic data from KNC. As a result, the data analysis shows that (a) Korean synesthesia conforms to Ullmann’s (1963) general scheme in the metaphoric mappings, (b) the predominant source domain is touch while the predominant target is hearing, which matches with Ullmann’s (1963) study as well, and (c) there could be a delicate cultural dependency, which means “taste” occupies a significant position together with “touch” in Korean synesthetic metaphors.

comes from the Ancient Greek σύν *syn*, “together”, and αἴσθησις *aisthēsis*, “sensation”. Basically, synesthesia refers to an involuntary neurological phenomenon where “sensory events in one modality take on qualities usually considered appropriate to another” (Marks, 1982, p. 15). To be more general, it means an experiential association of one sense domain with another, such as “sweet sound” and “cold color”. In linguistics, synesthesia is understood in terms of metaphor (Williams, 1976; Geeraerts, 2010), which means that a perceptual experience of one sense is described by lexical expressions associated with another. For example, “sweet sound” is linguistically synesthetic because the speaker expresses a perception of sound (“sound”) using a word related to taste (“sweet”), where “sound” becomes the target domain of the transfer and “sweet” is the source. The synesthetic metaphors were introduced by S. Ullmann (1963), where he proposed his theoretical framework of “hierarchical distribution” as a probable universal principle in the process of synesthetic mapping. In this light, the objective of the study is to test Ullmann’s (1963) theoretical framework using the synesthetic data from Korean National Corpus (KNC). Therefore, the research questions here are (a) what are the routes for Korean synesthetic transfers like?, (b) what are the predominant source and target domain for the transfers?, and (c) what are the universal and/or culture-specific aspects in the association?

## 1 Introduction

Synesthesia (also spelled synæsthesia or synaesthesia) has been an interesting research topic in diverse academic fields. The term synesthesia

## 2 Brief literature review

As the seminal work of synesthetic metaphors, Ullmann (1963), analyzing poetic writings of the nineteenth century in English, French, and Hungar-

ian, proposed his theoretical framework of “hierarchical distribution”, where he concluded three overall tendencies in synesthetic mappings. First of all, the majority of synesthetic transfers show the following direction: touch → heat → taste → smell → sound → sight. This transfers tend to move from the “lower” to the “higher” sensory domains, which is called “hierarchical distribution”. Second, based on the first tendency, the most frequent source domain of transfers is touch, the lowest level of sensation. Third, the most frequent target domain for synesthetic transfers is sound rather than sight.

Based on Ullmann’s (1963) study on data from poetry, Williams (1976) investigated the synesthetic transfer in daily language, namely, the historical change of meaning of synesthetic adjectives in daily English (together with some evidence from other Indo-European languages and Japanese as well). In sum, his results support Ullmann’s (1963) framework of “hierarchical distribution”, generalized as follows:

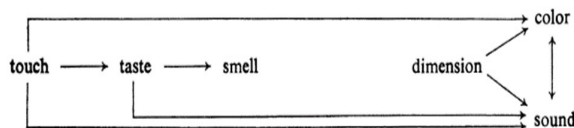


Figure 1. Synesthetic transfer route of Williams (1976)

Following Ullmann (1963) and Williams (1976), Yu (1992) applied their approaches to the data collected from Chinese literary and daily language. The conclusion of the research shows that Chinese synesthetic metaphors basically conform to their general schemes in metaphoric mappings. Yu (2003) also found almost the same results from the analysis of synesthetic data extracted from literary works written by current eminent Chinese novelist Mo Yan, examining synesthesia based on a cognitive perspective.

From the perspective of cognitive poetics, following Ullmann’s (1963) approach, Shen (1997) explored the directionality of transfer for synesthetic metaphors in Hebrew on the basis of the literary analysis of modern poetry and two experimental data. His findings reinforce Ullmann’s (1963) observation about the mapping in linguistic synesthesia. That is to say, the synesthetic expressions in the Hebrew language tend to map lower senses on to higher ones in their hierarchy. By way of the notion of “accessibility”, Shen

(1997) claims that the “low to high” transfer comes from the general cognitive constraints, where “a mapping from more ‘accessible’ or ‘basic’ concepts onto ‘less accessible’ or ‘less basic’ ones seems more natural and is preferred over the opposite mapping” (p. 51). Also, he points out that sight and sound are less accessible because they do not involve any direct contact with the perceived entity.

Recently, Strik Lievers (2015) reported a noticeable study about linguistic synesthesia by making use of corpora to investigate synesthetic transfers of English and Italian. Via a semi-automatic method for extracting synesthesiae from corpora, developed in Strik Lievers et al. (2013), she found large-scale data results and more clearly presented that the so-called principle of directionality just reflects the “frequency” of synesthetic connection types, adding a few interesting interpretations about the motivation of English and Italian synesthetic mappings.

### 3 Methodology

#### *Sensory domains*

Prior to the examination into synesthetic mappings in the linguistic text, sensory domains, or sensory modalities, are first to be designated. As a matter of fact, there is no agreement among scholars over how many sensory modalities there exist, and they can vary depending on researchers’ viewpoints and classificatory criteria (Strik Lievers et al., 2013; Strik Lievers, 2015). Most of synesthetic studies now follow the Aristotelian five-sense system of touch, taste, smell, hearing, and sight (cf. Cytowic, 1989; Shen, 1997; Strik Lievers 2015). Some of the studies, on the other hand, makes an adjustment to the above system. For instance, Ullmann (1963) separated “heat” from “touch”<sup>1</sup>, and Williams (1976) divided sight into two categories of “dimension” and “color”. Day (1996) is based on Ullmann’s (1963) taxonomy, while Yu (1992, 2003) follows Williams’s (1976). Lin and Hsien (2011) add “emotion” on the six senses of touch, temperature, taste, smell, hearing, and vision, and Zhao and Huang (2015) also take “emotion” into consideration along with the traditional five senses.

<sup>1</sup> However, Ullmann (1963) mentioned: “There is of course no harm in combining the two sets of data; actually this would only throw an even more glaring light on the general pattern”. (p. 278)

This study chooses the general Aristotelian sensory modes for broader reviews and comparisons.

**Data and corpus**

In this study, the synesthetic data will be collected from Korean National Corpus (KNC) well known as “21<sup>st</sup> century Sejong Project”. The 21<sup>st</sup> century Sejong Project is a comprehensive project aiming to build various kinds of language resources including Korean corpora, comparable to British National Corpus (BNC) (Aston & Burnard, 1998), and Korean electronic dictionaries. The KNC data basically have the raw corpora of modern Korean (written and spoken), North Korean, Korean used overseas, old Korean, and oral folklore literature. They include parallel corpora consisting of Korean and other languages such as English and Japanese, morph-tagged corpora, part-of-speech (POS)-tagged corpora, sense-tagged corpora, and a parsed corpus as well. Among them, the parsed corpus of modern written Korean will be selected for this study, because it is the only syntactically analyzed corpus and the contents are all composed of daily linguistic data. The Korean parsed corpus had been set up for 4 years from 2002 to 2006, and the size is 43,828 sentences (around 433,839 words) (NIKL, 2011).

**How to extract synesthetic metaphors from the corpus**

The study refers to Strik Lievers et al.’s (2013) methods to extract synesthetic data from KNC. The author firstly sets up the lexical items subdivided by five sensory modes respectively in terms of POS categorization of noun (N), adjective (A) and verb (V)<sup>2</sup>, and secondly, as for the synesthesia extraction from the corpus, a simplest method is applied that just lists all the sentences containing at least two perception-related words, given the fact that this simplest way can possibly collect the largest number of candidate sentences and the candidates will be affordable for the final manual checking because the corpus is not big. Finally, to sort out “true” synesthesiae, it is necessary to do a

<sup>2</sup> Regarding the POS matter of linguistic synesthesia, three parts of speech of noun, adjective and verb have to be taken into account because they are all able to engage in synesthetic connections (Strik Lievers et al., 2013). For example, “She has a golden [Adj/Source] voice [N/Target]”, “The flowers smell [V/Target] sweet [Adj/Source]” (Strik Lievers et al., 2013, p. 4).

handwork inspection of the extracted candidate output.

**4 Results and discussion**

**Sense-related word lists**

The total number of perception-related lexemes collected for this survey is 417 words. The summary is as follows:

	Touch	Taste	Smell	Sight	Hearing
N	31	15	28	68	54
A	52	31	8	47	6
V	12	8	12	25	20
Sub-total	95	54	48	140	80
Total	417				

Table 1. The distribution of sense-related words collected for the study

**Results**

The first finding is the whole result of synesthesia extraction from KNC, as summarized in Table 2. This data can show an overall outlook of corpus work on Korean synesthetic phenomena.

Total Corpus Sentences (TCS)	Extracted Positive Sentences (EPS)	True Positives (true synesthesiae) (TP)	TP / EPS (%)	TP / TCS (%)
43,828	1,250	100	8	0.23

Table 2. The total result of synesthesia extraction

The second is the overall synesthetic transfer route in Korean. It is directly concerning what the routes for Korean synesthetic transfers found in the corpus data are like. The result is displayed in the following figure:

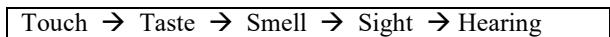


Figure 2. Overall synesthetic transfer route in Korean

The third one is the distribution of synesthetic mappings among sensory modes. This is practical informational data, which represent how frequent each mapping is and how many forward or backward transfers exist. The above result 2 is generalized from this data.

Target \ Source	Touch	Taste	Smell	Sight	Hearing	Total
Touch	0	3	3	11	20	37
Taste	1	0	8	9	15	33
Smell	0	0	0	1	2	3
Sight	2	1	4	0	13	20
Hearing	0	1	1	5	0	7
Total	3	5	16	26	50	100

Table 3. The distribution of Korean synesthetic mappings among sensory domains (TOKEN)<sup>3</sup>

### General discussion

The first issue to be discussed here is regarding the directionality of Korean synesthetic mappings. In a word, the result reported in this research at large conforms to the theory of “hierarchical distribution” by Ullmann (1963), as showed in Figure 2. The predominant source is touch, while the predominant target is hearing, as seen in Table 3, which also matches with the conclusion of Ullmann (1963). The above reports are confirmed again by Strik Lievers (2015) as well.

However, the universal tendency, as Strik Lievers (2015) noted, is not unidirectional but frequency-based. That is because some opposite transfers, or backward transfer types, are found, although the number of cases is remarkably low. More importantly, there could be found a delicate cultural dependency, or a subtle culture-based difference. In a closer observation on the finding data, a very noticeable point is detected in Korean synesthetic metaphor phenomena. It is with regard to the gustatory domain, taste, which works as a sec-

Target \ Source	Touch	Taste	Smell	Sight	Hearing	Total
Touch	0	2	3	9	14	28
Taste	1	0	7	8	12	28
Smell	0	0	0	1	2	3
Sight	2	1	4	0	10	17
Hearing	0	1	1	5	0	7
Total	3	4	15	23	38	83

<sup>3</sup> Table 4. The distribution of Korean synesthetic mappings among sensory domains (TYPE)

ond largest source modality of Korean synesthetic mappings investigated. The key point here is in that the difference of the proportion between the most and second frequent source sensory domains is very slight, as indicated in Table 5.

Touch	Taste	Sight	Hearing	Smell
37	33	20	7	3

Table 5. Korean source sensory domains in frequency-decreasing ordering (%)

This observation is comparable to Strik Lievers’s (2015) data in Table 6.

	Touch	Taste	Sight	Hearing	Smell
English	49.3	25.7	21.8	3.0	0.2
Italian	55.6	20.2	19.1	4.6	0.2

Table 6. English and Italian source sensory domains in frequency-decreasing ordering (%), adapted from Strik Lievers (2015)

On the other hand, the frequency of target modes in Korean synesthetic transfers is similar to the finding of Strik Lievers (2015). The comparison is displayed as follows:

	Hearing	Sight	Smell	Taste	Touch
Korean	50	26	16	5	3
English	52.3	28.0	12.4	5.3	2.1
Italian	50.2	42.5	3.8	3.0	0.2

Table 7. Target sensory domains in frequency-decreasing ordering in Korean, English, and Italian (%), merged with the data presented in Strik Lievers (2015)

Accordingly, this situation can imply that together with the tactile domain, touch, the sense of taste takes up a significant position in Korean or Asian cultural context, and so people in the cultural circle more often tend to describe something in terms of gustation or tactility, compared with western people. Such point of view can be strongly supported by Zhao and Huang (2015), who came to the following conclusion from their study on synesthetic metaphors in modern Chinese:



Figure 3. The hierarchy of synesthetic transfers among taste, touch, and smell in Chinese, excerpted from Zhao and Huang (2015)

## References

- Aristotle. *Poetics*. Translated by S. H. Butcher. (<http://www.gutenberg.org/files/1974/1974.txt>)
- Aston, G., & Burnard, L. (1998). *The BNC handbook: exploring the British National Corpus with SARA*. Capstone.
- Baron-Cohen, S. (n.d.). *Historical Perspective*. UK Synesthesia Association. Site: <http://www.uksynaesthesia.com/historical.htm>.
- Chung, I. (1997). *Synaesthetic Transfer of Korean Adjectives*. *Studies in Modern Grammatical Theories*, 11, 163-180. [in Korean]
- Cytowic, R. E. (1989). *Synesthesia: A Union of the Senses*. New York: Springer-Verlag.
- Cytowic, R. E. (1993). *The man who tasted shapes: A bizarre medical mystery offers revolutionary insights into reasoning, emotion, and consciousness*. New York: Putman.
- Day, S. (1996). *Synaesthesia and synaesthetic metaphors*. *Psyche*, 2(32), 1-16.
- Geeraerts, D. (2010). *Theories of lexical semantics*.
- Harris, R., & Taylor, T. J. (1997). *Landmarks in linguistic thought 1: The western tradition from Socrates to Saussure (Vol. 1)*. Psychology Press.
- Huang, C. R. (2015). *Towards a Lexical Semantic Theory of Synaesthesia in Chinese*. In *Keynote Speech in the 16th Chinese Lexical Semantics Workshop (CLSW-16)*. Beijing.
- Kang, B. & Kim H. (2004). *Sejong Korean Corpora in the Making*. Paper presented at the International Conference on Language Resources and Evaluation. Retrieved April 2, 2015, from <http://www.lrec-conf.org/proceedings/lrec2004/pdf/66.pdf>.
- Karwoski, T. F., Odbert, H. S., & Osgood, C. E. (1942). *Studies in synesthetic thinking: II. The role of form in visual responses to music*. *The Journal of general psychology*, 26(2), 199-222.
- Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. University of Chicago press.
- Lakoff, G., & Turner, M. (1989). *More than cool reasoning: A field guide to poetic metaphor*. University of Chicago press.
- Lee, S. (2015). *A study on synesthetic metaphor shown in Korean and Chinese advertisement from the viewpoint of cognitive linguistics*. [in Korean]
- Liddell, H. G., Scott, R., & Jones, H. S. (1940). *A Greek-English lexicon*.
- Lievers, F. S. et al. (2015). *Methods for the extraction of synaesthetic associations from corpora* (unpublished).
- Liu, H. & Huang, C. R. (2015). *The automatic extraction and mapping directionality of synaesthetic sentences of modern Chinese*. [in Chinese]
- Marks, L. E. (1982). *Synaesthetic perception and poetic metaphor*. *Journal of Experimental Psychology: Human Perception and Performance*, 8(1), 15.
- National Institute of Korean Language (NIKL). (2011). *21C Sejong Project final result (revised)*. <https://ithub.korean.go.kr>
- Park, G., (1978). *Synaesthetic mapping of poetic diction*. *Korean Language and Literature*, No.78. [in Korean]
- Ramachandran, V. S., & Hubbard, E. M. (2001). *Psychophysical investigations into the neural basis of synaesthesia*. *Proceedings of the Royal Society of London B: Biological Sciences*, 268(1470), 979-983.
- Sacks, Oliver (2007). *Musicophilia*. New York: Alfred A. Knopf.
- Shen, Y. (1997). *Cognitive constraints on poetic figures*. *Cognitive Linguistics*, 8(1), 33-72.
- Shin, S. (2005). *Automatic Pattern Extraction for Korean Sentence Parsing*. *Proceedings from the Corpus Linguistics Conference Series (Vol. 1, no. 1)*. University of Birmingham. Available on-line from <http://www.corpus.bham.ac.uk/PCLC>.
- Simner, J., & Hubbard, E. M. (Eds.). (2013). *Oxford handbook of synesthesia*. Oxford University Press.
- Stanford, W. B. (1936). *Greek metaphor: studies in theory and practice*. B. Blackwell.
- Stefanowitsch, A., & Gries, S. T. (Eds.). (2007). *Corpus-based approaches to metaphor and metonymy (Vol. 171)*. Walter de Gruyter.
- Strik Lievers, F. (2015). *Synaesthesia: A corpus-based study of cross-modal directionality*. *Functions of Language*, 22(1), 69-95.
- Strik Lievers, F., et al. (2013). *A methodology for the extraction of lexicalized synaesthesia from corpora*. Paper presented at ICL 19 (International Congress of Linguists). Geneva, Switzerland.
- Takada, M. (2008). *Synaesthetic metaphor: perception, cognition, and language*. VDM Publishing.
- Ullmann, S. (1938). *Synästhesien in der dichterischen Werken von Oscar Wilde*. *Englische Studien*, 72, 245-256.
- Ullmann, S. (1963). *The Principles of Semantics (2nd ed., 3rd Impression)*. Oxford: Basil Blackwell.
- Williams, J. M. (1976). *Synaesthetic Adjectives: A possible law of semantic change*. *Language*, 52, 461-478.
- Yoon, H. (1970). *Structurality of synesthetic metaphors*. *Korean Language and Literature*, No.49-50. [in Korean]

- Yu, N. (1992). A possible semantic law in synesthetic transfer. *SECOL Review*, 16(1), 20-40.
- Yu, N. (2002). Body and emotion: Body parts in Chinese expression of emotion. *Pragmatics & cognition*, 10(1), 341-367.
- Yu, N. (2003). Synesthetic metaphor: A cognitive perspective. *Journal of Literary Semantics*, 32(1), 19-34.
- Zhao, Q. & Huang, C. R. (2015). A corpus-based study on synaesthetic adjectives in modern Chinese.
- Zhao, Q., Huang, C. R., & Xu, H. (2015). Auditory Synaesthesia and Near Synonyms: A Corpus-Based Analysis of sheng1 and yin1 in Mandarin Chinese.