Cognition 150 (2016) 1–9

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Processing compound words: Evidence from synaesthesia

Jennifer L. Mankin [a,*], Christopher Thompson [b], Holly P. Branigan [c], Julia Simner [a,c]

[a] School of Psychology, University of Sussex, United Kingdom
[b] Institute for Computer Systems Architecture, School of Informatics, University of Edinburgh, United Kingdom
[c] Department of Psychology, University of Edinburgh, United Kingdom

ARTICLE INFO

Article history:
Received 28 November 2014
Revised 13 November 2015
Accepted 13 January 2016

Keywords:
Compound words
Synaesthesia
Word frequency
Semantic transparency
Lexical access
Dual-route model

ABSTRACT

This study used grapheme-colour synaesthesia, a neurological condition where letters evoke a strong and consistent impression of colour, as a tool to investigate normal language processing. For two sets of compound words varying by lexical frequency (e.g., *football* vs *lifevest*) or semantic transparency (e.g., *flagpole* vs *magpie*), we asked 19 grapheme-colour synaesthetes to choose their dominant synaesthetic colour using an online colour palette. Synaesthetes could then select a second synaesthetic colour for each word if they experienced one. For each word, we measured the number of elicited synaesthetic colours (zero, one, or two) and the nature of those colours (in terms of their saturation and luminance values). In the first analysis, we found that the number of colours was significantly influenced by compound frequency, such that the probability of a one-colour response increased with frequency. However, semantic transparency did not influence the number of synaesthetic colours. In the second analysis, we found that the luminance of the dominant colour was predicted by the frequency of the first constituent (e.g. *rain* in *rainbow*). We also found that the dominant colour was significantly more luminant than the secondary colour. Our results show the influence of implicit linguistic measures on synaesthetic colours, and support multiple/dual-route models of compound processing.

1. Introduction

Synaesthesia is a familial condition (e.g., Ward & Simner, 2005) where the perception of a stimulus in one modality triggers an automatic secondary sensation in another (e.g., Simner, 2012). Our study seeks to investigate natural language processing using *grapheme-colour synaesthesia*, where letters and numerals are perceived to have unique and consistent colours (e.g. *a* might be scarlet red or *7* might be leaf green; Rich, Bradshaw, & Mattingley, 2005; Simner et al., 2005; Ward, Simner, & Auyeung, 2005). Grapheme-colour synaesthetes also experience colours for whole words, and these colours are often systematically related to their synaesthetic colours for the component graphemes (Mills et al., 2002; Simner, Glover, & Mowat, 2006; Ward et al., 2005). For example, a synaesthete with a red *m* may also experience the whole word *man* as red as well (Mills et al., 2002). It is this linguistic aspect of whole-word colouring in grapheme-colour synaesthesia we explore in the present study, especially as it relates to the colouring of compound words (described further below).

Grapheme-colour synaesthesia is estimated to have a prevalence of about 1% in the general population (Simner, Mulvenna, et al., 2006) and to account for 35–45% of all cases of synaesthesia reported (Novich, Cheng, & Eagleman, 2011). Many aspects of the condition have been investigated in recent years, including its behavioural characteristics (e.g., Hubbard & Ramachandran, 2005; Ward et al., 2005), neurological roots (e.g. Rouw & Scholte, 2010; Sperling, Prvulovic, Linden, Singer, & Stirn, 2006) and associated advantages for cognition (e.g., Pfeifer, Rothen, Ward, Chan, & Sigala, 2014; Price, 2009; Ward, Thompson-Lake, Ely, & Kaminski, 2008). Of particular interest to the current paper, Simner (2007) suggested that there may be a special role for language as a synaesthetic inducer, since linguistic stimuli like words and graphemes are the triggers in 88% of the total reported cases of synaesthesia (Simner, Mulvenna, et al., 2006). This study seeks to use grapheme-colour synaesthesia to answer psycholinguistic questions about compound words and to provide a tool for exploring the mutual influences of synaesthesia on language and vice versa (for a review of this approach, see Cohen Kadosh & Henik, 2007; Simner, 2007). In particular, we ask what the synaesthetic colours for compound words can tell us about how such words might be stored in the mind for all people. Below, we first review previous evidence for linguistic influences in grapheme-colour synaesthesia,

* Corresponding author at: School of Psychology, University of Sussex, Brighton, United Kingdom.
E-mail address: J.Mankin@sussex.ac.uk (J.L. Mankin).

http://dx.doi.org/10.1016/j.cognition.2016.01.007
0010-0277/

and then we provide a brief overview of the psycholinguistic evidence to date for how compound words are processed in English.

1.1. Interaction of grapheme-colour synaesthesia and language

Many studies have already shown the close mutual influence that synaesthetic colour and language have on each other. By putting a symbol such as 5, ambiguous between *S* and *5*, in different linguistic contexts – that is, with bias towards a letter reading (MUSIC) or a number reading (3 4 5 6 7) – case studies showed that the synaesthetic colour experienced depends on the grapheme's linguistic meaning in context, and not simply its shape (Dixon, Smilek, Duffy, Zanna, & Merikle, 2006; Myles, Dixon, Smilek, & Merikle, 2003). Synaesthetes also show significant trends in the colouring of certain graphemes – for example, *a* is red more often than chance would predict (Rich et al., 2005; Simner, Glover, et al., 2006) and these trends are influenced by linguistic qualities like grapheme frequency. For example, high-frequency graphemes like *a* are likely to elicit higher frequency colour terms in English like red (Simner, Glover, et al., 2006; see also Emrich, Schneider, & Zedler, 2002). Later, a study in German showed that when the elicited synaesthetic colour was broken down into hue, saturation, and luminance (HSL), grapheme frequency was positively correlated with synaesthetic colour luminance and saturation (Beeli, Esslen, & Jäncke, 2007). The luminance effect was also replicated in English (Smilek, Carriere, Dixon, & Merikle, 2007). These studies clearly show that synaesthetic colour associations are not haphazard but systematic, and often based on linguistic qualities of the trigger.

The linguistic influences on grapheme-colour pairings are also seen in the way synaesthetes perceive colour for whole words. Grapheme-colour synaesthetes tend to report that words can have a combination of different colours (Mills et al., 2002), but as mentioned above, the colour of the first grapheme generally dominates the word in some way. For instance, having a blue *f* would mean a blue emphasis to the word *fan* (Baron-Cohen, Harrison, Goldstein, & Wyke, 1993), even though the colours for other letters in the word may also be perceived by that synaesthete. From synaesthete to synaesthete, this primary emphasis on the colour of words can come either by their first consonant (e.g., *fan* is the colour of *f*) or first vowel (e.g., *fan* is the colour of *a*), with the former being the most common (Simner, Glover, et al., 2006). Simner, Glover, et al. (2006) found that letters downstream in the word could influence colouring too; for example, in the word *ether*, the synaesthetically dominant colour of *e* was reinforced by a second *e* downstream in the word, evoking that colour more quickly and strongly than in a word like *ethos*, where the colour of *e* conflicted with the downstream *o* (Simner, Glover, et al., 2006).

These studies together reveal a complex but rule-based system of word colouring influenced by linguistic factors such as grapheme frequency, serial letter position, vowel/consonant status, grapheme repetition, and also by individual differences among synaesthetes. These linguistic influences in synaesthetic colouring also extend to non-alphabetic orthographies as well. We describe this here because it is possible to draw parallels with English compounding, the focus of our current paper. Hung, Simner, Shillcock, and Eagleman (2014) studied Chinese synaesthetes who experience synaesthetic colours for characters (i.e., the logographic writing units of Chinese). Hung et al. found that certain components of these characters, called *radicals*, influenced the colour of the character as a whole. For example, the character 櫻, meaning "cherry blossom", is a compound made up of the radicals 木, meaning "tree" (and providing semantic information for the whole compound), and 嬰, a character pronounced ying1 (providing the whole compound's phonetic pronunciation). In Hung et al.'s study, radicals on the right side of the compound (like 嬰 in 櫻) predicted the compound's overall luminance, whereas radicals on the left side (like 木 in 櫻) were marginally better for predicting its hue. Furthermore, semantic radicals on the left side of a compound, such as 木, marginally predicted saturation. This complex picture of how logographic radicals influence overall compound colouring may lead us to anticipate a similarly detailed situation in English compound colouring as well, and we explain this below.

1.2. Characteristics of compound words

In the current study, we look at synaesthesia in compound words in English, which are in some ways analogous to Chinese compound characters (but see Taft, Zhu, & Peng, 1999; Zhou, Marslen-Wilson, Taft, & Shu, 1999, for a discussion of their similarities and differences). Compound words in English are made up of two independent constituent words combined to make a new word, as in *rainbow* (i.e., *rain* + *bow*). These compounds are of special interest in lexical access research because their combined meanings and structure can be used to study how words are composed and represented in the mind (e.g., Taft & Forster, 1976). Several different types of theories have been proposed for how compounds are processed, which we test in our current study and so briefly review here.

Full-listing models of word processing propose that all words are stored in the mental lexicon as wholes, regardless of complexity. Lexical processing of compounds therefore consists of direct lookup of whole words in the lexicon (Butterworth, 1983). At the other extreme, *full-parsing* models claim that all complex words are decomposed into their constituents prior to lookup (Pinker, 1991; Stockall & Marantz, 2006; Taft, 1979, 1988, 2004). For example, a full-listing model would posit separate lexical entries for *rain, bow, and rainbow*, and the input *rainbow* would access that entry directly. A full-parsing model would posit that *rainbow* would first be obligatorily broken down into *rain* and *bow*, and those constituents would then be used to access the whole-word entry for *rainbow*. Combining the two are *dual-route models*, which suggest that both direct lookup and parsing routes work to process a word's representation. In particular, *parallel dual-route models* (Bertram & Hyönä, 2003; Schreuder & Baayen, 1995) propose that the two strategies race to the correct representation. More recently, dual-route models have been extended to *probabilistic multiple-route* models to account for information integrated from many sources during processing, including full forms, constituent words, morphological family size, and contextual and semantic cues (Kuperman, Bertram, & Baayen, 2008; Kuperman, Schreuder, Bertram, & Baayen, 2009).

We aim to provide data to test these models using the synaesthetic colours of compound words. We present our synaesthetes with compounds that vary on two linguistic features that are often used to test models of compound processing – word frequency and semantic transparency. Word frequency expresses how often a word occurs in a given language, and studies show that reading times decrease as a word's frequency increases (e.g., Ellis, 2002; Oldfield & Wingfield, 1965). Compounds can be quantified in terms of their overall compound frequency (e.g., the frequency of the word *rainbow* itself), but also by the frequencies of their constituents – for example, the frequencies of *rain* and *bow* independently. Frequency effects have been used often in compound-word research, the rationale being that if constituent frequencies influence how quickly a compound is processed, we would conclude that the compound has been decomposed in some fashion. For example, studies have shown that compound processing is facilitated if the first or second constituent is high frequency (Bien, Levelt, & Baayen, 2005) and in particular, if the high frequency element is the second constituent/head (Juhasz, Starr, Inhoff, & Placke, 2003; also Andrews, Miller, & Rayner, 2004; Inhoff, Starr, Solomon, & Placke, 2008). These studies point to a