Processing compound words

Citation for published version:
Mankin, J, Thompson, C, Branigan, H & Simner, J 2016, 'Processing compound words: Evidence from synaesthesia', *Cognition*, vol. 150, pp. 1–9. https://doi.org/10.1016/j.cognition.2016.01.007

Digital Object Identifier (DOI):
10.1016/j.cognition.2016.01.007

Published In:
Cognition

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
Available online 2 February 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 model where constituents within a compound are activated during processing and therefore suggest that compounds are decomposed into their constituents, as per the full-parsing or dual/multiple-route models. The constituents that are higher frequency are recognised more easily and, by extension, facilitate access to the compound as a whole.

However, there is also evidence that whole-word frequency influences response times independent of the frequency of the compound's constituents (Baayen, 2005). Eye-tracking reading studies have found significant reductions in gaze times for compounds with higher compound (i.e., whole-word) frequency, an effect that appears at least as early as the facilitation effects for the constituents (Andrews et al., 2004; also in Finnish: Pollatsek, Hyönä, & Bertram, 2000). Further eye-tracking investigations in compound reading have shown that whole-compound frequency has a significant effect on reading times, even before the second constituent has been fully identified (Kuperman et al., 2008; also in Dutch, Kuperman et al., 2009). Together with the facilitation effects of the constituents, this points to a combination of whole-word lookup and constituent decomposition, wherein the processing system makes use of all available strategies to arrive at the correct meaning (e.g., Libben, 2006). We might therefore expect to find a similar type of frequency effect at both constituent level and global level in the synaesthetic colouring of compound words.

The second linguistic characteristic this study considers is *semantic transparency*, or how clearly the meaning of a compound word is related to the meanings of its constituents. For example, *birdhouse* is a relatively transparent combination of the meanings of *bird* and *house*, but the compound *hogwash* is fully opaque in that it is related in meaning to neither *hog* nor *wash*. Does transparency affect whether compound words are mentally decomposed during language processing? Together, Sandra (1990) and Zwitserlood (1994) found differences between transparent and opaque compounds in a priming task, using semantic associates of the compound's constituents, e.g. *moon* for *Sunday*. Zwitserlood (1994) found priming effects with fully and partially transparent words (e.g., *birdhouse* and *jailbird*, respectively) but not with fully opaque compounds (e.g., *hogwash*). This provided evidence that opaque words may have less decomposition than transparent words.

However, the evidence on transparency has been mixed. Frisson, Niswander-Klement, and Pollatsek (2008) found no effect of transparency at all on eye movements in compound reading in English (see also Pollatsek & Hyönä, 2005). On the other hand, Juhasz (2007) reported a main effect of transparency in gaze durations (see also Marelli & Luzzatti, 2012, for Italian). In each case researchers were again seeking factors that might influence whether and when compounds are understood via decomposition. Of particular interest for the current study, MacGregor and Shtyrov (2013) found evidence that compounds may be decomposed differently dependent on their frequency. In an EEG study involving opaque compounds, they found that higher compound frequency elicited a stronger mismatch negativity component (MMN; known to index both lexical frequency and the congruence of semantic combinations) as opposed to low compound frequency. This indicates a high degree of lexicalisation (i.e., whole-word storage and processing) for high frequency compounds. Together, the studies above point to both whole-word access and decomposition strategies (the latter less so for opaque words) in the processing of compound words. Hence, Libben (1998, 2006) suggests that the language system may utilise all possible avenues of understanding a compound's meaning, including constituent processing and whole-word lookup.

1.3. Current study

The current study examines how word frequency and semantic transparency influence synaesthetic colouring of compound words. We consider both the number of synaesthetic colours triggered by different compound words (do they trigger one colour, or more than one?) as well as the nature of these colours (what is their saturation and luminance?). Our aim is to not only understand how linguistic features influence synaesthetic colours, but also to use synaesthesia to better understand models of compound processing. For words with low whole-word frequency (e.g., *lifevest*), we expect that the primary strategy for processing will be decomposition (Bien et al., 2005), which will therefore activate the constituents of compounds (e.g., *life* and *vest*). The activation of these two constituents may cause synaesthetes to be more likely to give low-frequency compounds two colours. High-frequency compounds, however (e.g., *football*), may be processed more directly via whole-word lookup (Andrews et al., 2004; Kuperman et al., 2008), so we expect a single synaesthetic colour will be more likely for these types of words. In summary, our prediction is that high-frequency compounds may be more likely to trigger one synaesthetic colour rather than two. This would provide support for a model of two different routes in compound processing, by which compounds are more likely to be decomposed in the mind if they are low (vs. high) frequency.

These predictions are partly inspired by an unpublished study by Kubitza (2006), who reported a case study with a single German synaesthete. Kubitza found that higher-frequency compounds in German were more likely to receive a single colour than low-frequency compounds. As this study did not ultimately appear in the literature, we attempt to first confirm this effect in a larger group of synaesthetes and, at the same time, see whether it extends to English. Furthermore, we also hypothesise that we may find an analogous effect regarding transparency. In transparent compounds (e.g., *birdhouse*), the meaning of the compound is directly related to the meanings of the constituents, and previous studies suggest this may lead to processing via decomposition (e.g. MacGregor & Shtyrov, 2013). If so, we predict that the activation of these constituents may lead to a higher likelihood of two synaesthetic colours. Conversely, the meanings of opaque compounds (e.g., *hogwash*) cannot be calculated from their constituents, and studies show that this discourages decomposition (e.g. Ji, Gagné, & Spalding, 2011), so we predict a higher likelihood that synaesthetes will give these compounds only one colour. This would support theoretical models of compound processing that propose two routes for lexical access, dependent on the semantic content of the compound (e.g., Zwitserlood, 1994).

In the second part of our experiment, we also examine more closely the precise nature of the colours that synaesthetes perceive for these words. We follow Beeli et al. (2007), Smilek et al. (2007), and Hung et al. (2014) in focusing on the saturation and luminance of synaesthetic colours, as this allows us to compare our results for compound words directly to previous findings in the synaesthesia literature (e.g., for graphemes; see above). We test whether, and under what circumstances, compounds might produce quantitatively different types of colours (e.g., colours with higher or lower luminance or saturation). For example, if whole-word frequency influences the nature of colours (e.g., if high whole-word frequency produces higher luminant colours), this would support full-listing models of lexical access by showing influences only at the level of the whole word. However, if colours are influenced by the frequencies of constituents (e.g., if high constituent-frequency produces higher luminant colours), this would show the influence of constituents within compounds and therefore support full-parsing models (or dual-route models if both types of frequency play a role). Finally, we also test whether transparent and opaque compounds produce different types of colours (again in their luminance or saturation). If, for example, transparent compounds are more likely to be decomposed during lexical access, we might find, say, additive luminances from two different constituents; this would support full-parsing models of lexical access by again suggesting decomposition during processing.

2. Method

2.1. Participants

Nineteen grapheme-colour synaesthetes (17 female, mean age 24.8, $SD = 11.3$) were recruited from the Sussex-Edinburgh Database of synaesthete participants and paid £12 for their participation. All participants were native English speakers and were confirmed to be genuine synaesthetes using the gold-standard behavioural test as follows. Synaesthetic colour associations are highly consistent, and synaesthetes can therefore be identified using a consistency test (Baron-Cohen, Wyke, & Binnie, 1987; Cytowic, 1989; Ward & Simner, 2003). Our diagnostic test was the online *Synesthesia Battery* at synesthete.org (see Eagleman, Kagan, Nelson, Sagaram, & Sarma, 2007, for methods). This test presents all 26 English graphemes three times in random order. For each grapheme, participants must choose their associated colour from a $256 \times 256 \times 256$ colour palette. The mean distance in colour space between the three colours given for each grapheme is converted into a standardised consistency score, with a score less than 1 indicating the high level of consistency characteristic of genuine synaesthesia. All 19 of our participants were below this required threshold (mean score = 0.62, $SD = 0.13$) and were therefore confirmed to have grapheme-colour synaesthesia.

2.2. Materials and procedure

Our core test materials were two sets of compound words. The first group of words varied incrementally by frequency (high to low), and the second varied by transparency (opaque to transparent). The first set ($n = 59$ compounds) were drawn from Janssen, Pajtas, and Caramazza (2011) and varied on lemma frequency[1] measured by the CELEX lexical database (Baayen, Piepenbrock, & van Rijn, 1993). These compounds were stress-initial, two-syllable, noun–noun compounds (e.g. *rainbow*). The second list ($n = 51$ compounds), which varied by semantic transparency, were taken from a study by Ji et al. (2011). In that study, transparency was rated on a scale from 1 ("totally opaque") to 7 ("totally transparent") by 36 raters, and the means of these ratings were the final transparency score for each compound. Our two wordlists contained no items with repeating consonants or vowels in the onset or nucleus of the constituent words (e.g. none such as *cr<u>o</u>ssb<u>o</u>w*). These words were removed because repeated letters have been shown to influence the synaesthetic colour of the whole word in a way not relevant to our current investigation (Simner, Glover, et al., 2006; Simner, Mulvenna, et al., 2006, see above). Table 1 lists the descriptive information about the wordlists and measures. In the frequency list, there was a marginal correlation between compound and second-constituent frequency ($r = .25$, $p = .054$); all other correlations were nonsignificant ($p > .46$). In the transparency list, there was no correlation between transparency rating and compound frequency ($p = .74$). Moreover, the items from Ji et al. (2011) were categorised as high or low transparency by Ji et al. and balanced between transparency conditions for lemma frequency (with a median split by transparency rating: $t(49) = -.018$, $p = .99$).

An online test was developed for the purposes of this experiment and participants were sent a link to this test via email. The test presented our frequency and transparency compounds separately, with the items randomised for each participant within each block. The order of the blocks was counterbalanced across participants. All target words were presented midway down the screen in bold (see Fig. 1). The participants were required to indicate whether each word had synaesthetic colour, and then chose that colour using a clickable colour palette. They were then asked whether the word had a second synaesthetic colour, and used a second colour palette to specify that colour (see Fig. 1). Therefore, participants could provide zero, one, or two colours for each compound.

Due to an oversight, the first four participants were given an option to skip the colour of any given word, which led to the loss of 47 responses (2.2% of the data overall). In the analysis, items that had been skipped were coded as uncoloured.

3. Results

In our study, we recorded two different dependent measures: the number of synaesthetic colours (zero, one, or two) that our synaesthetic participants provided for each compound and the nature of those colours as measured particularly by their saturation and luminance values. We will address the results of these two separate measures in different sections below, and within each section, we will take into account the manipulation of frequency and then transparency.

3.1. The number of synaesthetic colours for compound words

The majority of the words in both lists were given two colours: 827 out of 1121 (74%) in the frequency list and 720 out of 969 (74%) in the transparency list. In our analyses below, frequency and transparency are treated as continuous variables. However, for illustrative purposes only, Fig. 2 also divides our data into categories based on median splits of the frequency and transparency ratings, respectively.

We hypothesized that in the set of items that varied by lexical frequency, synaesthetes would be more likely to experience one colour (instead of two) for words of higher (vs. lower) frequency. Hence, we first analysed the frequency wordlist to investigate the effect of frequency on number of colours. We constructed a binomial linear mixed effects model, which predicted the likelihood of a two-colour versus one-colour response, including compound frequency and first and second constituent frequency as predictors (e.g. the frequencies of *rainbow*, *rain*, and *bow*) and random slopes to account for the random variation in participants and items. Zero-colour responses were excluded from the model, as skipped and uncoloured items were both recorded as having zero colours, and it would therefore be difficult to draw any conclusions about this type of response. Overall compound frequency was a significant predictor of number of colours (see Table 2): as compound frequency increased, so did the likelihood of a one-colour response. First and second constituent frequency were not found to significantly influence the model. Table 2 details the results of the linear mixed effects (LME) model, showing the influence of compound frequency on the likelihood of obtaining a two-colour response.

Our second hypothesis was that the number of colours would also be influenced by semantic transparency; higher transparency may make a two-colour response more likely because these compounds can easily be processed by splitting them into their constituents. However, this prediction was not supported by our data. There was no difference in the likelihood of one or two colours based on transparency ratings (see Fig. 2). Table 3 summarises the linear mixed effects model showing the

[1] Lemma frequency is the frequency of a word as it appears in all its inflexional variants (e.g., rainbow, rainbows, etc.) and this was the frequency measure available in the set of norms from which we drew our materials (Janssen et al., 2011).

Table 1
Means and standard deviations of the variables in each wordlist.

	Compound frequency		1st constituent frequency		2nd constituent frequency		Transparency rating	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Janssen et al. (frequency)	2.4	1.67	7.08	1.34	6.48	1.66	–	–
Ji et al. (transparency)	1.21	0.68	–	–	–	–	4.89	1.53

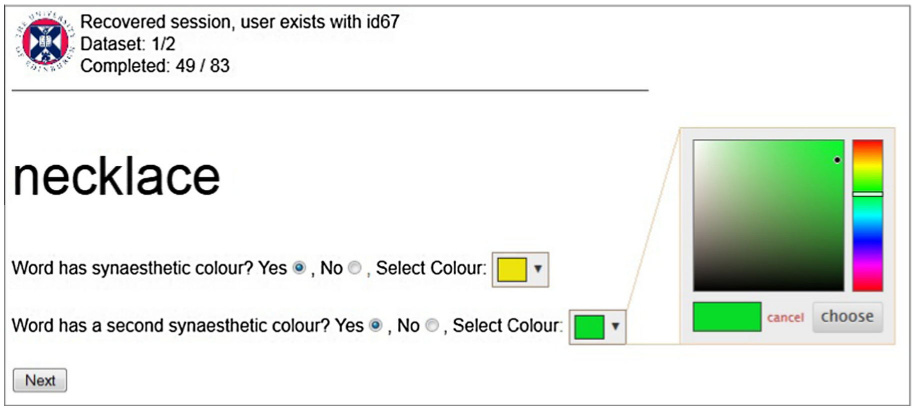

Fig. 1. The online word colour test. The test item is presented in bold letters (here *necklace*). Participants indicate whether the word has synaesthetic colour(s), then select those colour(s) using the colour palette (shown in its expanded form to right).

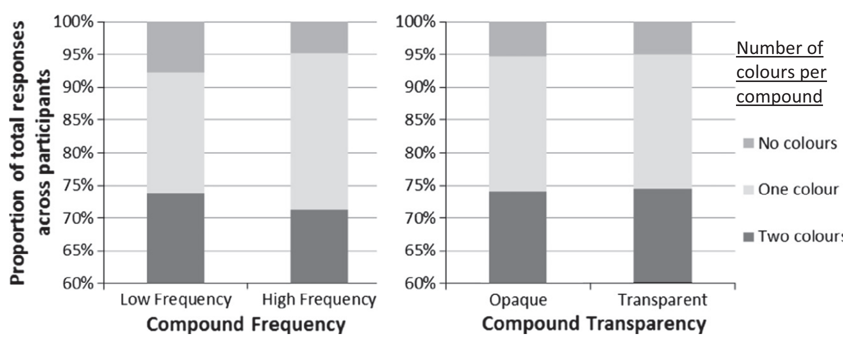

Fig. 2. The proportion of zero, one, or two colours, collapsed across participants. In the left panel, the set of items varying by compound frequency are divided into groups of low ($n = 30$) and high ($n = 29$) frequency. In the right panel, the set of transparency items are divided into opaque (i.e. low transparency rating; $n = 24$) and transparent (i.e. high transparency rating; $n = 25$) groups.

Table 2
LME model of frequency measures and number of colours.

Predictor	Estimate	*z*	Random variance (item)	Random variance (participant)	*p*
Intercept	2.33147	4.141	0.02898	5.30781	<.001
Compound frequency	−0.16060	−2.685			.007
1st constituent frequency	−0.02898	−0.399			0.69
2nd constituent frequency	0.05815	0.974			0.33

non-significant influence of semantic transparency on the likelihood of obtaining a two-colour response.

The linear mixed effects models above predict the binomial probability of a two-colour response. By transforming the log-odds into probabilities, we can represent the two models graphically as in Fig. 3. It is clear from the left panel (frequency model) that the probability of a two-colour response drops as frequency increases, which means a one-colour response becomes more likely at higher compound frequencies. However, the horizontal line in the right panel (transparency model) shows that the increase of transparency rating has no meaningful effect on the probability of obtaining a two-colour response.

To summarise, our analyses did find a significant effect of overall word frequency on the number of reported synaesthetic colours, but no effect of constituent frequency or semantic transparency.

3.2. The nature of synaesthetic colours for compound words

The next analyses examine the *nature* of synaesthetic colours. This investigation focused on saturation and luminance values, as these were expected to show systematic variation with the variables of frequency and transparency (Beeli et al., 2007; Simner, Glover, et al., 2006; Smilek et al., 2007).

Table 3
LME model of semantic transparency and number of colours.

Predictor	Estimate	z	Random variance (item)	Random variance (participant)	p
Intercept	3.32890	0.947	0.1565	10.6536	<.001
Compound transparency	−0.04041	0.814			.620

For each compound, there were up to four possible values: two possible colours per word, each having both saturation and luminance values. For example, participant 14 experienced the word *rainbow* with two synaesthetic colours, C1 and C2, with saturation and luminance values of 75, 61 (C1) and 79, 44 (C2) respectively. On the other hand, participant 1 experienced only one colour for this word, with saturation and luminance values of 33, 32 (C1). We calculated the average saturation and luminance values for the dominant and secondary colours elicited by each word across participants, which therefore gave us four mean values overall per compound (average saturation and luminance of both dominant and secondary colour). Looking first at the frequency manipulation, we constructed linear regression models predicting luminance and saturation values from (a) overall word frequency and the (b) first and (c) second constituent frequency. Of these models, only one showed a significant effect: we found a frequency effect on the luminance of the dominant colour of the word. Specifically, first constituent frequency significantly predicted dominant-colour luminance in a linear regression model (see Table 4). The relationship between dominant-colour luminance and first constituent frequency is depicted in Fig. 4. No other predictors approached significance in regression models (all $bs < 1.15$, all $ts < 1.67$, all $ps > .1$).

For the transparency wordlist, we chose a mixed repeated-measures ANOVA and divided the items into transparent and opaque groups taking the midpoint (4.0) on the rating scale as the point of division (transparent: $N = 17$, mean rating = 3.06, $SD = 0.64$; opaque: $N = 34$, mean rating = 5.80, $SD = 0.88$). We chose this simpler group design because it allows us to pursue an additional question: whether the average saturation and luminance values differed significantly between dominant and secondary colours. Since transparency was not found to influence number of colours in our first experiment, we suspected that we might not find an influence on the nature of those colours either. However, we could also investigate whether the two synaesthetic colours influenced each other, i.e. whether the saturation or luminance of the first, dominant colour was related to the saturation or luminance of the secondary colour in each word, aside from the effect of semantic transparency. Therefore, the mixed design ANOVA had two factors, each with two levels: transparency (transparent vs opaque) and colour dominance (dominant colour and secondary colour).

For the saturation values, the analysis showed no main effect of transparency, $F(1,49) = 1.2$, $p = .27$, or of colour dominance, $F(1,49) = 0.27$, $p = .61$, and no interaction between the two, $F(1,49) = .10$, $p = .92$. In other words, synaesthetic colours were equally saturated across transparent vs opaque compounds and across dominant vs secondary colours. For the luminance values, there was no main effect of transparency, $F(1,49) = 0.06$, $p = .8$, and no interaction, $F(1,49) = 2.01$, $p = .16$, but a marginally significant main effect of colour dominance, $F(1,49) = 3.61$, $p = .06$. Overall, the mean of the dominant colour luminance (45.5, $SD = 6.7$) was higher than that of the secondary colour luminance (42.5, $SD = 5.5$).

Table 4
Linear regression model predicting dominant colour luminance in frequency word set.

Predictor	b	t	p
Constant	33.98	5.61	.000
Compound frequency	.08	.60	.552
First constituent frequency	.38	3.05	.004
Second constituent frequency	−.09	−.70	.488

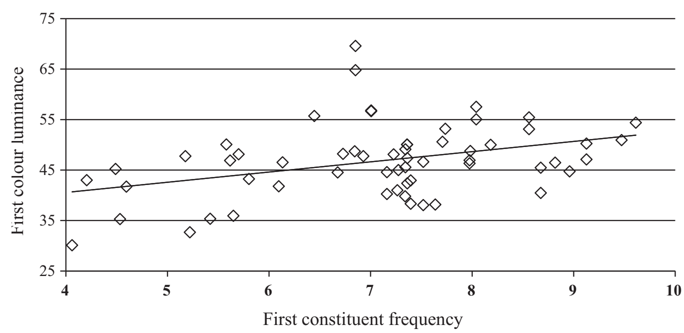

Fig. 4. Scatterplot and regression line showing the relationship between mean first colour luminance and first constituent frequency.

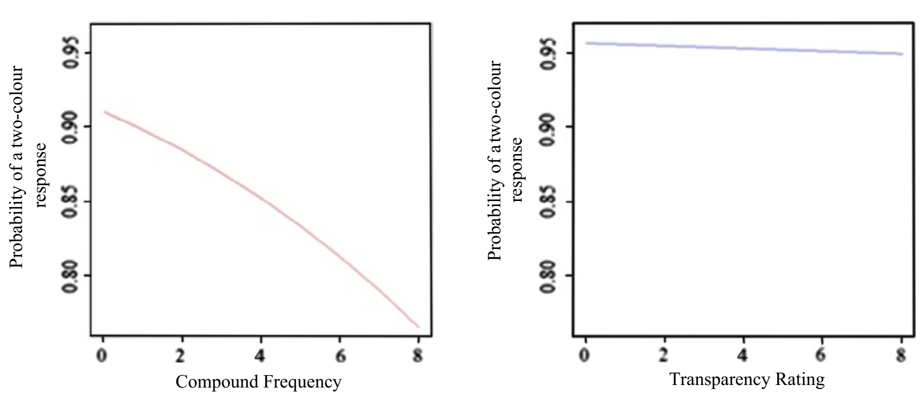

Fig. 3. The probability of two-colour responses for sets of words varying by frequency (left) and transparency (right).

Given this marginal effect of colour dominance in the transparent compounds, we decided to look for this same effect collapsing over both frequency and transparency word lists, to allow greater power. This colour dominance effect is a result of the contrast between two synaesthetic colours and not specifically linked to any of the linguistic features above. Therefore, we conducted two repeated-measures ANOVAs using all of the compounds from our experiment in both the frequency and transparency sets combined ($n = 110$). For this combined wordset, we tested mean luminance and saturation values in dominant versus secondary colours. Although, as above, there was no effect in saturation values ($F(1,109) = 1.596$, $p = .209$), there was a highly significant effect in luminance ($F(1,109) = 9.708$, $p = .002$). This indicates that the first-reported, dominant colour in a compound is significantly brighter (mean luminance = 46.2, $SD = 7.0$) than the secondary colour (mean luminance = 43.6, $SD = 6.4$).

To summarise, we found further evidence of a frequency effect on the nature of the synaesthetic colours evoked by compound words. Specifically, we found that the luminance of the dominant word colour was significantly predicted by the frequency of the first constituent in that compound. As for transparency, we confirmed a lack of influence of semantic transparency on the colouring of compounds. However, we did find a highly significant effect of colour dominance on luminance when both compound lists were combined.

4. Discussion

This study explored how linguistic features can affect synaesthetic colours, and how synaesthesia can be used as a tool to test hypotheses about psycholinguistic phenomena (see Cohen Kadosh & Henik, 2007; Simner, 2007). We focused on frequency measures in compound words (how often a compound or its constituents occur in English) and semantic transparency (how transparently the meaning of a compound is connected to the meanings of its constituents) as a way to test how compound words are processed and stored in the brain. We collected the synaesthetic colours for two lists of compound words and found the following results: (1) the likelihood of experiencing one synaesthetic colour (vs. two) increases with compound word frequency; (2) increasing first constituent frequency significantly predicts higher luminance in the dominant synaesthetic colour; (3) dominant synaesthetic colours are brighter (i.e. more luminant) than secondary colours; and (4) semantic transparency has no effect on the number of synaesthetic colours. We will examine the implications of each of these findings in turn.

Our finding that higher frequency compounds are more likely to be given a single colour than lower frequency compounds marries with the unpublished results of a single case-study reported by Kubitza (2006) in German, but here with a larger group of synaesthetes and in English. First, this result shows that psycholinguistic measures like word frequency do indeed influence synaesthetic colour responses, which can inform psycholinguistic theories. Specifically, our data suggest that high-frequency compounds would be more likely to be processed as wholes and would therefore be more likely to have a single synaesthetic colour. Although our frequency wordlist was not controlled for transparency, our results also showed that transparency has no appreciable effect on synaesthetic colour choice (see below), so our first finding seems to stem entirely from the frequency of the whole compound. We can extrapolate these results to provide evidence about how language processing occurs in non-synaesthetes as well as synaesthetes. The connection between single synaesthetic colours, high compound frequency, and lexicalisation of compounds suggests that even though compounds could be decomposed, those with high frequency are more quickly processed by direct lexical access (Kuperman et al., 2008). On the other hand, lower frequency compounds were more likely than high-frequency compounds to have two colours; this indicates that their constituents were more likely to be activated during processing, which in turn activated the two synaesthetic colours.

On the surface, our results speak against a *strictly* full-parsing model, in which all compounds would be broken down into their constituents preceding whole-word access (Stockall & Marantz, 2006; Taft, 1979, 1988, 2004). This would have predicted that constituents would always be activated (regardless of compound frequency) and therefore we should not have found any one-colour compounds at all. Although we did find a preponderance of two-colour compounds overall, we also found one-colour compounds, according to a frequency effect (i.e., higher whole-compound frequency predicts a greater likelihood of a single synaesthetic colour). This points to higher-frequency compounds, at least, being more likely to be directly retrieved as whole words. On the other hand, in a strictly full-listing model, which *does* posit direct access to whole compound words (e.g. direct access to *rainbow* without activating *rain* and *bow*; Butterworth, 1983), we would expect all compounds to be one-colour compounds, with no compound frequency effect at all. Our results are most compatible with dual/multiple-route models (e.g. Kuperman et al., 2009; Schreuder & Baayen, 1995), which would predict both one- and two-colour compounds, reflecting the use of both direct-lookup and decomposition strategies in processing, according to frequency. The compound frequency effect that we found for one- vs. two-colour compounds matches this model well, indicating that high-frequency compounds are indeed more likely to be directly accessed via their whole form while low-frequency compounds may be accessed only via their constituents.

We end this section by pointing to a recent refinement in the interpretation of full-parsing models (e.g. Juhasz, 2007; Taft, 2004; Taft & Ardasinski, 2006), which may be able to capture our data without a dual-route approach. Frequency is captured in these refined full-parsing models by the speed at which activation spreads from constituents to whole-word representations, with this being faster for high (vs. low) frequency compounds (Taft & Ardasinski, 2006; Taft & Nguyen-Hoan, 2010). Our own frequency finding (i.e., high frequency compounds tend to take a single colour, rather than two) could therefore be interpreted within these models *if* we stipulate that single colours arise when whole words are activated quickly.

As well as considering whole-word properties, we also looked at the frequencies of the constituents themselves (e.g., frequency of *rain* and of *bow*). The frequency of constituents exerted no influence on the number of synaesthetic colours. Although both first and second constituent frequency have been shown to influence how *quickly* a compound is processed in English (e.g., Bien et al., 2005; Inhoff et al., 2008), it may be that the one/two colour dependent measure of our own study was unable to capture this constituent effect. We did, however, find an effect of constituent frequency on the *nature* of the synaesthetic colours, which may be a more fine-grained way to tap into this type of effect. Specifically, compounds with first constituents that are encountered more often in English (e.g. high frequency *hand* in *handcuffs*) had a brighter dominant synaesthetic colour than compounds whose first constituent is seen less often (e.g. low frequency *cork* in *corkscrew*). Given the relationship between frequency and luminance in other areas of synaesthesia (e.g. in grapheme colours; Smilek et al., 2007), this may come as no surprise. Moreover, this influence of constituents on compound colouring is further support for full-parsing or dual-route models, both of which posit that whole compound words can be decomposed into constituents. Finally, the influence of first (rather than second) constituents here may reflect the first constituent's important role in processing in English compounds (Andrews et al., 2004) and also in other languages (e.g., Finnish; Pollatsek et al., 2000). We are now comparing the synaesthetic colours of compounds and constituents in follow-up studies in our lab.

In our final analyses of luminance we also found that colours reported as being "dominant" were more luminant than those described as "secondary". This luminance effect appears to reflect the prominent psychological status that the "dominant colour" has by definition. In other words, the fact that synaesthetes are able to identify which synaesthetic colour in a word is more "dominant" at all may well stem from that colour being overall brighter.

The remainder of our findings concerned compound transparency (cf. *birdhouse* vs *hogwash*), which had no significant effect on the number of elicited synaesthetic colours. Little research has explored the effects of word meaning in synaesthesia to date (but see Asano & Yokosawa, 2012; Gray et al., 2002) and our findings suggest semantic transparency does not play a role here. This was surprising, especially considering the analogous effect for word frequency. It may be possible that transparency is simply not a salient enough quality for it to influence synaesthetic colours, or that it may not be strong enough to overcome the known influences of grapheme frequency, colour term frequency, serial letter position, consonant/vowel status, and stress (Beeli et al., 2007; Simner, Glover, et al., 2006; Smilek et al., 2007; Ward et al., 2005) in addition to the frequency effect described in this current study. In summary, our data did not show differences in the number of synaesthetic colours for compounds that were transparent (e.g., *keyhole*) versus opaque (e.g., *hogwash*) and so we have no evidence that only the former are fully parsed into constituents. In both transparent and opaque compounds there was a very high probability of two colours for compounds. Again this finding is support for full-parsing or dual-route models, both of which posit that whole compound words can be decomposed into constituents. Indeed, the full-parsing model of Taft (2004; see also Juhasz, 2007; Taft & Ardasinski, 2006) suggests that every type of compound is necessarily decomposed *regardless of transparency*, thereby giving two different colours.

In conclusion, our study has shown that synaesthesia can be used to investigate questions about how words are stored in the brain and to evaluate existing theories of word processing. This initial confirmation of the influence of frequency on synaesthetic colour is a starting point to consider in more depth the questions of how meaning and language influence synaesthetes, and how synaesthesia can be used to gauge cognition in all of us.

Acknowledgements

This research was supported in part by a University of Sussex PhD research support grant to JLM. HPB was supported by a British Academy/Leverhulme Senior Research Fellowship. The research leading to these results has also received funding for author JS from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement n. [617678]. The authors would like to thank Dr. Martin Corley for his help with our analyses.

References

Andrews, S., Miller, B., & Rayner, K. (2004). Eye movements and morphological segmentation of compound words: There is a mouse in mousetrap. *European Journal of Cognitive Psychology, 16*(1–2), 285–311.

Asano, M., & Yokosawa, K. (2012). Synesthetic colors for Japanese late acquired graphemes. *Consciousness and Cognition, 21*(2), 983–993.

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). *The CELEX lexical database* {CD-ROM}.

Baayen, R. H. (2005). Storage and computation in the mental lexicon. In G. A. L. Jarema (Ed.), *The mental lexicon: Core perspectives* (pp. 81–104). Elsevier.

Baron-Cohen, S., Harrison, J., Goldstein, L. H., & Wyke, M. (1993). Coloured speech perception: Is synaesthesia what happens when modularity breaks down? *Perception, 22*(4), 419–426.

Baron-Cohen, S., Wyke, M. a., & Binnie, C. (1987). Hearing words and seeing colours: An experimental investigation of a case of synaesthesia. *Perception, 16*(6), 761–767.

Beeli, G., Esslen, M., & Jäncke, L. (2007). Frequency correlates in grapheme-color synaesthesia. *Psychological Science, 18*(9), 788–792.

Bertram, R., & Hyönä, J. (2003). The length of a complex word modifies the role of morphological structure: Evidence from eye movements when reading short and long Finnish compounds. *Journal of Memory and Language, 48*(3), 615–634.

Bien, H., Levelt, W. J. M., & Baayen, R. H. (2005). Frequency effects in compound production. *Proceedings of the National Academy of Sciences of the United States of America, 102*(49), 17876–17881.

Butterworth, B. (1983). Lexical representation. In B. Butterworth (Ed.). *Language production: Development, writing and other language processes* (Vol. II, pp. 257–294). London: Academic Press.

Cohen Kadosh, R., & Henik, A. (2007). Can synaesthesia research inform cognitive science? *Trends in Cognitive Sciences, 11*(4), 177–184.

Cytowic, R. E. (1989). *Synaesthesia: A union of the senses*. New York: Springer Verlag.

Dixon, M. J., Smilek, D., Duffy, P., Zanna, M., & Merikle, P. (2006). The role of meaning in grapheme-colour synaesthesia. *Cortex, 42*, 243–252.

Eagleman, D. M., Kagan, A. D., Nelson, S. S., Sagaram, D., & Sarma, A. K. (2007). A standardized test battery for the study of synesthesia. *Journal of Neuroscience Methods, 159*(1), 139–145.

Ellis, N. C. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition, 02*, 143–188.

Emrich, H. M., Schneider, U., & Zedler, M. (2002). *Welche Farbe hat der Montag?* Stuttgart: Hirzel.

Frisson, S., Niswander-Klement, E., & Pollatsek, A. (2008). The role of semantic transparency in the processing of English compound words. *British Journal of Psychology, 99*, 87–107.

Gray, J. A., Chopping, S., Nunn, J., Parslow, D., Gregory, L., Williams, S., ... Baron-Cohen, S. (2002). Implications of synaesthesia for functionalism. *Journal of Consciousness Studies, 9*(12), 5–31.

Hubbard, E. M., & Ramachandran, V. S. (2005). Neurocognitive mechanisms of synesthesia. *Neuron, 48*(3), 509–520.

Hung, W.-Y., Simner, J., Shillcock, R., & Eagleman, D. M. (2014). Synaesthesia in Chinese characters: The role of radical function and position. *Consciousness and Cognition, 24*, 38–48.

Inhoff, A. W., Starr, M. S., Solomon, M., & Placke, L. (2008). Eye movements during the reading of compound words and the influence of lexeme meaning. *Memory & Cognition, 36*(3), 675–687.

Janssen, N., Pajtas, P. E., & Caramazza, A. (2011). A set of 150 pictures with morphologically complex English compound names: Norms for name agreement, familiarity, image agreement, and visual complexity. *Behavior Research Methods, 43*(2), 478–490.

Ji, H., Gagné, C. L., & Spalding, T. L. (2011). Benefits and costs of lexical decomposition and semantic integration during the processing of transparent and opaque English compounds. *Journal of Memory and Language, 65*(4), 406–430.

Juhasz, B. J. (2007). The influence of semantic transparency on eye movements during English compound word recognition. In *Eye movements: A window on mind and brain* (pp. 373–389).

Juhasz, B. J., Starr, M. S., Inhoff, A. W., & Placke, L. (2003). The effects of morphology on the processing of compound words: Evidence from naming, lexical decisions and eye fixations. *British Journal of Psychology, 94*, 223–244.

Kubitza, A. (2006). Word frequency effect for synaesthetic word colour in German NN-compounds (case study). *Paper presented at the UKSA meeting*. London.

Kuperman, V., Bertram, R., & Baayen, R. H. (2008). Morphological dynamics in compound processing. *Language and Cognitive Processes, 23*(7–8), 1089–1132.

Kuperman, V., Schreuder, R., Bertram, R., & Baayen, R. H. (2009). Reading polymorphemic Dutch compounds: Toward a multiple route model of lexical processing. *Journal of Experimental Psychology: Human Perception and Performance, 35*(3), 876.

Libben, G. (1998). Semantic transparency in the processing of compounds: Consequences for representation, processing, and impairment. *Brain and Language, 61*(1), 30–44.

Libben, G. (2006). Why study compound processing? An overview of the issues. In G. Libben & G. Jarema (Eds.), *The representation and processing of compound words* (pp. 1–22). Oxford, UK: OUP Oxford.

MacGregor, L. J., & Shtyrov, Y. (2013). Multiple routes for compound word processing in the brain: Evidence from EEG. *Brain and Language, 126*(2), 217–229.

Marelli, M., & Luzzatti, C. (2012). Frequency effects in the processing of Italian nominal compounds: Modulation of headedness and semantic transparency. *Journal of Memory and Language, 66*(4), 644–664.

Mills, C. B., Viguers, M. L., Edelson, S. K., Thomas, A. T., Simon-Dack, S. L., & Innis, J. a. (2002). The color of two alphabets for a multilingual synaesthete. *Perception, 31*(11), 1371–1394.

Myles, K. M., Dixon, M. J., Smilek, D., & Merikle, P. M. (2003). Seeing double: The role of meaning in alphanumeric-colour synaesthesia. *Brain and Cognition, 53*(2), 342–345.

Novich, S., Cheng, S., & Eagleman, D. M. (2011). Is synaesthesia one condition or many? A large-scale analysis reveals subgroups. *Journal of Neuropsychology, 5*(2), 353–371.

Oldfield, R. C., & Wingfield, a. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology, 17*(4), 273–281.
Pfeifer, G., Rothen, N., Ward, J., Chan, D., & Sigala, N. (2014). Associative memory advantage in grapheme-colour synaesthetes compared to older, but not younger adults. *Frontiers in Psychology, 5*(696).
Pinker, S. (1991). Rules of language. *Science, 253*, 530–535.
Pollatsek, A., & Hyönä, J. (2005). The role of semantic transparency in the processing of Finnish compound words. *Language and Cognitive Processes, 20*(1–2), 261–290.
Pollatsek, A., Hyönä, J., & Bertram, R. (2000). The role of morphological constituents in reading Finnish compound words. *Journal of Experimental Psychology: Human Perception and Performance, 26*(2), 820–833.
Price, M. C. (2009). Spatial forms and mental imagery. *Cortex; A Journal Devoted to the Study of the Nervous System and Behavior, 45*(10), 1229–1245.
Rich, A. N., Bradshaw, J. L., & Mattingley, J. B. (2005). A systematic, large-scale study of synaesthesia: Implications for the role of early experience in lexical-colour associations. *Cognition, 98*(1), 53–84.
Rouw, R., & Scholte, H. S. (2010). Neural basis of individual differences in synesthetic experiences. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 30*(18), 6205–6213.
Sandra, D. (1990). On the representation and processing of compound words: Automatic access to constituent morphemes does not occur. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 42*(3), 529–567.
Schreuder, R., & Baayen, R. H. (1995). Modeling morphological processing. In L. B. Feldman (Ed.), *Morphological aspects of language processing* (pp. 131–154). Hillsdale, New Jersey: Lawrence Erlbaum.
Simner, J. (2007). Beyond perception: Synaesthesia as a psycholinguistic phenomenon. *Trends in Cognitive Sciences, 11*(1), 23–29.
Simner, J. (2012). Defining synaesthesia. *British Journal of Psychology (London, England: 1953), 103*(1), 1–15.
Simner, J., Glover, L., & Mowat, A. (2006). Linguistic determinants of word colouring in grapheme-colour synaesthesia. *Cortex; A Journal Devoted to the Study of the Nervous System and Behavior, 42*(2), 281–289.
Simner, J., Mulvenna, C., Sagiv, N., Tsakanikos, E., Witherby, S. a., Fraser, C., ... Ward, J. (2006b). Synaesthesia: The prevalence of atypical cross-modal experiences. *Perception, 35*(8), 1024–1033.
Simner, J., Ward, J., Lanz, M., Jansari, A., Noonan, K., Glover, L., ... Oakley, D. a. (2005). Non-random associations of graphemes to colours in synaesthetic and non-synaesthetic populations. *Cognitive Neuropsychology, 22*(8), 1069–1085.
Smilek, D., Carriere, J. S. a., Dixon, M. J., & Merikle, P. M. (2007). Grapheme frequency and color luminance in grapheme-color synaesthesia. *Psychological Science, 18*(9), 793–795.
Sperling, J., Prvulovic, D., Linden, D. E., Singer, W., & Stirn, A. (2006). Neuronal correlates of colour-graphemic synaesthesia: A fMRI study. *Cortex, 2*(42), 295–303.
Stockall, L., & Marantz, A. (2006). A single route, full decomposition model of morphological complexity: MEG evidence. *The Mental Lexicon, 1*(1), 85–123.
Taft, M. (1979). Recognition of affixed words and the word frequency effect. *Memory & Cognition, 7*(4), 263–272.
Taft, M. (1988). A morphological-decomposition model of lexical representation. *Linguistics, 26*, 657–667.
Taft, M. (2004). Morphological decomposition and the reverse base frequency effect. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 57*(4), 745–765.
Taft, M., & Ardasinski, S. (2006). Obligatory decomposition in reading prefixed words. *The Mental Lexicon, 1*(2), 183–199.
Taft, M., & Forster, K. I. (1976). Lexical storage and retrieval of polymorphemic and polysyllabic words. *Journal of Verbal Learning and Verbal Behavior, 15*(6), 607–620.
Taft, M., & Nguyen-Hoan, M. (2010). A sticky stick? The locus of morphological representation in the lexicon. *Language and Cognitive Processes, 25*(2), 277–296.
Taft, M., Zhu, X., & Peng, D. (1999). Positional specificity of radicals in Chinese character recognition. *Journal of Memory and Language, 40*(4), 498–519.
Ward, J., & Simner, J. (2003). Lexical-gustatory synaesthesia: Linguistic and conceptual factors. *Cognition, 89*(3), 237–261.
Ward, J., & Simner, J. (2005). Is synaesthesia an X-linked dominant trait with lethality in males? *Perception, 34*(5), 611–623.
Ward, J., Simner, J., & Auyeung, V. (2005). A comparison of lexical-gustatory and grapheme-colour synaesthesia. *Cognitive Neuropsychology, 22*(1), 28–41.
Ward, J., Thompson-Lake, D., Ely, R., & Kaminski, F. (2008). Synaesthesia, creativity and art: What is the link? *British Journal of Psychology (London, England: 1953), 99*(Pt 1), 127–141.
Zhou, X., Marslen-Wilson, W., Taft, M., & Shu, H. (1999). Morphology, orthography, and phonology reading Chinese compound words. *Language and Cognitive Processes, 14*(5–6), 525–565.
Zwitserlood, P. (1994). The role of semantic transparency in the processing and representation of Dutch compounds. *Language and Cognitive Processes, 9*(3), 341–368.