Semantic transparency in free stems: the effect of Orthography–Semantics Consistency in word recognition.

Marco Marelli

Center for Mind/Brain Sciences, University of Trento, Italy

Simona Amenta and Davide Crepaldi

Department of Psychology, University of Milano-Bicocca, Italy

Abstract

A largely overlooked side result in most studies of morphological priming is a consistent main effect of semantic transparency across priming conditions. That is, participants are faster at recognizing stems from transparent sets (e.g., *farm*) in comparison to stems from opaque sets (e.g., *fruit*), regardless of the preceding primes. This suggests that semantic transparency may be also consistently associated with some property of the stem word. We propose that this property might be traced back to the consistency, throughout the lexicon, between the orthographic form of a word and its meaning, here named *Orthography-Semantics Consistency* (OSC), and that an imbalance in OSC scores might explain the "stem transparency" effect. We exploited distributional semantic models to quantitatively characterize OSC, and tested its effect on visual word identification relying on large–scale data taken from the British Lexicon Project (BLP). Results indicated that (a) the "stem transparency" effect is solid and reliable, insofar it holds in BLP lexical decision times (Experiment 1); (b) an imbalance in terms of OSC can account for it (Experiment 2); and (c) more generally, OSC explains variance in a large item sample from BLP, proving to be an effective predictor in visual word access (Experiment 3).

Keywords: Orthography–Semantics Consistency, distributional semantic models, megastudies, visual word identification

Recent research on morphological processing has focused on the role played by semantic transparency in the recognition of derived words. Indeed, the meaning of a derived form can be more or less associated to the meanings of its constituent morphemes: words like *nameless, farmer, bakery* can be easily understood given the meaning of their roots (*transparent* words), whereas in words like *courteous, fruitful, cryptic* root meanings are not fully maintained (*opaque* words). How early semantic transparency comes to the stage during visual word processing has been the theoretical issue at the center of a long-standing debate (for a review, see Rastle & Davis, 2008). Priming was the main methodological tool adopted in addressing this question, with the assumption that, if the recognition of a root (e.g., *farm*) is made quicker by the previous presentation of a related derived form (e.g.,

farmer) in comparison to a control prime (e.g., *speaker*), this would mean that the root is accessed when processing the derived form. Typically, transparent and opaque derived primes are compared for their effectiveness in facilitating the identification of their (pseudo-)roots. In order to rule out any strategic effect related to prime awareness, the prime is often presented very shortly and preceded by an orthographic mask, making it virtually invisible at an explicit level (masked priming; Forster & Davis, 1984). Under these conditions, priming effect is regularly observed for both transparent and opaque prime-target pairs (that is, for both *farmer-farm* and *courteous-court*), although actual differences in the effect magnitude are still debated (Rastle, Davis, & New, 2004; Feldman, O'Connor, & Moscoso del Prado Martín, 2009).

Notwithstanding the large amount of data collected on this issue, a side effect often emerging in these studies has been largely overlooked. Indeed, a qualitative evaluation of the published results suggests that a *main effect of transparency* typically characterises these data as well, that is, target stems used in the transparent condition elicit quicker response times than target stems used in the opaque condition, *independently of prime type*. Table 1 reports average latencies for transparent and opaque pairs in all relevant studies, namely, those that (a) have stems as visually-presented target words; (b) employ lexical decision as task; (c) are run on native speakers of the language of interest; and (d) include both a transparent and an opaque condition in a between-target design. "Transparent" targets are systematically faster to recognize than "opaque" targets in English: the effect is quite strong in most of the studies, and even in those reporting smaller differences the effect direction is still consistent (with the only exception of Andrews & Lo, 2013). Russian, Dutch and Italian are on par with English. The only oddball seems to be French, where the difference between target stems in transparent and opaque conditions is not consistent (Longtin, Segui, & Hallé, 2003; Diependaele, Sandra, & Grainger, 2005).

Since the studies reported in Table 1 all involve the same (or very similar) experimental conditions, the qualitative observation that target stems in the transparent condition are generally easier to identify than target stems in the opaque condition can be assessed statistically in a meta-analysis considering those studies for which we were able to retrieve item means (and thus compute confidence intervals). Indeed, the forest plot reported in Figure 1 indicates that the best point estimate for the difference between "transparent" and "opaque" stems is 19 ms and its 95% confidence interval is 12–25.

	Language Investigated	Opaque Stems	Transparent Stems	Effect Size
Rastle et al. (2000), Exp. 1	English	612 ms	582 ms	31 ms
Devlin et al. (2004)	English	673 ms	639 ms	34 ms
Rastle et al. (2004)	English	616 ms	586 ms	30 ms
Lavric et al. (2007)	English	688 ms	666 ms	22 ms
Morris et al. (2007)	English	669 ms	648 ms	21 ms
Marslen-Wilson et al. (2008)	English	548 ms	531 ms	17 ms
McCormick et al. (2008), Exp. 4	English	627 ms	607 ms	20 ms
Rueckl & Aicher (2008), Exp. 1	English	648 ms	613 ms	35 ms
Rueckl & Aicher (2008), Exp. 2	English	667 ms	626 ms	41 ms
Feldman et al. (2009)	English	650 ms	617 ms	33 ms
Diependaele et al. (2011), Exp. 1	English	592 ms	589 ms	3 ms
Andrews & Lo (2013)	English	576 ms	579 ms	-3 ms
Diependaele et al. (2005), Exp. 1*	Dutch	629 ms	619 ms	10 ms
Diependaele et al. (2009), Exp. 1	Dutch	599 ms	584 ms	15 ms
Diependaele et al. (2009), Exp. 3*	Dutch	602 ms	583 ms	19 ms
Longtin et al. (2003), Exp. 1	French	629 ms	631 ms	-2 ms
Diependaele et al. (2005), Exp. 2*	French	623 ms	608 ms	15 ms
Marelli et al. (2013), Exp. 2	Italian	631 ms	594 ms	37 ms
Kazanina et al. (2008)	Russian	662 ms	643 ms	19 ms
Kazanina (2011), Exp. 2	Russian	679 ms	666 ms	13 ms

Table 1: Average response latencies in the transparent and opaque conditions of published visual masked priming experiments. *only RTs on visually presented prime-target pairs were considered.

In conclusion, it seems that experiments comparing transparent and opaque morphological priming do not only provide information on how (pseudo-)complex words (i.e., the primes) are analyzed morphologically, but also possibly reveal the existence of two groups of simple words (i.e., the stem targets) that are distinct for some property. Still, *prima facie*, it is not easy to identify which property might distinguish between words like *cheer, herb, poet, quiet, train* and words like *cheek, helm, pond, quest, trail*. This paper tests the hypothesis that these groups of words are characterized by how consistently each stem form is associated to its meaning, i.e., how informative any particular orthographic string is about the meaning of the word it identifies.

Before expanding on this idea, however, we need to assess the reliability of the "stem transparency" effect. Of course, there is a possibility that the difference observed is just a side effect of (a) some uncontrolled lexical variable, or (b) the presence of the primes. We cannot exclude (a) since targets for opaque and transparent sets were typically only matched for mean values (as opposed to distributions), and covariates were not always included in the statistical analyses. For what concerns (b), the priming effect was smaller for opaque than transparent pairs in many of the studies considered (e.g., Diependaele et al., 2005; Feldman et al., 2009; Kazanina, 2011; Marelli et al., 2013), a difference that obviously impacted on the collapsed mean latencies. For these reasons, in Experiment 1 we aimed at establishing the reliability of the effect by considering independent evidence from the British Lexicon Project (Keuleers, Lacey, Rastle, & Brysbaert, 2012).
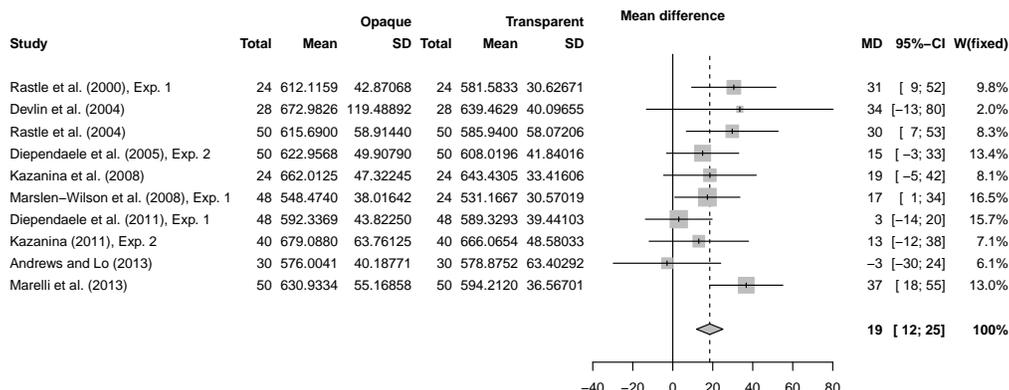
Study	Total	Opaque Mean	SD	Total	Transparent Mean	SD		MD	95%–CI	W(fixed)
Rastle et al. (2000), Exp. 1	24	612.1159	42.87068	24	581.5833	30.62671		31	[9; 52]	9.8%
Devlin et al. (2004)	28	672.9826	119.48892	28	639.4629	40.09655		34	[–13; 80]	2.0%
Rastle et al. (2004)	50	615.6900	58.91440	50	585.9400	58.07206		30	[7; 53]	8.3%
Diependaele et al. (2005), Exp. 2	50	622.9568	49.90790	50	608.0196	41.84016		15	[–3; 33]	13.4%
Kazanina et al. (2008)	24	662.0125	47.32245	24	643.4305	33.41606		19	[–5; 42]	8.1%
Marslen–Wilson et al. (2008), Exp. 1	48	548.4740	38.01642	24	531.1667	30.57019		17	[1; 34]	16.5%
Diependaele et al. (2011), Exp. 1	48	592.3369	43.82250	48	589.3293	39.44103		3	[–14; 20]	15.7%
Kazanina (2011), Exp. 2	40	679.0880	63.76125	40	666.0654	48.58033		13	[–12; 38]	7.1%
Andrews and Lo (2013)	30	576.0041	40.18771	30	578.8752	63.40292		–3	[–30; 24]	6.1%
Marelli et al. (2013)	50	630.9334	55.16858	50	594.2120	36.56701		37	[18; 55]	13.0%
								19	[12; 25]	100%

Figure 1. Forest plot showing a reliable difference between response times to stems used in trasparent vs. opaque conditions in masked priming experiments manipulating the semantic transparency of the derived forms.

Experiment 1 – Validating the "stem transparency" effect

In the present analysis, we aim at establishing the validity of the alleged "stem transparency" effect by excluding potential methodological and lexical confounder. In order to pursue this aim, we prepared an item set including those English words that were used as stem targets for either opaque or transparent primes in published priming studies; hence, we tested the difference between the transparent and opaque set on lexical decision latencies as extracted from the British Lexicon Project (BLP, Keuleers et al., 2012). Observing the effect under the relatively simple conditions of a pure lexical decision task will indicate that the phenomenon does not depend on the particular manipulations associated to the priming paradigm. Moreover, the present approach also allows to control the influence of several covariates, thus excluding alternative explanations based on other lexical variables.

Materials and Methods

Items were extracted from a concatenation of the set used by Rastle et al. (2000), Devlin et al. (2004), Rastle et al. (2004), Morris et al. (2007), Marslen-Wilson et al. (2008), and Andrews and Lo (2013), leading to a set including 335 stem words. *lock* and *port* were subsequently removed because they appeared in different sets in different studies: *port* was used as a target for the transparent prime *portable* in Morris et al. (2007) and as a target for the opaque prime *porter* in Marslen-Wilson et al. (2008), whereas the pair *locker–lock* was classified as opaque in Morris et al. (2007) and as transparent in Marslen-Wilson et al. (2008). Eight items were further removed because they were not included in the BLP. Therefore, the final set was comprised of 325 words, 157 of which were originally part of a set of transparent pairs (e.g., *cheer, herb, poet, quiet, train*) and 168 were originally part of a set of opaque pairs (e.g., *cheek, helm, pond, quest, trail*).

Response times (RTs) in lexical decision were extracted from the BLP. The SUBTLEX-uk database (van Heuven, Mandera, Keuleers, & Brysbaert, in press) was used to collect word frequency for each target item. The morphological annotation from CELEX

	Transparent set		Opaque set	
	Mean	SEM	Mean	SEM
Word frequency	13914	2762	11804	4519
Word FS	15.87	1.63	12.97	1.84
Word length	4.66	.06	4.41	.07

Table 2: Summary of the predictors for stems from the transparent and the opaque set.

	Estimate	Std. Error	t value	p value
Intercept	6.6351	0.0267	248.11	.0001
Word frequency	-0.0301	0.0024	12.42	.0001
Word FS	-0.0132	.0049	2.72	.0069
Word length	-0.0046	.0044	1.06	.2912
Transparency set	-0.0217	0.0074	2.95	.0034

Table 3: Results of the regression analyses on the lexical decision latencies extracted from the BLP.

(Baayen, Piepenbrock, & Gulikers, 1995) was exploited to compute morphological family size (hence, FS), that is, the type frequency of the morphologically complex words where each stem appears (De Jong, Schreuder, & Baayen, 2000). Table 2 reports a descriptive summary of the predictors in the opaque and the transparent set. Frequency and FS, along with word length (in letters), were included as covariates along with the "transparency" variable in a regression model. RTs, FS and frequency were log-transformed in order to obtain more Gaussian-like distributions. Once the model was fitted, outlying datapoints were removed on the basis of the model residuals (adopting 2.5 SD as a threshold). The model was then refitted to ensure that the analysis outcome was not determined by few influential outliers. The reported results are those of the refitted model.

Results

Targets from opaque sets elicited longer RTs ($mean = 581ms$, $SD = 56ms$) than targets from transparent sets ($mean = 553ms$, $SD = 47ms$). The difference is significant, even once the effects of the considered covariates are partialled out. Table 3 reports the results of the regression analysis.

Over and above the transparency effect, frequency and FS also have significant facilitatory effects on lexical decision latencies (the larger frequency and FS, the shorter the RTs). The effect of length is not significant, but this may be due to the extremely limited distribution of the variable in the dataset (80% of the items is either 4 or 5 letter long). Overall, the model fitted the observed data with an adjusted R-squared of .5081; predicted values were not correlated with the residuals ($r = -.0001$).

Discussion

The alleged effect of "stem transparency" received independent support from the analysis of the BLP data. Stems extracted from transparent sets are recognized faster than stems extracted from opaque sets, even in an experimental context unrelated to priming

techniques. The effect is also independent from possible mismatching in terms of frequency, length, or family size. These results confirm the validity and reliability of the phenomenon: in previous studies the item grouping based on the transparency of the derived forms also individuated distinct groups of stems.

Experiment 2 – An explanation for the effect

As anticipated in the Introduction, the hypothesis that we take up in this paper is that the difference described above is related to how reliable each word is as an orthographic cue for its meaning. Genuine morphological stems may also form opaque words – e.g., *invent* is a genuine stem in *invention* or *inventive*, but clearly does not contribute to the meaning of *inventory*. This phenomenon can be observed to different extent in the lexicon. At one extreme of the continuum, there are stems that always appear in words whose meaning is related to their own; in these cases, stems carry orthographic information that is consistently associated to a certain meaning, e.g., the orthographic chunk *widow* will be always associated to the *WIDOW* meaning, irrespective of the words it appears in (*widower, widowed, widowhood*). At another extreme, stems that mostly appear in opaque forms (e.g., *corn*), and therefore are found in words in which their meanings is not maintained (*corner, corny*), will not be very reliable orthographic cues for their semantics. As a result, the association between form and meaning in stems from the "opaque sets" will be potentially more difficult to learn, and hence weaker (see Andrews & Lo, 2013). In a sense, "opaque" stems are worse symbols than "transparent" stems, which may drive to slower response times.

One way to assess this hypothesis is to develop a measure that reflects how much a word is a reliable orthographic cue for its semantics, and then showing that stems in the opaque sets had lower scores than stems in the transparent sets in previous priming experiments. We call this new measure *Orthography–Semantics Consistency* (hence OSC), and we show in the present Experiment how it was defined (based on methods from distributional semantics) and how it explains indeed the "stem transparency" effect.

Materials and methods

We considered the same English item set employed in the previous analysis, including 157 items originally part of a set of transparent pairs and 168 items originally part of a set of opaque pairs, for a total of 325 items. We then collected all words starting with these items from a list including the top 30k most frequent content words (i.e., nouns, verbs, adjectives, adverbs) in a 2.8–billion corpus (detailed below), thus forming a family of "orthographic relatives" for each target stem. As an example, the family for the stem *whisk* includes *whisky, whiskey, whisker, whiskered*.

The next step was to compute a measure of semantic similarity between a stem and each of its orthographic relatives. In order to do so, we exploited methods borrowed from distributional semantics (Turney & Pantel, 2010), that have proven to be extremely effective in providing cognitively sound estimates of semantic association (e.g., LSA, Landauer & Dumais, 1997; HAL, Lund & Burgess, 1996) . This approach is based on the assumption that the meaning of a word can be approximated by the way that word co-occurs with other words in the lexicon. In a Distributional Semantic Model (hence, DSM) word meanings are

represented as vectors that are derived from these co-occurrences. The more two words tend to occur with the same set of other words (i.e., in similar contexts), the more their vectors will be close, the more their meanings will be considered to be similar. Geometrically, this amounts to measuring the cosine of the angle formed by the two vectors: the more similar the vectors, the smaller the angle between them, the higher their cosine. In the present study, we build a DSM by using a large part-of-speech tagged and lemmatized corpus, formed by a concatenation of the ukWaC (`http://wacky.sslmit.unibo.it/`), English Wikipedia (`http://en.wikipedia.org/`), and BNC (`http://www.natcorp.ox.ac.uk/`) corpora (about 2.8 billion words in total)[1]. We focused on the co-occurrences involving the top 30k most frequent content words (i.e., nouns, verbs, adjectives, adverbs), collected using a 5-word window. Raw counts were re-weighted using Positive Pointwise Mutual Information (Church & Hanks, 1990), and we reduced matrix dimensions by means of Non-negative Matrix Factorization (Arora, Ge, & Moitra, 2012), setting the number of dimensions of the reduced space to 350. These parameters were adopted because they were shown to produce high quality semantic spaces in previous studies (e.g., Bullinaria & Levy, 2007). For model implementation we relied on the freely available DISSECT toolkit (Dinu, The Pham, & Baroni, 2013).

Given a target word and the set of its k orthographic relatives, OSC was computed as the frequency-weighted average semantic similarity. In formal terms:

$$OSC(t) = \frac{\sum_{x=1}^{k} f_{r_x} * \cos(\vec{t}, \vec{r_x})}{\sum_{x=1}^{k} f_{r_x}}$$

Where t is the target word, r_x each of its k orthographic relatives, and f_{r_x} the corresponding frequencies extracted from the above described corpus. Since cosine values range from 0 to 1[2], the resulting OSC measure is a 0-to-1 score where values close to 0 identify words that are bad orthographic cues for their associated meanings, and values close to 1 indicate an almost perfect association between form and meaning.

The item *frisk* was not included in the semantic space and its OSC could not be computed; it was thus excluded from the following analysis. After having computed OSC for each target, we regressed it against log-transformed RTs extracted from the BLP, as for Experiment 1. Again, in accordance with Experiment 1, we also included in the regression analysis log-transformed frequency (from SUBTLEX-uk), family size (from CELEX), and length in letters.

Results

The density distributions of OSC in the two transparency sets is represented in Figure 2. The average OSC was significantly different in the transparent-set stems vís-a-vís the opaque-set stems ($t(322) = 7.41, p = .0001$), with the former showing larger OSC than the latter (.72 ± .22 vs. .50 ± .31 respectively). This OSC difference can explain the

---

[1]Since the reference corpora were POS-tagged, in the resulting DSM we obtained separate vectors for homographs with different grammatical class (e.g., a vector for the noun *run* and a vector for the verb *run*). When target items were ambiguous in relation to their grammatical class, they were assigned the one most frequently observed in the corpus in order to extract the corresponding vector.

[2]When all vector components are non-negative (as resulting from the settings of our DSM), the cosine is also non-negative.
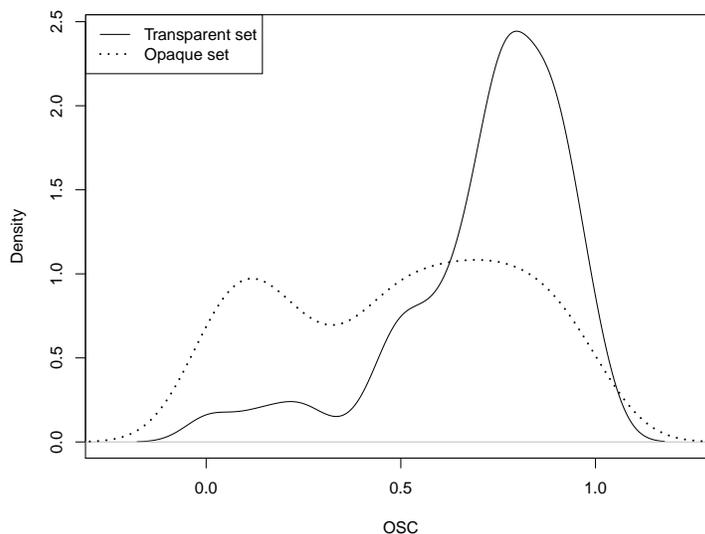
Figure 2. Density distribution of Orthography-Semantics Consistency in stems extracted from either transparent or opaque sets.

	Estimate	Std. Error	t value	p value
Intercept	6.6341	0.0266	249.84	.0001
Word frequency	-0.0282	0.0025	11.21	.0001
Word FS	-0.0147	.0048	3.06	.0024
Word length	-0.0031	.0045	0.71	.4826
OSC	-0.0461	0.0133	3.47	.0006

Table 4: Results of the regression analyses on the lexical decision latencies extracted from the BLP, when including OSC in place of transparency as predictor.

transparency effect in lexical decision latencies discussed above: indeed, if OSC is introduced in place of transparency in the regression model summarized in Table 3, its effect emerges as significant ($b = -0.046, t = 3.47, p = .0006$) and larger than the estimate previously observed for the transparency dummy variable ($b = -0.022$). This observation is confirmed by the Relative Importance Indicators extracted from a model including both predictors: normalized LMG (Lindeman, Merenda, & Gold, 1980; Kruskal, 1987) for transparency and OSC is, respectively, .05 and .11. The results of the model where transparency-set is substituted by OSC as a predictor are summarized in Table 4.

Along with OSC, also frequency and FS have significant facilitatory effects, whereas the effect of length is not significant. Overall, the model fitted the observed data with an adjusted R-squared of .5133; predicted values were not correlated with the residuals ($r = .0001$). The model with OSC provides a better fit to the data than the one reported in Experiment 1 (AIC -860.55 and -857.04, respectively); according to Wagenmakers and

Farrell (2004), $\Delta AIC = 3.51$ indicates that the OSC model is 5.78 times more likely to be the best one than the model including transparency as predictor.

Discussion

In this analysis, we have tested the effect of OSC in (a) distinguishing stems coming from either opaque or transparent sets in previous priming experiments; and (b) predicting lexical decision latencies for those items in the BLP. In both tests, OSC proved to have a significant effect in the expected direction: stems taken from transparent sets have significantly higher OSC than stems taken from opaque sets, and OSC has a facilitatory effect on RTs in lexical decision. This suggest that the bizarre phenomenon that inspired this study, that is, the main effect of semantic transparency on simple stem targets in priming experiments, may be explained by considering how much, in the whole lexicon, the orthographic information carried by the stem is consistent with its associated semantics. In conclusion, the grouping based on the semantic properties of the derived forms also identifies two sets of stems that are distinguishable for their level of Orthography-Semantics Consistency.

Once we consider the issue from this point of view, it is not surprising that OSC provides a better fit to RT data than the transparency predictor: the latter just happens to be a dichotomization of the former, and it is well known that, when dealing with naturally continuous variables, a continuous indicator has to be preferred to its dichotomized counterpart in terms of both statistical power and estimation accuracy (Cohen, 1983; Maxwell & Delaney, 1993). Concerning the difference in OSC distributions between transparent and opaque sets (Figure 2), it arguably reflects a selection bias in the item samples from the original studies. English derivational morphology is a productive system, leading to opaque words being naturally scarce; as a consequence, stems having at least an opaque form were most likely assigned to opaque sets in order to obtain as large samples as possible. Still, one single opaque form does not guarantee that OSC will be low, leading to the vaguely uniform OSC distribution observed for the opaque set. On the other hand, the transparent set is likely to include stems having *only* transparent forms (if they were associated also to opaque words, they would have been assigned to an opaque set); for this reason, the OSC distribution in the trasparent set is negatively skewed.

Why should OSC impact word processing? In order to account for their masked priming results, Andrews and Lo (2013) proposed that the association between an opaque derived word and its stem may be difficult to learn because of their similarity in form paired with their discrepancy in meaning. The present results suggest that this has not only consequences when both the derived word and the stem are explicitly used during the experiment, but also under simpler conditions when skilled readers are just presented with isolated stems: the knowledge that those stems have opaque and/or transparent derived forms is stored in the mental lexicon, and influences the way they are processed even when (pseudo-)morphological relatives are not involved by the experimental paradigm.

Experiment 3 – Generalization of the OSC effect

So far, OSC was shown to explain human performance on stems that happened to be used in previous priming experiments. Of course, before being able to make any general point about the role of OSC in visual lexical identification, we need to demonstrate that

	Mean	SEM	1st quartile	Median	3rd quartile	Asymmetry	Kurtosis
OSC	.69	.01	.56	.75	.88	-0.91	0.04
Word frequency	16119	1482	570	2515	10882	11.68 (-0.17*)	188.59 (-0.09*)
Word FS	14.13	1.07	4	7	14	24.46 (0.17*)	775.57 (1.08*)
Word length	5.26	0.03	4	5	6	0.34	-0.39

Table 5: Summary of the predictors considered in Experiment 3. *values obtained once the variable is log-transformed.

	Estimate	Std. Error	t value	p value
Intercept	6.5922	0.0109	602.89	.0001
Word frequency	-0.0308	0.0009	33.41	.0001
Word FS	-0.0041	0.0021	1.97	.0495
Word length	0.0035	0.0013	2.74	.0061
OSC	-0.0254	0.0066	3.84	.0002

Table 6: Results of the regression analyses on the lexical decision latencies extracted from the BLP for a large set of random words.

this effect holds in a much wider word sample that is independent from previous research. In our hypothesis, OSC is not an effect that can be limitedly observed in the two peculiar, extreme sets we have been considering so far, but rather something that regularly affects word processing in a continuous way. This is the issue that we take up in the present Experiment.

Materials and methods

For this analysis we considered a dataset of 1821 words, randomly sampled from the words included in both the semantic space described in Experiment 2 and the BLP database, and having at least one orthographic relative over and above itself (in order to exclude words with $OSC = 1$ that could have distorted the distribution of the variable of interest). OSC was defined following the same approach described in the Experiment 2. Again, for each target word we also collected frequency (from SUBTLEX-uk), family size (from CELEX), and length in letters. RTs in lexical decision were extracted from the BLP. Table 5 summarizes the distribution of the variables in the considered dataset.

OSC was not correlated with any of the other predictors (frequency: $\rho = .06$; FS: $\rho = -.08$; length: $\rho = .15$). RTs, FS and frequency were log-transformed in order to obtain more Gaussian-like distributions. The four predictors were tested in a regression model with RTs as dependent variable. The same procedure described for the previous experiments was followed.

Results

The results of the regression model are summarized in Table 6.

All the considered effects are significant. OSC, frequency and FS have facilitatory effects (larger values are associated to shorter RTs), whereas the effect of length is inhibitory

	Mean	SD	1st quartile	Median	3rd quartile	Confidence Interval	
Intercept	6.5922	0.0124	6.5839	6.5924	6.6008	6.5919	6.5926
Word frequency	-0.0309	0.0011	-0.0316	-0.0308	-0.0301	-0.0309	-0.0308
Word FS	-0.0041	0.0021	-0.0055	-0.0041	-0.0026	-0.0041	-0.0039
Word length	0.0035	0.0013	0.0027	0.0035	0.0044	0.0035	0.0036
OSC	-0.0255	0.0071	-0.0301	-0.0253	-0.0207	-0.0256	-0.0252

Table 7: Summary of the distributions of the effect parameters across 5,000 bootstrap samples.

(larger values are associated to longer RTs). Overall, the model fitted the observed data with an adjusted R-squared of .5222; predicted values were not correlated with the residuals ($r = .0001$).

The obtained model was validated in a bootstrap procedure with 5,000 resamples (see Wu, 1986). The resulting optimism indexes are small for both the R-squared and the mean standard error (.0023 and .0001 respectively), indicating that overfitting is extremely limited and thus confirming the reliability of the model. Table 7 summarizes the distributions of the effect parameters across the bootstrap samples.

Discussion

Even when considering a large set of items extracted from the BLP, the effect of OSC on lexical decision latencies is significant, and independent from frequency, FS, and length. This indicates that the effect of OSC is generally in place during word recognition, and not simply a bizarre consequence of the choice of particularly extreme item sets in the previous literature. The effect of OSC in such a large sample of words, along with the low correlations with the other predictors, suggests that the measure proposed in the present paper describes an important aspect, even if so far unexplored, of visual word recognition.

General Discussion

In the present paper we have investigated a curious side effect in priming experiments on morphological processing, namely that stems assigned to the transparent condition are identified more quickly than stems assigned to the opaque condition, regardless of any primes preceding them. This effect is puzzling because the difference between transparent and opaque conditions was exclusively related to the primes in those experiments, that is, there was nothing *a priori* that differentiated the stems in the two conditions; why *corn, fruit, whisk* (chosen as opaque targets because part of *corner, fruitful, whisker*) should be processed slower than *adore, farm, widow* (chosen as transparent targets because part of *adorable, farmer, widower*), even when the corresponding derived form is not presented to the participants? This question, springing from the casual observation of a bizarre yet empirically solid phenomenon, has led us to wonder about the semantic relations intercurring between a word and its orthographic relatives, and how this underexplored aspect may influence visual word processing. The outcome of this venture was the development of a new measure (OSC: Orthography-Semantics Consistency) quantifying the consistency of the carried orthographic and semantic information.

In a series of three experiments, we showed that (a) the "stem transparency" effect is solid and reliable, insofar it holds on BLP lexical decision times where no primes are

involved whatsoever (Experiment 1); (b) an imbalance in terms of Orthography–Semantics Consistency can account for this effect (Experiment 2); and (c) OSC explains variance in a large item sample from BLP that is independent of frequency, family size and word length (Experiment 3).

The OSC effect indicates that word processing is influenced by the relative distribution of form and meaning in the lexicon or, in other words, that the strength of the association between orthography and semantics contributes to determining how easily a word is recognized. The effect fits naturally well with learning models that see lexical and morphological effects as emerging at the interface between orthography and meaning (e.g., Plaut & Gonnerman, 2000; Baayen, Milin, Durđević, Hendrix, & Marelli, 2011). In these approaches, a word will be more difficult to process when part of opaque derived forms because of competition during learning: the same orthographic information (e.g., *bat*) is associated to several meanings in this case (e.g., *BAT, BATTERY, BATTLE, BATMAN*), with the result that it becomes a relatively unreliable cue for the associated semantic representation. Therefore, the result of the present study is a natural consequence of the learning process implemented in these models, and the OSC effect fits well with these kinds of architectures because it can be considered a proxy of the distributed nodes/weights linking orthography to semantics. It is also true, however, that, the same effect could be accounted for under a spreading–activation framework (e.g., Taft, 2004), although it is probably not an obvious prediction of most of these models. In these architectures, the effect would depend on competition between dissimilar semantic representations (e.g., *BAT* vs. *BATTLE*), activated at the same time by the corresponding lexical (or sub-lexical) units (e.g., *bat, battle*), that are in turn accessed because of the common orthographic input (e.g., *b, a, t*).

Being born from a side effect emerged in priming studies, one may wonder whether OSC could have any impact on the priming phenomenon itself. Could this effect be primarily driven by properties of the stems targets, as opposed to the derived primes? Indeed, if we hypothesize that priming effects can emerge for high-OSC words only, the facilitation observed for both transparent and opaque items will be explained by the OSC distributions in the target sets, since they both include words characterized by high levels of OSC (Figure 2). The more uniform distribution in the opaque set would also explain the much discussed variability in priming results for opaque forms (Rastle & Davis, 2008): item samples will be more or less likely to elicit a priming effect on the basis of their average OSC. As interesting as this idea may seem, we don't believe it to be fully supported by existing data. Järvikivi and Pyykkönen (2011) and Feldman, Kostić, Gvozdenović, O'Connor, and Moscoso del Prado Martín (2012) have convincingly shown that different priming effects are found for transparent vis-a-vis opaque primes in a *within-target design*, that is, even when OSC is kept constant. However, OSC may still modulate the *magnitude* of the priming effect, orthogonally to the properties of the derived primes. A preliminary analysis on the data reported by Rastle et al. (2004) indicates a mild, positive correlation between target OSC and priming effect size ($\rho = .19$). Although this correlation is very close to being significantly different from zero ($p = .0594$), further results, on purposely created sets, are clearly needed in order to draw a conclusion.[3]

---

[3]In a sample of similar size ($n = 113$) extracted from the Semantic Priming Project (Hutchison et al., 2013) no correlation between OSC and priming effect was found ($\rho = -.01$; $p = .8898$). This additional piece

The measure we proposed seemed to efficiently capture the Orthography-Semantics Consistency by computing the weighted average semantic similarity between a word (e.g., *whisk*) and the words that have similar orthographic onsets (e.g., *whisky, whisker*). Of course, the way we formalized OSC has not to be taken for granted and alternative implementations of the measure may be proposed, in particular with regards to the way we defined the relatives. For example, we may consider as an orthographic relative any string including the word, independently of its position. Although previous results has shown that the onset of a word is crucial for its orthographic processing (Jordan, Thomas, Patching, & Scott-Brown, 2003; White, Johnson, Liversedge, & Rayner, 2008), the same does not hold when semantics is at stake. Indeed, in English, crucial meaning information is often carried by the final portion of the word; this is the case, for example, for all prefixed words (e.g., *heat* in *reheat*) and most compounds (e.g., a *swordfish* is a *fish*, not a *sword*). The present version of the measure, that focuses only on word onsets (e.g., a *swordfish* is considered a relative for *sword*, but not for *fish*), is likely loosing a certain degree of information in this regard.

An even more radical approach would be to extract orthographic relatives by means of a continuous measure of orthographic similarity (e.g., Grainger & Whitney, 2004; Yarkoni, Balota, & Yap, 2008; Davis, 2010), in place of taking the whole word as a unique orthographic chunk. We could consider as relatives all the words that are above a certain threshold of similarity, irrespective of them beginning with the same string or not. If we adopt as similarity threshold a Levenshtein Distance (LD, Yarkoni et al., 2008) of 2, for example, we will extract as relatives of *boat* words like *moat* ($LD = 1$), *boot* ($LD = 1$), *boast* ($LD = 1$), *board* ($LD = 2$), *boost* ($LD = 2$), etc., over and above *boater* ($LD = 2$) that would have been already considered in the present version of the measure.

Finally, morphological family members may be considered in place of orthographic relatives – e.g., *dealer* would be a relative for *deal*, *fruitful* would be a relative for *fruit*, but *dialog* would not be a relative for *dial* because *dialog* is not a morphologically complex word). This approach would be closely related to that adopted in experiments showing that opaque family members do not contribute to the family size effect on word recognition (Bertram, Baayen, & Schreuder, 2000; De Jong et al., 2000; Moscoso del Prado Martín, Bertram, Häikiö, Schreuder, & Baayen, 2004). However, despite their theoretical association, family size and morphological OSC are quite different measures from a mathematical point of view. Whereas family size is a *discrete count* of how many semantically related members a word has in its morphological family, morphological OSC would be a *continuos estimate* of the meaning consistency within a morphological family. The actual count of family members does not influence directly OSC: being computed as an average between those very members, it is entirely possible to have words with identical OSC and very different family size, and vice-versa. In conclusion, the two measures, not being quantitavely associated, will arguably capture different word properties and it will be hence worth considering their combined effects on word recognition.

It is beyond the scope of this paper to investigate which of the above parameter settings may be better at capturing Orthography–Semantics Consistency. The crucial point

---

of evidence confirms that OSC is efficiently capturing a dimension encompassing *both form and meaning*: in fact, it is associated with morphological priming, but becomes irrelevant when purely semantic conditions are under exam.

that we want to make here, independently of any specific implementation, is that OSC does influence visual word access, even when target words are isolated stems. Importantly, we were able to show that this is the case using a measure that (a) is easily interpretable – it goes from 0, no consistency, to 1, perfect consistency; (b) can be computed automatically – one does not need to rely on human annotations; and (c) is also theory-independent – although its spirit is surely more in line with learning approaches than with more classic box–and–arrows models, the measure itself is completely bottom-up. In conclusion, OSC should be considered when implementing and developing models of visual word processing, and taken into the appropriate consideration when planning and running reading experiments.

References

Andrews, S., & Lo, S. (2013). Is morphological priming stronger for transparent than opaque words? it depends on individual differences in spelling and vocabulary. *Journal of Memory and Language*, *68*(3), 279 - 296.

Arora, S., Ge, R., & Moitra, A. (2012). Learning topic models–going beyond svd. In *Foundations of Computer Science (FOCS), 2012 IEEE 53rd Annual Symposium* (pp. 1–10).

Baayen, R. H., Milin, P., Durđević, D. F., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological review*, *118*(3), 438.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database (cd-rom)*. University of Pennsylvania, Philadelphia, PA: Linguistic Data Consortium.

Bertram, R., Baayen, R. H., & Schreuder, R. (2000). Effects of family size for complex words. *Journal of Memory and Language*, *42*, 390-405.

Bullinaria, J. A., & Levy, J. P. (2007). Extracting semantic representations from word co-occurrence statistics: A computational study. *Behavior Research Methods*, *39*(3), 510–526.

Church, K., & Hanks, P. (1990). Word Association Norms, Mutual Information, and Lexicography. *Computational Lingistics*, *16*, 22-29.

Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement*, *7*, 249–254.

Davis, C. J. (2010). The spatial coding model of visual word identification. *Psycholgical Review*, *117*, 713-758.

De Jong, N. H., Schreuder, R., & Baayen, R. H. (2000). The morphological family size effect and morphology. *Language and Cognitive Processes*, *15*, 329-365.

Devlin, J. T., Jamison, H. L., Matthews, P. M., & Gonnerman, L. M. (2004). Morphology and the internal structure of words. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(41), 14984–14988.

Diependaele, K., Duñabeitia, J. A., Morris, J., & Keuleers, E. (2011). Fast morphological effects in first and second language word recognition. *Journal of Memory and Language*, *64*, 344-358.

Diependaele, K., Sandra, D., & Grainger, J. (2005). Masked cross-modal morphological priming: Unravelling morpho-orthographic and morpho-semantic influences in early word recognition. *Language and Cognitive Processes*, *20*, 75–114.

Diependaele, K., Sandra, D., & Grainger, J. (2009). Semantic transparency and masked morphological priming: The case of prefixed words. *Memory & Cognition*, *37*(6), 895-908.

Dinu, G., The Pham, N., & Baroni, M. (2013). DISSECT: DIStributional SEmantics Composition Toolkit. In *Proceedings of ACL (System Demonstrations)* (pp. 31–36). Sofia, Bulgaria.

Feldman, L. B., Kostić, A., Gvozdenović, V., O'Connor, P. A., & Moscoso del Prado Martín, F. (2012). Semantic similarity influences early morphological priming in serbian: A challenge to form-then-meaning accounts of word recognition. *Psychonomic Bulletin & Review*, *19*(4), 668-676.

Feldman, L. B., O'Connor, P. A., & Moscoso del Prado Martín, F. (2009). Early morphological processing is morphosemantic and not simply morpho-orthographic: A violation of form-then-meaning accounts of word recognition. *Psychonomic Bulletin & Review*, *16*(4), 684–691.

Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of experimental psychology: Learning, Memory, and Cognition*, *10*(4), 680-698.

Grainger, J., & Whitney, C. (2004). Does the huamn mnid raed wrods as a wlohe? *Trends in Cognitive Sciences*, *8*(2), 58–59.

Hutchison, K., Balota, D., Neely, J., Cortese, M., Cohen-Shikora, E., Tse, C.-S., . . . Buchanan, E. (2013). The semantic priming project. *Behavior Research Methods*, *45*(4), 1099-1114.

Järvikivi, J., & Pyykkönen, P. (2011). Sub- and supralexical information in early phases of lexical access. *Frontiers in psychology*, *2*, 282.

Jordan, T. R., Thomas, S. M., Patching, G. R., & Scott-Brown, K. C. (2003). Assessing the importance of letter pairs in initial, exterior, and interior positions in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(5), 883–893.

Kazanina, N. (2011). Decomposition of prefixed words in russian. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(6), 1371-1390.

Kazanina, N., Dukova-Zheleva, G., Geber, D., Kharlamov, V., & Tonciulescu, K. (2008). Decomposition into multiple morphemes during lexical access: A masked priming study of russian nouns. *Language and Cognitive Processes*, *23*(6), 800-823.

Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The British Lexicon Project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods*, *44*(1), 287–304.

Kruskal, W. (1987). Relative importance by averaging over orderings. *The American Statistician*, *41*(1), 6–10.

Landauer, T., & Dumais, S. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychological Review*, *104*(2), 211-240.

Lavric, A., Clapp, A., & Rastle, K. (2007). Erp evidence of morphological analysis from orthography: A masked priming study. *Journal of Cognitive Neuroscience*, *19*(5), 866–877.

Lindeman, R. H., Merenda, P. F., & Gold, R. Z. (1980). *Introduction to bivariate and multivariate analysis.* Scott, Foresman Glenview, IL.

Longtin, C., Segui, J., & Hallé, P. (2003). Morphological priming without morphological relationship. *Language and Cognitive Processes*, *18*, 313-334.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical cooccurrence. *Behavioural Research Methods*, *28*, 203-208.

Marelli, M., Amenta, S., Morone, E. A., & Crepaldi, D. (2013). Meaning is in the beholder's eye: Morpho-semantic effects in masked priming. *Psychonomic Bulletin & Review*, *20*(3), 534-541.

Marslen-Wilson, W. D., Bozic, M., & Randall, B. (2008). Early decomposition in visual word recognition: Dissociating morphology, form, and meaning. *Language and Cognitive Processes*, *23*(3), 394–421.

Maxwell, S. E., & Delaney, H. D. (1993). Bivariate median splits and spurious statistical significance. *Psychological Bulletin*, *113*(1), 181–190.

McCormick, S. F., Rastle, K., & Davis, M. H. (2008). Is there a fetein fetish? effects of orthographic opacity on morpho-orthographic segmentation in visual word recognition. *Journal of Memory and Language*, *58*(2), 307–326.

Morris, J., Frank, T., Grainger, J., & Holcomb, P. J. (2007). Semantic transparency and masked morphological priming: An erp investigation. *Psychophysiology*, *44*(4), 506–521.

Moscoso del Prado Martín, F., Bertram, R., Häikiö, T., Schreuder, R., & Baayen, R. H. (2004). Morphological family size in a morphologically rich language: The case of Finnish compared to Dutch and Hebrew. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *30*, 1271–1278.

Plaut, D. C., & Gonnerman, L. M. (2000). Are non-semantic morphological effects incompatible with a distributed connectionist approach to lexical processing? *Language and Cognitive Processes*, *15*(4/5), 445-485.

Rastle, K., & Davis, M. H. (2008). Morphological decomposition based on the analysis of orthography. *Language and Cognitive Processes*, *23*, 942-971.

Rastle, K., Davis, M. H., Marslen-Wilson, W. D., & Tyler, L. K. (2000). Morphological and semantic effects in visual word recognition: A time-course study. *Language and Cognitive Processes*, *15*, 507-537.

Rastle, K., Davis, M. H., & New, B. (2004). The broth in my brother's brothel: Morpho-orthographic segmentation in visual word recognition. *Psychonomic Bulletin & Review*, *11*, 1090–1098.

Rueckl, J. G., & Aicher, K. A. (2008). Are corner and brother morphologically complex? not in the long term. *Language and Cognitive Processes*, *23*, 972-1001.

Taft, M. (2004). Morphological decomposition and the reverse base frequency effect. *The Quarterly Journal of Experimental Psychology*, *57A*, 745–765.

Turney, P. D., & Pantel, P. (2010). From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research*, *37*, 141-188.

van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (in press). Subtlex-uk: A new and improved word frequency database for british english. *The Quarterly Journal of Experimental Psychology*.

Wagenmakers, E.-J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, *11*(1), 192–196.

White, S. J., Johnson, R. L., Liversedge, S. P., & Rayner, K. (2008). Eye movements when reading transposed text: the importance of word-beginning letters. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(5), 1261–1276.

Wu, C. F. J. (1986, 12). Jackknife, bootstrap and other resampling methods in regression analysis. *The Annals of Statistics*, *14*(4), 1261–1295.

Yarkoni, T., Balota, D., & Yap, M. (2008). Moving beyond colthearts n: A new measure of orthographic similarity. *Psychonomic Bulletin & Review*, *15*(5), 971–979.