Lucia Busso*, Florent Perek and Alessandro Lenci

Constructional associations trump lexical associations in processing valency coercion

https://doi.org/10.1515/cog-2020-0050
Received May 14, 2020; accepted February 6, 2021;
published online March 11, 2021

Abstract: The paper investigates the interaction of lexical and constructional meaning in valency coercion processing, and the effect of (in)compatibility between verb and construction for its successful resolution (Perek, Florent & Martin Hilpert. 2014. Constructional tolerance: Cross-linguistic differences in the acceptability of non-conventional uses of constructions. *Constructions and Frames* 6(2). 266–304; Yoon, Soyeon. 2019. Coercion and language change: A usage-based approach. *Linguistic Research* 36(1). 111–139). We present an online experiment on valency coercion (the first one on Italian), by means of a semantic priming protocol inspired by Johnson, Matt A. & Adele E. Goldberg. 2013. Evidence for automatic accessing of constructional meaning: Jabberwocky sentences prime associated verbs. *Language & Cognitive Processes* 28(10). 1439–1452. We test priming effects with a lexical decision task which presents different target verbs preceded by coercion instances of four Italian argument structure constructions, which serve as primes. Three types of verbs serve as target: lexical associate (LA), construction associate (CA), and unrelated (U) verbs. LAs are semantically similar to the main verb of the prime sentence, whereas CAs are prototypical verbs associated to the prime construction. U verbs serve as a mean of comparison for the two categories of interest. Results confirm that processing of valency coercion requires an integration of both lexical and constructional semantics. Moreover, compatibility is also found to influence coercion resolution. Specifically, constructional priming is primary and independent from compatibility. A secondary priming effect for LA verbs is also found, which suggests a contribution of lexical semantics in coercion resolution – especially for low-compatibility coercion coinages.

***Corresponding author: Lucia Busso**, Aston Institute of Forensic Linguistics, Aston University, Birmingham, UK, E-mail: l.busso@aston.ac.uk
Florent Perek, Department of English and Applied Linguistics, University of Birmingham, Birmingham, UK, E-mail: f.b.perek@bham.ac.uk
Alessandro Lenci, Dipartimento di Filologia, Letteratura, e Linguistica, Università di Pisa, Pisa, Italy, E-mail: alessandro.lenci@unipi.it

Keywords: coercion; construction grammar; lexical and constructional semantics; priming

1 Introduction

In many languages, verbs are notoriously flexible in how they combine with their argument structure – especially in English. Consider for example sentences (1) and (2). In both examples, taken from real-life uses in contemporary English, verbs have been used creatively to construct a new coinage, with a different meaning from their prototypical one. In particular, two instances of typically intransitive verbs (*dance*, *dream*) are construed as transitive. Example (1) could be roughly paraphrased by 'he pushed me down the garbage chute by dancing/with a dance move', and example (2) by 'I'm wasting my life by only concentrating on dreams (and not reality)'.

(1) *He almost danced me right down the garbage chute* (*Friends*, season 4 episode 7)

(2) *People say I'm lazy dreaming my life away* (John Lennon, "Watching the wheels")

Mismatches of this kind between the typical environments a verb is used in, and its occurrence in a novel, creative use, have been often discussed under the name of valency coercion. Examples such as (1) and (2) above, and the oft-cited example from Goldberg (1995: 9), repeated as (3) below, have typically featured prominently among the early arguments for the need for a construction grammar approach, especially in the domain of argument structure.

(3) *He sneezed the napkin off the table.*

In contrast to earlier lexicalist approaches to argument structure (e.g., Pinker 1989), Goldberg (1995) argued that the aspects of interpretation that are missing from the verb in coercion examples such as (1)–(3) are more naturally attributed to the syntax itself rather than to verb polysemy, which would leave unexplained the productive nature of this phenomenon. In other words, general clause structures are directly paired with abstract semantic representations and are combined more or less freely with particular verbs. In cases of coercion as well as in the more 'regular' uses of verbs, the overall meaning of a clause results from the combination of the meaning of the verb with that of an argument structure construction; namely, in the case of (1)–(3), the notion that someone causes something to move in some way, contributed by the so-called Caused Motion Construction (Goldberg 1995, 2006).

With a few notable exceptions, most research on valency coercion has been done on English. However, as some studies indicate (e.g., Perek and Hilpert 2014), English might be unusual in the way it allows words to combine flexibly with syntactic constructions, and it remains to be seen whether similar coercion phenomena can be observed as extensively in other languages. This paper is part of a research effort aimed at investigating valency coercion in Italian (Busso et al. 2018, 2020), a language on which construction grammar studies, and studies of coercion in particular, are still rather scarce. Additionally, while valency coercion and its representation have received much attention at the theoretical and descriptive levels, its psycholinguistic effects on online sentence comprehension have been far less studied. This paper seeks to mend this gap by investigating the processing of valency coercion sentences in Italian, by means of a semantic priming experiment. Furthermore, the present study also brings evidence for the constructional approach in general, in that we find that constructional priming is primary with respect to lexical priming.

The aim of the study is to provide experimental data on the processing of the new, coerced meaning. The experiment consists in a choice lexical decision task that presents subjects with different target verbs preceded by coercion sentences, which serve as primes. Specifically, following coercion coinages we present participants with verbs that are associated with the prime in different ways: either verbs which are semantically similar to the overall construction (construction associates), or verbs that are similar to the mismatching verb (lexical associates), or verbs that are completely unrelated to either the construction or the verb used in the prime.

This paradigm allows us to investigate lexical and constructional associations in coercion processing. In fact, we hypothesize that the meaning of the main verb interacts with the general constructional meaning in the processing and elaboration of the coerced interpretation. Starting from this assumption, we address several research questions:

- Does coercion resolution involve both verb semantics and constructional meaning?
- Which element is more important in processing coercion sentences?
- Does the degree of semantic compatibility between the filler and the general construction affect coercion resolution in online sentence processing?

The paper is organised as follows. In the next section, we introduce the phenomenon of coercion in its various forms, and we discuss how valency coercion in particular has been treated in previous research. In Section 3, we describe our experiment, whose results are reported in Section 4. Section 5 offers some discussion of these results and a conclusion to our study.

2 Previous research on coercion

The flexibility with which verbs combine with their argument structures has interested linguists for decades and has received different theoretical accounts over the years. Generative linguistics and other similar frameworks (generally called *projectionist approaches*) claim that the syntactic structure of sentences vastly depends on the lexical properties of the verbs (or other predicates) that head them. In other words, the verb *projects* the morphosyntactic realization of its own argument structure (cf. Chomsky 1981; Levin and Rappaport Hovav 1996; Rappaport Hovav and Levin 1998). However, a number of psycholinguistic works since the mid-80s have presented an innovative hypothesis: Learners of a language use knowledge about the abstract semantic content associated with syntactic patterns to infer novel verbs' meaning (the so-called "Syntactic Bootstrapping" hypothesis) (*inter alia*, Gillette et al. 1999; Landau and Gleitman 1985). This idea has been taken further by many acquisition studies that collectively provided extensive evidence of the fact that speakers associate argument structures with abstract semantic content (*inter alia* Bencini and Goldberg 2000; Goldwater and Markman 2009; Kako 2006; Kashak and Glenberg 2000).

This claim is the core assumption of Construction Grammar (Goldberg 1995, 2006; Hilpert 2014). In Construction Grammar, the basic unit of language is considered to be the *construction*, a form-meaning pair generally defined as follows:

> Any linguistic pattern is recognized as a construction as long as some aspect of its form or function is not strictly predictable from its component parts or from other constructions recognized to exist. In addition, patterns are stored as constructions even if they are fully predictable as long as they occur with sufficient frequency (Goldberg 2006: 5).

In other words, constructions are abstract units with an autonomous meaning, which is independent from and combines with the semantics of the lexical items that it accommodates. Thus, the overall meaning of a linguistic expression is a combination of both lexical elements (or fillers) and the general construction. Fillers and constructions both contribute different levels of semantic interpretation, as fillers typically have a richer and more specific meaning than the semantic content of abstract constructions. That is, in general the abstract semantic information carried by the construction is redundant with the meaning of the verb, which is a more specific instantiation of the same general event encoded by the construction. Take for example the Caused Motion Construction (Goldberg 1995), which encodes the motion of a Patient caused by an Agent. In (4), the verb *to put* is perfectly in line with the general event encoded by the abstract construction.

(4) *The boy puts the book on the shelf.*

However, in cases such as (1)–(3) above, a conflict arises between the lexical filler (*dance*, *dream*, *sneeze*) and the specifications of the slot it occupies within a construction (caused-motion verb in the Caused Motion Construction). These types of mismatches between fillers and constructions are instances of a wider range of phenomena referred to as *coercion* in the literature (Audring and Booij 2016; Lauwers and Willems 2011; Michaelis 2003, 2004; Pustejovsky 1995). At the most general level, coercion refers to cases in which the semantic interpretation of a sentence requires adjustments in the typical meaning of its lexical items. Audring and Booij (2016) attempt to reconcile the various and sometimes very different uses of the term by distinguishing between three levels of coercion, according to the type of semantic processes it involves: coercion by selection, coercion by enrichment, and coercion by override.

> In coercion by selection, the resulting meaning is a part of the semantic repertoire of the coerced word to begin with. From this perspective, coercion works largely 'bottom-up', with only a light role for the context selecting one interpretation from a range of alternative readings. In coercion by enrichment, lexical semantics is preserved, but augmented in context. The 'adapter plug' represents a stronger 'top-down' influence, adding meaning to the utterance. In coercion by override, in turn, the contextual top-down force is strongest; it modifies, replaces, or removes properties of the coerced item (Audring and Booij 2016: 628).

Coercion effects are of particular interest to Construction Grammar, as they have been typically used to provide convincing arguments for the idea of constructions as symbolic units that can alter or override selected semantic features of their fillers (Goldberg 2019). In other words, "coercion by construction" (Michaelis 2003, 2004) results in an adjustment of the lexical meaning of the mismatching filler, in line with the general semantic content of the construction (Audring and Booij 2016; Busso et al. 2018). Michaelis (2003) shows how different types of constructions create coercion effects: nominal constructions (as in mass-count coercion, e.g., *There was apple all over his shirt*), aspectual constructions (such as the English progressive construction, e.g., *I'm believing every word he's saying*), and argument structure constructions (like the Caused Motion construction, e.g., *She sneezed the foam off the cappuccino*). In this paper, we are interested in particular in coercion created by the latter type of constructions, i.e., mismatches that arise when verbs and argument structure constructions combine in novel and flexible ways, which we call valency coercion (Perek 2015: 31).

Over the past few years, coercion has been investigated in a growing number of studies using a range of different approaches involving psycholinguistic, neurolinguistic and corpus-based methodologies (*inter alia* Baggio et al. 2010; Gries et al. 2010; Lukassek et al. 2017; Piñango et al. 2006). However, despite coercion being one of the cornerstones of Construction Grammar, the wide majority of the literature is concerned with English, which is remarkable in this respect, as it displays an extremely high filler flexibility for a large number of constructions (Perek and Hilpert 2014).[1] Not many other languages have been investigated with respect to coercion phenomena. In particular, Romance languages have been found to be typically *valency-driven* – i.e., their verbs typically combine with a narrower range of argument structure constructions than so-called construction-driven languages such as English (Perek and Hilpert 2014; Rostila 2014). Hence, it remains to be seen whether Romance languages may display a similar range of coercion phenomena and similar degrees of coercion effects as English, especially valency coercion. While some scholars have looked at coercion in other Romance languages (cf. for example Gonzalvez-Garcia 2007; Lauwers 2008, 2014), Italian is still scarcely investigated in this respect (cf. Jezek and Lenci 2007).

In a companion study to the present one, Busso et al. (2018, 2020) presented Italian speakers with coercion sentences instantiating various Italian argument structure constructions, i.e., sentences with verbs that are not normally used in the relevant constructions but nonetheless form a sensical combination with it, as opposed to "impossible" sentences, in which no semantic combination between verb and construction could conceivably be imagined. They found that Italian speakers do accept coercion sentences, although to a lesser extent than typical verb-construction combinations, and their tendency to do so varies widely according to the construction. Importantly, Busso et al. (2018) using a distributional semantic model (Lenci 2018) also finds that the acceptability of coercion sentences is affected by the semantic density of the construction (see also Perek 2016). That is, coercion sentences are judged more natural if the target construction is observed with verbs belonging to a few semantically dense classes or subclasses (i.e., a tight semantic distribution centered on a few verb classes). Construction frequency (both type and token) is only found to be

1 See for instance the following examples (from Perek and Hilpert 2014: 267)
a. *John sneezed the napkin off the table* (Caused motion).
b. *Mary poured John another whisky* (Ditransitive).
c. *Emeril sliced and diced his way to TV stardom* (*Way*-construction).
d. *The truck rumbled down the street* (Intransitive motion).
e. *Pat kissed Bill unconscious* (Resultative).
f. *Kate hit at the wasp* (Conative).

significant in distinguishing impossible and coerced sentences. In sum, distributional properties of constructions suggest that the semantic similarity of verbs occurring in coerced coinages should be considered in studying coercion resolution processes.

This finding lines up with a growing consensus that novel uses of verbs are more successful when the verb is similar to other verbs that typically occur in the target argument structure construction (Barðdal 2008, 2011; Langacker 1987; Suttle and Goldberg 2011). In fact, similarity to witnessed instances has been argued to be the most significant factor for licensing novel coinages (Bybee and Eddington 2006; Kalyan 2012). Since semantic similarity or (in)compatibility are gradable properties, this entails that coercion effects are gradable rather than binary. It is expected that the gradable nature of coercion will require different degrees of processing effort depending on compatibility (Gries et al. 2010; Yoon 2013, 2016, 2019).

Little is known about the psycholinguistic processing of valency coercion, compared to its semantic description, its formal representation, or the factors influencing its acceptability. In the present contribution, we propose the first online psycholinguistic study on Italian valency coercion, which investigates the different roles of verb, construction, and the compatibility between them by means of a priming experiment. While both aspectual coercion and complement coercion (i.e., sentences that require inferring unspoken semantic elements because of a mismatch between verb and complement, as in *He began the article* vs. *He wrote the article*) have received much attention in the psycholinguistic and neurolinguistics literature (cf. Baggio et al. 2010; Lukassek et al. 2017; Piñango et al. 2006), the processing of sentences with valency coercion is a widely under-studied area. A notable exception is a series of studies by Yoon (2013, 2016), who argues that coercion should be studied as a cognitive phenomenon. Using self-paced reading experiments, she shows that processing time of English sentences instantiating the ditransitive construction (Yoon 2013) and the sentential complement construction (Yoon 2016) is correlated with the degree of compatibility between the verb and the construction. These and other studies (Busso et al. 2018, 2020; Yoon 2019) suggest that coercion effects arise from both "top-down" (i.e., constructional) and "bottom-up" (i.e., lexical) processes and that semantic (in)compatibility plays a crucial role. Drawing from this previous research, we present in the next section the structure of our experiment, describing the materials, experiment design and procedure.

3 The experiment

3.1 Priming effects in language

The present study uses a priming experiment to investigate the processing of coercion sentences. In cognitive psychology, priming refers to the observation that prior exposure to a certain stimulus (the prime) influences behaviour with respect to another stimulus (the target), which is usually taken to mean that the two stimuli are somehow cognitively related. Priming is considered by many psycholinguists as a way to address questions pertaining to the cognitive organization of lexicon and grammar. Semantic priming arises when a prime stimulus enhances identification of a semantically related target (Johnson and Goldberg 2013; Lucas 2000; Sperber et al. 1979). Semantic priming involves categorical similarities and facilitates the processing of words: an activation of the lexical network leads to a spread-over effect to prime-related entries (Hare et al. 2009; McRae et al. 2005).[2] Extensive evidence is also found for syntactic priming (Bock 1986; Bock and Loebell 1990), whereby processing of a given syntactic structure makes it more likely for speakers to select the same structure over a competitor in a subsequent production task.

Different experimental procedures to investigate priming effects all share the same overarching structure, which allows the researcher to investigate the effect of a set of primes on a set of targets. First, the participant is presented with a prime stimulus on a screen. After a brief pause (of a few milliseconds), the prime is followed by either a target stimulus or a prompt for a target action (especially for experiments investigating neuro-motor responses). The participant is asked to perform a certain task upon witnessing the target stimulus (e.g., word completion, lexical decision) or to execute the target action. The response of the participant is then analysed according to different properties of prime and/or target, depending on the research question. Priming effects in psycholinguistic research are often assessed via experimental designs that feature a lexical decision task. In this task, participants are required to perform a decision involving the target stimulus. Most typically, the task consists in deciding whether the target is a word of a given language or not. Participants are told to respond both as quickly and accurately as possible and are scored on one or both of these two dimensions (i.e., Reaction Time and Accuracy).

2 Note that this approach to lexical processing should not be taken as literally describing what is supposed to happen in the brain at the neuronal level, but merely as an abstract and somewhat simplistic model of cognitive processing that is compatible with a spreading-activation network representation. We thank an anonymous reviewer for suggesting this necessary clarification.

In a priming study that is especially relevant to the present experiment, Johnson and Goldberg (2013) used Jabberwocky sentences as primes, i.e., sentences that are syntactically well-formed instances of an argument structure construction but contain non-sensical words, e.g., *She daxed him the norp*, which instantiates the English ditransitive construction. Using different kinds of verbs as targets, Johnson and Goldberg report effects of semantic priming: Jabberwocky sentences strongly prime verbs that closely correspond to its prototypical meaning (e.g., *give* and *hand* for the ditransitive construction), and to a lesser extent semantically related verbs that occur only marginally in the construction, if indeed at all (e.g., *transfer*). They take this to mean that the semantic representation of a construction is accessed in language processing, even when the specific lexical items in use do not instantiate the constructional meaning.

In this study, we use a similar design to investigate what aspects of coercion sentences are activated when these sentences are processed by speakers. As in Johnson and Goldberg's study, we employ a lexical decision task protocol in which reaction times are measured. For the purposes of this study, accuracy values were not used as a dependent variable, as our interest lies specifically in the time span required to access meaning. Moreover, all participants performed at ceiling (see Section 3.4 below).

We use coercion sentences with different constructions as primes, and we use verbs related to the construction in different ways as targets: *construction associated verbs*, *lexical associated verbs*, and *unrelated verbs* as control. Hence, our experiment essentially tests what kinds of targets are affected by these primes in lexical decision; we expect that this should reveal what kind of cognitive structures are activated when processing coercion sentences (and to what extent), depending on the relation between target and prime.[3] *Construction Associate* (CA) targets are verbs that are prototypically used in the construction, with a meaning that closely corresponds to that of the construction.[4] For example, *give* is a CA of the ditransitive construction. Similarly to what Johnson and Goldberg (2013) found for Jabberwocky sentences, we expect coercion sentences to prime CA verbs, relative to an unrelated control, because the semantic representation of the construction needs to be activated for coercion sentences to be interpreted. However, the role of the verb, relative to that of the construction, could not be tested by Johnson and Goldberg, precisely because they used

3 See below Section 3.2.2 for an in-depth explanation of target stimuli and related examples taken from the dataset.

4 For clarity, this and all other abbreviations subsequently introduced in this paper are summarised in a table in Appendix 1; we thank an anonymous reviewer for making this suggestion. This includes abbreviations of verb types (CA, LA, U), constructions (CM, DT, IM, VD), and dependent variables (RT, MAR).

nonce verbs, which were semantically empty and with which participants had, by definition, no prior experience. In our experiment, this is achieved by including *Lexical Associate* (LA) targets, i.e. verbs that are semantically similar to the verbs used in the sentences. By testing whether and to what extent LA verbs are primed, we are able to determine to what extent the semantic representation of the verb is activated when processing coercion sentences, compared to that of the construction. *Unrelated* (U) targets are verbs that are neither semantically similar to the construction nor to the verb in the prime sentence.

In the next section, we describe the stimuli used as primes and targets in the experiment and how they were selected. We then describe the general design of the experiment and the procedure used to test our participants.

3.2 Materials

3.2.1 Prime stimuli

Coerced instances of argument structure constructions were used as primes. Prime sentences belong to four Italian constructions chosen from previous research (Busso et al. 2018, 2020): Caused Motion (CM), Dative (DT), Intransitive Motion (IM) and Sentential or *Verba Dicendi* (VD), as shown in (5)–(8) below.

These particular constructions were chosen for a twofold reason. They are structures of intermediate schematicity (Barðdal 2008, 2013), and of different flexibility (or *coercibility*), as found in a previous experiment by Busso et al. (2018): in an acceptability rating task with nine different constructions, IM and VD constructions were found to be highly coercible. DT – on the other hand – was found to be the less acceptable when coerced. Statistical significance was not in evidence for CM.

For the present study, coercion stimuli were constructed using non-prototypical yet semantically compatible verbs for the target constructions, in line with our hypothesis:

(5)	*Il gatto*	*graffia via*	*la vernice*	*dalle sedie.*	[CM]
	'The cat	scratches off	the paint	from the chairs'	

(6)	*La donna*	*sbriciola*	*pane agli uccelli.*	[DT]
	'The woman	crumbles	bread to birds'	

(7)	*Il bambino*	*trotterella via*	*da scuola*	*felice.*	[IM]
	'The boy	trots away	from school	happy'	

Table 1: Verb classes and types used for the creative condition.

Constructions	Prototypical verb type	Coerced verb type
Caused motion (CM)	Causative (or force-exertion) verbs that encode movement. e.g.: *mettere* – 'to put'	Intransitive or not- prototypically transitive verbs that do not entail motion, but from which it can be presumed as an implication. e.g.: *Giulia starnutisce via il tovagliolo dal tavolo* – 'Giulia sneezes the napkin off the table'
Dative (DT)	Transitive verbs encoding a transfer event. e.g. *dare* –'to give'	Intransitive or not- prototypically transitive verbs that do not entail transfer, but from which it can be presumed as an implication. e.g.: *sorridere* – 'to smile'
Intransitive motion (IM)	Intransitive motion verbs e.g.: *andare* – 'to go'	Manner of motion verbs. e.g.: *strisciare* – 'to slither'
Verba dicendi (VD)	Verbs of saying or telling e.g.: *dire* – 'to say'	Sound emission verbs e.g.: *fischiettare che* … – 'to whistle that …'

(8) *Giovanni fischietta che verrà domani.* [VD]
'Giovanni whistles that he will arrive tomorrow'.

The coercion stimuli were selected from corpus data. The corpus of contemporary Italian *ItWac* (Baroni et al. 2009), accessed via the online corpus software *SketchEngine* (Kilgarriff et al. 2014), was used as a primary source.[5] Creative sentences (defined as *hapaxes* or rare occurrences for operational purposes) were mainly retrieved through CQL queries.[6] A minority of stimuli was retrieved by translating English coercion examples, or by casual encounters in everyday language by the first author, and subsequently checked on *ItWac* and/or Google. All selected stimuli are attested. A total of 68 stimuli were selected.

These sentences were edited appropriately to match the same structure, i.e., a subject in the third person singular and the main verb in the present tense (see examples above). Table 1 provides an overview of the constructions used in our stimuli, the prototypical verb types that are known to occur in them, and the verb types that were used in the coerced sentences we used as prime stimuli.

5 ItWac is a 1.5 billion words corpus of Italian texts collected from the Internet.

6 The queries searched for a given construction (e.g., CM) excluding all prototypical verbs to reveal uncommon uses. For example, a simplified and general query for the CM construction would be [tag = "N.*"][tag = "V.*" & lemma! = "mettere|tirare|rimuovere…."][]?[tag = PRE & lemma = *list of locative prepositions*][tag = "N."]. When necessary, the searches were refined by retrieving a frequency list of the query. It is important to note that the queries were not designed to be precise, but rather have a high recall, since the aim of such queries was to find creative uses of verbs rather than all instances of a given construction in the corpus.

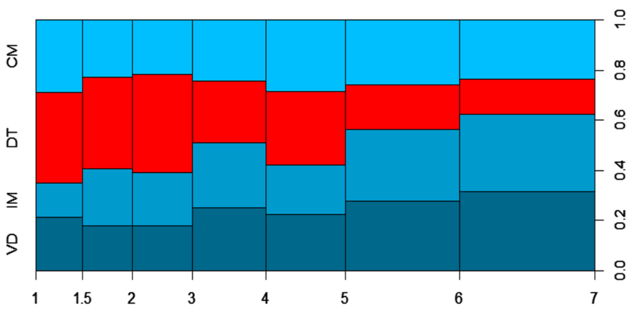

Figure 1: Mosaic plot of the distribution of acceptability ratings across the four constructions.

As we mentioned above, one of our main hypotheses is that compatibility influences coercion resolution. To include only stimuli with a partial compatibility (neither perfect, nor too low), we collected acceptability ratings, as the degree of compatibility also determines the degree of naturalness of the new formulation. Hence, the coercion prime stimuli were presented to 20 Italian native speakers, who were asked to rate the naturalness of the sentences on a 1–7 Likert scale. Stimuli were presented in random order in a Google form. Participants were presented all stimuli retrieved from the corpus.

Notably, stimuli of the DT construction received the lowest scores for naturalness overall, while IM and VD instances were perceived as more acceptable, consistently with what was found in previous research (Busso et al. 2020). Figure 1 visually represents the distribution of ratings for each Likert scale point. DT ratings are highlighted in red.

Only stimuli that received mean ratings between 2.5 and 6.5 were included. These thresholds were chosen to include an equal number of stimuli per construction, while at the same time removing stimuli that were judged as perfectly natural or nearly impossible. The range is skewed towards the high end of the Likert scale as an informal survey among participants revealed a similar skewness in judgments as well. In fact, by their own admission, the majority rated "perfectly acceptable" sentences as seven exclusively, while "completely unacceptable" sentences were rated as either one or two.

The final experiment set resulted in a total of 15 sentences for each construction (for a total of 60 sentences) in the specified interval (mean = 4.7, SD = 1.08). Stimuli were also normalized for character length, so that mean character lengths are similar across constructions (See Table 2).[7]

7 The only lexical items that were edited to normalize sentence length were those irrelevant to our research questions, i.e., we did not change the main verbs, but only complements, for example using longer or shorter proper nouns, or adding adverbial adjuncts (especially in the case of the IM construction, which, being intransitive, has naturally shorter instances).

Table 2: Descriptive statistics of character length for all stimuli in the dataset, for each construction: Caused Motion (CM), Dative (DT), Intransitive Motion (IM), *Verba Dicendi* (VD).

length				
	CM	**DT**	**IM**	**VD**
Mean	40.5	38.2	38.5	39.07
Std. dev.	2.5	2.8	2.4	1.1
C.I.	40.3–40.7	37.9–38.4	37.2–39.8	38.3–38.6

3.2.2 Target stimuli

Three types of target verbs were matched to the primes: construction associate (CA), lexical associate (LA), and unrelated (U) verbs. CAs are prototypical verbs occurring in the prime construction. The verbs also were retrieved through CQL queries of the target construction in *ItWac*. The 15 most frequently recurring verbs were selected with the "node forms" function on *SketchEngine* and then manually checked.

LAs are semantically similar to the main verb of a given prime sentence. They were chosen in a production task performed by 35 native Italian speakers. Participants were presented (in randomized order on a Google form) with the main verbs of the prime sentences in the infinitive form (e.g., the verb *graffiare*, 'to scratch', from example 2), and were asked to provide up to five related verbs. The most frequent productions were then used as LAs in the test.

Since LAs and CAs should prime different meaning components, the two sets of verbs should not be too similar. To assess the extent of similarity between LAs and CAs we employed the distributional measure of cosine distances.[8] We performed this control using the *SNAUT* interface (Mandera et al. 2017). We used the Italian model *WEISS* (Word-Embeddings Italian Semantic Spaces [Marelli 2017]) to compute cosine distances between pairs of verbs (CA – LA) in the dataset. We checked that pairwise cosine distances between LA and CA verbs were above the 0.6 threshold, which makes them well above the 0.5 midpoint of cosine distances, which are bounded between 0 and 1. Even though cosines are not linearly distributed, in practice a 0.6 distance threshold works well for our purposes.

Finally, U verbs serve as a control condition for the two aforementioned categories of interest. U verbs were also selected with *WEISS*. We relied on

8 Cosine distance (d) is a measure strictly related to cosine similarities (s): $s = 1 - d$. The conceptual difference is that cosine distance measures dissimilarity between words.

Table 3: Character length for target verbs: Construction Associates (CA), Lexical Associates (LA), Unrelated (U).

	CA	LA	U
Mean	8	8.5	8.1
Standard deviation	1.7	1.3	1.5
Confidence interval	7.8–8.2	8.4–8.6	8–8.2

distributional similarity to retrieve verbs that are completely unrelated to our stimuli both from a lexical and constructional perspective. Cosine dissimilarity allowed us to easily select verbs that are both semantically dissimilar and that appear in different constructional environments. We selected verbs with a cosine distance higher than 0.75 from both verbs in prime sentences and from CA targets. Verbs too were normalized for character length (Table 3).

To provide an example of the types of experimental targets in our dataset, target stimuli associated to the prime sentence in example (5) above (*Gianni fischietta che verrà domani*) are the three following verbs:

a. *dire* (to say, CA)
b. *canticchiare* (to hum, LA)
c. *invecchiare* (to age, U).

Table 4 summarizes the descriptive statistics of cosine distances for each target type. Unsurprisingly, LA targets are more similar to the verbs in prime sentences than CAs. Figure 2 visually represents the similarity of both targets for each construction.

Statistical significance between CA and LA was checked with a one-tailed two-sample *t*-test, which confirmed that CA distance is significantly higher than LA ($t = 9.910$, p-value < 0.0001). That is, LA verbs are significantly more similar to main verbs in coerced primes, while the CA condition consists of verbs that are more related to the general construction than to its main verb per se.

Table 4: Descriptive statistics of cosine distances of prime verbs to the three types of target verbs: Construction Associates (CA), Lexical Associates (LA), Unrelated (U).

	Distance CA	Distance LA	Distance U
Mean	0.76	0.48	0.83
Std. deviation	0.15	0.14	0.05
Minimum	0.4	0.25	0.75
Maximum	0.97	0.8	0.9

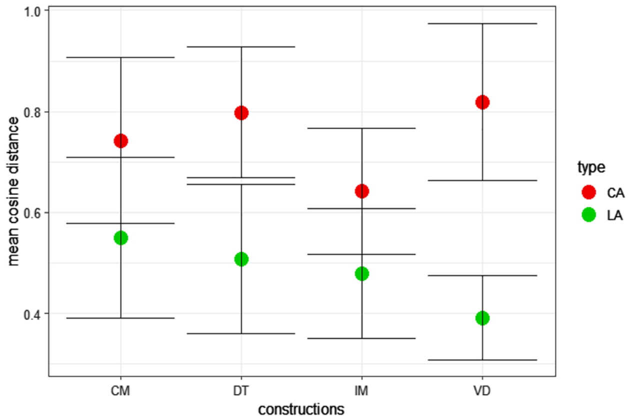

Figure 2: Mean cosine distances of prime verbs from construction associate targets (CA, in red) and from lexical associate targets (LA, in green), per each construction, with SD error bars.

3.2.3 Fillers

Beside primes and targets, a third important element to the design of a priming experiment is the incorporation of filler items, i.e., linguistic items that serve the purpose of distracting participants by hiding observable patterns of the critical items. Furthermore, by presenting fillers and primes in randomized order, participants are also prevented from picking up (subconsciously) information about stimuli distribution, which may lead to strategic responses based on their understanding of the experiment rather than the question of interest.

Therefore, a set of 20 filler sentences (filler primes) paired with 20 non-words (filler targets) were included. Filler primes were instances of different constructions than the ones in the stimuli. Each was paired with a non-word filler target created with the Italian version of *Wuggy* (Keuleers and Brysbaert 2010), "a pseudoword generator particularly geared towards making nonwords for psycholinguistic experiments". Non-words are used as filler targets given the nature of the lexical decision task itself: participants are asked to decide as quickly and accurately as possible if the string of characters appearing on the screen is an Italian verb or not, a task which only makes sense if some of these strings are indeed not actual words.

In creating non-words, the skewed percentage of the three Italian conjugations in the lexicon was maintained (I conj.: *-are*, 80%; II conj.: *-ere*, 10%; III conj.: *-ire*, 10%; data from Lubello 2016). Table 5 below reports the final composition of the complete experimental dataset.

Table 5: Final dataset.

Constructions	4
Prime sentences	60 (15 per construction)
Target verbs	180 (three per sentence)
Filler sentences + non-words	20

3.3 Experiment design and predictions

The experiment follows a between-subjects design. Three groups of participants were presented with the whole set of prime sentences, but in a different combination with target verbs. The order of appearance of the stimuli was randomised, but the matching of primes with targets was the same for all participants. Fillers and non-words were constant across groups.

The rationale behind this experiment design is that we expect reaction times to be significantly faster when participants are presented with congruent associations of prime sentence-target verb (either LA or CA) than when they are exposed to unrelated verbs. Additionally, we also hypothesize that coercion resolution is influenced by the dynamic interrelation of the meaning of verb and construction, but that constructional semantics is nonetheless more strongly activated when interpreting coercion sentences, and therefore primary in the interpretation process. Accordingly, CA target verbs should be more strongly primed by coercion sentences than LAs, and thus yield faster reaction times. Since we assumed that compatibility plays a role in the processing of coercion effects, we also expect sentence acceptability to influence reaction times.

3.4 Procedure

Stimulus presentation and data collection were controlled by means of the *PsychoPy* software package (version 1.90, the latest at the time of the experiment; Peirce 2007) running on an HP personal computer. In each trial, participants were asked to read the sentence on screen carefully. Visual presentation of the prime stimuli lasted 4000 ms. After a brief fixation cross (1000 ms), they were presented with one of the target verbs, or one of the non-word filler items, and had to decide whether it was an actual Italian word by pressing the left key (for a non-word) or right key (for a word).

Before the actual experiment, a short trial session (with unrelated constructions) was conducted, to familiarise participants with the task and procedure. For this reason, feedback was also provided during trials: after every choice, an appropriate message would appear to the participant: *Giusto!* ('Correct!') if they correctly identified a word or a non-word, *Ops! Sbagliato!* ('Oops! Wrong!') if they got it wrong. Trial data was not included in the analysis. The stimuli were presented in randomized order.

Reaction times from 39 Italian participants (13 per group) were collected (13 M, 26 F; mean age = 35.13, st. dev. = 13.68). For the purposes of this work, we do not consider the accuracy of response (as mentioned in Section 3.1). In fact, there is evidence in psychology that the two variables should not be assumed to measure the same underlying cognitive process (Santee and Egeth 1982; van Ede et al. 2012). Moreover, we are specifically interested in reaction time – which in our assumption approximates processing – and accuracy would not be informative for our aim, since in the experimental design the lexical choice task is not only trivial, but was also helped by a trial section. Accuracy rates are in fact at ceiling (at or near 100% accuracy) for all the variables of interest. Accuracy rates are slightly lower for filler (i.e., non word) recognition (see Table 6 below). Since fillers were excluded from the analyses however, this does not impact our findings.

All participants were adult native speakers of Italian, with normal or corrected-to-normal vision. Three participants were left-handed, but a preliminary inspection of the data did not reveal any significant errors in the task. All participants were either enrolled in a University course or already in possession of (at least) a bachelor's degree. Geographical origin was not considered nor controlled for.

4 Results and analysis

Following Baayen and Milin (2010) and Whelan (2008), the reaction time data (henceforth: RT) was subjected to mild a-priori screening and trimming of outliers.

Table 6: Accuracy rates per target type.

Target type	Errors	Accurate responses	Accuracy rate	St. dev.
CA	1	779	99.9%	0.004
LA	1	779	99.9%	0.004
U	2	778	99.7%	0.006
Fillers	23	757	97%	0.018

Table 7: Descriptive statistics of the data, per construction: Caused Motion (CM), Dative (DT), Intransitive Motion (IM), *Verba Dicendi* (VD), and per target verb type: Construction Associates (CA), Lexical Associates (LA), Unrelated (U).

General			
	CA	**LA**	**U**
Mean	1,045	1,175	1,356
St. dev.	420.0	482.8	517.0
CM			
Mean	1,038.1	1,141.9	1,142.2
St. dev.	406.2	449.4	738
DT			
Mean	1,064.7	1,117.2	1,386
St. dev.	435.7	494.3	586.7
IM			
Mean	1,098.6	1,227.7	1,301.7
St. dev.	597.2	541.7	493
VD			
Mean	1,025.2	1,290.5	1,467.4
St. dev.	388.7	572.9	601.5

This inspection allows to clean the data of observed data points which are not a result of the process of interest. Following Luce (1986), RTs shorter than 100 ms and longer than 3500 ms were considered errors. However, to account for the inevitable temporal lag caused by measuring RTs with a keypress, an initial 50 ms latency was considered. We therefore eliminated all RTs faster than 150 ms and slower than 4000 ms. In fact, RT longer than 4000 ms could reflect low-familiarity or a subject's distraction, rather than lexical access.

Table 7 reports descriptive statistics of all data and for each construction. For clarity, RTs are reported in milliseconds and not in logarithmic scale.

The relations among the variables of interest are graphically explored in the figures below. The violin plot in Figure 3 illustrates the differences in data distribution between the three target types. Similarly to box plots, violin plots depict data distribution, with the addition of the probability density of the data at different values (smoothed by a kernel density estimator). The black line in the middle of the box plot is the median value. From the plot in Figure 3, RTs for U verbs appear to be longer than for the other two conditions, and RTs for LAs are longer than for CAs. Most notably, this holds true for all our constructions.

Our hypotheses thus seem to be supported by the data. In the next section, we provide a detailed statistical analysis of these interactions.

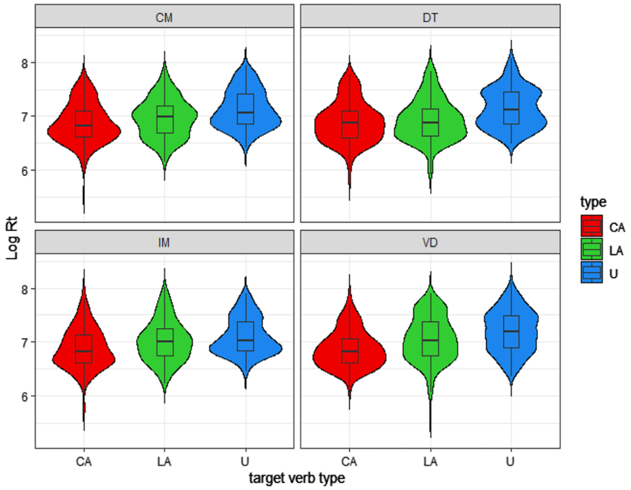

Figure 3: Violin plot of the distribution of (log) RTs for the three verb types in each of the four constructions.

4.1 Priming effects

To assess statistical significance, a linear mixed effects model was fitted to the data, using the R package *lmerTest* (Kuznetsova et al. 2017).[9] Beside the predictors (verb type and construction), we also included corpus frequencies (normalized for 1 M words) of both target verbs and verbs in the prime stimuli. The frequencies were retrieved from the ItWac corpus for consistency.

Model selection was performed via Likelihood Ratio Test, as implemented by the R package *afex* (Singmann 2020).[10] Both predictors of interest (verb type and construction) were significant, as was their interaction. However, both target and prime frequencies were not, and we therefore decided to exclude them from the final model (verb type: chi sq. = 294.7, $p < 0.0001$; construction: chi sq. = 9.36, $p < 0.05$; verb type:construction: chi sq. = 21.37, $p < 0.005$; prime frequency: chi sq. = 1.83, $p > 0.1$; target frequency: chi sq. = 2.1, $p > 0.1$).We report below (in *R* syntax) the formula of the model.

9 We are grateful to Bodo Winter and one anonymous reviewer for useful comments and suggestions on the model.

10 R syntax formula for the maximal model: *log RT ~ verb type + construction + verb type:-construction + prime frequency + target frequency + (1 + verb type + construction | subject) + (1 + verb type | item).*

$$LogRT \sim verb\,type + construction + verb\,type * construction + (1 + verb\,type + construction \,|\, subject) + (1 + verb\,type \,|\, item)$$

Contrasts were sum-coded for the factor "construction"; the factor "type" was releveled to have U as a reference level, so that RTs of CA and LA target verbs would be compared to the ones of the unrelated verbs. Fixed effect results of the model, complete with R^2 values, confidence intervals (CI), and standardised beta coefficients – i.e., the measure of how many standard deviations the dependent variable changes per standard deviation increase in the predictor – were obtained with the *SjPlot* package (Lüdecke 2020) (Table 8).

Results from the statistical model confirm that CA and LA reaction times significantly differ from the intercept, i.e., U verbs. To check whether the difference between the two target verb types is also significant, we fitted a pairwise post-hoc comparison of the factor "verb type" using the R package *emmeans* (Lenth et al. 2019). The analysis compares estimated marginal means of the model, i.e., computes a prediction of the mean at each point of the reference grid (a grid of all combinations of reference levels). Estimated marginal means (EMMs) are defined as equally weighted means of these predictions at specified margins. All the interactions turned out significant (Table 9). This suggests that participants were faster in recognizing verbs associated with the general semantic content of

Table 8: Fixed effects table of the model. In order, it presents results for verb type, construction, and the interaction between the two.

Predictors	***Estimates***	***Std. beta***	***CI***	***Standardized CI***	***p***
(Intercept) U	**7.15*** **	0.37	7.07 to 7.23	0.17 to 0.58	<0.001
CA	**−0.27*** **	−0.69	−0.31 to −0.22	−0.81 to −0.58	<0.001
LA	**−0.16*** **	−0.40	−0.20 to −0.11	−0.52 to −0.28	<0.001
CM	−0.01	−0.02	−0.05 to 0.03	−0.12 to 0.08	0.690
DT	0.00	0.00	−0.04 to 0.04	−0.09 to 0.10	0.954
IM	**−0.05** **	−0.12	−0.08 to −0.01	−0.21 to −0.02	0.01
VD	**0.05** **	0.13	0.01 to 0.09	0.03 to 0.24	0.01
CA: CM	0.00	0.01	−0.05 to 0.06	−0.14 to 0.15	0.935
LA: CM	−0.02	−0.05	−0.07 to 0.03	−0.18 to 0.08	0.455
CA: DT	0.01	0.04	−0.04 to 0.07	−0.11 to 0.18	0.616
LA: DT	**−0.06** **	−0.16	−0.11 to −0.01	−0.29 to −0.03	0.01
CA: IM	0.05+	0.12	−0.01 to 0.10	−0.02 to 0.26	< 0.1
LA: IM	**0.07** **	0.18	0.02 to 0.12	0.06 to 0.31	0.005
CA: VD	**−0.06* **	−0.16	−0.12 to −0.01	−0.31 to −0.02	<0.05
LA: VD	0.01	0.03	−0.04 to 0.06	−0.10 to 0.15	0.679
Marginal R^2/Conditional R^2 0.087/0.560					

Table 9: Pairwise contrasts of the factor "verb type": Construction Associates (CA), Lexical Associates (LA), Unrelated (U).

Contrast	Estimate	SE	*t*. ratio	*p*-value
U – CA	**0.269*** **	0.023	11.8	<0.0001
U – LA	**0.156*** **	0.023	6.67	<0.0001
CA – LA	–**0.114*** **	0.019	–6.04	<0.0001

the construction than verbs associated with the lexical meaning of the verb of the sentence. We interpret this finding as supporting our hypothesis that processing of coercion sentences is influenced by both lexical and construction meaning, but constructional meaning seems to be primary. This result is more clearly represented in Figure 4, which plots fixed effects estimates of the model for each target type.[11]

The effect of verb type on RTs varies according to the construction. Significant effects are found for the interaction of LA and DT. LA RTs are significantly faster than average (–0.06), whereas no effect is found for CA RTs (0.01). We cautiously interpret this finding as supporting our claim that compatibility is an important factor in successful coercion resolution. In fact, speakers consistently rated DT as less coercible than other constructions (see Section 2 and Busso et al. 2020), which may suggest that for non-felicitous resolutions, lexical priming is significantly

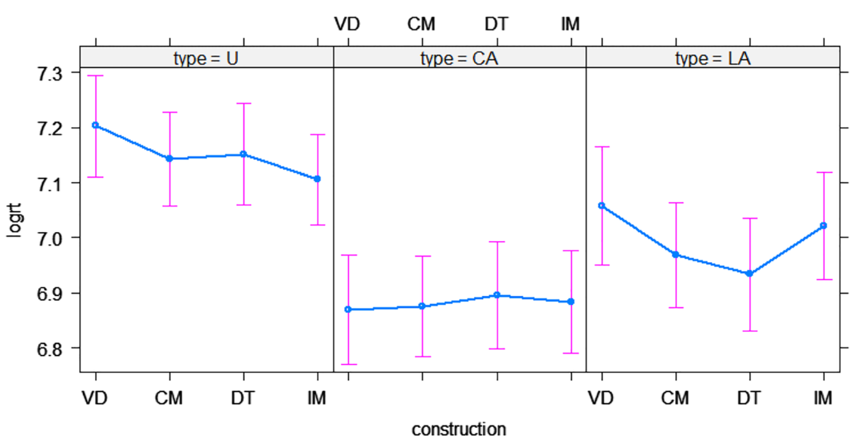

Figure 4: Fixed effects divided for target type, with 95% CI error bars.

11 Estimates were extracted from the model with the R package *effects* (Fox 2003).

faster than average. In other words, participants did not generalize beside the main verb of DT coercion stimuli: lexical semantics prevails over constructional semantics because of the difficulty to coerce creative DT sentences into a cohesive meaning. For IM, instead, both RTs are relatively slow, with CA being faster (0.05) than LA RTs (0.07); in this case, constructional meaning is better primed. The VD construction displays a clear preference of constructional meaning, with CA RTs being significantly faster (−0.06) than LA (0.01). In other words, both constructions show significant constructional priming. As in the case of DT, this finding suggests that higher naturalness – i.e., compatibility – leads to a predominant effect of construction over lexical priming.

4.2 Compatibility effects

We have so far determined that the processing time of coercion is influenced by both lexical and constructional semantics, with construction emerging as the primary factor, and that CA priming effects seem to prevail for instances of constructions which were previously rated by speakers as relatively acceptable. However, whether and how the processing time of coercion effects is influenced by the degree of compatibility between verb and construction remains yet to be answered.

As mentioned earlier (see Section 2.1), we approximate compatibility with the acceptability judgements collected for data selection. We use acceptability as a proxy for semantic compatibility as Yoon (2013 and 2016) showed that the parameters of acceptability, semantic compatibility and frequency of usage are closely correlated. According to Yoon, verbs that are more compatible with a target construction are "used more frequently, processed more rapidly, and judged more acceptable when they are used in the construction" (Yoon 2016: 77). Hence, we consider coercion coinages with a high acceptability rating as being more compatible as well.

To render the continuous predictor of mean acceptability ratings (henceforth: MAR) more interpretable, the variable was centred. Centering is a linear transformation that subtracts the mean to each data point to render continuous predictor variables more interpretable. As the mean is "centered" on zero, values above mean will be positive and values below will be negative (Winter 2019).

Figure 5 fits a linear regression line to (log) RTs and MAR. Error bands represent the standard error of mean.

The *y* axis represents the mean RTs of each stimulus in each of the three conditions, for ease of visualization. That is, we average RTs across participants for each sentence in each condition (i.e. we computed the mean of the log-transformed

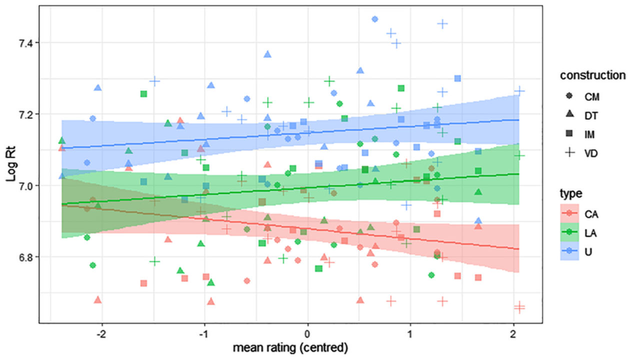

Figure 5: Relationship of RTs as a function of MAR (mean acceptability rating) for each stimulus for the three target types: Construction associates (CA), lexical associates (LA), unrelated (U).

RT). Hence, we have 60 datapoints, one for each experimental stimulus. The *x* axis represents the mean acceptability rating (again, averaged over participant) for each stimulus. The shape of the data points (i.e. the stimuli) depends on the construction used in each stimulus: instances of CM are plotted as circles, instances of DT are triangles, etc.

We can see increasingly slower RTs in the LA condition (i.e., a higher y value) as the acceptability of the prime sentences increases (i.e., a higher x value). The RTs for the CA condition instead become faster the more natural the stimuli are. With very unnatural stimuli (bottom left of the graph) the two conditions largely overlap. As primes become more acceptable – and hence with a higher compatibility – the two conditions gradually diverge. In other words, the data seems to suggest that priming effects for coercion structures are in an inverse relation with their MAR: Coercion sentences with high MAR prime CA verbs more strongly than LA verbs, but as MAR decreases, RTs to LA verbs become shorter, and LA verbs are increasingly primed as strongly as CA verbs.

We test this hypothesis – namely that MAR of prime sentences differently affects RTs for CA and LA targets – by fitting a second mixed effect model. We fit a second model instead of adding predictors to the first one (see Section 4.1) for a twofold reason: firstly, to avoid three-way interactions in the same model (construction * verb type * MAR), and secondly to avoid adding too many predictors to the model. As we are dealing with a relatively limited dataset in this exploratory study (each prime/target combination is evaluated by 13 individuals), high-order interactions and a high number of predictors should be avoided in such a context (Field et al. 2012).

As predictors, we set MAR (averaged for each stimulus over participants) in interaction terms with target verb type. The continuous predictor of MAR is centred. For the factor predictor of verb type, U was chosen as a reference level. As a dependent variable, we used log-transformed RTs. For model selection, we use LRT as in the previous model. We built a maximal model with random slopes for type, and included frequency (prime and target) as a fixed effect. Since the model did not converge, we removed random slopes. The model selection process revealed that the interaction of verb type and mean acceptability rating is significant. Since prime and target frequency do not reach significance, we excluded them from the final model (Verb type: chi sq.: 295.21, $p < 0.001$; MAR: chi sq.: 1.72, >0.1; freq_prime: chi sq.: 1.13, $p > 0.1$; freq_target: chi sq.: 2.25, $p > 0.1$; verb type:MAR: chi sq.: 4.87, $p = 0.05$). Table 10 reports fixed effects and R^2 values for the model, whose formula in *R* syntax is as follows:

$$\text{Log}\, RT \sim verb\, type * MAR + (1\,|\,subject) + (1\,|\,item)$$

Results corroborate the positive influence of MAR on target verbs: as MAR increases, LA targets RTs increase as well. In other words, stimuli with a low semantic compatibility between verb and general construction (i.e., stimuli with low MAR) elicit faster RT for LA targets than stimuli with a higher degree of compatibility (i.e., higher MAR). Interestingly, no discernible effect of MAR on CA targets is found. This suggests that constructional priming is not affected by compatibility, whereas lexical priming is. Figure 6 plots the effects of the model.

We interpret these findings as reinforcing our hypothesis that coercion resolution depends on a combination of both verb and constructional semantics. Lexical semantic specifications of the verb are dominant when the required coercion is stronger, that is, when the constructional meaning is particularly distant from the preferred verb usage and imposes a deeper change in the verb meaning to produce the new coinage. In other words, for highly compatible coinages (i.e., with a high MAR), verb semantics is only minimally adapted to

Table 10: Fixed effects of the second model. It presents estimates for the predictor of 'verb type', 'MAR', and the interaction between the two.

Predictors	***Estimates***	***Std. beta***	***CI***	***Standardized CI***	***p***
(Intercetpt)	**7.16*****	0.36	7.07 to 7.24	0.15 to 0.58	**<0.001**
CA	**−0.27*****	−0.69	−0.30 to −0.24	−0.76 to −0.62	**<0.001**
LA	**−0.16*****	−0.39	−0.18 to −0.13	−0.46 to −0.32	**<0.001**
MAR	−0.01	−0.02	−0.02 to 0.01	−0.07 to 0.04	0.6
CA * MAR	0.01	0.02	−0.02 to 0.03	−0.05 to 0.10	0.497
LA * MAR	**0.03***	0.08	0.00 to 0.05	0.01 to 0.16	**<0.05**
Marginal R^2/Conditional R^2 0.080/0.514					

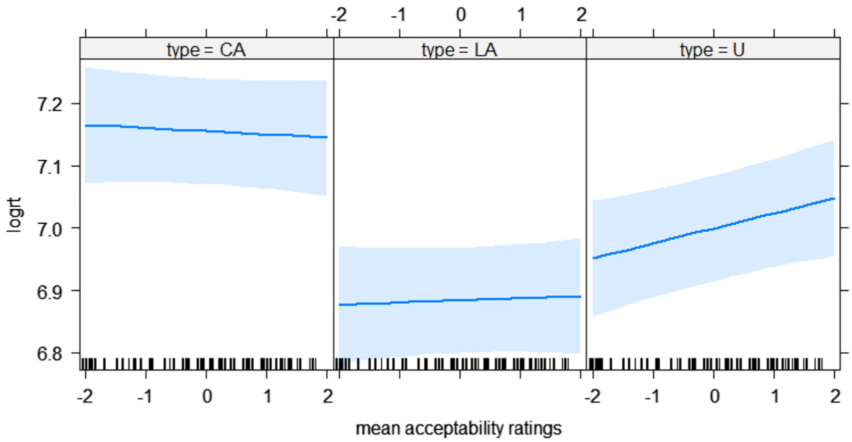

Figure 6: Fixed effects of the second model, with 95% CI error bands.

requirements of the constructional meaning. The mismatch is easily resolved by imposing constructional meaning upon the new sentence. As compatibility decreases, coercing coinages using almost exclusively constructional meaning becomes more difficult, and the verb meaning becomes more dominant.

For example, the processing of VD sentences such as example (7), *Giovanni <u>fischietta</u> che verrà domani* 'Giovanni whistles that he will arrive tomorrow', will require a minimal contribution of the lexical semantics of *fischiettare* 'to whistle', as the specifications of verb and general construction do not differ too much from each other. Therefore, the effect of the construction overrides that of lexical semantics, and the mismatch is easily resolved using the top-down influence of constructional meaning. Instead, for low-compatibility DT coinages such as example (5), *La donna <u>sbriciola</u> pane agli uccelli* 'The woman crumbles bread to the birds', the semantic specifications of construction and verb are more distant. When coercing the sentence into a new meaning, the verb meaning becomes more dominant, because lower compatibility makes it harder to discern a coherent combination of verb and construction.

5 Discussion and conclusions

In this paper, we have presented one of the first psycholinguistic experiments on valency coercion, and the first one that investigates the phenomenon in Italian. The findings of this study provide important evidence into the nature of constructional meaning that carries theoretical significance for construction grammar in general, well beyond the specific language investigated in the experiment. In fact, we have presented psycholinguistic evidence that suggests that coercion processing requires a dynamic interaction of lexical and constructional meaning, with constructional meaning being primary. Particularly, we hypothesize – building on previous research – that coercion processing involves both verb and constructional semantics, and that its successful resolution depends on the degree of compatibility between the main verb and the general construction (Busso et al. 2018, 2020; Perek and Hilpert 2014; Rostila 2014; Yoon 2016). These hypotheses follow a recent line of research that advocates the gradable nature of coercion and the importance of filler-construction dynamic interrelations (Goldberg 2006; Johnson and Goldberg 2013; Suttle and Goldberg 2011; Yoon 2013, 2016). To test our hypotheses, two mixed effect models were fitted to the experimental data. The overall findings validated our initial assumptions.

The first model found a statistically robust priming effect for both constructional and lexical meaning. Constructional priming (i.e., priming of CA targets) was stronger overall than priming effects for lexical associated (LA) targets. These results suggest that coercion resolution relies on the dynamic interaction between verb and constructional meaning, with a predominant role played by constructional semantics. Moreover, the DT construction – which was found to be difficultly coercible – shows significantly faster-than-average reaction times (RTs) for LA targets. Vice versa, easily coercible constructions (IM and VD) display faster RTs for CA targets. This finding points to an influence of compatibility over priming effects, and hence coercion processing.

To test the effect of compatibility, a second model incorporated averaged acceptability judgments for each sentence as a predictor in interaction with target type. In fact, following recent research such as the papers by Yoon (2013, 2016), we approximate naturalness of the coinage – and hence compatibility – with the sentence ratings provided by speakers. We found that the degree of compatibility of prime stimuli for each construction significantly affected RTs to target verbs: as naturalness increases, RTs for LA targets increase as well. In other words, when compatibility is high, the resolution of coercion effects does not heavily rely on verb semantics but rather on the overall construction, and priming of LAs target verbs decreases (as RTs increase). When compatibility is low, speakers rely more on lexical semantics. This finding corroborates our initial hypothesis that compatibility influences coercion resolution.

In sum, the present work finds that constructional priming is primary across stimuli and trump effects from lexical semantics: RTs are overall significantly faster for CA verbs, meaning that prior exposure to a coerced instance of a construction *X* elicits a priming effect on a verb strongly associated with *X*. Moreover, constructional priming is also found not to be significantly influenced by compatibility, which is consistent with our hypothesis. Constructional priming is hence primary over lexical priming and independent from the naturalness of the coercion prime.

However, in line with our initial assumptions, we also found that coercion resolution also elicits a secondary priming effect for verbs associated with the mismatching filler-verbs. This finding suggests a contribution of lexical semantics in coercion resolution. We find further evidence to this claim in our finding that lexical priming is stronger for low compatibility constructions, suggesting that lexical semantics of the main verb is particularly relevant when the semantic specifications of the verb diverge significantly from those of the coercing construction. Results align with a growing trend in the current literature that increasingly emphasizes how coercion resolution depends on the interaction of different elements, such as filler-construction compatibility, or pragmatic inferences (see also Leclerq 2019; Mitkovska 2019; Yoon 2019).

The primary effect of constructional priming that we found in our data constitutes substantial evidence for the cognitive linguistic tenet that learners are sensitive to statistical information above the level of individual verbs (Perek and Goldberg 2017; Wonnacott et al. 2008). Additionally, this is also in line with usage-based and Construction Grammar principles, as constructions are only partially productive, as they can be extended to new coinages only with a limited range of items (Bowerman 1988; Goldberg 1995, 2019; Pinker 1989; Suttle and Goldberg 2011). The acceptability of new coinages, that is, is limited by the constraints of the general constructions to which novel uses of verbs must conform (Ambridge et al. 2018; Boyd and Goldberg 2011). Different types of constraints have been discussed in the literature. The most widely accepted are statistical pre-emption (or blocking) and coverage, which both consider the level of the verb and that of the construction. Statistical pre-emption assumes that novel formulations (i.e., verb + construction) compete with other similar formulations that might be expected to occur in a given context. Coverage, instead, refers to the degree to which attested instances (i.e., verbs) fill the semantic or phonological space that includes the target instance (see Goldberg 2019 for an in-depth discussion of both). More generally, we could say that constraints on constructional flexibility recognise an important role for the compatibility of fillers and construction. Cross-linguistic experimental support to statistical pre-emption and coverage is still an understudied area to date.

In this study we provided support to the validity of Construction Grammar and substantiate the claim of (valency) coercion being a gradable phenomenon. In processing novel coinages, the top-down “coercing” influence of the construction interacts with verb-level lexical semantics to produce a successful resolution of coercion effects – and hence a semantically sensical coinage.

The present paper leaves the field open for much needed additional research. Studies with a higher number of participants are needed to strengthen our findings, and more studies on under-investigated languages are needed to provide further information on the cross-linguistic nature of coercion phenomena.

Moreover, a further in-depth investigation of the interrelation between fillers and constructions could shed more light on coercion effects and their resolution.

List of abbreviations

Abbreviation	Meaning	Explanation
CA	Constructional associate (verbs)	Prototypical verbs for a given construction in the primes. e.g., *dire* ('to say') is a CA of the VD construction
LA	Lexical associate (verbs)	Verbs semantically similar to the main verbs of the prime stimuli. e.g., *canticchiare* ('to hum') is LA to the verb *fischiettare*
U	Unrelated (verbs)	Verbs completely different from either the main verb and the general construction of the prime, used as a baseline for comparing CA and LA. e.g., *invecchiare* (' to grow old') is U for the sentence *Gianni fischietta che verrà domani*
CM	Caused motion (construction)	An Agent (subject) directly causes the movement of the Theme (direct object) from a location A to a location B along a path specified by a Locative phrase. e.g., *Giacomo tossisce via vino dal bicchiere* ('Giacomo coughs wine off the glass')
DT	Dative (construction)	An Agent (the subject), transfer (literally or metaphorically) a Theme (direct object) to a Recipient. e.g., *Il bambino scalcia il giocattolo all'amico* ('The boy kicks the toy to his friend')
IM	Intransitive motion (construction)	A Theme (subject) moves along a Path (Locative phrase) to a different point in space. e.g., *Lorenzo sfreccia via dall'ospedale* ('Lorenzo speeds away from the hospital')
VD	Verba dicendi (construction)	An Agent (subject) communicates a Theme (relative clause). e.g., *Gianni fischietta che verrà domani* ('Gianni whistles that he will come tomorrow')
RT	Reaction time	Measure (in ms) of the time employed by subjects to perform the lexical decision task
MAR	Mean acceptability ratings	Mean measure of acceptability of coercion stimuli, used as a dependent variable as a proxy for compatibility.

References

Ambridge, Ben, Libby Barak, Elizabeth Wonnacott, Colin Bannard & Giovanni Sala. 2018. Effects of both preemption and entrenchment in the retreat from verb overgeneralization errors: Four reanalyses, an extended replication, and a meta-analytic synthesis. *Collabra: Psychology* 4(1). 23.

Audring, Jenny & Geert Booij. 2016. Cooperation and coercion. *Linguistics* 54(4). 617–637.
Baayen, Harald & Petar Milin. 2010. Analyzing reaction times. *International Journal of Psychological Research* 3(2). 12–28.
Baggio, Giosuè, Travis Choma, Michiel van Lambalgen & Hagoort Peter. 2010. Coercion and compositionality. *Journal of Cognitive Neuroscience* 22(9). 2131–2140.
Barðdal, Jóhanna. 2008. *Productivity: Evidence from case and argument structure in icelandic* (Constructional Approaches to Language), vol. 8. Amsterdam: John Benjamins.
Barðdal, Jóhanna. 2011. Lexical vs. structural case: A false dichotomy. *Morphology* 21(3–4). 619–654.
Barðdal, Jóhanna. 2013. Construction-based historical-comparative reconstruction. In Thomas Hoffman & Graeme Trousdale (eds.), *The Oxford handbook of construction grammar*, 438–457. Oxford: Oxford University Press.
Baroni, Marco, Silvia Bernardini, Adriano Ferraresi & Eros Zanchetta. 2009. The WaCky wide web: a collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation* 43(3). 209–226.
Bencini, Giulia & Adele E. Goldberg. 2000. The contribution of argument structure constructions to sentence meaning. *Journal of Memory and Language* 43(4). 640–651.
Bock, Kathryn. 1986. Syntactic persistence in language production. *Cognitive Psychology* 18(3). 355–387.
Bock, Kathryn & Helga Loebell. 1990. Framing sentences. *Cognition* 35(1). 1–39.
Bowerman, Melissa. 1988. The ‘no negative evidence’ problem: How do children avoid constructing an overly general grammar? In John A. Hawkins (ed.), *Explaining language universals*, 73–101. Oxford: Basil Blackwell.
Boyd, Jeremy K. & Adele E. Goldberg. 2011. Learning what not to say: The role of statistical preemption and categorization in a-adjective production. *Language* 87. 55–83.
Busso, Lucia, Alessandro Lenci & Florent Perek. 2020. Valency coercion in Italian. *Construction and Frames* 12(2). 171–205.
Busso, Lucia, Pannitto Ludovica & Alessandro Lenci. 2018. Modelling Italian construction flexibility with distributional semantics: Are constructions enough? In Elena Cabrio, Alessandro Mazzei & Fabio Tamburini (eds.), *Proceedings of the fifth Italian conference on computational linguistics (CLiC-it 2018)*, 68–74. Torino: Accademia University Press.
Bybee, Joan L. & David Eddington. 2006. A usage-based approach to Spanish verbs of “Becoming”. *Language* 82(2). 323–355.
Chomsky, Noam. 1981. *Lectures on government and binding: The Pisa lectures*. Berlin: De Gruyter.
Field, Andy, Jeremy Miles & Zoë Field. 2012. *Discovering statistics using R*. New York: Sage Publications.
Fox, John. 2003. Effect displays in R for generalised linear models. *Journal of Statistical Software* 8(15). 1–27.
Gillette, Jane, Gleitman Henry, Lila Gleitman & Anne Lederer. 1999. Human simulations of vocabulary learning. *Cognition* 73(2). 135–176.
Goldberg, Adele E. 1995. *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.
Goldberg, Adele E. 2006. *Constructions at work: The nature of generalization in language* (Oxford Linguistics). New York: Oxford University Press.
Goldberg, Adele E. 2019. *Explain me this*. Princeton: Princeton University Press.
Goldwater, Micah B. & Arthur B. Markman. 2009. Constructional sources of implicit agents in sentence comprehension. *Cognitive Linguistics* 20(4). 273–278.

González-García, Francisco. 2007. 'Saved by the reflexive': Evidence from coercion via reflexives in verbless complement clauses in English and Spanish. *Annual Review of Cognitive Linguistics* 5(1). 193–238.
Gries, Stephan Th., Beate Hampe & Doris Schönefeld. 2010. Converging evidence II: More on the association of verbs and constructions. In Sally Rice & John Newman (eds.), *Empirical and experimental methods in cognitive/functional research*, 59–72. Stanford, CA: CSLI.
Hare, Mary, Micheal Jones, Caroline Thomson, Sarah Kelly & Ken McRae. 2009. Activating event knowledge. *Cognition* 111(2). 151–167.
Hilpert, Martin. 2014. *Construction grammar and its application to English*. Edinburgh: Edinburgh University Press.
Jezek, Elisabetta & Alessandro Lenci. 2007. When GL meets the corpus: A data-driven investigation of semantic types and coercion phenomena. In Pierrette Bouillon, Laurence Danlos & Kyoko Kanzaky (eds.), *Proceedings of the 4th international workshop on generative approaches to the Lexicon (May 10–11, 2007)*, 1–9. Paris: France.
Johnson, Matt A. & Adele E. Goldberg. 2013. Evidence for automatic accessing of constructional meaning: Jabberwocky sentences prime associated verbs. *Language & Cognitive Processes* 28(10). 1439–1452.
Kako, Edward. 2006. The semantics of syntactic frames. *Language & Cognitive Processes* 21(5). 562–575.
Kalyan, Siva. 2012. *Similarity in linguistic categorization: The importance of necessary properties*. Berlin: De Gruyter Mouton.
Kaschak, Michael P. & Arthur M. Glenberg. 2000. Constructing meaning: The role of affordances and grammatical constructions in sentence comprehension. *Journal of Memory and Language* 43(3). 508–529.
Keuleers, Emmanuel & Marc Brysbaert. 2010. Wuggy: A multilingual pseudoword generator. *Behavior Research Methods* 42(3). 627–633.
Kilgarriff, Adam, Vít Baisa, Bušta Jan, Miloš Jakubíček, Vojtěch Kovář, Michelfeit Jan, Pavel Rychlý & Vít Suchomel. 2014. The sketch engine: Ten years on. *Lexicography* 1(1). 7–36.
Kuznetsova, Alexandra, Per B. Brockhoff & Rune H. B. Christensen. 2017. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* 82(13). 1–26.
Landau, Barbara & Lila R. Gleitman. 1985. *Language and experience: Evidence from the blind child*. London: Harvard University Press.
Langacker, Ronald W. 1987. *Foundations of cognitive grammar: Theoretical prerequisites*. Stanford: Stanford University Press.
Lauwers, Peter. 2008. The nominalization of adjectives in French: From morphological conversion to categorial mismatch. *Folia Linguistica* 42(1–2). 135–176.
Lauwers, Peter. 2014. Between adjective and noun: Category/function mismatch, constructional overrides and coercion. In Simone Raffaele & Francesca Masini (eds.), *Current issues in linguistic theory*, vol. 332, 203–226. Amsterdam: John Benjamins.
Lauwers, Peter & Dominique Willems. 2011. Coercion: Definition and challenges, current approaches, and new trends. *Linguistics* 49(6). 1219–1235.
Leclercq, Benoît. 2019. Coercion: A case of saturation. *Constructions and Frames* 11(2). 270–289.
Lenci, Alessandro. 2018. Distributional models of word meaning. *Annual Review of Linguistics* 4. 151–171.
Lenth, Russell, Henrik Singmann, Jonathon Love, Buerkner Paul & Maxime Herve. 2019. emmeans: Estimated marginal means, aka least-squares means. Available at: https://CRAN.R-project.org/package=emmeans.

Levin, Beth & Malka Rappaport Hovav. 1996. Lexical semantics and syntactic structure. In Shalom Lappin (ed.), *The handbook of contemporary semantic theory*, 487–507. Oxford: Blackwell.
Sergio Lubello (ed.). 2016. *Manuale di linguistica italiana* (Manuals of Romance linguistics 13). Berlin: De Gruyter.
Lucas, Margery. 2000. Semantic priming without association: A meta-analytic review. *Psychonomic Bulletin & Review* 7(4). 618–630.
Luce, Paul A. 1986. *Neighborhoods of words in the mental Lexicon. Research on speech perception*. Speech Research Laboratory, Psychology Department, Indiana University. Technical Report No. 6.
Lüdecke, Daniel. 2020. *sjPlot: Data visualization for statistics in social science*. R package version 2.8.3. Available at: https://CRAN.R-project.org/package=sjPlot.
Lukassek, Julia, Prysłopska Anna, Robin Hörnig & Claudia Maienborn. 2017. The semantic processing of motion verbs: Coercion or underspecification? *Journal of Psycholinguistic Research* 46(4). 805–825.
Mandera, Paweł, Emmanuel Keuleers & Marc Brysbaert. 2017. Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *Journal of Memory and Language* 92. 57–78.
Marelli, Marco. 2017. Word-embeddings Italian semantic spaces: A semantic model for psycholinguistic research. *Psihologija* 50(4). 503–520.
McRae, Ken, Mary Hare, Jeffrey Elman & Ferretti Todd. 2005. A basis for generating expectancies for verbs from nouns. *Memory & Cognition* 33(7). 1174–1184.
Michaelis, Laura. 2003. Headless constructions and coercion by construction. In Elaing Francis & Laura Michaelis (eds.), *Mismatch: Form-function incongruity and the architecture of grammar*, 259–310. Stanford, CA: CSLI.
Michaelis, Laura A. 2004. Type shifting in construction grammar: An integrated approach to aspectual coercion. *Cognitive Linguistics* 15(1). 1–67.
Mitkovska, Liljana. 2019. Types of modality in South Slavic stative reflexive-dative constructions. *Slověne = Словьне. International Journal of Slavic Studies* 8(2). 260–287.
Peirce, Jonathan W. 2007. PsychoPy—psychophysics software in Python. *Journal of Neuroscience Methods* 162(1–2). 8–13.
Perek, Florent. 2015. *Argument structure in usage-based construction grammar: Experimental and corpus-based perspectives*. Amsterdam: John Benjamins.
Perek, Florent. 2016. Using distributional semantics to study syntactic productivity in diachrony: A case study. *Linguistics* 54(1). 149–188.
Perek, Florent & Adele E. Goldberg. 2017. Linguistic generalization on the basis of function and constraints on the basis of statistical preemption. *Cognition* 168. 276–293.
Perek, Florent & Martin Hilpert. 2014. Constructional tolerance: Cross-linguistic differences in the acceptability of non-conventional uses of constructions. *Constructions and Frames* 6(2). 266–304.
Piñango, Maria Mercedes, Aaron Winnick, Rashad Ullah & Edgar Zurif. 2006. Time-course of semantic composition: The case of aspectual coercion. *Journal of Psycholinguistic Research* 35(3). 233–244.
Pinker, Steven. 1989. *Learnability and cognition: The acquisition of argument structure*. Cambridge, MA: MIT Press.
Pustejovsky, James. 1995. *The generative Lexicon*. Cambridge, MA: MIT Press.

Rappaport Hovav, Malka & Beth Levin. 1998. Building verb meanings. In Miriam Butt & Wilhelm Geuder (eds.), *The projection of arguments: Lexical and compositional factors*, 97–134. Stanford, CA: CSLI Publications.

Rostila, Jouni. 2014. Inventarisierung als Grammatikalisierung: produktive Praepositionalobjekte und andere grammatikalisierte Linking-Muster. In Lasch Alexander & Ziem Alexander (eds.), *Grammatik als Netzwerk von Konstruktionen: Sprachwissen im Fokus der Konstruktionsgrammatik.* (Sprache Und Wissen 15), 97–116. Berlin: de Gruyter.

Santee, Jeffrey & Howard Egeth. 1982. Do reaction time and accuracy measure the same aspects of letter recognition? *Journal of Experimental Psychology: Human Perception and Performance* 8(4). 489–501.

Singmann, Henrik, Bolker Ben, Jake Westfall, Frederik Aust, S Mattan, Ben-Shachar, Søren Højsgaard, John Fox, Michael A. Lawrence, Ulf Mertens, Jonathon Love, Russell Lenth, Rune Haubo Bojesen Christensen. 2020. afex: Analysis of factorial experiments. R package version 0.28-0. Available at: https://CRAN.R-project.org/package=afex.

Sperber, Richard D., Charley McCauley, Ronnie D. Ragain & Carolyne M. Weil. 1979. Semantic priming effects on picture and word processing. *Memory & Cognition* 7(5). 339–345.

Suttle, Laura & Adele E. Goldberg. 2011. The partial productivity of constructions as induction. *Linguistics* 49(6). 1237–1269.

van Ede, Freek, Floris de Lange & Eric Maris. 2012. Attentional cues affect accuracy and reaction time via different cognitive and neural processes. *Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 32(30). 10408–10412.

Whelan, Robert. 2008. Effective analysis of reaction time data. *The Psychological Record* 58(3). 475–482.

Winter, Bodo. 2019. *Statistics for linguists: An introduction using R*. London: Routledge.

Wonnacott, Elizabeth, Elissa L. Newport & Michael K. Tanenhaus. 2008. Acquiring and processing verb argument structure: Distributional learning in a miniature language. *Cognitive Psychology* 56(3). 165–209.

Yoon, Soyeon. 2013. Correlation between semantic compatibility and frequency: A usage-based approach. *Linguistic Research* 30(2). 243–272.

Yoon, Soyeon. 2016. Gradable nature of semantic compatibility and coercion: A usage-based approach. *Linguistic Research* 33(1). 95–134.

Yoon, Soyeon. 2019. Coercion and language change: A usage-based approach. *Linguistic Research* 36(1). 111–139.

Supplementary Material: The online version of this article offers supplementary material (https://doi.org/10.1515/cog-2020-0050).