COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 45 (2021) 13050

ISSN: 1551-6709 online
DOI: 10.1111/cogs.13050

Neglected Factors Bearing on Reaction Time in Language Production

Tobias Scheer, Fabien Mathy

Université Côte d'Azur, CNRS, Bases Corpus Langage (BCL), France

Received 20 May 2020; received in revised form 27 August 2021; accepted 29 August 2021

Abstract

The input to phonological reasoning are alternations, that is, variations in the pronunciation of related words, such as in *electri[k]* - *electri[s]-ity*. But phonologists cannot agree what counts as a relevant alternation: the issue is highly contentious despite a research record of over 50 years. We believe that the experimental setup presented may contribute to this debate based on a kind of evidence that was not brought to bear to date. Our experiment was thus designed to distinguish between alternations where phonological computation plays no role, and those where it has contributed to language production. The design manipulates two factors that to date have not been considered in experimental studies of language production: linguistic complexity and alternation specificity. The former is understood as extra processing demands issued by two types of linguistic activity, morphosyntactic and phonological. Our results show that reaction time latencies are longer when participants are burdened with both morphosyntactic and phonological tasks than when they carry out just a morphosyntactic task, and they are still shorter in absence of both types of demands. These results allowed us to address alternation specificity, that is, the fact that different alternations (within a language or across languages) may be driven by different production routines (an idea that is consensual among linguists but underdeveloped in the psycholinguistic literature). Our study shows that four different alternations in French produce alternation-specific signatures in reaction time latencies. These findings may thus redound to the advantage of psycholinguistic studies by identifying two new factors, as well as produce results that speak to the linguistic quarrels.

Keywords: Alternation specificity; Language production; Linguistic complexity; Morphosyntax; Phonology; Reaction time

---

Correspondence should be sent to Tobias Scheer, 24 av. des Diables Bleus, F-06300 Nice, France. E-mail: scheer@unice.fr

1. Introduction

1.1. Linguistic quarrels and two factors in language production

Despite a research record of over half a century, what constitutes a relevant phonological alternation is still controversial among linguists. A typical case in point is the variable pronunciation of the stem *electric*, whose last segment is pronounced [k] in *electri[k]*, but [s] in *electri[s]-ity*: whether or not phonological computation is responsible for the k-s alternation is contentious. Sahin, Pinker, Cash, Schomer, and Halgren (2009) have designed a protocol that, although not made for this purpose, can help phonologists decide whether language production in a given alternation involves simple lexical access, additional concatenation of pieces, or a true online phonological computation on top of that. Our experiment based on this protocol followed the hypothesis that the extra processing activity involved in the two latter options produces longer response times.

Our results also speak to psycholinguistic issues. They suggest that in past and current experimental studies of language production in the area of phonology, two factors have been neglected that may play a role: alternation specificity and linguistic complexity. We believe that moving them up on the agenda in experimental design, results and data interpretation may redound to the discipline's advantage.

1.2. Alternation specificity

An alternation is a situation where a given item has two or more realizations. An example from morphology is the definite masculine article in Italian, which is realized as either *il* or *lo* depending on phonological context: *lo* occurs when the following word begins with *s* followed by another consonant (*lo studio* "the study") or a geminate (*lo gnomo* [lo ɲɲomo] "the gnome"), while *il* is found elsewhere (*il treno* "the train"). In phonology, [k] and [s] realize the stem-final consonant of the same stem, such as in English *electri[k]* and *electri[s]-ity*. *Il* and *lo* alternate, and so do [k] and [s].

Processing routines (in production or perception) are alternation specific when two distinct alternations produce different behavior in response to the same experimental conditions. This would be the case if, for example, in English the k-s alternation were not produced by the same processing mechanism as, say, the stem-final g-zero alternation in *sign* [-n] - *si[gn]ature*. Processing routines may also be language specific. This would be the case if, for example, the mechanism managing the English k-s alternation were not identical to, say, the one that drives the same k-s alternation in French (*électrique* [k] "electric" - *électri[s]-ité* "electricity").

Studies which are concerned with only a single alternation thus do not speak to alternation specificity. For instance, Miozzo and Caramazza's (1999) work on the influence that an intervening adjective has on definite article selection in Italian (*lo sgabello* "the stool" vs. *il piccolo sgabello* "the small stool" vs. *il piccolo treno* "the small train") is not about alternation specificity since only one alternation is considered (*il-lo*).

We study alternation specificity because, as we explain below, it is viewed as an established fact on the linguistic side, where the debate is more fine-grained: it is the interpretation of individual alternations that is disputed. Our linguistically motivated focus appears to be understudied in the psycholinguistic literature, though, which typically assumes either implicitly or explicitly that processing routines are the same across alternations (Section 1.2.1). Given the above definition of alternation specificity, we could not identify any psycholinguistic study concerned with this issue.

1.2.1. Psycholinguistic literature

The idea that processing routines in language production are alternation-general, that is, the same for all alternations of a language, or even universal, that is, the same for all languages, is quite widespread. Regarding language specificity, Caramazza, Miozzo, Costa, Schiller, and Alario (2001: 209) write that "current models of language production have emphasized the language-universal aspects of the process; that is, the details of the processing routines (processing levels, units, time course of the computations) are always thought to be identical across languages."

The universalist position may be illustrated (among others, see the discussion in Section 4.5.2) by Bürki, Frauenfelder, and Alario (2015). Studying French liaison in the indefinite article *un* "a" (whose *-n* is pronounced before vowel-initial words as in *un* [ɛ̃n] *accident* "an accident" but not before consonant-initial words as in *un* [ɛ̃] *café* "a coffee"), the authors conclude that *un* has two lexical recordings, with (/ɛ̃n/) and without (/ɛ̃/) the *-n* (rather than one single lexical recording /ɛ̃n/ whose *-n* is then deleted in appropriate environments). In the last sentence of the article, the authors generalize from this particular alternation to all cases of systematic variation in French and in language as such: the results "suggest that systematic variation is better accounted for in terms of multiple abstract word form representations than in terms of contextual rules."

Such a step can only be taken on the assumption that all alternations in all languages have the same workings. If, say, the alternation between t and zero that is observed in the root of the verb *partir* "to leave" in French (*il part* [paʁ] "he leaves" vs. *nous part-ons* [paχt-ɔ̃] "we leave") were due to a different production routine that issues different processing demands, the way liaison works in *un* does not allow us to draw any conclusion on how the alternation in *partir* is managed.

Based on a cross-linguistic study of the production of determiners in noun phrases (NPs), the above mentioned contribution by Caramazza et al. (2001) was the starting point for cross-linguistic work on closed versus open class items, showing that language-specific routines known from perception also exist in production (Alario & Caramazza 2002, Alario, Ayora, Costa, & Melinger, 2008, Foucart, Branigan, & Bard, 2010, Janssen & Caramazza 2003, Spalek, Bock, & Schriefers, 2010).

While this is evidence for language-specific production routines, it does not speak to alternation specificity: open and closed lexical classes are not alternations since they are not surface realizations of the same item (unlike *il* and *lo*, which are surface realizations of the definite article).[1] As was mentioned, alternation-specific studies compare the behavior produced by distinct alternations within a given language in response to the same experimental conditions.

Beyond the above mentioned article by Bürki et al. (2015), the listing–decomposition debate is a relevant example for the poor consideration of alternation specificity. The more general question here is whether the pieces that the analyst can identify in a morphologically complex item such as *dark-ness* are really stored independently in long-term memory (decomposition), or rather as one single piece (listing), or as either (depending on lexical frequency). Relevant literature includes Bürki et al. (2015), Bürki, Ernestus, and Frauenfelder (2010), Butterworth (1983), Janssen, Bi, and Caramazza (2008), Levelt (2001), Levelt, Roelofs, and Meyer (1999), Pinker (1991), Caramazza, Laudanna, and Romani (1988), Hay (2001), and Schreuder and Baayen (1995). Informed surveys appear in Bölte, Zwitserlood, and Dohmes (2004) and Cohen-Goldberg (2013). Based on this literature where evidence from specific processes in particular languages is adduced, the meta-analysis by Cohen-Goldberg (2013) concludes that in the competition of listing and decomposition, the latter is massively supported, even though some evidence for the former exists. This is interpreted as a competition which can only have one winner. The idea that listing and decomposition may co-occur across languages (language specificity) or within a given language (alternation specificity) and that this is not self-contradictory in any way appears to be left unconsidered (see the discussion in Section 4.5.2).

1.2.2. Linguistic literature

In the linguistic literature, alternation specificity is an established fact and there is a broad consensus that each and every alternation needs to be studied in its own right (see also Section 4.5.2). The interpretation of individual alternations is disputed, though, and has produced ever inconclusive controversies: given, for example, *electric- electricity*, are we facing 1° suppletion, 2° allomorphy, or 3° morphophonology (these analytic options familiar in linguistic quarters will be explained shortly)? All attempts at establishing a set of criteria (called evaluation measure or evaluation metrics) undertaken since the 1970s were unsuccessful (Goyvaerts, 1981; Kiparsky, 1974, see the survey for 3° morphophonology by Bermúdez-Otero & McMahon 2006, pp. 383ff). On the syntactic side, the issue is known as *lexicalism* (Williams, 2007).

Today, probably all phonologists agree that a pair such as *eye–ocular* represents two independent lexical entries (but this has not always been the case, Lightner, 1978). On the other end of the scale, linguists will also typically agree that alternations produced by inflectional morphology such as the progressive voice assimilation that occurs in English plural formation (/rock/ + /z/ → *rock*-[s]) are true instances of 3° phonological computation. There is a large middle ground, though, where probably each and every alternation may and will be disputed: the question is alternation specific. A particularly prominent item located in this gray zone is the alternation between [k] and [s] that was used for the purpose of illustration above (*electri[k] - electri[s]ity*), called *velar softening* (Chomsky & Halle 1968, pp. 219ff, 426f; Hooper, 1975, pp. 544f; Kiparsky, 1982, pp. 40f; Halle & Mohanan 1985; Harris, 1994, pp. 21ff; Kaye, 1995, pp. 312, 328; Coleman, 1995, pp. 375ff; Halle, 2005; McMahon, 2007, see the surveys in Hayes, 1995 and Green, 2007, pp. 172ff).

In order to be able to characterize phonological computation as it really occurs in the human mind (and ultimately in the brain), we thus need to "set limits to the space of possible interactions between phonology, morphology, and the lexicon: in particular, […] [we] must ascertain the proper division of labor between storage and computation," as Bermúdez-Otero (2012: 8) puts it. Therefore, the question of the purview of phonology is what Bermúdez-Otero and McMahon (2006: 389) call the Gordian knot of the discipline. In the current situation, the competition among theories is significantly biased: a theory that accounts for velar softening in the phonology cannot be compared to a theory considering that velar softening has nothing to do with phonology. The set of things to be explained is not the same, and significantly diverges at the scale of a language, let alone phonology as such. Before theories can compete, the question what a true phonological phenomenon is thus needs to be addressed.

In other words, phonology as a field of scientific inquiry is currently in a position where its very input is unclear: what phonological reasoning is fed with most often depends on the gut feeling of each analyst, based on their theoretical inclination or intuitive views on what counts as a relevant alternation. Unsurprisingly enough, competing theories built on these wildly varying sets of empirical material significantly diverge—not because of the theorizing itself but because of the uncertain input.

This situation is well illustrated by a key debate regarding the aforementioned listing–decomposition issue. Following a modular approach to the organization of the mind (Fodor, 1983), Pinker (1991) holds that the past tense of regular verbs in English (*walk* - *walked*) is typically decomposed (with the pieces $\sqrt{}$*walk-* and *-ed* combined through computation), while the past tense of irregular verbs is always listed, that is, stored as one single piece (*hold* - *held*). The alternative connectionist position where a single mechanism produces both regular and irregular past tense forms is taken by Rumelhart and McClelland (1986). The debate is recapitulated in Pinker and Ullman (2002a, b) and McClelland and Patterson (2002a, b). In our terms, it opposes an alternation-specific (Pinker et al.) and an alternation-unspecific (McClelland et al.) perspective. The former follows the view held on the linguistic side whereby no evidence withstands the coexistence of listing and decomposition (even in a given language and concerning the same grammatical form, past tense). In this case, the two alternations at hand are called regular versus irregular, that is, involving a suffixal (*-ed* added) versus a stem-internal (modification of the stem vowel) management of past tense.

1.3. Reaction time and decomposition into individual linguistic tasks

Like alternation specificity, linguistic complexity is a factor that to date plays no role in language production studies as far as we can see.

The psycholinguistic literature has identified different linguistic components in language production (lexical access, morphosyntactic, phonological, and phonetic encoding, see Section 1.4), but to the best of our knowledge, the effect on reaction time (RT) latencies that their presence or absence provokes in a given production task has not been studied.[2] Among the factors known to bear on reaction time latencies, prominent items are lexical frequency, predictability, and lexical class (content vs. function words) (see Bell, Brenierc, Gregoryd, Girande, & Jurafsky, 2009: 92f). Starting with Oldfield and Wingfield (1965), a large body of converging evidence both in speech perception and language production supports this insight (Indefrey, 2011, Indefrey & Levelt 2004, Jescheniak & Levelt 1994).

Experimental evidence that linguistic complexity plays a role in language production was adduced by Sahin et al. (2009). Based on electroencephalogram (EEG) data and consistent with the psycholinguistic evidence mentioned, the authors found that language production involves three distinct and sequentially successive processing steps: 1° lexical access (retrieval of the pieces involved from long-term memory, at about 200 ms from stimulus onset), 2° concatenation of these pieces (and morphosyntactic tagging thereof, at about 320 ms), and 3° phonological adjustment (at about 450 ms). While 1° of course is always performed, 2° and 3° may or may not be carried out depending on the task assigned: word repetition is done in the absence of linguistic computation (only 1° involved), while the additional task of concatenation and morphosyntactic tagging adds processing demand (1° and 2° involved), and the presence of a phonological task on top of that makes processing still more linguistically complex (1°, 2,° and 3° involved).

In sum, Sahin et al.'s (2009) results suggest that the different components which contribute to language production may or may not be active in a given production task. That is, lexical access, morphosyntactic, and phonological processing each add extra processing burden; the latter two may or may not be active in a given production. Linguistic complexity then is the sum of the processing burden generated by the three individual tasks: when only lexical access occurs in a production (1+0+0), linguistic complexity is lower than in case both lexical access and morphosyntactic (but no phonological) activity is present (1+1+0). Finally, when phonological processing is carried out on top of the preceding (1+1+1), linguistic complexity peaks.

Based on this linguistic complexity hierarchy, our hypothesis is that language production in a specific linguistic configuration may be burdened with more or less of these processing demands, and that this not only leaves traces in relevant event-related potentials (ERPs) at a particular point in the time course, but also impacts the time that is needed to carry out the overall task. That is, language production will be faster when only 1° is performed; it will take a little longer when 2° is added but not 3°, and it will take even more time when all three tasks are necessary.

Note that our hypothesis may turn out to be unfounded: there is no a priori reason that experimental conditions which produce distinct ERPs in the EEG modality also translate into variable reaction times. It could be the case that the extra processing demands generated by each condition "heat up" the neural and cognitive system by requiring more energy to be put to use, but leave the time course of events unaffected. In this case, the amplitude of relevant ERPs would show the extra neural activity, but no increased RT would occur.

Testing the hypothesis that the ERPs at hand leave an RT footprint, we thus set out to provide support for Sahin et al.'s EEG-based results coming from the behavioral modality. If successful, our study based on simple RT analysis (instead of the more demanding EEG) should adduce evidence for a factor bearing on RT latencies in language production which to the best of our knowledge is undocumented thus far.[3]

1.4. Sequence of events upon language production

There is a broad agreement in the experimental literature and the theoretically oriented generative (chomskyan) approach to language regarding the workings of language production (although both approaches are based on quite different assumptions and are often critical of one another). The former holds that after conceptualization (i.e., the definition of what the speaker wishes to express), a lexical access retrieves so-called lemmas containing morphosyntactic information (Sahin et al.'s 1° above), which is then used for grammatical encoding (Sahin et al.'s 2°, called concatenation above), which in turn precedes phonological encoding (Sahin et al.'s 3°, phonological processing), and finally phonetic encoding (Bock, 1982; Cohen-Goldberg, 2013; Indefrey, 2011; Levelt et al., 1999). The same sequence of operations—first lexical access, then concatenation, and then phonology—is the scenario assumed in generative grammar, known as the inverted T model (Chomsky, 1965, pp. 15ff, see Boeckx & Uriagereka 2007 for an overview).

Sahin et al. (2009) have suggested that this broad scenario is correct. Their results show that there are at least three linguistically distinct processes that can be separated in time and space in the brain: first lexical access occurs, then concatenation is performed, and then phonology is operated. The study was performed using intracranial electrophysiology, that is, by recording local field potentials from neuron populations using electrodes implanted (for clinical evaluation) in language-related brain regions of three patients with epilepsy; the patients silently repeated words or pronounced inflected versions of noninflected stimuli.[4]

One of the linguistic phenomena used as stimuli by Sahin et al. was English plural formation where the plural morpheme /-z/ (spelt *-s*) is attached to roots and devoices when the root-final consonant is voiceless as in *rock-s* [ɹɔk-s]. That the plural marker is voiced in its stored form is shown by the fact that it is voiced after vowels, that is, in the absence of a conditioning consonant (e.g., *tree-s* [tɹii-z]). Sahin et al.'s experimental setup had three conditions: 1° Read, 2° Null Inflect, and 3° Overt Inflect. The 1° Read condition requires participants to repeat a word upon visual stimulus presentation (instruction "repeat word," stimulus appears on the screen in spelling: *rock*); this production involves lexical access, but no concatenation (i.e., morphosyntax) or phonological computation since the target word is not part of a grammatical structure (i.e., a sentence).

The Null Inflect condition inserts the same word into a grammatical structure: a carrier sentence visible before stimulus onset enforces the production of the singular form. For example, *that is the* __ is followed by the stimulus *rock*. This task requires concatenative computation even if the singular inflection is not visible on the result produced (2° Null Inflect): concatenation (or grammatical encoding) not only combines morphemes or words, but also syntactic features. In our case, it will tag the lexical item *rock* with grammatical attributes, such as number (singular), case (nominative), and so on. There is no such grammatical tagging in the 1° Read condition. Note that in the 2° Null Inflect condition, however, no phonological activity is carried out since upon the production of *rock*, there is no phonological modification of the form that was retrieved from long-term memory.

Finally, the 3° Overt Inflect condition requires visible (overt) inflection: the carrier sentence *those are the* __ induces the pronunciation of the plural form *rocks* when the stimulus *rock* appears on the screen. This requires the participant to concatenate the plural morpheme /-z/, which is then adapted to the last consonant of the stem: phonological computation turns the lexical /-z/ into the surface [-s] that appears in *rock-s* [ɹɔk-s] (progressive voice assimilation).

This experimental setting is summarized in Table 1.

Table 1
Experimental conditions in Sahin et al. (2009) and corresponding ERPs observed

Conditions	Lexical access	Concatenation	Phonology	Task
Read	Yes	—	—	repeat: __ [rock]
Null Inflect	Yes	Yes	—	that is the __ [rock]
Overt Inflect	Yes	Yes	Yes	those are the __ [rocks]
ERP	∼200 ms	∼320 ms	∼450 ms	

Sahin et al.'s (2009, p. 446) results revealed neural activity that is distinct in temporal order: lexical access produced a significant neural response (ERP) at ∼200 ms (all three conditions 1°–3°), concatenative activity provoked an ERP at ∼320 ms (2° Null Inflect and 3° Overt Inflect), and phonological computation induced an ERP at ∼450 ms (only 3° Overt Inflect).

1.5. Linguistic situation corresponding to Sahin et al.'s experimental conditions

As was mentioned, for the past 50 years or so (starting with Kiparsky, 1968–1973), a central question in grammatical theory that linguists argue about is whether the pieces (morphemes) that may be identified by the analyst are really the ones that are stored in long-term memory and computed online by the cognitive system of present-day natives. Despite constant attempts for decades, any arguments made have failed to decide the issue.

The word pair *electric* - *electricity* represents a prototypical example: the phonologically relevant fact is the alternation between [k] (in *electri*[k]) and [s] (in *electri*[s]*ity*). Linguistically speaking, analytical options are (at least) threefold: 1° suppletion, 2° allomorphy, and 3° morphophonology. 1° Suppletion requires no computation of any kind: *electricity* is stored as a whole (just as *electric*) and upon production is retrieved from long-term memory as one single chunk exactly like monomorphemic *cat* or *table* would be. 2° Allomorphy on the other hand does require computation, which, however, is not phonological in kind: the suffix selects one of two root allomorphs that are both stored (/electrik/ and /electris/). The pieces chosen are then concatenated (*electri*[s] + *-ity*). Finally, the 3° morphophonological option does involve phonological computation: only /*electrik*/ and /*-ity*/ are stored in long-term memory. After retrieval, they are first concatenated (morphosyntax) and then /k/ is turned into [s] before /i/ by a phonological process. That is, unlike in the two previous options discussed, *electric* and *electricity* share a common lexical form, /electrik/.

The pronunciation of *electricity* may thus correspond to the following situations when the linguistic and the production perspective are correlated. In option 1° suppletion, the only thing that happens is the retrieval of a lexical item from long-term memory without any additional linguistic calculation; this describes Sahin et al.'s 1° Read condition where only lexical access is performed without any further computation. In option 2° allomorphy, *electricity* is produced by performing a lexical access as before, but on top of that a morphosyntactic operation is carried out that decides which root allomorph is appropriate to precede *-ity*. This corresponds to Sahin et al.'s 2° Null Inflect condition where lexical access is associated with morphosyntactic computation. Finally, in option 3° morphophonology, the two preceding actions occur, that is, lexical access and concatenation of the root *electri[k]*- and *-ity* (the

Table 2
Experimental conditions in Sahin et al. (2009) and linguistic correlates (analytical option) thereof

	Conditions and analytical options	Lexical access	Concatenation	Phonology	Task
a.	Read ling. suppletion	Yes	—	—	repeat: __ [rock]
b.	Null Inflect ling. allomorphy	Yes	Yes	—	that is the __ [rock]
c.	Overt Inflect ling. morphophonology	Yes	Yes	Yes	those are the __ [rocks]

latter representing morphosyntactic computation), but in addition, phonological processing is carried out since the root-final k needs to be turned into s. This corresponds to Sahin et al.'s 3° Overt Inflect condition, where exactly these successive processing steps are taken: lexical access, concatenation, and phonology.

Our idea is that Sahin et al.'s protocol is precisely designed to isolate just these three steps: lexical access alone (1° Read, in linguistic terms suppletion), lexical access plus morphosyntactic, but without phonological computation (2° Null Inflect, in linguistic terms allomorphy), and all three actions lexical access, morphosyntactic, and phonological computation applied successively (3° Overt Inflect, linguistically speaking morphophonology).

This situation is summarized in Table 2, where Sahin et al.'s experimental conditions and the corresponding linguistic analyses are superposed.

1.6. Relating ERPs and reaction time

Key to the implementation of our study is the fact that Sahin et al.'s experimental conditions 1° Read, 2° Null Inflect, and 3° Overt Inflect match the three analytical options that linguists use (1° suppletion, 2° allomorphy, and 3° morphophonology). This correspondence is shown in Table 2. We hypothesize that, for instance, the absence of a delayed response for a given alternation in condition 3° Overt Inflect, as compared to its presence in other alternations in the same experimental context, means that the production of this alternation does not involve specific phonological computation.

Our study is also designed to show that Sahin et al.'s results can be corroborated in the behavioral modality (voice key-measured reaction time), noninvasively, and using stimulus sets from French (Section 2.1.1).

2. Materials and methods

2.1. Four alternations from French

2.1.1. Description

The experiments that we conducted are based on four alternations from French, each representing a particular phonological alternation (like k-s in *atomique* [k] - *atomi*[s]-*ité*).

Table 3
Stimulus sets: four different alternations

	Read			Null Inflect			Overt Inflect		
Alternation	carrier	stimulus	target	carrier	stimulus	target	carrier	stimulus	target
1. float. C adj.	–	grand	grand	ils sont__	grand	grands	elle est __	grand	grande
2. float. C conjug.	–	part	part	il __	partir	part	nous __	partir	partons
3. schwa	–	achète	achète	il __	acheter	achète	nous __	acheter	achetons
4. *–ique*	–	atomique	atomique	ils sont __	atomique	atomiques	on a parlé de __	atomique	atomicité

Following Sahin et al.'s (2009) experimental setup that involves the three conditions 1° Read, 2° Null Inflect, and 3° Overt Inflect (Section 1.4), Table 3 shows carrier sentences (that participants saw before stimulus onset), stimulus and target words (which participants were expected to produce given the carrier sentence) used for each alternation in our experiment (the complete set of stimuli and target words is provided in Supplementary Materials).

Alternations (1) float. C adj. and (2) float. C conjug. involve the presence or absence of a root-final consonant, either in masculine/feminine forms of adjectives (alternation 1: *grand* [gʁɑ̃] "big, masc." vs. *grande* [gʁɑ̃d] "big, fem.") or in verbal paradigms (alternation 2: *il part* [paʁ] "he leaves" vs. *nous part-ons* [paχt]ons "we leave"). Such consonants that alternate with zero are called floating consonants in the phonological literature.

Alternation (3) schwa is found in verbal paradigms, where [ɛ] occurs before word-final consonants as in *achète* [aʃɛt] "s/he buys" and when followed by a schwa [ə] (which may or may not be pronounced) as in *achètera* [aʃɛtəʁa] / [aʃɛtχa] "s/he will buy"), while schwa (which may or may not be pronounced) occurs when followed by a full vowel (i.e., different from schwa) as in *acheter* [aʃəte] / [aʃte] "to buy (inf.)." Note that the pronunciation of schwa is optional in French, and that its presence or absence is transparent to the alternation at hand (schwa/zero - ɛ).

Finally, alternation (4) *-ique* is the French equivalent of English velar softening, where [k] and [s] alternate: the former occurs in adjectives (*historique* [istoʁik] "historic"), while the latter is found in nominalized forms thereof that are produced by adding the suffix *-ité* (*historicité* [istoʁis-ite] "historicity").

As shown in Table 3, in each case, the carrier sentence enforced a specific production in the 2° Null Inflect and 3° Overt Inflect conditions. The 2° Null Inflect carrier leads to a form that is grammatically processed and thus bears grammatical structure (such as plural in alternation (1) float. C adj. *ils sont grands* "they (masc.) are big"), which, however, leaves no phonetic trace in the output (the *-s* marking plural in spelling is not pronounced: *grands* = [gʁɑ̃]). In the 3° Overt Inflect condition, things are as before except that the grammatical marker now is overt, that is, pronounced: *grande* = [gʁɑ̃d] with a final [-d] marking the feminine.

2.1.2. Stimulus = target, stimulus ≠ target

In Sahin et al.'s experimental setup, for both alternations examined the stimulus was the same throughout the three experimental conditions: for example, *rock* for the singular-plural paradigm (targets: 1° Read *rock*, 2° Null Inflect *rock*, 3° Overt Inflect *rocks*). At the same time, stimulus and target words were phonetically identical in both 1° Read and 2° Null Inflect: *rock*. Sahin et al.'s setup is thus characterized by this double identity "stimuli identical through experimental conditions and stimuli = target in 1° Read and 2° Null Inflect."

The four French alternations discussed do not all allow for this double identity. While alternations (1) float C adj. and (4) *-ique* do (Table 3), alternations (2) float C conjug. and (3) schwa do not. For the two latter, obeying double identity would mean to choose the target word of the 1° Read condition for being the stimulus: in the examples under (3) *part* and *achète*. These forms, however, are not the citation forms that speakers are likely to expect when identifying a word: in French, the citation form for verbs is the infinitive. Hence, if *part* and *achète* were the stimuli for 2° Null Inflect and 3° Overt Inflect, speakers would be exposed to these, rather than to the citation form, when prompted to fill in the blank of the carrier sentence. Note that this issue does not arise in English where the citation form is phonetically identical with the target form of the 1° Read condition (*rock*).

For this reason and because we wanted to contrast a number of different alternations (testing alternation specificity is a main goal of our study), we chose to depart from double identity for the two alternations mentioned. This move produced a situation where for these alternations the stimulus and target words are phonetically distinct in 2° Null Inflect (float C conjug.: *partir* - *part*, schwa: *acheter* - *achète*). In comparison to the two double-identity alternations, this opens the door for a potential extra processing burden when participants go from the stimulus to the target. We do not control for this factor, but by hypothesis and given Sahin et al.'s results where lexical access occurs uniformly at about 200 ms no matter what the stimulus type (presence or absence of a carrier sentence, see Section 4.2), the path from the stimulus to the identification of the lexical item and its retrieval from long-term memory does not appear to depend on the target word. Rather, the target word is produced on the grounds of what was loaded from long term into working memory, and possible further extra processing (morphosyntactic and phonological).[5]

Finally, another consequence of the situation described is that Sahin et al.'s experimental setup is only fully observed for two out of four alternations tested. We are, therefore, not claiming that our experiment is a strict behavioral replication of Sahin et al.: rather, it is crucially inspired by their experimental setup and, if conclusive, may offer a behavioral corroboration of their findings.

2.1.3. Linguistic analysis

The four alternations have aroused more or less debate in the phonological literature: are the pronunciations that involve an overt alternation the result of simple storage of the item as a whole (1° suppletion), of 2° allomorphy or 3° morphophonological computation? Alternation

(3) schwa has received attention in Schane (1968a, p. 30ff), Dell (1973a, pp. 198ff), Charette (1991), Tranel (1987, 1988, p. 172ff), and Morin (1988).

The floating consonants in adjectives (alternation 1) and conjugation (alternation 2) concern alternations of a consonant with zero, which in the phonological literature gave rise to the controversy regarding so-called protective schwas. The presence of these items at the end of words was held to be the reason why a common underlying form that ends in a consonant is pronounced with this consonant on the surface. On this take, the common lexical entry for both *chaud* [ʃo] "warm, masc." and *chaude* [ʃod] "warm, fem." is /ʃod/. In the masculine form, the stored item is subject to a deletion rule (phonological computation) that eliminates the final -d: the result is [ʃo]. In the feminine form, though, a feminine marker /-ə/ (protective schwa) is added and the string /ʃod-ə/ is submitted to phonological computation. The consonant deletion rule does not apply because it only targets word-final consonants, and the result is [ʃod]. Note that the feminine schwa itself is not pronounced either (in reference varieties of the language at least): it is itself subject to a deletion rule at the end of the derivation.

This analysis (Dell, 1973b; Schane, 1968b, 1972) was said to be abstract. It was called into question by so-called concrete analyses (Tranel, 1981, pp. 248ff, 251ff), which argue for an allomorphic perspective without common lexical entry for consonant-zero alternations in conjugation (*il part* - *nous par[t]ons*). That is, the root *part-* has two stored items (allomorphs), /par-/ and /part-/, which are selected according to the suffix.

Finally, alternation (4) *-ique* does not seem to have produced relevant studies thus far: it is the French version of the aforementioned English velar softening (and actually its diachronic precursor since it was imported into English by the Norman Conquest), which has triggered a massive body of controversial literature (Section 1.2.2).

2.2. Implementation of the experimental setup

The linguistic properties of the four alternations discussed in Section 2.1.1 conditioned the setup of the stimulus list for each data set. Stimulus word selection is discussed at greater length in Section 2.5.

General experimental constraints were another relevant factor that imposed restrictions on the number of lexical items submitted to participants: given their limited sustained attention over long periods, we did not wish to exceed 40 min of participant involvement per session.

Note that two of the three major factors that have been documented to weigh on RTs in the literature (Bell et al., 2009, Section 1.3), predictability from context and lexical class (content vs. function words), are excluded by definition given our experimental setup. All stimulus and target words are content words, thus excluding a bias due to the content–function word difference. Regarding predictability, the carrier sentences certainly allow the participant to predict the major category and sometimes gender of the target word (this is the raison d'être of the carrier: *elle est* __ "she is __" prompts a feminine adjective), but within the lexical array thus defined nothing allows the participant to predict any further property of the stimulus or the target word.

The third factor, lexical frequency (reported in Supplementary Materials), was not controlled for in stimulus selection since this would have conflicted with stimulus selection criteria demanded by linguistic properties (see Section 2.5). Instead, we evaluate the possible impact of cumulative frequency on our data (covariate in a mixed model in Section 3.1.2, discussed in Section 4.3).

2.3. Participants

Forty healthy students from the Université Côte d'Azur aged 25 on average ($sd = 7$) participated voluntarily, after signing an informed consent form. They were all French natives and had not spent more than 12 months in a row abroad. There was some geographical variation regarding the place of birth of participants and where they were raised, but ensuing dialect variation did not impact the four alternations studied which are known to be robust across dialects.

2.4. Procedure

The display time for instructions was set to 650 ms. In the 1° Read condition, participants were visually presented with the instruction *répétez* __ ("repeat __"), followed by a fixation cross (display time 1100 ms) and the stimulus word (display time 250 ms). Then, a second fixation cross was displayed for 1500 ms during which the participant was instructed to pronounce the target word aloud. The same protocol was displayed in the 2° Null Inflect and 3° Overt Inflect conditions, except that the relevant carrier sentence appeared instead of the instruction.[6]

Hence, in the 1° Read condition, *répétez* ___was followed, for instance, by the stimulus *électrique*, which required the response *électrique* [elɛktχik] (absence of morphosyntactic or phonological processing).

In the 2° Null Inflect condition, there were two different carrier sentences depending on the stimulus set, as shown in Table 3: *ils sont* __ ("they are __") and *il* __ ("he __"). That is, *ils sont* __ followed by *électrique* required the response *électriques* [elɛktχik], which involved morphosyntactic processing but no phonology: the target word of the 1° Read and 2° Null Inflect conditions is phonetically identical. As was discussed in Section 2.1.2, in some cases, the correct response in the 2° Null Inflect condition was phonologically distinct from the stimulus word. For instance, regarding alternation (2) float C conj., the carrier sentence *il* __ followed by the stimulus word *partir* required the answer *part* [paʁ]. Note that this is the bare root without any overt suffix (*-ir* in *part-ir* is the infinitive marker) where the final *-t* is not pronounced because it is floating: it is only pronounced, for example, in 1pl *nous part-ons* [paχt-ɔ̃] (which is elicited in the 3° Overt Inflect condition). That is, in the 2° Null Inflect condition, the target word *part* [paʁ] (different pronunciation with respect to the stimulus word) and the target word *électriques* [elɛktχik] (identical pronunciation with respect to the stimulus word) are identical in terms of the grammatical operations required: in both cases, a lexical item is retrieved from long-term memory and morphosyntactically processed, but remains phonologically unmodified.

Finally, there were three different carrier sentences in the 3° Overt Inflect condition in order to enforce the relevant category of the lexical item, as shown in Table 3: *on a parlé de* __ ("people talked about __") when the target word was a noun (stimulus *électrique* requires the response *électricité*), *elle est* __ ("she is __") when the target word was the feminine form of adjectives (stimulus *grand* elicited *grande*) and *nous* __ ("we__") in case a vowel-initial verbal inflection marker was needed (stimuli *partir*, *acheter* produced the responses *partons*, *achetons*, respectively).

The experiment was implemented in E-Prime. A list of 154 words (see Section 2.5 and Supplementary Materials) was subjected to the three experimental conditions (Read, Null Inflect, and Overt Inflect), for a total of $3 \times 154 = 462$ trials. The 462 trials were randomized for each subject. Response times were collected using a vocal key CHRONOS synchronized with E-Prime, after the visual onset of the stimulus word.

The experiment was preceded by a warmup, including 60 words, which did not belong to the experimental lists, with the 30 first trials run at a slow pace (8 s per trial before the second fixation cross), and the 30 last trials run at the normal pace (2 s per trial before the second fixation cross). Precise instructions were given before the warmup. The instructions indicated that there could be different inflections of the words during the experiment when required to complete the sentences, and several simple examples were provided.

For each participant, the experimental session lasted 40 min (for a total of 60 min including two breaks, briefing and debriefing). These 40 min were cut into three equal runs separated by two breaks (the participant pressed the space bar to continue when ready after a break). To match parameters of another similar functional magnetic resonance imaging (fMRI) study conducted in our lab (a study in which the stimuli where jittered randomly), each inter-trial interval was set to 1296 ms.

2.5. Stimuli

The 154 words in the Overt Inflect condition fall into 68 float C adj. items, 68 float C conj. items, 5 schwa items, and 5 *-ique* items. The list of these 146 items (including cumulative frequency as well as token frequency of the stimulus and the target) appears in Supplementary Materials. The eight remaining trials (~5% of the stimuli) were fillers used to try out new words for future experiments, which were not under scrutiny in the present study.

Stimuli were selected only according to linguistic criteria, which are explained below. These led to variations of lexical frequency. The averages of cumulative frequency (used in the study, see Section 3.1.1) for stimuli words are as follows: (1) float C adj. = 140.3, (2) float C conj. = 452.9, (3) schwa = 215.7, and (4) *–ique* = 28.7 (number of occurrences per million words).

The numeric disproportion between (1) float C_adj. / (2) float C conjug. on the one hand (68 items each) and (3) schwa / (4) *-ique* on the other hand (5 items each) is due to both linguistic reasons and cumulative frequency. For one thing, it grossly reflects the disproportion that is found in the French lexicon: the database (extracted from an electronic dictionary) from which stimuli were chosen contains 29 words (or actually word pairs, e.g., *historique* - *historicité*) for alternation (4) *-ique*, 94 items for alternation (3) schwa, 128 items for alternation (3) float. C conjug., and 374 items for alternation (1) float C adj.

The high number of stimuli for the two alternations that involve floating consonants also stems from the wish to represent the numerous different floating consonants that occur. In *chaud* [ʃo] - *chaude* [ʃod] "hot, masc., fem.," d is concerned, but ʃ is floating in *frais* [fχɛ] - *fraîche* [fχɛʃ] "fresh, masc., fem.," g in *long* [lɔ̃] - *longue* [lɔ̃g] "long, masc., fem.," kt in *distinct* [distɛ̃] - *distincte* [distɛ̃kt] "distinct, masc., fem.," t in *lent* [lɑ̃] - *lente* [lɑ̃t] "slow, masc., fem.," s in *bas* [ba] - *basse* [bas] "low, masc., fem.," and so on.

The low number of stimuli for alternation (3) schwa is also due to a further linguistic caveat. For some words, there is an alternative way of inflecting the verbs at hand: *app[ə]ler* "to call, inf." can only be *app[ɛ]lle* in 3sg, but aside from 3sg *béqu[ɛ]te*, the verb *béqu[ə]ter* "to pick (with a beak), to eat" also (in fact typically) produces 3sg *becte* [bɛkt]. Still for other verbs, the [ɛ] form will be highly unlikely for speakers to be produced or accepted (*déchiqueter* "to tear apart," 3sg *?déchiquète*, regular 3sg *déchicte*), and there will also be cases where speakers do not know what to do. For *briqueter* "to build of bricks," for example, both candidates 3sg *briquète* and 3sg *bricte* appear awkward, and these verbs are then simply defective (i.e., speakers will avoid producing the forms in question). Out of the 94 schwa-e items on our record, a good deal is unusable for this reason: forms without [ɛ] (*il becte* [bɛkt]) would be of no use for the experiment (and would presumably involve extra processing time caused by the competition between the [ɛ] form and the vowel-less alternative). In order to avoid their production, only unambiguous items were included.

2.6. Data selection

The 17,520 productions (40 participants $\times$ 3 conditions $\times$ 146 words) were checked exhaustively, and any production that was not expected given the stimulus (e.g., blank, other word produced) was manually excluded. We only kept unexpected cases as valid cases when words were produced with an elided indefinite article (for instance, "d'électricité" instead of "électricité"). Three participants were excluded from the results (one participant who did not understand the instructions, two participants who scored below 80% correct; the remaining subjects scored at least 97.7% correct).

All RTs were reanalyzed using CHRONSET (Roux, Armstrong, & Carreiras, 2017), an automated method for detecting speech onset. The distribution of the resulting RTs (the onset of vocalization after onset of the stimulus word, detected by CHRONSET) presented a positive skewness. The non-normality was mainly due to the right tail of the distribution. To reduce skewness, we excluded RTs smaller than 30 ms and higher than 1500 ms as they were considered outliers (overall, less than 1% of the data). This left us with 14,440 trials to be analyzed. The characteristics of the overall corrected distribution of RTs were the following: $M = 382$ ms, $sd = 215$, and skewness = 1.42.

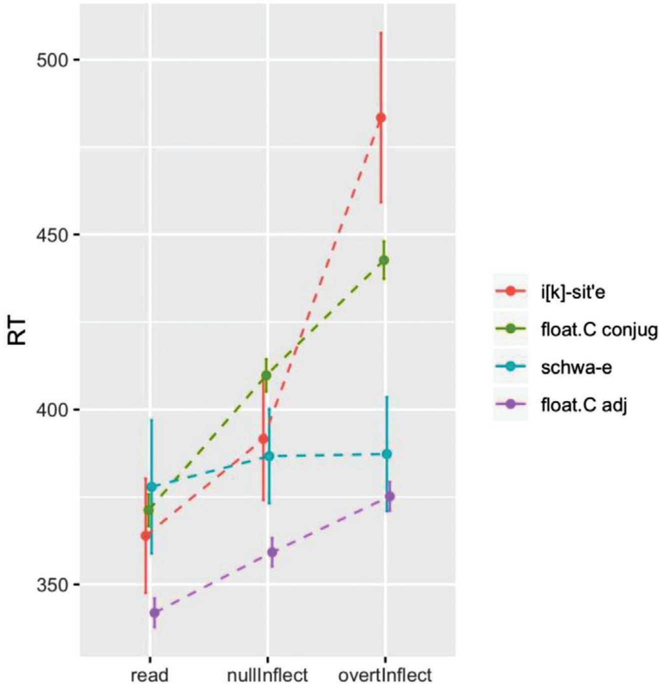
3. Results

3.1. 1° Read < 2° Null Inflect < 3° Overt Inflect

3.1.1. Effect of condition

The mean RTs for the four alternations under scrutiny across the three conditions appear in the figure in Table 4; relevant descriptive statistics are provided in Table 5.

Table 4
Mean RTs as a function of alternation (float C adj., float C_conjug., schwa, -*ique*) and condition (1° Read, 2° Null Inflect, and 3° Overt Inflect)



Note: Error bars are +/– one standard error.

Table 5
Mean RTs for the 1° Read, 2° Null Inflect, and 3° Overt Inflect conditions across alternations, followed by standard deviations (SD), standard errors (SE), and +/– 95% confidence intervals (CI) around the mean value

Alternation	Condition	*n*	Mean	SD	SE	CI
float C adj.	Read	2337	342	200	4	8
float C adj.	Null inflect	2347	359	197	4	8
float C adj.	Overt inflect	2252	375	196	4	8
float C conj.	Read	2365	371	222	5	9
float C conj.	Null inflect	2173	410	215	5	9
float C conj.	Overt inflect	1999	443	240	5	11
schwa	Read	169	378	248	19	38
schwa	Null inflect	171	387	176	13	27
schwa	Overt inflect	160	387	206	16	32
-*ique*	Read	175	364	216	16	32
-*ique*	Null inflect	171	392	228	17	34
-*ique*	Overt inflect	121	483	266	24	48

Note: *n* is the number of trials across participants (Section 2.5 explains why we chose to unbalance the number of cases per alternation).

Table 6
Likelihood ratio tests comparing successive increasingly complex models

Models	Df	AIC	LogLik	Chi-sq	Df-diff	*p*
(1 | Subject)	3	20,103	–10,049			
(1 | Stimulus)	4	19,364	–9678	741	1	***
Condition	6	19,188	–9588	180	2	***
Alternation	9	19,152	–9567	42	3	***
Interaction	15	19,136	–9553	28	6	***
Cumulative Frequency	16	19,131	–9550	7.3	1	**
(1 + Condition | Subject)	21	19,026	–9492	5	5	***
(1 + Condition + Alternation | Subject)	36	19,013	–9470.5	43	15	***
(1 + Condition | Stimulus)	41	18,984	–9450.9	38	5	***

Note: The first column indicates the variables successively entered in the model. Df-diff is the number of degrees of freedom (Df) differing between two successive models. AIC is the Akaike information criterion (a lower value is best). Significance levels are * = $p < .05$; ** = $p < .01$; *** = $p < .001$.

Our main analysis was based on a linear mixed effects analysis of the log of the RTs as the dependent variable, including participants and stimulus words as random factors. We introduced in the model Condition (1° Read, 2° Null Inflect, and 3° Overt Inflect) and Alternation (float C adj., float C conjug., schwa, *-ique*) as fixed factors. The model was thought to test the interaction between these two main factors. The model allowed each participant and each stimulus word to have an independent intercept. We also added random slopes for Condition and Alternation by participant and stimulus in the most specific model, with the idea that participants could show individual differences when producing words in the different conditions and that some stimuli could trigger larger effects.

The cumulative frequency of the words was introduced in the model as a covariate. It was calculated per million words from the *Lexique* French database (New, Pallier, Ferrand, & Matos, 2001). The measure used in the analyses was the natural log of the cumulative frequency of each word, corresponding to the sum of the frequencies of all derived and inflected words of the morphological family of each experimental word.

The mixed-effects model was run in R (R Development Core Team, 2007) using the lme4 package version 1.1-8 (Bates, Mächler, Bolker, & Walker, 2015). The model was fit using restricted maximum likelihood estimation. Table 6 shows the comparison between the successive models, with the chi-square value equaling twice the difference between the log-likelihood of the two compared models, and *p* values indicating whether there was a significant gain in using a more complex model. The last model with 41 degrees of freedom was our final model, including Condition and Alternation as fixed factors (including the interaction term), cumulative frequency as a covariate, and participants (with random slopes for Condition and Construction by participant) and by-stimulus-words random slopes for Condition, as the two random factors with adjustments to the intercepts.

Table 6 shows the summary of the model. To give a simpler view of the fixed effects, we ran a Type III analysis of variance using Satterthwaite's method based on the anova procedure for the model. The best model (Df = 41) revealed a significant effect of Condition ($F(2{,}148) =$

Table 7
Estimates of the fixed effects for the model Df = 41 for the prediction of the log(RT)

	Estimate	Std. Error	df	t value	$p(>\|t\|)$
(Intercept)	5.67E+00	5.46E-02	37.2		
Null Inflect	6.86E-02	2.64E-02	94.3	2.60	.011*
Overt Inflect	1.31E-01	2.79E-02	98.1	4.69	***
schwa	5.37E-02	4.06E-02	202	1.32	.19
float C conj.	8.46E-02	1.75E-02	98	4.84	***
–ique	5.84E-02	3.77E-02	301	1.55	.12
Fq	-2.34E-05	7.78E-06	1257	–3.01	.003**
Null Inflect: schwa	4.72E-02	7.46E-02	222.4	0.63	.53
Overt Inflect: schwa	-2.93E-02	8.00E-02	180.0	–0.37	.71
Null Inflect: float C conj.	7.42E-02	2.76E-02	236.4	2.68	.008**
Overt Inflect: float C conj.	1.05E-01	2.97E-02	196.2	3.53	***
Null Inflect: *–ique*	3.32E-02	7.43E-02	225.2	0.45	.66
Overt Inflect: *–ique*	1.83E-01	8.34E-02	210.7	2.20	.03*

Note: Significance levels are * = p < .05; ** = p < .01; *** = p < .001.

16.6, $p < .001$), a significant effect of Alternation ($F(3,106) = 20.3$, $p < .001$), a significant interaction ($F(6,181) = 3.2$, $p = .005$), and a significant effect of the covariate cumulative Frequency ($F(1,1257) = 9.1$, $p < .003$), with an AIC of 18,984. The complete list of the estimates is given in Table 7.

Note: significance levels are * = p < .05; ** = p < .01; *** = p < .001.

Also note that the length of the read instruction and the carrier phrases varied. The instruction in the 1° Read condition was a single word (*répétez*), while the carrier sentences for the other conditions were either longer or shorter and differed also across alternations. For instance, the carrier for the (4) *–ique* construction in the 3° Overt Inflect condition was much longer (*on a parlé de* __) than all the others, and interestingly, this construction in that condition was the one with longest RTs. It is thus possible that the longer carrier presented at the same latency of 650 ms had some carryover effects, that is, created additional processing demands. To rule out any possible effect of instruction length, we added this covariate as a fixed effect in our mixed model, but this new factor was not found significant when implemented in the best model we previously obtained, and it worsened the AIC value.

The mean RTs for the conditions 1° Read, 2° Null Inflect, and 3° Overt Inflect (all four alternations collapsed) were significantly distinct in the way predicted by Sahin et al. (2009): 1° Read < 2° Null Inflect < 3° Overt Inflect. The respective RTs were 358, 384, and 408 ms with standard errors equal to 3.0, 3.0, and 3.3 (*SDs* were, respectively, 213, 207, and 222).

We consider this to be our main result: our experiment matches the findings of Sahin et al. (2009) in a noninvasive behavioral modality and based on French stimuli falling into four different alternations.

3.1.2. *Effect of cumulative frequency*

The stimuli of our experiment were not controlled for lexical (token or cumulative) frequencies: their selection was based only on linguistic criteria (see Section 2.5).

Looking at the cumulative frequency of the target words for each condition (Read, Null Inflect, and Overt Inflect), it appears that in the Read condition, no significant effect of cumulative frequency is observed ($F(1,22.8) = 2.27$, n.s.) and no specific contrast either (we ran a Games–Howell Post-Hoc Test, similar to Tukey's test in its formulation but without assuming equal variances and sample sizes).

In the Null Inflect condition, a significant effect of cumulative frequency is observed ($F(3,18) = 6.77$, $p < .005$), and running Games–Howell Post-Hoc Test on specific comparisons showed only one significant contrast between *-ique* and float C. adj alternations ($p = .002$).

Finally, in the Overt Inflect condition, a significant effect of cumulative frequency was observed ($F(3,12.7) = 5.52$, $p = .01$), and post-hoc specific comparisons showed that the float C. adj alternation was significantly different from each of the three other alternations (all ps $< .001$).

These results show that cumulative frequency plays a role in RT latencies, which, however, cannot be reduced to this factor since the influence of Condition documented in Section 3.1.1 is also relevant. The coexistence of both factors is discussed in Section 4.3.

3.2. *Alternation-specific signatures*

A further result concerns the idiosyncratic behavior of individual alternations: the significant effect of Alternation shows that the time needed to carry out the task varied significantly among the four different alternations. This indicates that different grammatical alternations may generate quite distinct processing demands for the cognitive system.

The interaction reported in Table 6 calls for studying each phenomenon separately. We, therefore, ran separate mixed models for each alternation using the same factors as in our best model (Df = 41), except for the Alternation factor itself. We do not report the detailed statistics but in short, the models revealed that the factor Condition was significant within all alternations, except schwa. That is, the schwa alternation does not show any significant effect across conditions. This is confirmed below by the pairwise comparisons we conducted. Also, interestingly, the factor cumulative frequency was only found significant ($p < .05$) within float C conj.

In order to characterize the individual signature of alternations more precisely, we conducted pairwise comparisons among the three conditions within each alternation: (1° Read vs. 2° Null Inflect), (1° Read vs. 3° Overt Inflect), and (2° Null Inflect vs. 3° Overt Inflect). By applying the Benjamini–Hochberg procedure to control for false discover rate, we found eight significant pairwise comparisons among the 12 pairwise comparisons (three conditions times four constructions). This is shown in Table 8.

The summary of these results in Table 9 affords a better representation of the individual signature produced by each alternation.

Table 8
Pairwise comparisons within the Alternation factor obtained from the model ($Df = 41$) obtained using lmer(logRT ~ Condition * Alternation + Fq.cumul + (1 + Condition + Alternation | Subject) + (1 + Condition | stimulus), data)

Alternation	Conditions compared	Estimate	Std. Error	df	*t* value	*p*	
(1) float. C adj.	1°–2°	–0.066	0.025	65.3	–2.70	8.81E-03	*
	1°–3°	–0.129	0.026	70.2	–4.88	6.31E-06	*
	2°–3°	–0.062	0.019	115.2	–3.36	1.04E-03	*
(2) float. C conjug.	1°–2°	–0.143	0.029	129.7	–4.88	3.12E-06	*
	1°–3°	–0.236	0.031	134.5	–7.67	3.12E-12	*
	2°–3°	–0.093	0.019	128.8	–4.87	3.18E-06	*
(3) schwa	1°–2°	–0.116	0.087	261.6	–1.33	1.85E-01	
	1°–3°	–0.102	0.092	248.2	–1.11	2.70E-01	
	2°–3°	0.0146	0.067	133.7	0.22	8.27E-01	
(4) *-ique*	1°–2°	–0.101	0.064	91.5	–1.57	1.19E-01	
	1°–3°	–0.311	0.075	98.3	–4.13	7.57E-05	*
	2°–3°	–0.210	0.071	163.2	–2.97	3.46E-03	*

Note: The asterisk in the last column indicates significance of the *p* value after applying the Benjamini–Hochberg procedure.

Table 9
Individual signature of the four alternations

Alternation	1° Read versus 2° Null	2° Null versus 3° Overt	1° Read versus 3° Overt
float C adj.	effect	effect	effect
float C conj.	effect	effect	effect
-ique	–	effect	effect
schwa	–	–	–

As may be seen, three distinct patterns occur: two alternations, float C adj. and float C conj., show significant effects for all three comparisons; the difference between 1° Read and 2° Null Inflect is not significant for *-ique*, but the two other comparisons do produce a significant effect for this alternation; finally, none of the three comparisons reaches significance for schwa.

4. Discussion

4.1. Summary and limits

Our study suggests the existence of a factor weighing on RT latencies that does not seem to have received much attention thus far: linguistic complexity, which translates as increasing processing demands when a word is merely retrieved from long-term memory (1° Read), when in addition it undergoes morphosyntactic computation (2° Null Inflect), and in case on top of the two preceding operations, it requires phonological computation (3° Overt Inflect). The idea that this cumulative calculus translates into increasing burden in online processing is supported by the fact that Sahin et al.'s (2009) EEG-based data show an actual increased amplitude of neural activity for each condition in case it is involved in the production (relevant ERPs).

We are aware that these first results are only indicative and preliminary in kind. In order to strengthen the evidence that there are significant differences among alternations, pairwise comparisons among alternations (1° Read of alternation A compared to 1° Read of alternation B, etc.) would need to be carried out, but a more refined material (thus probably a smaller sample of words to balance alternation types) should be in use for controlling the lexical frequency of the words. Also, in order to strengthen the evidence that there are significant differences within alternations in a given condition, future studies could take advantage of the variances observed within alternations, in a particular condition. For instance, a noteworthy observation is that the standard error within the *-ique* alternation is about five times as large as in the two float C alternations. This could indicate that the cognitive demand is relatively homogeneous among words in the float C alternations, thus meaning that the different words we selected all trigger a similar phonological computation. However, the alternation *–ique* that produces a higher variance of RT might not systematically involve phonological computation. A more refined analysis could take advantage of these heterogeneous variances observed across alternations, for instance, using a Bayesian analysis word-by-word. Such an analysis could help diagnose which cognitive operations are the best hypothesis for the production of specific words given the likelihood of response times for each condition. A Bayesian analysis would be useful to compute the plausibility that a given word is produced based on a true phonological computation, but lexically specific distinctions lie beyond the scope of the present study, which focuses on alternation specificity. Moreover, we believe this tentative analysis could be fruitful in future studies based on material offering a perfect control for lexical frequency.

A reviewer pointed out a further limitation of the experimental protocol: even if participants have seen and processed the carrier sentence before the stimulus appears on the screen, it could be the case that when the stimulus is available, that is, during RT recording, they silently read the carrier anew in order to fill in the hole. The carrier is variable in length, though, ranging from one (*il* __) over two (*ils sont* __) to five syllables (*on a parlé de* __). The processing time needed for those variable carrier lengths may thus be a factor in the overall RT recorded and may have biased our results. This is correct and future studies will want to control for the length of the carrier sentence.

However, globally, our results appear to offer a corroboration of Sahin et al. in the behavioral modality. This is not trivial in itself, but also recall from Section 1.3 that the RT effect observed, which follows Sahin et al.'s predictions, was not necessarily to be expected: in the EEG modality, the effect of each of the three conditions translates as a distinct peak in neural activity (ERP). It could have been the case that the extra processing demands generated by each condition "heat up" the neural and cognitive system by requiring more energy to be put to use, but leave the time course of events unaffected. In this case, the amplitude of relevant ERPs would show the extra neural activity, but no increased RT would occur. This is not the case, though: in the present task, extra neural activity seems to go hand in hand with extra processing time.

Our results also prompt alternation specificity as a factor in RT latencies. The linguistic literature discussed in Section 1.2.2 unanimously works on the assumption that the presence or absence of phonological processing is alternation specific: for about 50 years, the field has desperately (and unsuccessfully) been trying to identify a diagnostic that is able to tell whether or not phonological processing is contributing to the production of alternation X, Y, or Z. On the other hand, alternation specificity as a factor in RT latencies appears to have received no attention in the psycholinguistic literature on language production (see Section 4.5.2). Our study may thus be pioneering in carrying over what linguists believe is an established fact into the experimental realm.

4.2. Instructions: "repeat" versus carrier sentence

An issue when comparing 1° Read on the one hand and 2° Null Inflect / 3° Overt Inflect on the other was that the latter two involve a carrier sentence which is absent in the former. The absence of a carrier in 1° Read does not mean that there is no instruction, though: like the other conditions, 1° Read is introduced by an instruction, "repeat," that needs to be processed. All instructions produce processing demands and hence it is not the presence (2° Null Inflect and 3° Overt Inflect) versus absence (1° Read) of instruction-born processing that needs to be considered, but rather a putative difference between a "repeat" instruction and a carrier-based instruction.

The assumption here was that instructions produce comparable processing demands in perception. That is, preparing for the task at hand by processing the instruction is akin to a categorization task. Depending on the instruction, participants can rapidly decide which task is to come, mostly based on visual information.

Also, an influence of perception on RT latencies would only occur if the former were carried over into the latter, that is, if perception were still running after the onset of RT measurement. In our setup (see Section 2.4), this is when the stimulus appears on the screen, that is, after the instruction appeared (display time of instructions was 650 ms, followed by a fixation cross shown for 1100 ms). Previous studies on visual encoding have shown that visual processing in a categorization task can be achieved in 150 ms after stimulus onset at the most for complex images (e.g., Rayner, Smith, Malcolm, & Henderson, 2009, Thorpe, Fize, & Marlot, 1996). It is thus plausible that both perception and reading of the instruction are completed before stimulus onset, and hence that the RT latencies measured remain unaffected.

EEG-based evidence supports this conclusion: Sahin et al. (2009) have documented that there is a uniform ERP at about 200 ms shared by the three experimental conditions, which they interpreted as representing lexical access. If this interpretation is correct, lexical access, that is, the fact of retrieving the target word from the mental lexicon, occurs at the same point in time for all three conditions. This means that the presence versus absence of a carrier sentence has no influence on the latency of lexical access, and thus on further events that may follow (conditions 2° and 3°).

4.3. A hitherto undocumented factor weighing on processing time

Our study suggests the existence of a factor weighing on reaction time in language production that does not seem to have received attention thus far: linguistic complexity. The model we set up for statistical analysis (Section 3.1) showed both an effect of cumulative frequency (covariate) and linguistic complexity (Condition). Given that cumulative frequency is a robust factor in RT latencies, its bearing on our results comes as no surprise. However, detected effects cannot be reduced to this factor alone since we also show evidence for the influence of the factor Condition (Section 3.1.1). In this article, we do not ambition to tease apart the exact proportions of the two factors at play: our goal is to document the fact that linguistic complexity can weigh on RT latencies, in addition to basic frequency effects.

What we call linguistic complexity is based on the distinction of three specific tasks that are sequentially executed (1° lexical access > 2° concatenation > 3° phonology) according to the established view on language production in both psycholinguistic and generative quarters (see Section 1.4). Being able to look into different epochs of the overall processing time through EEG, Sahin et al. (2009) documented the fact that processing at stages 2° and 3° may show significant discrepancies in the demand on neural activity (the amplitude of Sahin et al.'s ERPs). These discrepancies correlate with the presence or absence of extra processing demands controlled for by the event-related experimental protocol: 1° Read (only lexical access, but no concatenation or phonology), 2° Null Inflect (lexical access and concatenation, but no phonology), and 3° Overt Inflect (all three tasks).

Controlling for extra processing demands, this protocol was applied to the behavioral modality, hypothesizing that extra tasks will translate into increased RT latencies. Our results show that this is indeed the case, precisely in the way expected given Sahin et al.: 1° Read took significantly less time than the more complex 2° Null Inflect, which in turn was carried out faster than the most complex 3° Overt Inflect (see Section 3.1.1). It appears to be especially noteworthy that significant differences in RT latencies between 1° Read and 2° Null Inflect occur, while target words pronounced by the participants are phonetically identical.

Unlike EEG-based measurements, of course the behavioral modality cannot look inside the reaction time window to determine what happens in the time elapsed between stimulus onset and pronunciation. However, the behavioral response can evidence *that* additional processing demands, which are absent in a control condition, have burdened processing and produced a longer reaction time.

4.4. Previous evidence for an impact of grammatical processing on RT (1° vs. 2°)

The general psycholinguistic literature on the processing of morphologically complex words (Baayen, Milin, Filipović Đurđević, Hendrix, & Marelli, 2011; Hay, 2001; Hay & Baayen 2005; Schreuder & Baayen 1995, see the survey specifically for language production in Bölte et al., 2004) is focusing on a number of issues, including listing versus decomposition, deterministic versus gradient processing, family size effects, relative frequency effects of roots and affixes (Hay, 2001), or productivity and regularity effects of affixes (on which more below). Studies that evaluate the impact of linguistic complexity on production through the comparison of items that do and do not involve specific linguistic tasks (morphosyntactic as in our 2° Null Inflect, phonological as in our 3° Overt Inflect) do not appear to be high ranking on the agenda, though.

The fact that extra grammatical processing (our 2° Null Inflect) increases RT latencies when compared to its absence (our 1° Read) was on occasion observed in earlier work, both regarding production and perception.

On the production side, Jescheniak and Levelt (1994, experiment 5) manipulated the presence or absence of grammatical encoding by having participants repeat either just the word displayed (our 1° Read) or the word together with its definite article (our 2° Null Inflect). They found that on a subsequent gender decision task (press a button for either fem. / masc. or neuter, the language tested being Dutch), the frequency effect disappeared in the latter case, that is, when the participants had previously accessed gender information of the noun (producing the definite article requires accessing this information). In the former case where participants did not previously access gender information, the frequency effect remained upon later gender decision, though.

Hence, mere word repetition does not entail accessing gender information, although this information is present in the lexical entry of the word that is accessed. This is consistent with Sahin et al. (2009) and our own findings, all indicating that a mere repetition task does not lead to grammatical processing: the only process that happens is lexical access. It is, therefore, unsurprising that extra morphosyntactic processing (2° Null Inflect), which is absent from mere repetition (1° Read), adds to the overall reaction time.[7]

On the perception side, based on a lexical decision task in Italian where participants had to say whether a given stimulus is or is not an existing word, Caramazza, Laudanna, and Romani (1988) showed that morphological decomposability impacts RT: stimuli that are decomposable (i.e., where substrings are existing morphemes) require significantly longer RT than stimuli that cannot be decomposed (i.e., where no substring is an existing morpheme). For example, the 2sg imperfective of the first conjugation verb *cantare* "to sing" is *cant-avi*. The authors created a pseudoword by adding the second conjugation 2sg imperfective suffix -evi (*cadere* "to fall" - *cad-evi*) to the first conjugation root *cant-*. The result, *cant-evi*, does not exist but is made of two morphemes that do exist in the language and that speakers may recognize: it is decomposable. By contrast in a pseudoword like *canzovi*, no substring is an existing morpheme: there is no root *canz-*, and there is no suffix *-ovi*.

Caramazza et al.'s (1988) results show that participants take significantly more time to decide that the decomposable type *cant-evi* does not exist (and also produce a higher error rate) than to arrive at the same conclusion for the nondecomposable type *canzovi*. They conclude that decomposition is performed in perception whenever possible, and that this extra processing (as compared to items where nothing can be decomposed) translates into longer RT.

Caramazza et al.'s (1988) study thus documents the impact of morphosyntactic activity (decomposition) on RT. In perception, extra morphosyntactic activity (our 2° Null Inflect) *de*composes a stimulus. In production, it *pieces together* different items coming from long-term memory. In perception, the presence of extra morphosyntactic activity is opposed to its absence when the stimulus is not decomposable, while in production, the contrast is with items that have not undergone any piecing-together because they are the only thing that was retrieved from long-term memory (our 1° Read).

The overall picture is thus consistent for all evidence mentioned coming from RT-based studies in perception (Caramazza et al., 1988) and production (our own results), from the frequency-based study in production by Jescheniak and Levelt (1994) and from EEG-based data (Sahin et al., 2009): morphosyntactic processing is absent when speakers merely repeat (production) or cannot decompose (perception) a stimulus, and in case it is performed produces increased RT latencies and neural activity (documented by relevant ERPs).

The evidence mentioned is about an extra processing burden created by morphosyntactic activity (1° vs. 2°). We are unaware of studies that investigate the specific burden added by phonological activity (3°).

4.5. *Alternation-specific signatures*

4.5.1. *Status and possible linguistic interpretation*

Our results show that all alternations do not behave in the same way (see Section 3.2): two of the four French patterns tested, (1) float C adj. and (2) float C conjug., produce significant effects for all three pairwise comparisons of Condition, that is, 1° versus 2°, 1° versus 3°, and 2° versus 3°. One alternation, (4) *-ique*, exhibits significant differences for the two latter, but not for the former. The fourth alternation, (3) schwa, shows no significant differences between any pair of conditions.

For reasons pertaining to statistical inquiry and experimental design explained in Section 3.2, our findings regarding alternation-specific signatures in RT are only preliminary, though: they need to be confirmed in future work. But what they suggest (Table 9), linguistically speaking, is that (1) float C adj. and (2) float C conjug. stand a good chance to be all online-processed (every step on the 1°–2°–3° ladder requires extra processing time), that is, decomposed (1° vs. 2°: made of two separate lexical items that are pieced together in online processing) and with online involvement of phonological computation (2° vs. 3°). By contrast, (3) schwa shows no effect for any of the steps, or for the 1°–3° comparison, suggesting that there is no specific online processing performed: items incline toward full listing (i.e., independent lexical storage of alternants in *(il) appelle - appeler*, etc.) and thus no specific phonological computation.

Finally, (4) *-ique* shows only a trend for the 1°–2° step, but a significant difference for the 2°–3° step (as well as for the 1°–3° comparison). We believe that the trend for the 1°–2° step may be linked to the small number of words that constitute this alternation. Using only five words is most probably responsible for the large standard errors observed for all three conditions of this alternation when compared to the two alternations that contained 68 words each (Table 4). It may thus be the case that the 1°–2° step turns out to be significant in further studies if the number of words contained in this class is increased. Finally, the comparison with (3) schwa is interesting in the sense that this alternation also contained only five words and as expected also produced important standard errors (Table 4). The difference is that the close RT among the three conditions did not reach significance anywhere. We thus believe that for (4) *-ique* the absence of significance in the 1°–2° step may be an artifact of the small number of words tested, while the absence of significance for (3) schwa all through is linguistically meaningful. But of course this interpretation needs to be ascertained in future studies, which should increase the number of words (pace the linguistic reservations exposed in Section 2.5): power analyses based on the current experiment should be conducted in order to reduce the discrepancy regarding the sample size of target words. These analyses will allow for a better detection of possible effects within classes.

4.5.2. Alternation specificity in the psycholinguistic literature on language production

We believe that the nonuniform behavior of different alternations is a result in itself in the context of the psycholinguistic literature on language production where we could not find any study inquiring on (or having as a by-product) alternation specificity. Recall from Section 1.2 that in the experimental context, an alternation-specific study compares the behavior produced by distinct alternations in response to the same experimental protocol.

Alternation specificity offers a contrasting view to the universalist position discussed in Section 1.2.1: according to this perspective, results gained from a specific phenomenon in a particular language may be generalized to all relevant phenomena of that language, or to language as such (this position was instantiated by Bürki et al., 2015). We believe that alternation specificity is a valuable contribution to the psycholinguistic discussion given its current absence and the fairly widespread universalist view, of which more instances are discussed below.

Bürki et al. (2010) study listing versus decomposition in cases of free variation (French *fenêtre* "window" pronounced with *[fən]être* or without *[fn]être* the initial schwa) and conclude that their experimental result "provides evidence that words with regular pronunciation variation, such as schwa words in French, are represented in the (production) lexicon with at least two lexemes, which requires modifications of current abstractionist and exemplar-based models" (p. 434). The authors thus generalize from one particular alternation in a specific language to language production as such.

Regarding the listing versus decomposition debate (see Section 1.2.1), literature overviews report that experimental evidence massively inclines toward the latter. Hence, Bien, Baayen, and Levelt (2011: p. 684) state that "[t]here is converging evidence that the production of complex words involves access to the constituent morphemes," and Cohen-Goldberg (2013: p. 1039) says that "[d]espite some variation in the findings, there is strong evidence that morphologically complex words are represented in a morpheme-based format (and perhaps, in addition, as whole words) in production."

This situation may then be interpreted as a binary yes-no competition between listing and decomposition: if the latter is massively supported, all morphologically complex words in all languages are decomposed, that is, stored and processed piecemeal. This is the position taken by Cohen-Goldberg (2013) in his theory of postlexical processing (the operations that determine how a plurimorphemic string will be pronounced, our 3° Overt Inflection), heterogeneity of processing hypothesis (HPH): since decomposition is general, morpheme boundaries are always present when phonological (and phonetic) processing occurs.

"Based on the evidence described earlier, the process of assembling post-lexical representations appears to leave boundary symbols and structural weaknesses at morpheme boundaries in the postlexical representation. This leads to the first prediction of the HPH, which is that postlexical processes will apply more strongly to tautomorphemic phonemes than heteromorphemic phonemes" Cohen-Goldberg (2013: 1042).

Linguistic approaches, such as Chomsky and Halle (1968), have also taken this position (all boundaries are always present upon phonological computation), but then need to implement alternation-specific behavior in terms of distinct boundaries. There is a massive literature on so-called class 1 versus class 2 morphology in English (see the overviews in Bermúdez-Otero & McMahon 2006 and Scheer, 2011, §141). Among other things, class 1 suffixes (typically of Romance origin) are stress-shifting, as for example *-al* in *parént-al*, as compared to the unsuffixed form *párent*. By contrast, class 2 suffixes (typically of Germanic origin) as for example *-hood* in *párent-hood* are stress-neutral: they do not modify the stress location of the unsuffixed form *párent*. On Chomsky and Halle's analysis, class 1 suffixes come with a + boundary (*parent+al*), while class 2 suffixes bear a # boundary (*parent#hood*). Phonological computation (postlexical processing in Cohen-Goldberg's terms) then produces the alternation-specific surface stress patterns by taking into account the two different boundaries (# blocks the application of the stress-shifting rule).

An alternative approach to class 1–class 2 morphology, represented by Kaye (1995), is based on the insight that the affixation of class 1 affixes produces plurimorphemic strings that are indistinguishable from monomorphemic strings (class 1 boundaries are invisible to phonological computation); by contrast, class 2 affixes alter the regular course that phonology would have taken in a monomorphemic string (class 2 boundaries are visible to phonological computation). Hence, given penultimate stress in English (for the sake of exposition: stress assignment is actually more complex than that), *párent* follows the rule just as much as *parént-al* (both have penultimate stress), the class 1 boundary thus being irrelevant (or invisible, or ignored in the phonology). *Párent-hood* is violating the rule, though, since it does not bear penultimate stress, and this is because of the class 2 affix (or class 2 boundary), which does impact phonological computation.

It thus appears to be a linguistic fact that whether or not phonological processing takes into account morphological boundaries is alternation specific (see also Section 1.2.2). This can be either implemented as the systematic presence of all boundaries in the input to phonological computation (which then takes some boundaries into account, but not others, Chomsky & Halle 1968), or in a privative way whereby only those morphological boundaries are present in the input to phonological computation that will actually have an effect on it (Kaye, 1995). For reasons of cognitive workload parsimony (why would you insert an item into the string that undergoes phonological computation when it does no labor?), the latter solution is favored in current syntactic phase theory (Chomsky, 2001) that carries over privativity into phonology (D'Alessandro & Scheer 2015, see the overview in Scheer, 2011, §756).

4.6. Conclusion

We believe that closer consideration of alternation specificity in experimental design and data interpretation when language production is studied may offer interesting perspectives. Just as much as the experimental probing of hypotheses and controversies that occur in linguistic quarters when it comes to the question whether a given alternation is (in linguistic terms) suppletive (lexicalized, i.e., listed as one piece), allomorphic (decomposed but without phonological computation contributing in production), or morphophonological (decomposed with production involving phonological computation).

Notes

1 The same goes for the study of Bien, Baayen, and Levelt (2011), which is concerned with deverbal adjectives and inflected verbs in Dutch: distinct lexical classes are not alternations; they do not represent alternative realizations of the same item.

2 Note that we talk about the presence versus absence of tasks for the sake of exposition. Rather, what Sahin, Pinker, Cash, Schomer, and Halgren (2009), or our study for that matter, measure is the extra activity generated by a specific linguistic task demanded by an experimental condition that modulates a baseline activity, which is present anyway. See note 7 for more detail.

3 Sahin et al.'s (2009, SOM Supporting Online Material pp. 4, 10) own experimental setup included a key-press task to insure that the patients really complied with the general task: upon stimulus, they had to silently pronounce a word and then press a key. Trials were only taken into account if the key-press fell into the window of response time allotted (1750 ms). Sahin et al. made explicit that key-presses were not designed for analytic input, in particular because patients were informed that they were not used to evaluate speed or accuracy, and that in fact "[t]here was no direct measure of reaction time […] in this task" (SOM p. 4). Nevertheless, Sahin et al. (2009, SOM p. 10) analyzed the key-press data from their three patients. In two patients, there was a statistically significant difference in the time elapsed between stimulus onset and key press when comparing 1° Read and 2° Null Inflect, and 1° Read and 3° Overt Inflect. There was no significant difference between 2° Null Inflect and 3° Overt Inflect, though (in one of the two patients, the reaction time of the former was even longer than the one of the latter). In the third patient, there was a limited statistical significance between 2° Null and 3° Overt Inflect, but not between 1° Read and 2° Null Inflect or 1° Read and 3° Overt Inflect.These nonuniform results illustrate the fact that Sahin et al.'s key-press data do not properly measure reaction time in production. Also, the medical condition limited Sahin et al. to only three participants, which is not a relevant number allowing them to generalize the results to the population at large. The present study is based on 40 participants whose real reaction times were measured. As we will see, results are more fine-grained given this method.

4 In Sahin et al.'s setup, all "pronunciation" is covert (inner speech), that is, does not involve speech organs: the participant merely produces the word silently in

their mind. This is to avoid noise in the EEG recording generated by speech organ activity.

5 But the issue is of course real. Interestingly, there is no trace of a putative extra burden created by a phonetic difference between stimulus and target words in our results: as discussed in greater detail in Section 3.2, one of the two alternations with this difference, (2) float C conjug., shows effects for all comparisons of the three experimental conditions, while the other (3) schwa shows no effect for any comparison. The two alternations where stimulus and target words are identical, (1) float C adj. and (4) *-ique*, either side with the alternation that shows effects in all three comparisons (float C adj.), or is in an intermediate position where only two comparisons produce a significant effect (*–ique*). Were there an extra processing burden due to the phonetic difference between stimulus and target words, one would expect the two alternations concerned to behave in a similar fashion—but this is not the case.

6 The type and display duration of successive screens follow Sahin et al.'s (2009) settings.

7 In this context, note that certain phonological tasks always take place no matter what. This is the case of stress placement, for example (in languages such as English and French where stress is not lexical): no utterance (or word) can be pronounced without being stressed. Hence, stress assignment should occur in all our experimental conditions, including when the word is merely repeated (1° Read). This means that the three experimental conditions implemented in our study do not measure the presence versus absence of phonological activity, but rather the extra activity generated by a specific phonological task demanded by 3° Overt Inflect that adds on to the baseline activity (including stress assignment), which occurs anyway.

Funding

This work has been supported by the French government, through the UCAJEDI Investments in the Future project managed by the National Research Agency (ANR) with the reference number ANR-15-IDEX-01, project *From Mind to Brain*.

Conflicts of interest

None.

Ethical approval

Participants in the experimental study granted informed consent and their privacy rights were protected. Ethical standards consistent with APA requirements were approved by INRIA's Comité Opérationnel d'Evaluation des Risques Légaux et Ethiques (COERLE, no 2018-004).

References

Alario, F.-X., Ayora, P., Costa, A., & Melinger, A. (2008). Grammatical and nongrammatical contributions to closed-class word selection. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 960–981.

Alario, F.-X., & Caramazza, A. (2002). The production of determiners: Evidence from French. *Cognition*, *82*, 179–223.

Baayen, R. H., Milin, P., Filipović Đurđević, D., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review*, *118*, 438–481.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.

Bell, A., Brenierc, J. M., Gregoryd, M., Girande, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, *60*, 92–111.

Bermúdez-Otero, R. (2012). The architecture of grammar and the division of labor in exponence. In J. Trommer (Ed.), *The morphology and phonology of exponence* (pp. 8–83). Oxford: OUP.

Bermúdez-Otero, R., & McMahon, A. (2006). English phonology and morphology. In B. Aarts & A. McMahon (Eds.), *The handbook of English linguistics* (pp. 382–410). Oxford: Blackwell.

Bien, H., Baayen, R. H., & Levelt, W. J. M. (2011). Frequency effects in the production of Dutch deverbal adjectives and inflected verbs. *Language and Cognitive Processes*, *26*, 683–715.

Bock, J. K. (1982). Toward a cognitive psychology of syntax. *Psychological Review*, *89*, 1–47.

Boeckx, C., & Uriagereka, J. (2007). Minimalism. In G. Ramchand & C. Reiss (Eds.), *The Oxford handbook of linguistic interfaces* (pp. 541–573). Oxford: OUP.

Bölte, J., Zwitserlood, P., & Dohmes, P. (2004). Morphology in experimental speech production. In T. Pechmann & C. Habel (Eds.), *Multidisciplinary approaches to language production* (pp. 431–471). Berlin: De Gruyter.

Bürki, A., Ernestus, M., & Frauenfelder, U. H. (2010). Is there only one "fenêtre" in the production lexicon? On-line evidence on the nature of phonological representations of pronunciation variants for French schwa words. *Journal of Memory and Language*, *62*, 421–437.

Bürki, A., Frauenfelder, U. H., & Alario, F. X. (2015). On the resolution of phonological constraints in spoken production: Acoustic and response time evidence. *Journal of the Acoustical Society of America*, *138*, EL429–EL434.

Butterworth, B. (1983). Lexical representation. In Butterworth, B., (Ed.), *Language production. Vol. 2. Development, writing and other language processes* (pp. 257–294). London: Academic Press.

Caramazza, A., Laudanna, A., & Romani, C. (1988). Lexical access and inflectional morphology. *Cognition*, *28*, 297–332.

Caramazza, A., Miozzo, M., Costa, A., Schiller, N., & Alario, F.-X. (2001). A crosslinguistic investigation of determiner production. In E. Dupoux (Ed.), *Language, brain, and cognitive development* (pp. 209–226). Cambridge, MA: MIT Press.

Charette, M. (1991). *Conditions on phonological government*. Cambridge: Cambridge University Press.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, N. (2001). Derivation by phase. In M. Kenstowicz (Ed.), *Ken Hale: A life in language* (pp. 1–52). Cambridge, MA: MIT Press.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. Cambridge, MA: MIT Press.

Cohen-Goldberg, A. M. (2013). Towards a theory of multimorphemic word production: The heterogeneity of processing hypothesis. *Language and Cognitive Processes*, *28*, 1036–1064.

Coleman, J. (1995). Declarative lexical phonology. In J. Durand & F. Katamba (Eds.), *Frontiers in phonology. Atoms, structures, derivations* (pp. 333–382). London & New York: Longman.

D'Alessandro, R., & Scheer, T. (2015). Modular PIC. *Linguistic Inquiry*, *46*, 593–624.

Dell, F. (1973a). "e muet", fiction graphique ou réalité linguistique? In S. Anderson & P. Kiparsky (Eds.), *A Festschrift for Morris Halle* (pp. 26–50). New York: Holt, Rinehart & Winston.

Dell, F. (1973b). *Les règles et les sons* (2nd edition). Paris: Hermann.

Fodor, J. (1983). *The modularity of the mind. An essay on faculty psychology*. Cambridge, MA: MIT-Bradford.

Foucart, A., Branigan, H. P., & Bard, E. G. (2010). Determiner selection in Romance languages: Evidence from French. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 1414–1421.

Goyvaerts, D. (1981). Introduction. In D. Goyvaerts (Ed.), *Phonology in the 1980's* (pp. 1–26). Ghent: Story-Scientia.

Green, A. D. (2007). *Phonology limited*. Potsdam: Universitätsverlag Potsdam.

Halle, M. (2005). Palatalization/velar softening: What it is and what it tells us about the nature of language. *Linguistic Inquiry*, *36*, 23–41.

Halle, M., & Mohanan, K. (1985). Segmental phonology of modern English. *Linguistic Inquiry*, *16*, 57–116.

Harris, J. (1994). *English sound structure*. Oxford: Blackwell.

Hay, J. B. (2001). Lexical frequency in morphology: Is everything relative? *Linguistics*, *39*, 1041–1070.

Hay, J. B., & Baayen, R. H. (2005). Shifting paradigms: Gradient structure in morphology. *Trends in Cognitive Sciences*, *9*, 342–348.

Hayes, B. (1995). On what to teach undergraduates: Some changing orthodoxies in phonological theory. In I.-H. Lee (Ed.), *Linguistics in the morning calm 3* (pp. 59–77). Seoul: Hanshin.

Hooper, J. (1975). The archi-segment in natural generative phonology. *Language*, *51*, 536–560.

Indefrey, P. (2011). The spatial and temporal signatures of word production components: A critical update. *Frontiers in Psychology*, *2–255*, 1–16.

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*, 101–144.

Janssen, N., Bi, Y., & Caramazza, A. (2008). A tale of two frequencies: Determining the speed of lexical access in Mandarin Chinese and English compounds. *Language and Cognitive Processes*, *23*, 1191–1223.

Janssen, N., & Caramazza, A. (2003). The selection of closed-class words in noun phrase production: The case of Dutch determiners. *Journal of Memory and Language*, *48*, 635–652.

Jescheniak, J. D., & Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Language, Memory, and Cognition*, *20*, 824–843.

Kaye, J. (1995). Derivations and interfaces. In J. Durand & F. Katamba (Eds.), *Frontiers of phonology* (pp. 289–332). London & New York: Longman. Also in SOAS Working Papers in Linguistics and Phonetics 3, 1993. 90–126.

Kiparsky, P. (1968–1973). How abstract is phonology? In O. Fujimura (Ed.), *Manuscript circulated since 1968 and published 1973 in: Three dimensions of linguistic theory* (pp. 5–56). Tokyo: TEC.

Kiparsky, P. (1974). On the evaluation measure. In A. Bruck, R. Fox, & M. L. Galy (Eds.), *Papers from the parasession on natural phonology* (pp. 328–337). Chicago, IL: Chicago Linguistic Society.

Kiparsky, P. (1982). Lexical morphology and phonology. In I.-S. Yang (Ed.), *Linguistics in the morning calm* (pp. 3–91). Seoul: Hanshin.

Levelt, W. J. M. (2001). Spoken word production: A theory of lexical access. *Proceedings of the National Academy of Sciences*, *98*, 13464–13513.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1–75.

Lightner, T. (1978). Generative phonology. In W. O. Dingwall (Ed.), *A survey of linguistic science* (2nd edition, pp. 1–32). Stamford, CT: Greylock.

McClelland, J. L., & Patterson, K. (2002a). Rules or connections in past-tense inflections: What does the evidence rule out? *Trends in Cognitive Sciences*, *6*, 465–472.

McClelland, J. L., & Patterson, K. (2002b). "Words *or* Rules" cannot exploit the regularity in exceptions. *Trends in Cognitive Sciences*, *6*, 464–465.

McMahon, A. (2007). Who's afraid of the vowel shift rule? *Language Sciences*, *29*, 341–359.

Miozzo, M., & Caramazza, A. (1999). The selection of determiners in noun phrase production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 907–922.

Morin, Y. C. (1988). De l'ajustement du schwa en syllabe fermée dans la phonologie du français. In P. Verluyten (Ed.), *La phonologie du schwa français* (pp. 133–202). Amsterdam & Philadelphia: Benjamins.

New, B., Pallier, C., Ferrand, L., & Matos, R. (2001). Une base de données lexicales du français contemporain sur Internet: Lexique. *L'année Psychologique*, *101*, 447–462.

Oldfield, R. C., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, *17*, 273–281.

Pinker, S. (1991). Rules of language. *Science*, *253*, 530–535.

Pinker, S., & Ullman, M. T. (2002a). Combination and structure, not gradedness, is the issue. *Trends in Cognitive Sciences*, *6*, 472–474.

Pinker, S., & Ullman, M. T. (2002b). The past and future of the past tense. *Trends in Cognitive Sciences*, *6*, 456–463.

R Development Core Team (2007). Eye movements and visual encoding during scene perception. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Rayner, K., Smith, T. J., Malcolm, G. L., & Henderson, J. M. (2009). Eye movements and visual encoding during scene perception. *Psychological Science*, *20*, 6–10.

Roux, F., Armstrong, B. C., & Carreiras, M. (2017). Chronset: An automated tool for detecting speech onset. *Behavior Research Methods*, *49*, 1864–1881.

Rumelhart, D. E., & McClelland, J. L. (1986). On learning the past tenses of English verbs. In J. L. McClelland & D. E. Rumelhart (Eds.), *Parallel distributed processing: Explorations in the microstructures of cognition, Vol. 2* (pp. 216–271). Cambridge, MA: MIT Press.

Sahin, N. T., Pinker, S., Cash, S. S., Schomer, D., & Halgren, E. (2009). Sequential processing of lexical, grammatical, and phonological information within Broca's area. *Science*, *326*, 445–449.

Schane, S. (1968a). *French phonology and morphology*. Cambridge, MA: MIT Press.

Schane, S. (1968b). On the abstract character of French "e muet". *Glossa*, *2*, 150–163.

Schane, S. (1972). The hierarchy for the deletion of French "e muet". *Linguistics*, *82*, 63–69.

Scheer, T. (2011). A guide to morphosyntax–phonology interface theories. *How extra-phonological information is treated in phonology since Trubetzkoy's Grenzsignale*. Berlin: Mouton de Gruyter.

Schreuder, R., & Baayen, R. H. (1995). Modeling morphological processing. In L. Feldman (Ed.), *Morphological aspects of language processing* (pp. 131–154). Hillsdale, NJ: Erlbaum.

Spalek, K., Bock, K., & Schriefers, H. (2010). A purple giraffe is faster than a purple elephant: Inconsistent phonology affects determiner selection in English. *Cognition*, *114*, 123–128.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*, 520–522.

Tranel, B. (1981). The treatment of French liaison: Descriptive, theoretical and methodological implications. In H. Contreras & J. Klausenburger (Eds.), *Proceedings of the 10th Anniversary Symposium on Romance Languages* (pp. 261–281). Seattle, WA: University of Washington.

Tranel, B. (1987). Floating schwas and closed syllable adjustment in French. In W. Dressler, H. Luschützky, O. Pfeiffer, & J. Rennison (Eds.), *Phonologica 1984* (pp. 311–317). London: Cambridge University Press.

Tranel, B. (1988). A propos de l'ajustement de e en français. In P. Verluyten (Ed.), *La phonologie du schwa français* (pp. 89–131). Amsterdam & Philadelphia: Benjamins.

Williams, E. (2007). Dumping lexicalism. In G. Ramchand & C. Reiss (Eds.), *The oxford handbook of linguistic interfaces* (pp. 353–381). Oxford: OUP.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Supporting Information