MAKE YOUR MOUTH *AGUA*: IDIOMS AND THE INTEGRATED LEXICON HYPOTHESIS[*]

Luis López[a], Irati de Nicolás[b], Rodi Laanen[c], Charlotte Pouw[d], José Sequeros-Valle[e] & M. Carmen Parafita Couto[c,f]
[a]University of Illinois Chicago, [b]University of Chicago, [c]Leiden University, [d]University of Amsterdam, [e]University of Nebraska - Omaha, [f]Universidade de Vigo

ABSTRACT. Recent work argues that a bilingual linguistic system is fully integrated in one competence whole and does not consist of two separate, autonomous structures as is commonly assumed (see Alexiadou et al 2015, Alexiadou et al 2018, Goldrick et al 2016, Grimstad et al 2014, López 2020, Riksem 2017). Here, we explore the organization of the lexicon within the integration hypothesis using data based on idioms and code-switching. The working hypothesis is that if the lexicons of a bilingual person are integrated in the sort of grammatical architecture presented in López (2020), based on Distributed Morphology (Marantz 1997), one should be able to code-switch within the idiom and retain the idiomatic meaning. After a pilot study with a community of Papiamentu-Dutch bilinguals, we tested this hypothesis with two communities of bilingual code-switchers: Basque-Spanish, English-Spanish. The task consisted of choosing a plausible meaning out of three choices for a range of sentences, some of which included code-switched idioms. Our results, by-and-large, show that code-switching does not destroy the integrity of the idiom, thus providing empirical evidence for the integrated lexicon hypothesis as well as the validity of Distributed Morphology as a theory of grammar.

Keywords. lexicon, code-switching, bilingualism, integrated hypothesis, distributed morphology

RESUMEN. Unos trabajos recientes argumentan que un sistema lingüístico bilingüe está completamente integrado en una unidad de competencia única y que no consiste en dos estructuras separadas y autónomas, como se suele suponer (véase Alexiadou et al 2015, Goldrick et al 2016, López 2020, Riksem 2017). En este trabajo exploramos la organización del léxico dentro de la hipótesis de la integración usando unos datos basados en las expresiones idiomáticas y el cambio de códigos. La hipótesis de trabajo es que si los léxicos de una persona bilingüe están integrados en el tipo de modelo presentado en López (2020), basado en la Morfología Distribuida (Marantz 1997), uno debería poder cambiar de código dentro de la expresión idiomática y retener el significado idiomático. Después de hacer un estudio piloto de bilingües papiamento-holandés, investigamos esta hipótesis con

dos comunidades de bilingües que cambian códigos: español-vasco y español-inglés. La tarea consistió en escoger un significado plausible de entre tres posibilidades en un conjunto de oraciones que incluían expresiones idiomáticas con cambio de código. Nuestros resultados, en líneas generales, muestran que el intercambio de código no destruye la integridad de la expresión idiomática, lo cual nos da evidencia empírica a favor de la hipótesis del léxico integrado, así como la validez de la Morfología Distribuida como teoría de la gramática.

Palabras clave. Léxico, cambio de código, bilingüismo, hipótesis de la integración, morfología distribuida

1. Introduction

The assumption that being bilingual entails having two lexicons (and two grammars) in one head is commonsense and largely unquestioned in linguistic studies of bilingualism.[1] This view of bilingualism has recently been challenged within different analytical frameworks, specially within a soft-constraint approach (Goldrick et al 2016, Hsin and Légendre 2019) as well as within exoskeletal models of the language faculty (see Alexiadou et al 2015, Alexiadou et al 2018, Grimstad et al 2014, Riksem et al 2019, among others). Within this second tradition, López (2020) presents empirical arguments that bilinguals are in possession of a unique linguistic competence – which includes a unique Lexicon and a unique externalization system (Phonetic Form or PF). The model in López (2020) is couched within Distributed Morphology (henceforth DM) (Halle and Marantz 1993, Marantz 1997 et seq.). In this model, there is no entity that we may call a lexicon in traditional terms – rather, lexical knowledge is distributed in three lists (see also Jackendoff 1997 et seq.), each of which houses a distinct subset of the properties that we normally associate with a morpheme. Thus, DM is opposed to theories of linguistic architecture that assume the existence of a Lexicon where lexical units are stored with their semantic, syntactic, and phonological properties.

This article shows that idiomatic expressions – syntactic structures with a non-compositional meaning – provide an interesting database to test the integrated hypothesis as well as the grammatical architecture put forth in DM. In sum, we show that code-switched idiomatic expressions like 'make your mouth *agua*' (=to be mouth-watering) are interpreted as idioms by bilingual speakers. This result supports Integrationism against Separationism and DM against Lexicalism.

In section 2 we present our theoretical framework, we show how it is deployed for the study of code-switching, we discuss alternatives and formulate our main hypothesis. Thus, section 2 sets up the groundwork for the survey work discussed in sections 3 through 5.

2. Background

2.1. Distributed morphology and the bilingual lexicon

The model assumed in this article is pictured in (2). The input to the computational system is a list of roots and functional categories without phonetic content (called List 1). The computational system builds complex structures – words and phrases - out of items from List 1. At some point, it branches into two computations, one leading to the external sensori-motor systems (S-M), the other

---

[1] However, psycholinguists have claimed there is an expanded lexicon for bilinguals for some time (cf. Kroll et al 2015).

to the Conceptual-Intentional systems (C-I). On the right branch, a list of realizational rules (List 2) matches phonological matrices to syntactic terminals. On the left branch, a third list (List 3) assigns meanings to labeled structures (List 3 is not discussed further in this article).

(2) *Distributed Morphology framework* (see Harley and Noyer 2000 i.m.a.)

C-I

S-M

[PAST]

[PAST] ← ☺ Ø ‖ {hit, cut…} ___

[PAST]←→/əd/ ‖ [ALVEOLAR]__

[PAST]←→ /t/ [-VOICE] __

…

List 2

[PAST]

[SUBJUNCTIVE]

[PLURAL]

[NEG]

√123

…
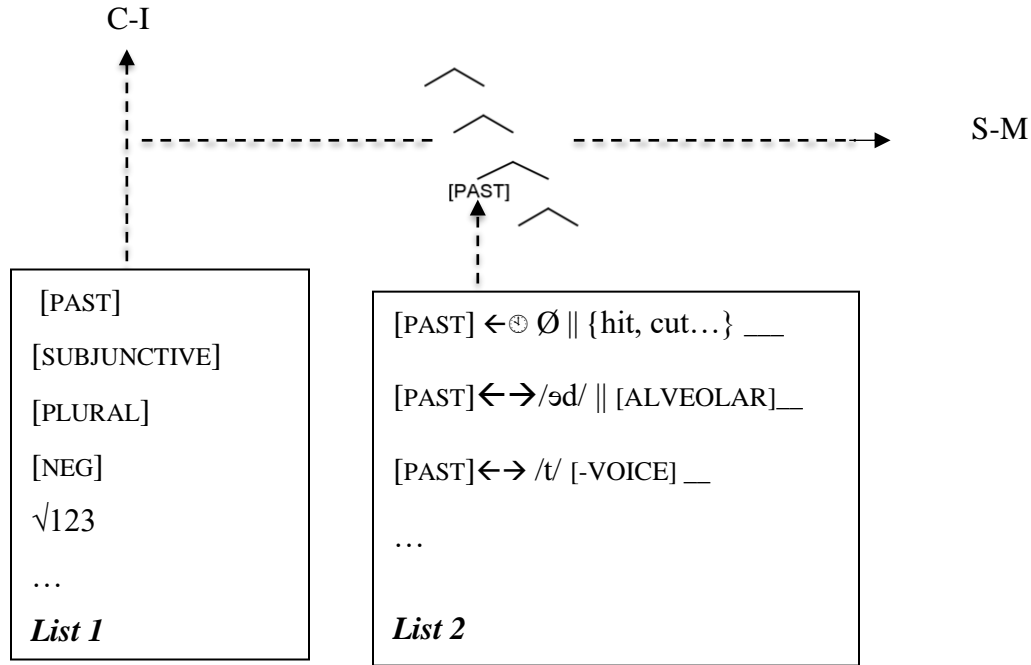
List 1

López (2020) argues that bilinguals do not have separate List 1, 2 and 3, one for each language, but rather, integrated lists. The List 1 of a bilingual encompasses all the lexical roots as well as the functional categories that comprise the bilingual's linguistic competence. Likewise, all the List 2 items are stored together.

2.2. The alternative spell-out hypothesis

The integrated hypothesis within the DM system adopted here leads to a very natural consequence, proposed in López (2020): many words that appear to be duplicated in the two languages are *alternative spell-outs* of the same root: 'planet' and *planeta*, 'air' and *aire*, 'water' and *agua*, etc. Although not logically required, we take it that this assumption is plausible and would lead to an organization of a bilingual's vocabulary that would eliminate a considerable amount of redundancy.

The alternative spell-out organization is certainly the case for most nouns and relational adjectives (*económico* 'economic', *francés* 'French'), and many qualifying adjectives (*alto* 'tall', *inteligente* 'intelligent'). It is probably the case also for some verbs such as *estudiar* 'study' and some prepositions, although the semantics of the latter two classes of words often do not match cross-linguistically and therefore spell-out duplication must be limited.

If the alternative spell-out hypothesis is correct, the bilingual Spanish-English speaker may have only one root in List 1 for many pairs of nouns and adjectives. Assume this is the case for 'dog' and *perr(o)*. In example (4), this root is identified with the arbitrary numerical index √45 (using the theory of List 1 argued for in Harley 2014) to indicate that a root does not have any

meaning in itself and its only work is to bind a slice of conceptual structure (an item from List 3) with a spell out rule in List 2. The root √45 has two possible spell-outs in List 2: /perr(o)/ and /dog/. The word *perr(o)* is spelled out when √45 is selected by a *n* with a gender feature while 'dog' spells out √45 when there is no gender in the structure. The root, in combination with the morpheme that categorizes it, is matched to the concept DOG (in small caps) in List 3, which corresponds to the ordinary *canis familiaris*:

(4)   List 1:   root: √45
        List 2:   √45 ←→ /perr-/ || _____ *n*[gender]
                       √45 ←→ /dog/
        List 3:   [√45 + n] : DOG

We refer the reader to López (2020) for details of the proposal.[2]

2.3. Idioms, MDM and Code-Switching [3]

There are two features of idiomatic expressions that are of interest. Some idioms are connected to the compositional meaning in some vague form and their meaning can be easily inferred with minimal context. Among our examples (5), (6) and (7), (6) is probably the most transparent, (7) is more opaque than (6), and (5) is fully opaque. The second feature of idioms that we are interested in is that many idiomatic expressions are lexically rigid, i.e., you cannot change the lexical items involved without destroying the idiomatic meaning (Nunberg et al, 1994). Taking (5) as an example, 'kick the bucket' can mean 'die', but 'kick the can' or 'kick the bottle' only have their literal meanings. In this article, we are primarily interested in rigid idioms.

(5) Kick the bucket   #the can   #the bottle   #the glass
(6) Kick the can down the road #the bottle   #the bucket   #the pail
(7) Shoot the breeze   #the rain   #the wind   #the draught

Within DM, it is possible to view idioms not as lexical items but as syntactically compositional phrases and displace the idiosyncratic meaning to List 3 (Marantz 1997). Additionally, McGinnis (2002) shows that inner aspect, a syntactically based semantic feature, is also compositional within idioms. Within DM assumptions, the verb phrase 'kick the bucket' is built out of two roots and the functional category [definite], as shown in (8a). A representation like (8a) with only three components for (5) is in fact a simplification, since lexical roots are listed without specification

---

[2] We would like to emphasize that what we are working on is a theory of the lexicon as a component of the grammatical competence of (bilingual) speakers. As such, it is not meant to compete with or contribute to the many psycholinguistic studies on the bilingual lexicon. These studies concern themselves with processing and production, which are orthogonal to our purposes. (for general overviews and references, see Fernández and Cairns, 2017, Papafrogou, Trueswell and Gleitman 2022).

[3] The literature on the grammar of idioms in monolingual contexts is rich and there is no hope we could summarize it here. The reader is referred to the survey article Espinal and Mateu (2019) and references therein; Bruening (2020) includes an extensive and very useful list of different types of idioms in English. Idioms are the foundation of Construction Grammar, see the contributions in Hoffman and Trousdale (2013). There are also numerous papers on idioms from a psycholinguistic perspective: Titone et al (2019) provide a useful introduction to the topic and Titone et al (2015) a discussion of bilingual idioms.

for category and therefore need to combine with a categorizer, which is not represented in (8a). Other possible grammatical features are not represented either.

(8b) shows the structure built out of the three items in (8a). (8c) shows the spell out rules that apply to the syntactic terminals and finally (8d) tells us the meaning of the full structure (see Marantz 1997):

(8) a. List 1: {√23, √32, def}

    b.


```
                         VP
                        /  \
                     √23    DP
                           /  \
                       [def]    √32
```
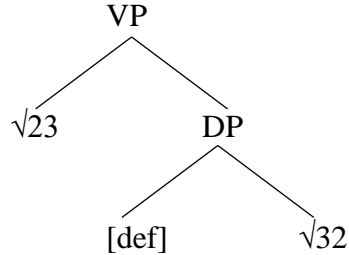
    c.   List 2: √23 ←→ /kick/
                 √32 ←→ /bucket/
                 [def] ←→ /ðə/

    d.   List 3: VP ←→ DIE

We suggest that idiomatic expressions provide an excellent playground to test the integrated hypothesis. Assume the English word 'bucket' and the Spanish *cubo* (which means 'bucket') are indeed two forms of spell out for the same lexical root and matching concept in List 3. If so, then they should be exchangeable in the idiomatic expressions. This reasoning entails that (9), (10) and (11) could be interpreted idiomatically by a Spanish-English bilingual:

(9)   Kick the *cubo* (=bucket)
(10)   Kick the *lata* (=can) down the road
(11)   Shoot la *brisa* (=breeze)

At this point, it is crucial to recall that 'kick the bucket' is a rigid idiom, one cannot change the lexical components without doing away with the idiomatic meaning, as shown in (5). Thus, if 'bucket' and *cubo* are indeed interchangeable, then they must be exponents of the same root.

(12) represents the structure of the idiom 'kick the bucket/*cubo*' within distributed assumptions in a bilingual grammar. The items in List 1 and List 3 are identical in relevant respects to what we see in the monolingual representation (8), as is the syntactic structure.[4] The only difference is that the bilingual will have two alternative ways to spell out √32:

---

[4] As an anonymous reviewer points out, the structure that yields *cubo* is not identical in every respect to the one that yields *bucket*. For once, the *n* of *cubo* has a gender feature that spells out as *o*. It follows that what is code-switched in these examples is not just the root but the root and *n*. This fact does not alter the lay out of the hypothesis. The discussion in the main body of this paper simplifies matters where it is inconsequential.

(12)   a.   List 1: {√23, √32, def}

        b.

```
                VP
               /  \
            √23    DP
                  /  \
                def   √32
```
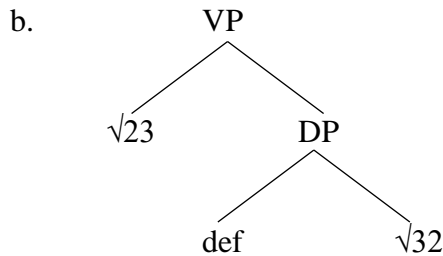
        c.   List 2: √23   ←→ /kick/
                            √32   ←→ {/bucket/, /cub(o)/}
                            [definite] ←→ /ðə/

        d.   List 3: VP ←→ DIE

    There are two crucial hypotheses embedded in (12). The first one is straightforward: the idiomatic meaning is built on the phrase structure shown in (8b) and (12b), which is populated by abstract roots and bundles of features. The interpretation in List 3 is sensitive only to the List 1 items and the way they are composed, not their phonetic form. It follows that the idiomatic meaning is preserved as long as we have the structure in (12b) - regardless of how the syntactic terminals are spelled out. This conclusion yields the second hypothesis and is the focus of this article: if both 'bucket' and *cub(o)* can spell out √32, then either of them could be spelled out and yield the idiomatic meaning. This is what we set out to test in this article. See figure 1 for initial anecdotal evidence that the hypothesis may hold water.

Figure 1: Facebook caption that shows the use of 'make your mouth agua' (agua = 'water')

    If the data in (9) – (11) turn out to be interpretable to bilingual speakers with the idiomatic meaning, the Separationist hypothesis represented in (1) could not account for it in a direct manner

without additional stipulations: within Separationism, 'bucket' and *cubo* are two different lexical items living in separate lexicons. If idioms like the ones in (9) – (11) turn out to be grammatical for bilingual speakers, the Separationist scholar would have to posit a rule like the following:

(13)   Assume a bilingual person with lexicons L1 and L2 corresponding to each of the languages.
         Assume a w1 in L1 which is a crucial component of an idiom i.
         Assume further that a w2 in L2 is semantically equivalent to w1.
         Then you can substitute w2 for w1 in i while retaining the idiomatic meaning.

    This is not only a complicated rule but also arbitrary, because there is no particular reason why it should exist. If we need to posit analogical rules like (13) to maintain two separate lexicons, one should wonder about the wisdom of Separationism itself. On the other hand, the Integrated model incorporates the effect of rule (13) organically without further assumptions.
    Our proposal affects our theory of the bilingual lexicon because we ask whether we should regard it as divided in two halves, each corresponding to what we usually refer to as a language, or whether it is integrated.

2.4. Bilingual idioms and grammar architecture
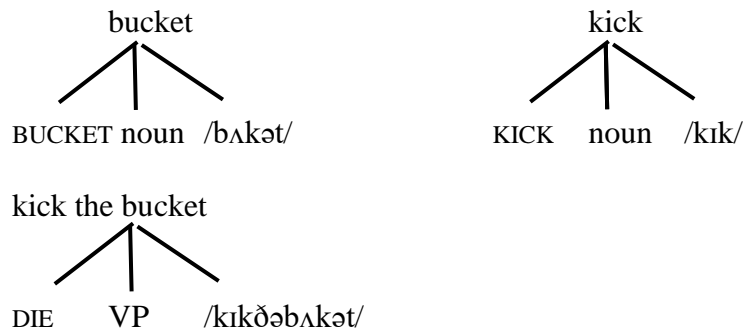
    Our proposal also affects how we understand the architecture of language more generally, because we also ask whether we should assume that there exists a traditional lexicon or a distributed system. In fact, the view of idioms adopted in this article is not the mainstream one. Traditional generative grammar assumes that our linguistic competence includes a lexicon in which *lexical items* are stored with their morphosyntactic, semantic and phonological information:

(14)   Lexicon: word: dog

                              DOG   noun   /dog/

    These lexical items provide the input to the computational system(s). This approach is referred to as *Lexicalism* (see for instance a textbook introduction like Spencer 1991 and, for recent discussion of current lexicalist theories, Audring 2022). Within Lexicalism, the semantic unpredictability of idioms leads to the conclusion that they are also stored in the lexicon like regular lexical items or constructions (Nunberg et al 1994, Goldberg 2006). Thus, in this view, the lexicon of an English speaker would include items such as the following:

(15)   bucket   kick

         BUCKET noun   /bʌkət/   KICK   noun   /kɪk/

    kick the bucket

         DIE   VP   /kɪkðəbʌkət/

The contrast between Lexicalism and DM is that the latter acknowledges that idioms are syntactically compositional even if they are not semantically so; non-compositionality is then displaced to the organization of List 3 (Marantz 1997 et seq).

Regarding the bilingual architecture, the predictions generated by Lexicalism regarding (9)-(11) are distinct from those generated by DM. The lexicalist view of the lexicon entails that the idiom 'kick the bucket' is listed in the lexicon of a bilingual, while forms such as 'kick the *cubo*' or even '*patear* the bucket' would not be listed and there is no available mechanism to compositionally build (9)-(11). Consequently, (9) – (11) would be rejected by bilinguals. Keller (2020) acknowledges the difficulties that mixed idioms present for the Lexicalist Hypothesis: "One might assume that a string of words which appears as a unit on the level of conceptual representation should be barred from internal language mixing in order to preserve the exact meaning or pragmatic function of the unit." (Keller 2020: 200). She further claims (2020:217) that the idiomatic expression must maintain its lexical integrity – including phonetics - in the original language although functional items can be built from the other language. In effect, she realizes that Lexicalism predicts that examples like (9), (10) and (11) should be unacceptable as idioms. Thus, the current article may provide evidence for DM over lexicalist architectures while simultaneously providing evidence for the integrated framework of the bilingual lexicon.

2.5. Hypothesis

To sum up this introduction, this is the hypothesis that we set out to test in this paper, stated informally:

(16)   Hypothesis:
        Let i be an idiom in L1.
        Let w1 be a word in L1 and a crucial component of i1.
        It should be possible to substitute w2, a word of L2, for w1 and maintain the idiomatic
        meaning in the grammar of an L1/L2 bilingual person.

(16) includes some vocabulary that presupposes an understanding of bilingualism that we do not subscribe to: the existence of a linguistic competence divided between an L1 and an L2, with their separate lexicons etc. We re-state (16) formally in DM terms as follows:

(17)   The choice of a List 2 item does not alter the interpretation of an idiom.

In a way, assuming DM makes (17) trivial, since it follows from DM that all List 2 items are independent of semantic processes. In order to test the hypothesis, we undertook to find out how bilingual speakers would interpret code-switched idioms. We operated under the assumption that identifying a set of bilingual speakers who maintained idiomatic interpretations in switched idioms would provide support for the Integrated Hypothesis.

We designed a small pilot survey with 17 Papiamentu-Dutch speakers (nine females), all of them early bilinguals. We took the two idiomatic expressions listed in (18) and we replaced the last word for the equivalent in the other language, In (18a), a Dutch expression, we used the Papiamentu word *manga* instead of *mouw*, in (18b), a Papiamentu expression, we used the Dutch word *schaduw* instead of the Papiamentu *sombra*:

(18)   a.   De aap komt uit de mouw.   (manga)
                    The monkey comes out the sleeve
                    'The truth is revealed'

         b.   Awendia bo no por konfia nib o sombra   (schaduw)
                    Nowadays you cannot trust even your shadow
                    'Nowadays you cannot trust anyone.'

    Participants were asked to choose one of three options: (i) a continuation that built on the idiomatic meaning; (ii) a reasonable continuation that did not build on the idiomatic meaning and (iii) none of the above
We found that 15 out of 17 participants chose the idiomatic interpretation with the switched word. We take this to suggest that in the lexicon of these speakers the words *manga* and *mouw* are exponents of the same root. Encouraged by this result, we engaged the two broader surveys that we report on in the following sections.[5]

    The rest of this article is organized as follows. Section 3 presents the survey that we carried out with the Basque-Spanish bilingual group. Section 4 presents the English-Spanish group. Section 5 concludes in light of the previous theoretical discussion. Detailed information about the complete surveys can be found in the appendices.[6]

3. Basque/Spanish Survey

3.1. Participants

    22 early Basque/Spanish bilinguals (8 male, age 18-56), born and raised in areas around Bilbao and schooled in Basque up to grade 12, took part in the study. Proficiency was determined via self-reports[7] (Table 1). All the participants had some higher education (which could be in Basque, Spanish or English).

Table 1: Average self-ratings (5 point scale).

	Basque	**Spanish**
	M (SD)	*M (SD)*
Reading	4.47 (0.51)	4.73 (0.56)
Writing	4 (0.88)	4.63 (0.68)
Speaking	4 (0.88)	4.68 (0.67)
Understanding	4.36 (0.49)	4.78 (0.41)
Overall	4.21 (0.21)	4.71 (0.95)

---

[5] An attempt to administer a broad scale survey in the Papiamentu community failed for lack of access during the pandemic.

[6] Appendices are available via https://osf.io/9hzvs/?view_only=5ef9817d4c924525a7b56fd7559cb65f

[7] See Edele, Seuring, Kristen & Stanat, 2015 and Li & Zhang, 2021 regarding self-reports as a complementary proficiency measure. Additionally, van Osch 2019 shows that self-reports correlate significantly with other measures of proficiency such as the DELE (Diploma Español de Lengua Extranjera) or lexical decision tasks.

3.2. Task and Stimuli

Participants were asked to decide on the meaning of an idiom out of three possible choices (see Appendix A for an English version of the instructions). We provided participants with idiomatic expressions containing code-switching. We selected four Spanish idiomatic expressions and four Basque idiomatic expressions, selected in consultation with two speakers from the community, see Table 2). After the main task, we tested participants' knowledge of the idioms. Additionally, we made sure the idioms were specific of a particular language (idioms used in both languages were not included).

Table 2: Idioms selection for Spanish (#1-4) / Basque (#5-8) Survey.

#	Idiomatic expressions	Idiomatic meaning
1	*Írsele la olla* go.refl.dat the pot	to get mad
2	*Estar de mala leche* be of bad milk	to be in a bad mood
3	*Dar en el clavo* give on the nail	to nail something
4	*Importar un pimiento* care one pepper	not to care
5	*Adarra jo* horn play	to make fun
6	*Kopet ilun jarri* forehead dark put	to get mad
7	*Arkakusoak hil* fleas kill	to hit someone
8	*Larru gorritan egon* skin in red be	to be naked

The noun in each of these expressions was replaced with the Basque or Spanish counterpart and a context was added for each sentence.[8] In total, we had four Basque idioms with a code-switched Spanish noun and four Spanish idioms with a code-switched Basque noun. (19) is an

---

[8] An anonymous reviewer asks why we replaced nouns rather than verbs or other categories. As we discuss in section 2.2, verbs tend to have idiosyncratic meanings cross-linguistically and therefore they are less likely to instantiate the duplicated List 2 vocabulary items that we look for to test the Integrated Hypothesis.

example of a code-switched idiom. We replaced the word *leche* 'milk' with *esne*- the Basque equivalent. See Table 3 for all tokens (See also Appendix B for a complete list of stimuli).

(19)   Hoy   he   visto a Jon   que parecía que estaba   de mala *esne*
        today have.1 seen ACC Jon   that seemed that was   of bad milk
        'Today I saw Jon, who seemed to be in a bad mood.'

Table 3: Idioms selection for Basque/Spanish Survey (Spanish (#1-4) / Basque (#5-8))

#	Idiomatic expression	Idiomatic meaning	Noun replaced
1	*Ayer* *en las txosnas a* *Mikel se* *le* *fue* *la* lapiko yesterday in the txosnas acc Mikel refl dat went the pot 'Yesterday in the txosnas, Mikel got angry.'	to get mad	Spanish *'olla'* = pot
2	*Hoy* *he* *visto a* *Jon, que parecía que* *estaba de* today have.1 seen acc Jon, who seemed that was of *mala* esne. bad milk 'Today I saw Jon, who seemed to be in a bad mood.'	to be in a bad mood	Spanish '*leche'* = milk
3	*Con el* opari *de Idoia* *he* *dado* *en el* iltze with the present of Idoia have.1 given in the nail 'I nailed it with Idoia's present'	to nail something	Spanish '*clavo'* = nail
4	*No* *sé* *cuánto* *gana Ane* *y* *me* *importa un* piper not know.1 how much earns Ane and dat.1 care one pepper 'I don't know how much Ane gets paid and I don't care.'	not to care	Spanish *'pimiento'* = pepper
5	Ayer Maitek *cuernoa* jo zidan yesterday Maite.erg horn play AUX 'Yesterday Maite made fun of me'	to make fun	Basque 'adarra' = horn
6	Atzo gauean Telmo *frente* ilun jarri zen yesterday night Telmo forehead dark put AUX 'Yesterday night, Telmo got mad.'	to get mad	Basque 'kopet' = forehead
7	Atzo Landerri *pulgak* hil zizkioten yesterday Lander.dat fleas kill AUX 'Yesterday, Lander got beaten up.'	hit someone	Basque 'arkakuso' = flea
8	Amets gaizto bat izan dut, *piel* gorritan nengoen dream bad one have AUX, skin. red was 'I had a nightmare, I was naked.'	to be naked	Basque 'larru' = skin

We presented participants with the code-switched idiom and three possible continuations: the actual idiomatic meaning, a non-idiomatic meaning drawn from a plausible semantic field, and none of the above. Participants were asked to choose the most logical continuation among these three options. An example is found in (20):

(20)   Hoy   he   visto a   Jon   que parecía que estaba   de mala *esne*
        today have.1 seen   acc   Jon   that seemed that was   of bad milk
        'Today I saw Jon, who seemed to be in a bad mood.'

Options:

 a) (<u>Non-idiomatic meaning):</u> Jon venía de la *denda*
           'Jon was coming from the store'
 b) (<u>Idiomatic meaning):</u> Jon estaba enfadado
           'Jon was mad'

 c) None of the above

We also included twelve distractors (Appendix C) consisting of eight sentences that were syntactically and semantically compositional as well as four sentences that included "fake idioms", phrases that could look like an idiom except for a misleading item. For instance, where the expression in Spanish says "no tener pies ni cabeza" (not to have heads or feet=not to have rhyme or reason) we changed it to "no tener pies ni dedos" (not to have feet or fingers). We used fake unilingual and bilingual idioms. Fake idioms were deemed necessary to control for possible anticipatory effects: it was possible that participants might check the idiomatic meaning after recognizing the first part of the phrase (for anticipatory effects in a reading task see Titone et al. 2015). The order of presentation of the stimuli was pseudorandomized.

3.3. Procedure

The stimuli were presented in written form (Qualtrics). First, participants completed the code-switching Forced Choice task. The instructions were code-switched so they would get into code-switching mode (González-Vilbazo et al, 2013). Before the task, they saw some trial questions that followed the same format as the questions of the task. After the code-switching task, participants were presented with the same idioms monolingually. This was done to ensure the participants knew the meaning of the idioms. Finally, they completed a background questionnaire (Appendix D).

3.4. Results

In this section we present the results for the total responses. We looked for correlations to any of the background factors controlled for in the survey: gender, living location, age, level of education, attitudes towards code-switching. We found none (see Appendix E).

  Figure 2 shows the results for the Spanish and Basque monolingual idioms. Participants assigned the idiomatic interpretation 100% of the time for the Spanish idioms. With Basque idioms, participants assigned the idiomatic interpretation 81% of the time. As for the other 19%, participants chose 'none of the above'.
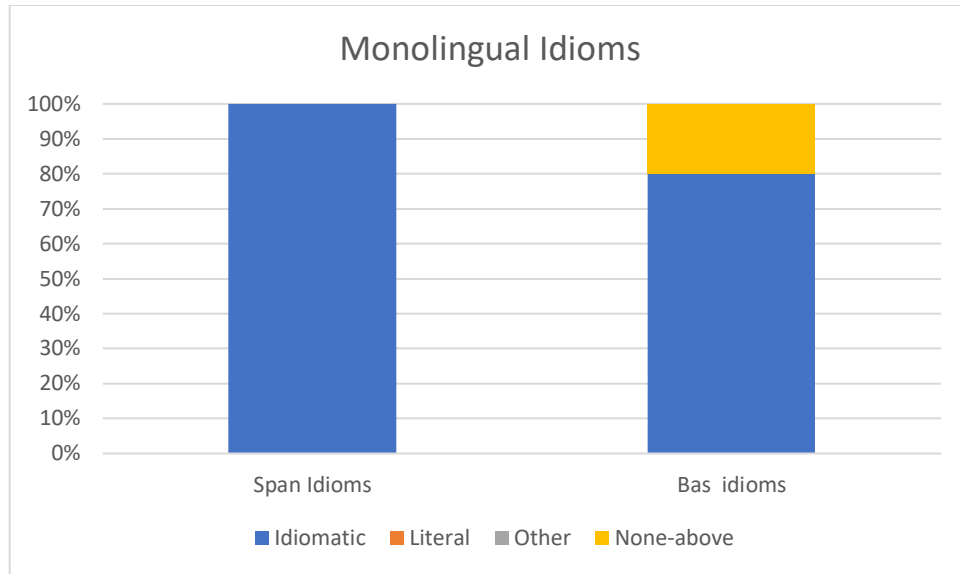
Figure 2: Descriptive results for the monolingual idioms.

Figure 3 shows a comparison between monolingual and code-switched idioms. Spanish monolingual idioms (column 1, same as column 1 in Figure 2) received an idiomatic interpretation 100% of the time (n = 88 / 88) while Spanish idioms with a switch to Basque (column 2) received an idiomatic interpretation ~97% of the times (n = 86 / 88). The same figure also shows that Basque monolingual idioms (column 3, same as column 2 in Figure 2) received an idiomatic interpretation ~81% of the time (n = 72/ 88) while Basque idioms with a switch to Spanish (column 4) receive an idiomatic interpretation ~62,5% of the time (n = 55 / 88). The Fisher Exact Test calculations show that the difference among the two conditions is significant ($\chi2 = 0.006$, $p < .05$).

Figure 3: Monolingual vs. code-switched idioms in Spanish-Basque

3.5. By-Item, Pairwise Comparisons

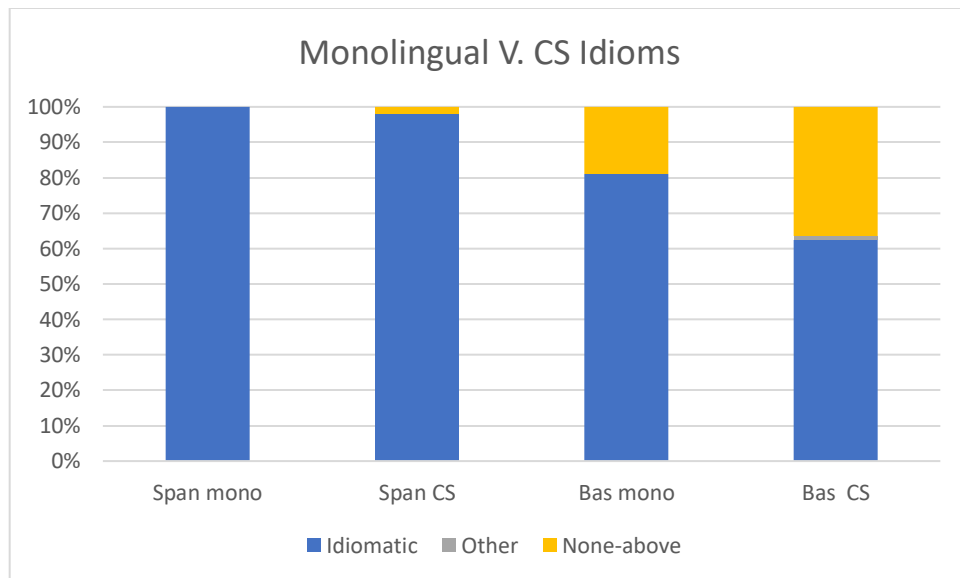As shown in Figure 3, there is more variation with Basque idioms than with Spanish idioms. In order to further examine this, we looked at each idiom individually and we compared each code-switched version with their monolingual counterpart. The results are presented in Table 4, which shows the percentage choice of the idiomatic interpretation.

The Spanish idioms show the expected results: all the participants assigned the idiomatic interpretation to the monolingual expression and almost all the participants assigned the idiomatic interpretation to the code-switched idioms (with the exception of 'Se le fue la lapiko''he lost his pot', which was interpreted as idiomatic 90% of the time).

There is more variation in the Basque expressions, observing two main patterns. For items 5 and 8, there is a significant difference between the code-switched item and the monolingual item. For items 6 and 7, there is no significant difference between the code-switched and monolingual expressions; however, it seems that not all the participants know the meaning of the Basque monolingual idiom - the percentage of instances in which the monolingual idiomatic expressions are understood as idiomatic is low (72.72% and 55.54% respectively). Thus, we decided to analyze the data again and remove the answers of participants who did not know the meaning of the Basque idiom. These results are presented in the next section.

Table 4: Pairwise comparisons by-item.

#	BILINGUAL	%	Sig.?	%	MONOLINGUAL
1	*Írsele la* **lapiko** 'to lose his pot' Idiom.: to get mad	90.90		100	*Írsele la* **olla** 'to lose his pot' Idiom.: to get mad
2	*Estar de mala* **esne** 'to be of bad milk' Idiom.: to be in a bad mood	100		100	*Estar de mala* **leche** 'to be with bad milk' Idiom.: to be in a bad mood
3	*Dar en el* **iltze** 'to give on the nail' Idiom.: to nail something	100		100	*Dar en el* **clavo** 'to hit the nail' Idiom.: to nail something
4	*Importar un* **piper** 'to care a pepper' Idiom.: not to care	100		100	*Importar un* **pimiento** 'to care a pepper' Idiom.: not to care
	OVERALL	98%		100%	OVERALL
5	*Cuernoa* **jo** 'play the horn' Idiom.: to make fun	72.72	*	100	**Adarra** jo 'to play the horn' Idiom.: to make fun
6	*Frente* **ilun jarri** 'to put your forehead dark' Idiom.: to get mad	72.72		72.72	**Kopet** ilun jarri 'to put your forehead dark' Idiom.: to get mad
7	*Pulgak* **hil** 'to kill the fleas' Idiom.: to hit someone	36.36		55.54	**Arkakusoak** hil 'to kill the fleas' Idiom.: to hit someone
8	*Piel* **gorritan egon** 'to have the skin red' Idiom.: to be naked	68.18	*	100	**Larru** gorritan egon 'to have the skin red' Idiom.: to be naked
	OVERALL	62%		81%	OVERALL

3.6. Individual Variation

Since some of the participants did not know the meaning of some Basque idioms, we decided to look at each idiom and participant individually and remove the responses of the participants that did not know the meaning of the monolingual idiom. These are presented in Tables 5 and 6. Table 5 presents the number and percentage of participants who chose the idiomatic option for the monolingual items. We also include the individual results, item by item in Appendix F.

Table 5: Percentage of participants who choose the idiomatic meaning for the monolingual idioms.

#	Item	Participants	%	Overall
	MONOLINGUAL			
1	*Írsele la olla* 'to lose his pot' Idiom.: to get mad	22 out of 22	100.00	
2	*Estar de mala leche* 'to be of bad milk' Idiom.: to be in a bad humor	22 out of 22	100.00	
3	*Dar en el clavo* 'to give on the nail' Idiom.: to nail something	22 out of 22	100.00	100%
4	*Importar un pimiento* 'to care a pepper' Idiom.: not to care	22 out of 22	100.00	
5	**Adarra jo** 'to play the horn' Idiom.: to make fun	22 out of 22	100.00	
6	**Kopet ilun jarri** 'to put your forehead dark' Idiom.: to get mad	16 out of 22	72.72	
7	**Arkakusoak hil** 'to kill the fleas' Idiom.: to hit someone	12 out of 22	54.54	81%
8	**Larru gorritan egon** 'to have the skin red' Idiom.: to be naked	22 out of 22	100.00	

Table 6 presents the number and percentage of participants who chose the idiomatic option for the bilingual items, out of those who chose the idiomatic option in the monolingual version.

Tables 5 and 6 show that participants regarded all Spanish idioms as expected: participants assigned the idiomatic meaning to almost all the code-switched idioms and to all the monolingual counterparts. For the Basque idioms, after removing the responses of the participants who did not know the meaning of the monolingual items, the proportion of the idiomatic interpretation for the bilingual expressions increased from 62% (Table 4) to 70.83% (Table 6).

Table 6: Percentage of participants who choose the idiomatic meaning for the code-switched idioms.

#	**CODE-SWITCHED** Item	Participants	%	Overall
1	*Írsele la* **lapiko** 'to lose his pot' Idiom.: to get mad	20 out of 22	90.90	
2	*Estar de mala* **esne** 'to be of bad milk' Idiom.: to be in a bad humor	22 out of 22	100.00	
3	*Dar en el* **iltze** 'to give on the nail' Idiom.: to nail something	22 out of 22	100.00	97.72%
4	*Importar un* **piper** 'to care a pepper' Idiom.: not to care	22 out of 22	100.00	
5	*Cuernoa* **jo** 'to play the horn' Idiom.: to make fun	16 out of 22	72.72	
6	*Pulgak* **hil** 'to kill the fleas' Idiom.: to hit someone	13 out of 16	81.25	
7	*Piel* **gorritan egon** 'to have the skin red' Idiom.: to be naked	7 out of 12	58.33	70.83%
8	*Frente* **ilun jarri** 'to put your forehead dark' Idiom.: to get mad	15 out of 22	68.18	

3.7. Interim Discussion

To sum up the results presented in this section, we conclude the following: regarding the Spanish items, as there is no significant difference between the Spanish monolingual and the Spanish code-switched items (Figure 2, columns 1-vs-2), we can conclude that participants interpret these code-switched items as idiomatic. In other words, participants interpret something like '*está de mala* esne' with the same idiomatic interpretation as '*está de mala leche*' that is, 'she/he is angry'. This result confirms our hypothesis.

The Basque items require more discussion. First, we notice that there is always a cline in the acceptability of code-switched idioms vis-à-vis their monolingual counterparts, a cline that becomes significant for items 5 and 8. The general cline can be interpreted as follows. In an independent project (Pouw et al, in progress) of Dutch/Papiamentu code-switching, we have found out that bilingual speakers who readily accept code-switching of a noun in post-verbal position tend to reject a similar code-switching at the beginning of a sentence. We surmise that speakers are more likely to accept (insertional) code-switching after the sentence structure has been put together but are less so when the sentence has not been fully defined yet. An inspection of the examples in Table 6 reveals that the code-switched Spanish nouns are all at the beginning of the

sentence. This probably has led our participants to some difficulty accepting the sentence itself and therefore some difficulty extracting the idiomatic meaning.

Another important methodological consideration is the translation of words such as *larru* which means 'skin', but it can also be translated as 'leather' and thus, this can also interfere in their interpretation. Turning back to our original hypothesis, it is very likely the case that *larru* and *piel* are not exponents of the same root but rather two independent roots. This accounts for item 8.

We also analyzed each participant individually to see if there was any variation that could explain our results. We did not find any clear pattern.

3.8. The fake idioms

The Spanish idiom *ni pies ni cabeza* literally 'neither feet nor head', idiomatically meaning 'incoherent' was altered for our survey as *ni pies ni dedos* 'neither feet nor fingers' in unilingual Spanish and *ni pies ni atzamar* in bilingual Spanish-Basque. We found that 100% of our participants identified the unilingual idiom correctly and 33% misinterpreted the fake unilingual idiom. Therefore, anticipatory effects did occur in this task. We also found that almost 49% misinterpreted the code-switched idiom, a percentage that is significantly lower than the lowest real Spanish code-switched idiom (at 90%). We conclude that anticipatory effects exist in this task and code-switching enhances the anticipatory effects. However, there is a gap between the effect that can be attributed to anticipation (as shown with fake idioms) and the acceptance of real idioms. This gap is what can be accounted for with our integrated hypothesis.

For other fake idioms, we encountered the problem discussed above that many of our participants did not know the Basque idioms and therefore we could not reach any conclusions.

4. English/Spanish Survey

4.1. Participants

30 highly proficient English-Spanish early bilingual speakers (25 female), all of them of Mexican ancestry, participated. 28 lived in the United States at the time of testing, while the other two lived in Mexico. The average age was 29.93 (*SD* = 5.69; *Range* = [19-42]). There were 22 college students and 8 college graduates. Their overall self-reported proficiency was over 4.00 (SD = 1.20) in Spanish and over 4.47 (SD = 0.68) in English.

4.2. Task, Stimuli, and Procedure

The overall design of the Spanish/English Forced Choice Task was the same as for the Basque/Spanish survey (Appendix F presents the background questionnaire). Table 8 shows the idiomatic expressions used in the Spanish/English survey.

Table 8: Idioms selection for English (#1-4) and Spanish (#5-8) Survey.

#	Idiomatic expressions	Idiomatic meaning
1	It was raining cats and dogs	It was raining a lot
2	Peter lost his marbles	Peter got furious
3	Straight from the horse's mouth	it came from a reliable source of information
4	We paid an arm and a leg	we paid too much
5	*Hablando del rey de Roma* Speaking of-the king of Rome	to talk about someone right before the person arrives
6	*Sin pelos en la lengua* Without hairs on the tongue	to talk clearly, with no regards for political correctness
7	*Echar la casa por la ventana* Trow.inf the house by/through the window	to spend too much money
8	*La gota que derramó el vaso* The drop that spill.past the glass	the event that caused a chain reaction

Table 9 presents the eight idiomatic expressions included in the task as well as the replacements we made (see Appendix G for the stimuli, Appendix H for the distractors and Appendix I for the background questionnaire.)

Table 9: Idioms selection for English/Spanish Survey.

#	Idiomatic expressions	Idiomatic meaning	Noun replaced
1	Raining *gatos y perros* cats and dogs	raining a lot	Spanish *perros y gatos* = cats and dogs
2	Peter lost his *canicas* marbles	to get furious	Spanish *canicas* = marbles
3	Straight from *la boca del caballo* the mouth of-the horse	coming from a key source of information	Spanish *la boca del caballo* = the horse's mouth
4	We paid *un brazo y una pierna* an arm and a leg	We paid too much	Spanish *un brazo y una pierna* = an arm and a leg
5	*Hablando del* king of Rome Speaking of-the	to talk about someone right before the person arrives	English 'king of Rome' = Original Spanish '*rey de Roma*'
6	*Sin pelos en la⁹* tongue Without hairs on the	to talk clearly, with no regards for political correctness	English 'tongue' = Original Spanish '*lengua*'
7	*Echar la casa* by the window Trow.inf the house	to spend too much money	English 'by the window' = Original Spanish '*por la ventana*'
8	*La gota que derramó* the glass The drop that spill.3rd.past	the event that caused a chain reaction	English 'the glass' = Original Spanish '*el vaso*'

4.3. Group Results

We found no correlation to any of the background factors controlled for in the survey: gender, living location, age, level of education, or attitudes towards code-switching (see Appendix E).

Figure 4 shows the results for the Spanish and English monolingual idioms. As observed, participants assigned the idiomatic interpretation ~84% of the time for the Spanish idioms (n = 201 / 240), a 'none of the above' interpretation ~17% of the time (n = 39 / 240), and one individual instance of the 'Other' interpretation (n = 1 / 240). When the idioms were in English, participants assigned the idiomatic interpretation ~87% of the time (n = 206 / 240), a 'none of the above' interpretation ~12% (n = 32 / 240), and two instances of the 'Other' interpretation (n = 2 / 240). The 'Literal' option was never chosen.

---

⁹ We maintained the feminine article *la* for *Sin pelos en la tongue* as a direct translation of *Sin pelos en la lengua*, maintaining the feminine gender in the code-switched idioms as in the original, monolingual idiom.
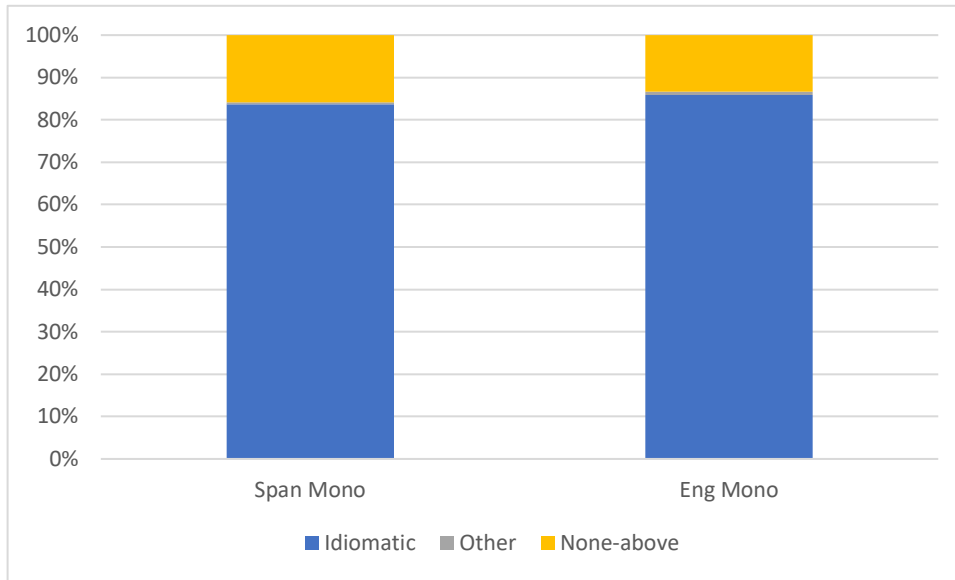
Figure 4: Descriptive results for the monolingual items in the Spanish-English language pair

Figure 5 shows a comparison between monolingual idioms and idioms that include a code-switch. Spanish monolingual idioms (column 1) receive an idiomatic interpretation ~84% of the time (n = 201 / 240) while Spanish idioms with a switch to English (column 2) receive an idiomatic interpretation ~68% of the time (n = 82 / 120), and the Fisher Exact Test calculations show that the difference among the two conditions is significant ($\chi2 < 0.01$, p < .05). The same figure also shows that English monolingual idioms (column 3) receive an idiomatic interpretation ~85% of the time (n = 206 / 240) while English idioms with a switch to Spanish (column 4) receive an idiomatic interpretation ~78% of the time (n = 92 / 118), and the Fisher Exact Test calculations show that the difference among the two conditions is not significant ($\chi2 = 0.07$, p > .05).
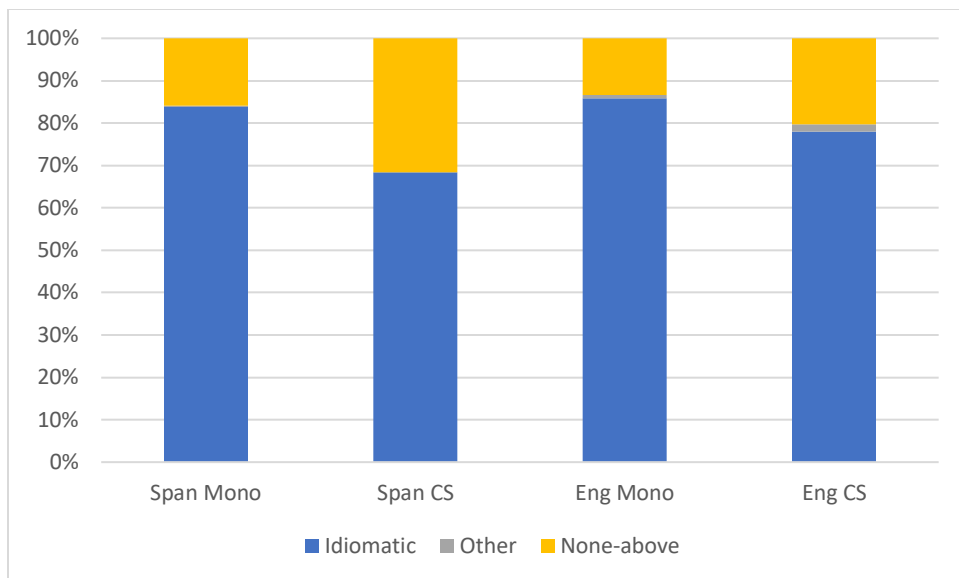
Figure 5: Descriptive comparison of monolingual vs. bilingual idioms in the Spanish-English language pair.

4.4. By-Item, Pairwise Comparisons

Table 10 shows the pairwise comparisons between the monolingual and the bilingual version of each of the surveyal items:

Table 10: Pairwise comparisons by-item. Percentages represent the proportion of idiomatic interpretations.

#	BILINGUAL	%	Sig.?	%	MONOLINGUAL
1	**Raining *gatos y perros*** 'Raining cats and dogs' Idiom.: Raining a lot	96.428		100	**Raining cats and dogs** Idiom.: Raining a lot
2	**Peter lost his *canicas*** 'Peter lost his marbles' Idiom.: To get furious	50		73.333	**Peter lost his marbles** Idiom.: To get furious
3	**Straight from *la boca del caballo*** 'Straight from the horse's mouth' Idiom.: Coming from a key source of information	73.333		90	**Straight from the horse's mouth** Idiom.: Coming from a key source of information
4	**We paid *un brazo y una pierna*** 'We paid an arm and a leg' Idiom.: It was overpriced	93.333		96.666	**We paid an arm and a leg** Idiom.: It was overpriced
	OVERALL				**OVERALL**
5	***Hablando del* king of Rome** 'Speaking of the king of Rome' Idiom.: To talk about someone right before the person arrives	86.666		86.666	***Hablando del rey de Roma*** 'Speaking of the king of Rome' Idiom.: To talk about someone right before the person arrives
6	***Sin pelos en la* tongue** 'Without hairs on the tongue' Idiom.: To talk clearly, with no regards for political correctness	70		86.666	***Sin pelos en la lengua*** 'Without hairs on the tongue' Idiom.: To talk clearly, with no regards for political correctness
7	***Echar la casa* by the window** 'To throw the house by/throw the window' Idiom.: To spend too much money	53.333	*	86.666	***Echar la casa por la ventana*** 'To throw the house by/throw the window' Idiom.: To spend too much money
8	***La gota que derramó* the glass** 'The drop that spilled the glass' Idiom.: The event that caused a chain reaction	63.333		73.333	***La gota que derramó el vaso*** 'The drop that spilled the glass' Idiom.: The event that caused a chain reaction
	OVERALL				**OVERALL**

Table 10 reveals that most of the participants who accepted an idiomatic meaning for a monolingual phrase also accepted the same idiomatic meaning for the code-switched version. We discuss the outlier item 7 in section 3.6.

4.5. Individual Variation

Table 11 presents the number and percentage of participants who chose the idiomatic option for the monolingual items. We can see that the idioms in rows 2 and 8 were known by only three quarters of our participants; we return to this point in section 3.6.

Table 11: Percentage of participants who chose the idiomatic meaning for the monolingual idioms.

#	**MONOLINGUAL** Item	**Participants**	**%**	**Overall**
1	**Raining *gatos y perros*** 'Raining cats and dogs' Idiom.: Raining a lot	30 out of 30	100.00	
2	**Peter lost his *canicas*** 'Peter lost his marbles' Idiom.: To get furious	22 out of 30	73.33	90.00%
3	**Straight from *la boca del caballo*** 'Straight from the horse's mouth' Idiom.: Coming from a key source of information	27 out of 30	90.00	
4	**We paid *un brazo y una pierna*** 'We paid an arm and a leg' Idiom.: It was overpriced	29 out of 30	96.67	
5	*Hablando del* **king of Rome** 'Speaking of the king of Rome' Idiom.: To talk about someone right before the person arrives	26 out of 30	86.67	
6	*Sin pelos en la* **tongue** 'Without hairs on the tongue' Idiom.: To talk clearly, with no regards for political correctness	26 out of 30	86.67	83.33%
7	*Echar la casa* **by the window** 'To throw the house by/throw the window' Idiom.: To spend too much money	26 out of 30	86.67	
8	*La gota que derramó* **the glass** 'The drop that spilled the glass' Idiom.: The event that caused a chain reaction	22 out of 30	73.33	

In Table 12 we show how many participants who assigned the idiomatic meaning to the monolingual idiom did so again to the bilingual idiom. In sum, of those participants who did know a given monolingual idiom, they also gave the code-switched equivalent item an idiomatic interpretation over 75% of the time. Exceptions were items #2 and #7, which are discussed in section 3.6.

Table 12: Percentage of participants who chose the idiomatic meaning for the code-switched idioms.

	CODE-SWITCHED			
#	**Item**	**Participants**	**%**	**Overall**
1	**Raining *gatos y perros*** 'Raining cats and dogs' Idiom.: Raining a lot	27 out of 30	90.00	
2	**Peter lost his *canicas*** 'Peter lost his marbles' Idiom.: To get furious	13 out of 22	59.09	83.18%
3	**Straight from *la boca del caballo*** 'Straight from the horse's mouth' Idiom.: Coming from a key source of information	21 out of 27	77.78	
4	**We paid *un brazo y una pierna*** 'We paid an arm and a leg' Idiom.: It was overpriced	28 out of 29	96.55	
5	*Hablando del* **king of Rome** 'Speaking of the king of Rome' Idiom.: To talk about someone right before the person arrives	24 out of 26	92.31	
6	*Sin pelos en la* **tongue** 'Without hairs on the tongue' Idiom.: To talk clearly, with no regards for political correctness	20 out of 26	76.92	78.00%
7	*Echar la casa* **by the window** 'To throw the house by/throw the window' Idiom.: To spend too much money	16 out of 26	61.54	
8	*La gota que derramó* **the glass** 'The drop that spilled the glass' Idiom.: The event that caused a chain reaction	18 out of 22	81.82	

Appendix J presents the individual results, item by item.

4.6. Interim Discussion

The lack of a significant difference between monolingual and code-switched items in Figure 5, columns 3-vs-4, shows that our participants give English expressions an idiomatic meaning, despite the inclusion of a switch to Spanish. This aligns with the Integrated Hypothesis. Additionally, there is a significant difference between monolingual and code-switched items in Figure 5, columns 1-vs-2 – those columns that refer to the Spanish idioms. This significant difference requires some discussion because it may lead one to the erroneous conclusion that, while a monolingual Spanish item such as *Hablando del rey de Roma* 'Speaking of the king of Rome' receives the idiomatic meaning 'We were just talking about him,' the code-switched item

Hablando del king of Rome does not receive that idiomatic interpretation. This impression is dispelled when looking at the pairwise comparisons in Table 10. Table 10 shows that all items received a high percentage of idiomatic interpretation, with no significant differences between the monolingual and the code-switched versions of the same item. We argue that this pattern shows that code-switching does not delete the idiomatic meaning of an idiomatic expression. The only exception to this pattern is item #7. We come back to this item below.

Table 11 shows that the participants generally have the proposed items in their idiom repertoire. Similarly, Table 12 shows that the idiomatic meaning is maintained despite the appearance of code-switching. There are, however, two aspects of the individual results that are worth discussing in more detail:

First, participants present higher overall percentages of an idiomatic interpretation for English items (both in monolingual English and when there is a switch to Spanish) than for Spanish items (also in both monolingual Spanish and when there is a switch to English): We claim that this is the result of a general slight English dominance of the group of participants (M = -0.32 [-0.69 – 0.05]). This phenomenon is parallel to what we saw in the Basque-Spanish survey, where Spanish dominance led to higher knowledge of idiomatic expressions.

Second, the code-switched item #2 *Peter lost his canicas*[10] 'Peter lost his marbles' (meaning 'to get furious') and #7 *Echar la casa by the window* 'To throw the house by the window' (meaning 'to spend too much money') obtained a much lower percentage of idiomatic interpretation (~60%) in comparison to the rest of code-switched items (~76% / ~92%). Here we would like to venture some reasons for this discrepancy. In item #2 *Peter lost his canicas*, the translation of 'marbles' as 'canicas' was a bad choice, since our bilingual speakers are probably more familiar with other vocabulary items for 'marbles' such as *bolita, boliche, chibola,* etc. Alternatively, given their English dominance, they may not have a "Spanish" version of the word other than 'marbles'. As for item #7 *echar la casa by the window*, a better English translation of *por la ventana* in *echar la casa por la ventana* might have been *through the window* or *out the window* instead of *by the window* ('To throw the house through the window,' meaning 'To spend too much money'). Therefore, we argue that the original significant difference was only the result of our own mistake in translating the item 'Echar la casa *by the window*' in its code-switched version. Note, however, that both *Peter lost his canicas* and *Echar la casa through the window* still receive an idiomatic interpretation most of the time. When all is said and done, the resulting patterns align with the Integrated Hypothesis.

4.7. Fake idioms

Unilingual English fake idioms were misinterpreted as real idioms 23% of the time. Bilingual English fake idioms were also interpreted as real idioms 26% of the time and the difference is not significant (Chi-square = 0.1642, p-value >.05). With real code-switched idioms, the lowest acceptance we have is 59%.

Unilingual Spanish idioms are misinterpreted as real idioms 26%, a percentage that increases substantially when the idiom is code-switched: 53%. The difference between fake bilingual idioms and real idioms is significant (Chi-square = 8.8889, p-value <.05).

---

[10] The code-switched item 'Peter lost his *canicas*' did NOT receive a significantly lower rate of idiomatic meaning in comparison to its monolingual counterpart 'Peter lost his marbles.' However, as shown in Tables 4 and 5, this idiom received a low rate of idiomatic meaning in comparison to the other English idioms with a Spanish switch, which is why we prefer to clarify what may possibly have happened.

The English-Spanish survey confirms that there are anticipatory effects, and these are more noticeable in code-switching (even if only slightly so). However, there is a significant gap between the effects that can be accounted for by means of anticipation (as revealed by the fake idioms) and the results that we get with true idioms. This gap is what can be accounted for with the integrated hypothesis.

5. Conclusions.
    The main goal of this paper is to test the following hypothesis:

(20)   The choice of a List 2 item does not alter the interpretation of an idiom. (=17)

This hypothesis emerges from a framework that has the following main properties:
(i) Integrationism: the I-language of a multilingual person constitutes a unique competence system. Among other consequences, this assumption entails that the lexicon of the bilingual is fully integrated.
(ii) Distributed Morphology: The traditional lexicon is distributed in three lists. List 1 consists of numerical indices and functional features and is the input to the syntactic system. List 2 is a set of rules of exponence for syntactic terminals and List 3 consists of slices of conceptual structure.
    From this framework the possibility arises that many words in a bilingual lexicon that appear to be duplicated – say, *planeta* and 'planet' – are in fact alternative exponents for the same item in List 1 and List 3. If this is the case, then rigid idioms should maintain the idiomatic meaning even if one or more words from the idiom are code-switched. This is because code-switching would only involve choosing the alternative exponent for one List 1 and List 3 item.
    We explored our hypothesis (20) by means of one pilot and two surveys carried out in three different bilingual communities. By and large, (20) was confirmed to the extent that in an overwhelming majority of cases, bilingual speakers were comfortable choosing an idiomatic meaning in a code-switched phrase. There were some instances in which bilingual speakers did not accept the code-switched idiom, but we were able to account for most of them. We also found that acceptance of the code-switched idiom was greater when the idiom was in the language in which our participants felt more proficient, a datum for which we do not have an immediate account.
    We would like to emphasize that our results are incompatible with a view of the lexicon based on the Lexicalist hypothesis or a Separationist view of bilingualism. The Lexicalist hypothesis posits that idioms are stored in the lexicon as individual units. If that is the case, then replacing one of the words of the idiom with another word should derail the idiomatic meaning. Since this is not the case with the bilingual idioms explored in this paper, Lexicalism does not receive support from our study. However, DM, which hypothesizes three separate lists in lieu of a traditional lexicon, can provide a framework that accounts for bilingual idioms.
    Additionally, our results do not support the received idea that a bilingual person is in possession of two lexicons (and, *a fortiori*, of two separate systems of linguistic competence). If there were two lexicons, then a rigid idiom could not accept a word from a different lexicon, as pointed out by Keller. Let us elaborate on this. One could consider that the results could be accounted for with two separate lexicons if the subjects, upon encountering the switch, would back-track, find the equivalent word in the first lexicon and then interpret the idiom on the basis of an understanding that Word 1 in Lexicon 1 = Word 2 in Lexicon 2. However, an account based on two separate lexicons that have such expansive ad hoc interconnections between them runs the

risk of becoming a messy version of the Integrated Hypothesis – a simple application of Occam's Razor should lead to prefer the Integrated Hypothesis. Moreover, the two-lexicon system expanded to include connectivity rules should raise the question of why such rules should exist. Instead, the integrated system leads to *expect* that a rigid idiom may accept an equivalent word in the "other" language because this possibility is organically integrated in the hypothesis.

Luis López
School of Cultural Studies, Literatures and Linguistics
University of Illinois Chicago
601 South Morgan MC 315
Chicago, IL 60607
USA
luislope@uic.edu

Irati de Nicolás
Department of Romance Languages and Literatures
University of Chicago
1050 E. 59th Street
Chicago, IL 60637
idenicolassaiz@uchicago.ed

Rodi Laanen
Leiden Institute of Advanced Computer Science
Leiden University
Niels Bohrweg 1
2333 CA Leiden
The Netherlands
r.laanen@umail.leidenuniv.nl

Charlotte Pouw
Institute for Logic, Language and Information
University of Amsterdam
Carolina MacGillavrylaan 3158
1098 XK Amsterdam
The Netherlands
c.m.pouw@uva.nl

José Sequeros-Valle
Foreign Languages and Literature
University of Nebraska-Omaha
6001 Dodge Street ASH-O
Omaha, NE 68182
USA
jsequeros-valle@unomaha.edu

M. Carmen Parafita Couto

Leiden University Center for Linguistics
Reuvensplaats 3-4, 110
2311 BE Leiden
The Netherlands
&
Language Variation and Textual Categorisation
Universidade de Vigo
Spain
m.parafita.couto@hum.leidenunv.nl & mariacarmen.parafita@uvigo.gal

Ethics and consent
The UIC office for the protection of human subjects determined that the research reported in this paper is exempted from the federal regulations for the protection of human subjects. The specific exemption category under 45 CFR 46.101(b) is: 2. Every participant signed a consent form before beginning the survey.

References

Alexiadou, A., and T. Lohndal. 2018. Units of language mixing: A cross-linguistic perspective. *Frontiers in Psychology* Volume 19, Article 1719. https://doi.org/10.3389/fpsyg.2018.01719

Alexiadou, A., T. Lohndal, T. Åfarli and M. B. Grimstad. 2015. Language mixing: A Distributed Morphology approach. In Thuy Bui and Denis Ozildiz (eds) *Proceedings of NELS*. Amherst, Mass: GLSA Publications, pp 25-38.

Audring, J. 2022. Advances in Morphological Theory: Construction Morphology and Relational Morphology. *Annual Review of Linguistics* 8: 39-58. https://doi.org/10.1146/annurev-linguistics-031120-115118

Bruening, B. 2020. Idioms, Collocations and Structure. *Natural Language and Linguistic Theory* 38: 365-424. https://doi.org/10.1007/s11049-019-09451-0

Espinal, M. T. and J. Mateu. 2019. Idioms and Phraseology. In *Oxford Research Encyclopedia, Linguistics.* https://doi.org/10.1093/acrefore/9780199384655.013.51

Davies, M. 2006. *A Frequency Dictionary of Spanish*. New York, NY & London: Routledge. https://doi.org/10.4324/9780203415009

Edele, A., Seuring, J., Kristen, C. and Stanat, P. 2015. Why bother with testing? The validity of immigrants' self-assessed language proficiency. *Social Science Research*, 52, 99-123. https://doi.org/10.1016/j.ssresearch.2014.12.017

Fernández, E. and H. Smith Cairns (eds.). 2017. *The Handbook of Psycholinguistics*. Boston: Wiley-Blackwell. https://doi.org/10.1002/9781118829516

Goldberg, A. 2006. *Constructions at Work*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199268511.001.0001

Goldrick, M., M. Putnam and L. Schwartz. 2016. Coactivation in bilingual grammars: A computational account of code mixing. *Bilingualism: Language and Cognition* 19: 857-876. https://doi.org/10.1017/S1366728915000802

González-Vilbazo, K. and L. López. 2012. Little v and parametric variation. Natural Language & Linguistic Theory, 30:1, 33-77. https://doi.org/10.1007/s11049-011-9141-5

González-Vilbazo, K., L. Bartlett, S. Downey-Giménez, S. Ebert, J. Heil, B. Hoot and S. Ramos. 2013. Methodological considerations in code-switching research. *Studies in Hispanic and Lusophone Linguistics* 6: 119-138. https://doi.org/10.1515/shll-2013-1143

Grimstad, M. B., T. Lohndal and T. Åfarli. 2014. Language mixing and exoskeletal theory: A case study of word-internal mixing in American Norwegian. *Nordlyd* 41: 213-237, special issue on Features edited by M. Krämer, S. Ronai and P. Svenonius. University of Tromsø – The Arctic University of Norway. https://doi.org/10.7557/12.3413

Halle, M. and A. Marantz. 1993. Distributed morphology and the Pieces of inflection. In K. Hale and S. J. Keyser (eds.) The View from Building 20. Cambridge, MA: MIT Press, pp 111-176.

Harley, H. 2014. On the identity of roots. *Theoretical Linguistics* 40:3/4, 225-276. https://doi.org/10.1515/tl-2014-0010

Harley, H. and R. Noyer. 2000. Formal versus encyclopedic properties of Vocabulary: Evidence from nominalizations. In Bert Peeters (ed) *The Lexicon-Encyclopedia Interface*. Elsevier Press, pp 349-374. https://doi.org/10.1163/9780585474465_014

Hoffman, T. and G. Trousdale. 2013. The Oxford Handbook of Construction Grammar. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195396683.001.0001

Hsin, L. and G. Legendre. 2019. Strong Integration in Bilingual Grammar, Formalized. *Linguistic Approaches to Bilingualism* 9(3): 427-467. https://doi.org/10.1075/lab.17040.hsi

Jackendoff, R. 1997. The Architecture of the Language Faculty. MIT Press. https://doi.org/10.1163/9789004373167_003

Keller, M. 2020. Compositionality: Evidence from Code-Switching. In E. Piirainen, N. Filatkina, S. Stumpf and C. Pfeiffer (eds) *Formulaic Language and New Data*. Berlin: De Gruyter, pp 197-224. https://doi.org/10.1515/9783110669824-009

Kroll, J., P. Dussias, C. Bogulski and J. Valdés Kroff. 2012. Juggling Two Languages in One Mind: What Bilinguals Tell Us About Language Processing and its Consequences for Cognition. *Psychology of Learning and Motivation* 56: 229-262. https://doi.org/10.1016/B978-0-12-394393-4.00007-8

Kroll, J. F., P. Dussias, K. Bice, and L. Perrotti. 2015. Bilingualism, Mind, and Brain. *Annual Review of Linguistics* 1: 377-394. https://doi.org/10.1146/annurev-linguist-030514-124937

Levelt, W.J.M. 1989. *Speaking: From Intention to Articulation.* Cambridge, Mass.: Massachussetts Institute of Technology Press.

Levelt, W.J.M. & A. Meyer .2000. Word for word: Multiple lexical access in speech production. European Journal of Cognitive Psychology 12: 433–452. https://doi.org/10.1080/095414400750050178

Li, M., and X. Zhang. 2021. A meta-analysis of self-assessment and language performance in language testing and assessment. *Language Testing*, 38: 189-218. https://doi.org/10.1177/0265532220932481

López, L. 2018. Questions on data and input to GEN. *Bilingualism, Language and Cognition* 19: 889-890. https://doi.org/10.1017/S1366728916000079

López, L. 2020. Bilingual Grammar. Toward an Integrated Model. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108756181

MacSwan, J. 2000. The architecture of the bilingual language faculty: evidence from intrasentential code switching. *Bilingualism, Language and Cognition* 3:1, 37-54. https://doi.org/10.1017/S1366728900000122

Marantz, A. 1997. No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. In Alexis Dimitradis et al. (eds.) *Proceedings of the 21st Penn Linguistics Colloquium.* Philadelphia: University of Pennsylvania. Upenn Working Papers in Linguistics, pp 221-225.

McGinnis, M.J. 2002. On the Systematic Aspect of Idioms. *Linguistic Inquiry* 32: 665-672. https://doi.org/10.1162/ling.2002.33.4.665

Meisel, J. 2019. *Bilingual Children. A Guide for Parents*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781316850329

Myers-Scotton, C. 1993. *Duelling Languages*. Oxford: Clarendon Press.

Nunberg, G., I. Sag and T. Wasow. 1994. Idioms. *Language* 70: 491-532. https://doi.org/10.2307/416483

Papafragou, A, J.C. Trueswell and L.R. Gleitman (eds.) 2022. *The Oxford Handbook of the Mental Lexicon.* Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198845003.001.0001

Pouw, C., I. De Nicolás, R. Laanen, L. López, M.C. Parafita Couto, J. Sequeros,. In progress. *The role of INFL in code-switching*.

Riksem, B., M.B. Grimstad, T. Lohndal and T. Åfarli. 2019. Language mixing within verbs and nouns in American Norwegian. *Journal of Comparative Germanic Syntax*: 189-209. https://doi.org/10.1007/s10828-019-09109-6

Senaldi, M., Wei, J., Gullifer, J. W., & Titone, D. 2022. Scratching your tête over language-switched idioms: Evidence from eye-movement measures of reading. Memory & cognition, 50(6), 1230–1256. https://doi.org/10.3758/s13421-022-01334-x

Spenser, A. 1991. *Morphological Theory*. Cambridge, MA: Blackwell.

Titone, D., G. Columbus, V. Whitford, J. Mercier and M. Libben. 2015. Contrasting Bilingual and Monolingual Idiom Processing. In R.R. Heredia and A.B. Cie´slicka (eds) *Bilingual Figurative Language Processing* Cambridge University Press, pp 171-2-7. https://doi.org/10.1017/CBO9781139342100.011

Titone, D., K. Lovseth, K. Kasparian, M. Tiv, R. Jamieson. 2019. Are figurative interpretations of idioms directly retrieved, compositionally built, or both?. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale* 73: 216-230. https://doi.org/10.1037/cep0000175

van Osch, B. 2019. Vulnerability in heritage speakers of Spanish in The Netherlands: An interplay between language-internal and language/external factors. Doctoral dissertation, University of Amsterdam, The Netherlands.

Volterra, V. and T. Taeschner. 1978. The acquisition and development of language by bilingual children. *Journal of Child Language* 5: 311-326. https://doi.org/10.1017/S0305000900007492

Zeng, T., Branigan, H. P., & Pickering, M. J. 2020. Do bilinguals represent between-language relationships beyond the word level in their lexicon? *Journal of Neurolinguistics* 55: 100892. https://doi.org/10.1016/j.jneuroling.2020.100892