



**HAL**  
open science

# A phraseological and phonological analysis of don't: A stay abroad perspective

Amanda Edmonds, Elisa Sneed German, Pascale Leclercq

► **To cite this version:**

Amanda Edmonds, Elisa Sneed German, Pascale Leclercq. A phraseological and phonological analysis of don't: A stay abroad perspective. *Lexis. Journal in English Lexicology*, 2021, 18. hal-03494512

**HAL Id: hal-03494512**

**<https://hal.science/hal-03494512v1>**

Submitted on 26 Jan 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License



**Lexis**

Journal in English Lexicology

**18 | 2021**

**Lexical learning and teaching**

---

## A phraseological and phonological analysis of *don't*: A stay abroad perspective

Amanda Edmonds, Elisa Sneed German and Pascale Leclercq

---



### Electronic version

URL: <https://journals.openedition.org/lexis/5718>

DOI: 10.4000/lexis.5718

ISSN: 1951-6215

### Publisher

Université Jean Moulin - Lyon 3

### Electronic reference

Amanda Edmonds, Elisa Sneed German and Pascale Leclercq, "A phraseological and phonological analysis of *don't*: A stay abroad perspective", *Lexis* [Online], 18 | 2021, Online since 18 December 2021, connection on 21 December 2021. URL: <http://journals.openedition.org/lexis/5718> ; DOI: <https://doi.org/10.4000/lexis.5718>

---

This text was automatically generated on 21 December 2021.



Lexis is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

---

# A phraseological and phonological analysis of *don't*: A stay abroad perspective

Amanda Edmonds, Elisa Sneed German and Pascale Leclercq

---

## Introduction

- <sup>1</sup> For language teachers and learners alike, lexical learning is often defined as the learning of a new word and its definition. Much second language (L2) research on lexical learning has thus sought to quantify the number of individual words for which a learner knows a definition, thereby providing an indication of the size of the learner's lexicon (e.g., Batista & Horst [2016]; Meara & Milton [2003]). Although this constitutes an important aspect of lexical learning, the learning of new words involves more than mastering the connection between a form and its meaning. According to Nation [2001: 27], for example, word knowledge involves receptive and productive knowledge with regards to a word's form, its meaning and its use. Recent research has begun to pay more attention to the multifaceted nature of lexical learning (e.g., González-Fernández & Schmitt [2020]) and specifically to aspects beyond the form-meaning connection, such as the acquisition of polysemy (Crossley, Salsbury & McNamara [2010]), the development of lexical sophistication (Kyle & Crossley [2015]), and phraseological development (Edmonds & Gudmestad [2021]; Siyanova-Chanturia & Spina [2020]). The current article contributes to this body of research with an analysis that focuses on phraseological development in L2 English over the course of an academic year spent in a target-language environment. In particular, we focus on how one form – *don't* – is used by five learners, identifying one specific phraseological use involving this word and exploring whether phonetic realization differs as a function of the phraseological versus non-phraseological status of the string. Given the longitudinal nature of our corpus, we also explore how the use and phonetic realization of strings involving *don't* may evolve over time, interpreting these results in light of the stay abroad experience. Data come from the PROLINGSA project,<sup>1</sup> for which five

francophone university-level English majors were followed over the course of 12 months, including an academic year spent in either Ireland or England through the Erasmus+ program. Each participant engaged in an oral semi-guided interview on five occasions, resulting in a corpus of approximately 12 hours of total talk and more than 68,800 words produced by the five L2 speakers.

## 1. Literature review

- 2 We begin by offering a concise overview of the field of phraseology and introducing the sub-type of phraseological unit that is central to the current analysis: conventional expressions<sup>2</sup> (Bardovi-Harlig [2009]). We then review research on phraseology in spoken language, with a focus on the phonetic and prosodic realization hypothesized to be associated with phraseological units. The final section addresses research on the development of phraseological competence during a stay abroad as seen through analyses focusing on conventional expressions.

### 1.1. Phraseology and phraseological units

- 3 The field of phraseology focuses on word combinations, covering a vast spectrum of lexical sequences, including collocations, phrasal verbs, idioms, lexical bundles, and conventional expressions. Although this area of research is characterized by a wide variety of definitions, identification criteria and terms, phraseological studies have in common the fact that they focus on lexical co-occurrence patterns that are generally unpredictable (e.g., *take a nap* but not *do a nap*) and that have become conventionalized for a given speech community. Granger & Paquot [2008] offer a clear overview of research on phraseology, and identify two approaches to studying phraseological units. On the one hand, phraseological (or traditional) approaches consider “phraseology as a continuum along which word combinations are situated, with the most opaque and fixed ones at one end and the most transparent and variable ones at the other” Granger & Paquot [2008: 28]. Thus, on one end of the spectrum, we find pure idioms (e.g., *by and large*), which are both syntactically non-compositional and semantically opaque and fixed in form, whereas certain collocations (e.g., *carry out/conduct research/a study/an experiment*) are situated on the opposite end, given that they are transparent and allow more variation. A large number of phraseological units lie between these two poles, including many conventional expressions (e.g., *How do you do?, I'm just looking*), which are relatively fixed in form but transparent in meaning and syntax. The identification of phraseological units in such approaches is top-down in nature in that the researcher relies on linguistic criteria (e.g., opacity, fixedness, phonological coherence, discursive function, etc.) to identify sequences. On the other hand, what Granger & Paquot call “distributional” approaches to phraseology observe language in use in order to subsequently identify recurring lexical units. In such research, phraseological units are thus frequently co-occurring word combinations (where “frequent” can reflect raw frequency or strength of association measures). Distributional approaches adopt a broader vision of phraseology than in traditional analyses, as the units identified as phraseological do not always correspond to predefined linguistic categories. This can be seen in the following recurrent two-word sequences identified in learner essays by Granger & Bestgen [2014: 235], none of which would be considered phraseological in a

traditional, top-down approach: *of the, in the, to the*. Regardless of the approach adopted, many L2 researchers interested in lexical development have called attention to the fact that the learning of phraseological patterns is often slow (Schmitt [2013]), and authors such as Ellis, Simpson-Vlach, Römer, O'Donnell & Wulff [2015] have highlighted the importance of massive contact with the target language (during a stay abroad, for example) for the development of phraseological competence.

- 4 In the current study, these two approaches to phraseology will be combined. A distributional approach will be used in the first part of our analysis, in which we seek to establish a complete picture of usage patterns involving *don't* in the PROLINGSA corpus. A phraseological approach guides the second part of our analysis, in which we use the criterion of discursive function to identify which instances of the string *I don't know* constitute phraseological uses. *I don't know* was the most frequent string involving *don't* in this corpus, and uses identified as phraseological are argued to be examples of conventional expressions (Bardovi-Harlig [2009]). Conventional expressions have a special position within the spectrum of phraseological units, insofar as these expressions – unlike idioms, collocations, phrasal verbs, etc. – are not always phraseological. In other words, strings identified as conventional expressions can constitute a phraseological (or conventional) way of saying something in certain situations but not in others, depending on the communicative function associated with the string. De Cock [1998] provides a clear example of a conventional expression in her discussion of the string *you know*. When this sequence is used to express its literal meaning, as in example 1(a) taken from our corpus, it is fulfilling a referential function and is not phraseological. However, this same string can also be used as a discourse marker, defined by Pichler [2009: 561] as “linguistic items or expressions that have little or no referential meaning but serve multiple pragmatic functions in the interpersonal and textual domains of discourse.” When used as a discourse marker, *you know* is a conventional expression and, thus, phraseological (see 1b).

- (1) a. I don't know if **you know** her (Y, interview 1)  
 b. as much as French so but **you know** I I I feel more comfortable to express my ideas in French so (Y, interview 1)

- 5 In the case of our analysis of *don't*, we follow researchers such as Aijmer [2009], Baumgarten & House [2010], and Pichler [2009] in considering that *I don't know* can be used phraseologically as a conventional expression (often acting as a discourse marker), just as it can be used non-phraseologically to communicate the literal lack of knowledge about something.

## 1.2. Phraseology and spoken language

- 6 Although phraseological units are part of any language system, some scholars suggest that such units are more frequent in (especially spontaneous) spoken than in written language (e.g., De Cock [2000]; Ellis *et al.* [2015]). The reasons for this difference lie with the fact that the use of phraseological units is hypothesized to “ease processing problems”, because using them “buy[s]’ processing time while other computation proceeds, enabling us to plan ahead for the content of what we are going to say, as well as the linguistic form” (Skehan [1998: 40]). If Skehan – but also Pawley & Syder [1983: 192], Wood [2002: 7] and others – consider that the use of phraseological units “buys processing time”, it is because they consider phraseological units to benefit from

facilitated processing, either because such strings may be stored whole in the mental lexicon (see Wray [2002]), or because the activation of such strings, by dint of being used frequently, has become automated, leading to what Lin [2010] refers to as holistic processing (see also Bybee [2002]). The ease of processing documented for phraseological units comes from experimental studies (which generally show faster response times for the processing of phraseological versus non-phraseological strings, see Conklin & Schmitt [2012], but also from studies of the phonological characteristics of phraseological units. Indeed, it has been suggested that phraseological units have specific phonetic and prosodic characteristics because of their holistic or automatic nature. In particular, many researchers expect that phraseological units are less likely to be interrupted by a pause or other dysfluencies (Raupach [1984]), that phraseological units should tend to align with tone unit boundaries (Lin & Adolphs [2009]), and that both articulation rate and phonetic reduction rate may be higher in phraseological units (Strik, Hulsbosch & Cucchiarini [2010]). The sequence *I don't know* has been subject to both phonetic and prosodic analyses in order to verify these expectations.

- 7 Bybee & Scheibman [1999] and Scheibman [2000] analyzed the same 225-minute corpus of conversations among six speakers of American English. Both studies examined the hypothesis that certain strings involving *don't* may be undergoing grammaticalization. The authors argued that the “use of elements in sequence strengthens their syntagmatic relations. Elements very frequently used together fuse” (Bybee & Scheibman [1999: 578]). For this reason, the authors focused on frequent word strings involving *don't* and explored the possibility that high frequency strings would show phonetic “fusion”. Bybee & Scheibman showed that the majority of the 138 occurrences of *don't* in their corpus were used with pronominal subjects and, in particular, with *I* (63.77%), and although 26 different verbs were found in combination with *don't*, approximately half of all tokens were accounted for with three verbs: *know* ( $n = 39$ ), *think* ( $n = 20$ ), and *have* ( $n = 9$ ). They then classified all tokens into four different categories as a function of how *don't* was pronounced: (a) full-stop consonant and full vowel, (b) reduced consonant and full vowel, (c) reduced consonant and reduced vowel, and (d) reduced vowels only. Their analysis revealed that frequent strings tended to be associated with greater phonetic reduction, what Bybee and colleagues later refer to as “special reduction” (Bybee, File-Muriel & Napoleão de Souza [2016]). For *I don't know*, they moreover noted that the few cases involving no or less reduction were generally literal (i.e., referential non-phraseological) uses of this string.
- 8 In her 2009 study, Pichler focused on one of the findings reported by Bybee & Scheibman [1999], namely that (referential vs. pragmatic) function may influence phonetic realization of *don't*. Pichler analyzed more than 35 hours of speech collected from 36 speakers in Berwick-upon-Tweed in Northern England, and identified 380 unbound uses of *I don't know* (meaning instances where this string does not introduce a direct object or a subordinate clause). Tokens were analyzed both with respect to their pragmatic function and their phonetic realization. The author coded for four different functions, three of which would be considered conventional expressions insofar as the string is used as a discourse marker, whereas the fourth (referential uses) corresponds to what we have termed non-phraseological uses. For the phonetic analysis, the author classified each token into one of four categories: full variants (*I don't know*), semi-reduced variants (*I dono*), reduced variants (*I dunno*), and the local variant (*I divn't knaa*). Results confirmed findings from Bybee & Scheibman's work on American English, as

referential uses of *I don't know* strongly favored the use of the full variant, whereas pragmatic uses (i.e., where *I don't know* acted as a conventional expression) were most often realized with a reduced variant.

- 9 Finally, in the only study that has focused on L2 speakers, Lin & Adolphs [2009] examined the production of the string *I don't know why* from a phraseological perspective. Both Lin [2010] and Lin & Adolphs state that within phraseology, phonological coherence refers to a variety of different characteristics, including the general lack of dysfluencies or pauses within a phraseological unit or alignment with tone unit boundaries. However, whereas such coherence has been empirically addressed in numerous studies looking at children acquiring their first language (e.g., Peters [1983]), little research exploring whether phraseological units enjoy greater phonological coherence among adults speaking either their first or second language currently exists. Lin & Adolphs address this gap, with a study focusing on the most frequent 5-word sequence – *I don't know why* – in an interview corpus made up of between 3 and 5 interviews with 17 different Chinese learners of English during their stay abroad in Britain. In total, this string was used 56 different times by the learners. The authors performed an acoustic analysis on the occurrences, and found that 55% corresponded perfectly to a tone unit, providing some initial evidence of alignment with tone unit boundaries.

### 1.3. Conventional expressions and the L2 immersion experience

- 10 The acquisition of conventional expressions by L2 learners is thought to be facilitated by contact with target-language interactions (particularly immersion experiences) for two principal reasons. First, as already mentioned, immersion experiences are hypothesized to be particularly beneficial for the development of phraseological competence in general (Ellis *et al.* [2015]), insofar as large amounts of input are necessary for learners to have sufficient contact with many phraseological patterns. Second, given the potential for numerous and varied interactions with target-language speakers, it is not surprising that interactional (i.e., pragmatic, discursive, sociolinguistic) aspects of language use have been shown to benefit from a stay abroad, particularly when the stay is long (see Bardovi-Harlig & Bastos [2011]; Charkova & Halliday [2011]; Geeslin & Long [2014]). In this section, we focus on research having explored learners' competence with conventional expressions during an immersion experience, beginning with studies that have taken a general approach to conventional expressions before looking at research focused on a subset of these sequences, namely discourse markers.
- 11 How a stay abroad may impact the acquisition of conventional expressions has been explored by several authors, using a variety of elicitation techniques: multiple-choice tasks (Roever [2006]), discourse completion tasks (Barron [2007]), written interpretation and completion tasks (Kecskes [2000]), written essays (Li & Schmitt [2009]), aural recognition and oral production tasks (Bardovi-Harlig & Bastos [2011]). These studies all demonstrated a positive impact of contact with a target-language community on the mastery of conventional expressions. For example, Bardovi-Harlig & Bastos report on data from 122 L2 English speakers studying in the United States and 49 native speakers. In this project, data were collected both on the ability to recognize conventional expressions and use them in appropriate contexts. In their analysis,



Bardovi-Harlig & Bastos explored the role of L2 proficiency (four different levels), length of stay in the United States (range: 1-18 months), and the intensity of interaction (based on self-reported contact with native speakers and television watching) on the scores obtained on the two tasks. A logistic regression analysis of the results from the receptive task showed that only intensity of interaction was a significant predictor: Higher levels of reported contact with English favored higher scores on the recognition task. The analysis of the production task showed two significant factors: Higher proficiency and greater intensity of contact favored higher scores on the production task. Both of these results underscore the importance of contact with the target language for the acquisition of conventional expressions.

- 12 Other researchers have focused specifically on sets of discourse markers used by non-native speakers in an immersion context (Fuller [2003]; Hellerman & Vergun [2007]; Liao [2009]; Sankoff *et al.* [1997]; Tavakoli [2018]). In general, these researchers were interested in documenting the variety and the frequency of discourse markers used, as well as the (pragmatic and discursive) functions that they fulfill in native and non-native productions. Of particular interest for the current study is Tavakoli [2018], which offers a view of change in discourse marker use over the course of four weeks spent abroad. Forty B2-level non-native speakers of English studying at a university in the UK took part in the study, for which they participated in a monologue and a dialogue task in weeks six and ten of their intensive English class. At week ten, results showed an overall significant increase in all discourse markers in the monologue task, and a significant increase in longer (i.e., at least two-word) discourse markers in the dialogue task. These findings suggest that, at least in terms of quantity, even a short stay abroad may lead to an evolution in discourse marker use. That said, other authors have reported that even advanced non-native speakers tend to use discourse markers less frequently than do native speakers (e.g., Fuller [2003]; Sankoff *et al.* [1997]), and that individual learners may show the tendency to (over)rely on a small set of discourse markers, using them repetitively (see Fuller [2003]). Finally, echoing the findings from Bardovi-Harlig & Bastos [2011], authors such as Hellerman & Vergun [2007] and Liao [2009] found evidence that “the more contact NNSs have with the target language culture, the more likely they will use DMs [discourse markers] in their spoken discourse” (Liao [2009: 3]).
- 13 Taken together, previous research using elicitation tasks has shown that a stay abroad tends to facilitate performance with respect to the ability to recognize and produce conventional expressions. Research on the oral production of a subset of conventional expressions (namely, discourse markers) by stay-abroad learners has reported that they tend to underuse such expressions when compared to native speakers, but that their use may increase over the course of a stay abroad. Moreover, results from studies using both elicitation tasks and oral corpora suggest that intensity of contact with the target language may be an important factor in the acquisition of this pragmatic resource. With respect to the specific sequence *I don't know*, no previous study has explored how its use may evolve over the course of an immersion experience. We do know, however, that this conventional expression is frequently used by native speakers (see Baumgarten & House [2010: 1186]), and that its phonetic realization may differ as a function of phraseological status. However, most of the studies having explored the phonetic realization of *I don't know* did not conduct an acoustic analysis, relying instead on a general classification of tokens into broad categories. Moreover, no analysis of phonetic reduction in this string involving non-native speakers has been carried out.



For the current study, we aim to contribute to research on the development of conventional expressions during a stay abroad, relying on both a phraseological and phonetic analysis to respond to two research questions:

1. What distributional patterns characterize the use of the sequence *don't* by five francophone learners of English before, during, and after a stay in an English-speaking environment?
2. For the sequence *I don't know*, how does its use as a discourse marker and its phonetic realization evolve over the stay abroad?

## 2. The current study

### 2.1. The corpus

- 14 For this study, we analyzed data from the PROLINGSA (Linguistic progress during study abroad) corpus, which we compiled between June 2018 and June 2019. Our goal in creating this corpus was to collect case studies with which it is possible to investigate linguistic development in L2 English by Francophone speakers over the course a stay abroad in an English-speaking environment. Five francophone students enrolled in either their first or second year of a degree course in Applied Foreign Languages and slated to spend the following academic year as an Erasmus+ student in either Ireland or England agreed to contribute to the corpus. This meant that each student met with a researcher on five occasions over the course of one year. The first meeting took place in June 2018, prior to the participants' departure to their Erasmus location. The following three interviews (in November, February, and March) took place while the participants were abroad. The final interview was carried out after the participants' return to France, in June 2019. Each interview followed a set protocol of questions, which were intended to lead to extended talk on the part of the students. The questions focused on a range of topics, and while most aimed to elicit narrative talk (e.g., Describe your first day in X; What was your best social experience of the last two months?), participants were also asked to express their opinions on personal (e.g., Do you think living in X has had an influence on your personality? If so, how?) and news events (e.g., What do you think about Brexit?). There were large differences in interview length, with the shortest interview lasting 12:18 (C, interview 3) and the longest 56:11 (Y, interview 2), for a total of approximately 12 hours of recorded interactions.
- 15 Details concerning the five participants are provided in Table 1. Three of the participants (A, C, M) were female, and two were male (N, Y). All reported speaking French as a native language, although Y grew up speaking Turkish with his mother (he stated that he was French dominant). For their degree program, students had to specialize in two foreign languages. English was one of these languages for all five students, and their second language of specialty was either Arabic, Chinese, Italian or Spanish. During the first meeting with the participants, they completed a version of the Oxford Quick Placement Test which targeted grammatical and lexical knowledge in order to provide a general measure of proficiency. Using the scoring conversion provided with the test, the five participants scored either within the lower intermediate (A, C) or advanced range (M, N, Y) at the beginning of the project. Four of the five participants spent their Erasmus+ year in Ireland, whereas Y spent his in

England. Finally, we include the number of words produced by each participant over the course of the five interviews. It is noteworthy that Y (the most talkative participant) produced more than 4.5 times as many words as the least talkative participant (C).

Table 1. Background details on participants

Participant	Gender	Age (Pre- sojourn)	Languages (specialization)	L1	Oxford Quick Placement Test		Country	# of words
					Pre-sojourn	Post- sojourn		
A	F	19	English + Italian	French	Lower intermediate	Upper intermediate	Ireland	13,512
C	F	18	English Chinese +	French	Lower intermediate	Upper intermediate	Ireland	6,118
M	F	19	English Spanish +	French	Advanced	Advanced	Ireland	7,167
N	M	19	English Chinese +	French	Advanced	Advanced	Ireland	13,232
Y	M	19	English + Arabic	French + Turkish	Advanced	Advanced	England	28,790

## 2.2. Data coding

- <sup>16</sup> All interviews were video-recorded and subsequently transcribed in CLAN using the CHAT transcription conventions (MacWhinney [2000]). Each transcription was verified and corrected by a second transcriber. For the current analysis, we searched each transcript for the strings *don't* and *dunno*.<sup>3</sup> In addition to noting from which interview and participant each occurrence came, we also coded all tokens for their usage patterns, allowing us to respond to our first research question. In particular, we coded for both the subject and verb used with *don't*. To respond to our second research question, we carried out additional coding on all instances of *I don't know*, coding tokens for their phraseological status and phonetic realization. Beginning with phraseological status, after examining the data, we focused on the use of *I don't know* as one type of conventional expression, namely as a discourse marker. In these uses, *I don't know* fulfills either speech management functions (e.g., floor yielding, coherence), interactional functions (e.g., hedging, politeness), or both (Aijmer [2009]). In 2(a), M uses *I don't know* to preface her turn, which appears to serve to hedge her response (see Aijmer [2009]). In 2(b), N uses the same sequence to close his turn and yield the floor to the interviewer. Finally, in 2(c) we provide a typical example from Y's interviews, in which he uses *I don't know* liberally in its speech management function, namely as a

placeholder or hesitation phenomenon. Finally, participants A and especially Y made use of longer versions of this discourse marker, namely *I don't know* followed by an interrogative word: *I don't know why, I don't know what, I don't know how* (see 2d).<sup>4</sup>

- (2) a. INT1: ok so what did you imagine? M: **I don't know** I imagine more of a big party (Interview 4)
- b. N: I did my best and hopefully it's good **I don't know** (Interview 5)
- c. Y: uh but uh yeah **I don't know** at the uh **I don't know** it's maybe like attachment or uh (Interview 3)
- d. INT1: she's had more practical practical translation than you've had Y: yes than yeah the Italian group then any other **I don't know why** (Interview 3)

- 17 In carrying out the phonetic coding, we observed the spectrogram for each token of *I don't know* using Praat (Boersma & Weenink [2021]). The phonetic realization coding focused on two measurements of length that we hypothesized may be sensitive indicators of phonetic reduction, with greater reduction expected to be reflected in shorter duration of segments. First, we examined the initial consonant of *don't* and measured overall length in milliseconds. This measurement included the closure and, if present, the burst. For cases in which the initial consonant was not pronounced (*I \_unno*), the duration measurement of the consonant was 0. We took similar measurements for vowel length, measuring length from the end of the initial consonant (either the end of the closure or the end of the burst, if present) to the onset of the word *know* as based on spectral characteristics, namely the abrupt spectral change corresponding to the closure of the oral tract that is a hallmark of nasal consonants. Our vowel measurement thus includes the oral portion of the vowel and any subsequent nasalization of the vowel prior to the following oral closure. Our corpus does not include any instances of *don't know* in which the final /t/ of *don't* is realized. The non-pronunciation of the vowel (*I d\_nno*) was assigned a length of 0ms.

### 2.3. Data analysis

- 18 Given the nature of the data from the PROLINGSA corpus (i.e., five case studies), our analysis relies on detailed descriptions of usage patterns and does not make use of inferential statistics. To address our first research question, we use a bottom-up, distributional approach to explore the variety of sequences involving *don't* used by each participant in each interview. The inventories for participants are detailed, highlighting recurrent patterns in the uses of *don't* and potential changes across time. For the second research question, we focus on the most frequent string involving *don't*: *I don't know*. We first explore the distribution of *I don't know* tokens over time as a function of phraseological status, opposing discourse marker uses to other uses. We then use data visualization techniques (using the ggplot2 package in RStudio) to view trends in the phonetic realization results for the two most frequent users of *I don't know*, participants A and Y. For this analysis, we look for evidence of phonetic reduction in the realization of *don't* as a function of (a) whether the string *I don't know* was used as a discourse marker or not and (b) time.

### 3. Results

#### 3.1. Characterizing uses of *don't*

- 19 A total of 734 instances of *don't* were identified in the transcripts. In order to characterize the usage patterns involving *don't*, we first examined the subjects used with this verb form. We found that the majority (87.7%, 644/734) of tokens was used with the first-person singular subject *I*. We then looked at the number of different verbs used with *don't* by each participant (see Table 2). Frequent recycling of certain *don't* + V combinations is visible when a high number of total *don't* tokens is paired with a low number of different verbs. Such frequent recycling is consistent with the idea of *don't* participating in phraseological units (in a distributional approach to phraseology). With this in mind, the data in Table 2 illustrate clear individual differences in the use of *don't*. On one end of the spectrum, Y makes repetitive use of a small number of verbs (visible in his frequent use of *don't* with a small number of different verbs), whereas C shows little to no repetition of *don't* + V combinations within a given interview. The ratio of *don't* tokens to different verbs for the remaining three participants lies between these two poles. When we examine these data with respect to evolution over time, we do not see a consistent picture of linear change in the ratio of different verbs to overall *don't* tokens. Thus, whereas N and C show little change over time, M, Y, and A all show the most repetition either at the end of their stay abroad (interview 4 for both M and Y) or after the return to France (interview 5 for A).

Table 2. Distribution of verbs used with *don't* by speaker and interview

Speaker	Interview				
	1	2	3	4	5
A					
Total <i>don't</i> tokens	20	23	33	22	56
Total different verbs	6	8	10	7	10
C					
Total <i>don't</i> tokens	2	10	5	5	5
Total different verbs	2	6	3	4	3
M					
Total <i>don't</i> tokens	10	7	18	23	13
Total different verbs	5	6	6	5	4
N					
Total <i>don't</i> tokens	12	45	13	33	31

Total different verbs	5	19	6	13	12
Y					
Total <i>don't</i> tokens	35	72	76	71	106
Total different verbs	5	14	9	6	16

20 We then looked at which *don't*+V strings were frequently used, and noted an abundance of *don't know* sequences, amounting to 424 in total. The next most frequent verbs appear much less frequently: *have* ( $n = 66$ ) and *think* ( $n = 41$ ). We thus decided to focus the second part of our analysis on *don't know* sequences. Almost all of these tokens correspond to the lexical combination *I don't know*, which occurred 410 times, thus accounting for 55.86% of all instances of *don't*. Looking at individual speakers (Table 3), Y is far and away the highest user of this sequence: He alone produces 240 tokens. This contrasts with C, who only produces 8. Table 3 also provides information concerning use over time, showing that the number of *I don't know* tokens generally increases over the course of the five interviews.

Table 3. Distribution of *I don't know* tokens and other *don't* tokens by speaker and interview

Speaker	Interview					Total
	1	2	3	4	5	
A						
<i>I don't know</i>	12	10	13	10	37	82
Other <i>don't</i> tokens	8	13	20	12	19	72
C						
<i>I don't know</i>	1	2	1	1	3	8
Other <i>don't</i> tokens	1	8	4	4	2	19
M						
<i>I don't know</i>	4	1	10	16	7	37
Other <i>don't</i> tokens	6	6	8	7	7	34
N						
<i>I don't know</i>	3	8	7	12	13	33
Other <i>don't</i> tokens	9	37	6	21	18	90
Y						

<i>I don't know</i>	28	42	51	52	67	240
Other <i>don't</i> tokens	7	30	24	19	29	109

### 3.2. Analysis of *I don't know*

- 21 For this second part of the study, we carried out two parallel analyses regarding the use of *I don't know*. The first and third authors identified the function of each occurrence, distinguishing between phraseological uses (in the form of discourse markers) and all other uses. At the same time, the second author analyzed the phonetic realization of a subset of the tokens. In this way, the two parts of the analysis were independent of each other. Beginning with the analysis of function, we found that of the 410 instances of *I don't know* in the PROLINGSA corpus, 233 functioned as discourse markers (56.83%). When broken down over time, we find that before the students' departure, 47.92% of their *I don't know* tokens were discourse markers. During the stay abroad, discourse markers were found to account for 55.56% (interview 2), 50% (interview 3), and 49.45% (interview 4) of all *I don't know* tokens. At the last interview, after the students' return to France, we observed a strong increase in the proportion used as discourse markers: 70.08% of these strings were discourse markers.
- 22 However, these global trends mask uses that are clearly different from one individual to the next. In Table 4, we provide the number of *I don't know* tokens used as a discourse marker at each interview for each participant. We have also indicated the rate of use per 1,000 words, allowing comparison across participants and interviews. Beginning with trends across participants, these data show that although Y produced the highest absolute number of *I don't know* tokens, his relative (per 1,000 words) frequency of use is often similar to that of other participants (e.g., A at interviews 1 and 5, M at interview 4). This discrepancy between absolute and relative frequency is explained by Y's talkativeness, as he produced (at least) two times as many words as the other participants. The relative numbers thus allow us to observe that the participants who use *I don't know* as a discourse marker the most often are Y, A, and M; discourse marker uses are present in N's interviews, but at lower relative rates, and C makes only marginal use of *I don't know* as a discourse marker. Turning to trends across interviews, Y is the only participant to show linear increase in the relative use of *I don't know* as a discourse marker over the five interviews. For A and N, discourse marker use is highest in interview 5, followed by interview 1, with the lowest relative use of *I don't know* as a discourse marker visible for the three interviews conducted while the participants were abroad. For M, the final three interviews show higher discourse marker use than in interview 1, with the highest use being observed in interview 4. Finally, no evolution can be observed for C, who produced only two tokens.

Table 4. Uses of *I don't know* as a discourse marker

	A		C		M		N		Y	
Interview	#	per 1K	#	per 1K	#	per 1K	#	per 1K	#	per 1K

1	5	3.92	1	0.91	2	1.91	2	1.50	13	3.72
2	5	1.25	–		1	0.72	3	0.68	26	3.77
3	4	1.49	–		6	4.58	2	0.98	29	4.52
4	6	3.26	1	1.01	8	5.58	1	0.48	29	5.95
5	26	7.02	–		6	3.00	9	2.65	48	6.75

- 23 For the final portion of our analysis, we examined whether the phonetic realization of *I don't know* varied (a) with respect to the function fulfilled (discourse marker vs. other) and (b) over time. We thus subjected all *I don't know* tokens produced at interviews 1, 3, and 5 by the highest users – A ( $n = 62$ ) and Y ( $n = 146$ ) – to a phonetic analysis whose aim was to quantify possible phonetic reduction (for both the initial consonant and the vowel of *don't*). In both cases, we hypothesized that more reduced forms would have a shorter duration. Results are summarized in Table 5 and visually represented in Figures 1 and 2 for participant A and in Figures 3 and 4 for participant Y. In each of the four figures, results are presented separately by interviews and by phraseological status along the x-axis. The y-axis corresponds to the measurements (in milliseconds) of either the vowel or the initial consonant of *don't*. Each occurrence of *I don't know* corresponds to one dot, and box-plot overlays have been added to facilitate interpretation.
- 24 Beginning with A, we observe that when *I don't know* fulfills the function of a discourse marker, *don't* tends to include both a shorter vowel and initial consonant. Moreover, there appears to be less variability in the values recorded for discourse marker uses. This is visible both in the lower standard deviations provided in Table 5, and in the tighter grouping of data points seen in both Figures 1 (for vowel length) and 2 (for consonant length). Overall, it thus appears that A produces greater phonetic reduction of *don't* when she uses *I don't know* as a discourse marker as opposed to with other functions. Looking now to the data over time, this function-based distinction shows signs of strengthening: We see widening gaps between mean vowel and consonant duration (Table 5) and we observe less overlap in the interquartile ranges (Figures 1 and 2) in interviews 3 and 5 versus interview 1. However, these observations are tentative, as the figures show that these patterns are in part due to outliers present in the third and fifth interviews.

Table 5. Vowel and consonant length of *don't* (in ms) as a function of time and phraseological status

	A		Y	
Interview	Vowel length M (SD)	Consonant length M (SD)	Vowel length M (SD)	Consonant length M (SD)
1				
Discourse marker	9.95 (3.00)	5.08 (1.96)	4.90 (2.71)	2.53 (1.78)



Other	12.65 (6.76)	5.76 (3.43)	5.13 (2.37)	2.95 (1.91)
3				
Discourse marker	7.72 (2.83)	4.97 (4.00)	3.27 (2.10)	4.10 (1.68)
Other	16.03 (12.76)	9.06 (4.21)	3.47 (2.29)	4.89 (3.38)
5				
Discourse marker	11.20 (4.24)	6.00 (4.30)	2.53 (2.29)	5.53 (2.77)
Other	15.53 (4.99)	10.87 (9.54)	2.64 (3.53)	6.25 (2.44)

Figure 1. Vowel length of *don't* as a function of time and phraseological status (participant A)

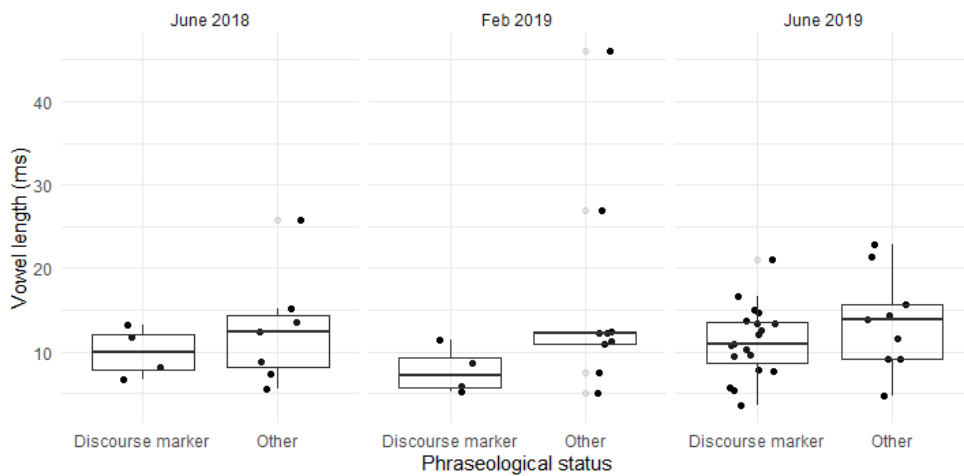
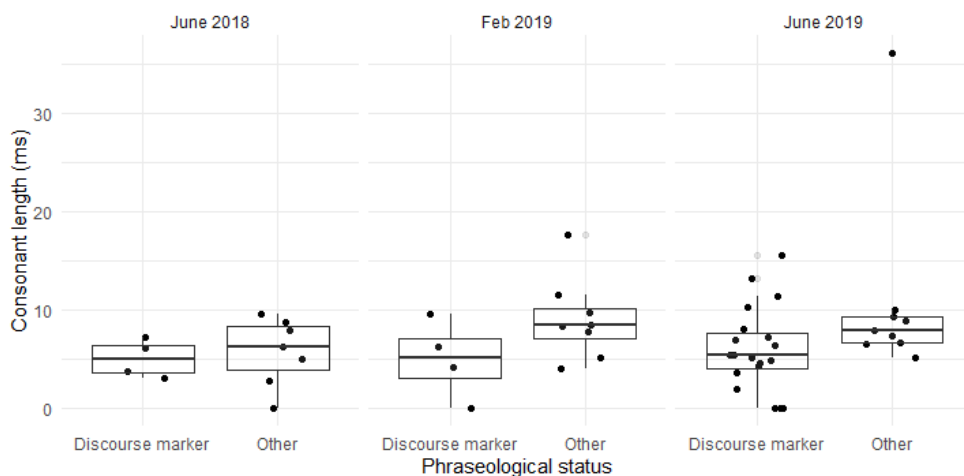


Figure 2. Length of initial consonant of *don't* as a function of time and phraseological status (participant A)



25 The findings for participant Y tell a different story. We observe that although the mean durations reported in Table 5 for *don't* when used as a discourse marker are always

numerically lower than the means found for other uses of *I don't know*, the differences are very slight. Moreover, standard deviations for discourse marker and other uses are similar. Taken together, this suggests little difference in phonetic reduction as a function of phraseological status before, during, and after an academic year abroad. If no change is seen over time in the phonetic realization of discourse marker versus other uses of *I don't know*, we do observe evolution at a more macro-level. More specifically, and as is clearly visible in Figure 3 (vowel duration), Y shows a greater tendency to reduce the vowel in *don't* as the year abroad progresses, regardless of the phraseological status of the expression *I don't know*. Moreover, extreme cases of vocalic reduction, whereby the vowel of *don't* is completely elided (resulting in *I d\_nno*), also become more frequent over time: Whereas only two such tokens are visible in interview 1 (7.1% of all *I don't know* tokens), 23.5% ( $n = 12$ ) of *I don't know* tokens at interview 3 and 34.4% ( $n = 23$ ) at interview 5 show vowel elision. This tendency towards greater vocalic reduction is accompanied – somewhat surprisingly – by the opposite trend in consonant duration: Y's pronunciation of the initial consonant of *don't* increases steadily in length over the period studied (see Figure 4). These contrasting patterns are also visible in Table 5, where the average vowel length for *don't* at Time 1 is cut in half by Time 5, whereas the average length of the initial consonant concomitantly doubles, for both discourse marker and other uses of *I don't know*.

Figure 3. Vowel length of *don't* as a function of time and phraseological status (participant Y)

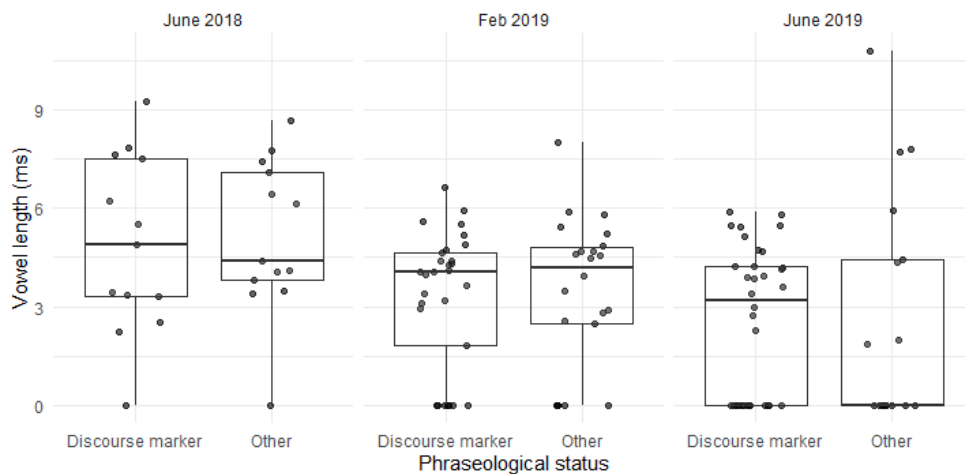
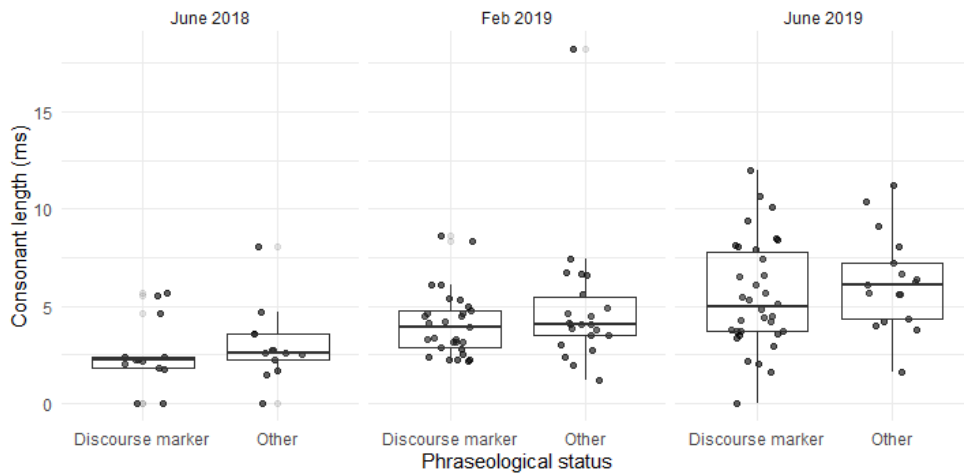


Figure 4. Length of initial consonant of *don't* as a function of time and phraseological status (participant Y)



## 4. Discussion and conclusions

- 26 In this study, we examined lexical development during a stay abroad through the lens of phraseology. More specifically, we focused on the usage patterns involving the sequence *don't* exemplified in a series of interviews conducted with five francophone learners of English over the course of a year. This led us to focus on one specific sequence – *I don't know* – which we argued can be considered to be a conventional expression. This means that the same surface string can be either phraseological (in this case, a discourse marker) or non-phraseological (e.g., referential). In addition to investigating overall distribution of *don't* and functions associated with the uses of *I don't know*, we heeded Lin's [2010: 174] call to push investigations of spoken phraseology “a step beyond listing spoken lexical bundles as if they are merely textual in nature.” Although Lin is referring specifically to the need to investigate the prosody of spoken phraseology, her observation is relevant for all aspects of pronunciation. In what follows, we interpret the three components of this study (overall distribution patterns, functional distinctions in the use of *I don't know*, and the phonetic realization of *don't* when part of the sequence *I don't know*), contextualizing them with respect to past research and reflecting on potential change in phraseology over the course of a stay abroad.
- 27 Our first research question focused on identifying overall usage patterns involving the sequence *don't*. Like Bybee & Scheibman [1999], we identified all instances of *don't* in our corpus ( $n = 734$ ), before examining the subjects and verbs with which it combines. In comparing the results across the two studies, the same trends are visible, although the distributional patterns in the PROLINGSA corpus are more extreme than those reported by Bybee & Scheibman. With regards to subjects used with *don't*, the American speakers recorded by Bybee & Scheibman used the first-person singular pronoun with 63.77% of all *don't* tokens. The pronoun *I* was also found to be dominant in the PROLINGSA corpus, although to a greater extent insofar as 87.7% of *don't* tokens occurred with *I*. The stronger presence of *I* in PROLINGSA doubtless reflects differences in the types of speech elicited. Although both studies are based on dialogues, the PROLINGSA interview protocol involved questions that encouraged participants to talk

about themselves, whereas Bybee & Scheibman report on “natural conversation.” Moving on to verbs used with *don't*, we found that the three most frequent were *know* ( $n = 424$ ), *have* ( $n = 66$ ), and *think* ( $n = 41$ ). These three verbs were also the most frequent in Bybee & Scheibman’s analysis. However, out of the 124 tokens they analyzed (Bybee & Scheibman [1999: 582]), *know* only occurred 39 times, whereas *have* and *think* occurred 9 and 20 times, respectively. This leads us to two observations. First, whereas the same three verbs are most frequent in both corpora, their combined frequency corresponds to 54.84% (68/124) of all *don't* tokens in the Bybee & Scheibman corpus, but to 72.34% (531/734) in the PROLINGSA corpus. Second, although *know* is the most frequent verb in both analyses, *don't know* is much more dominant in the PROLINGSA corpus: On its own, *don't know* accounts for more than 57% of all *don't* tokens (vs. 31% in Bybee & Scheibman’s corpus). These comparisons demonstrate that recurrent combinations are even more prevalent in the five PROLINGSA participants’ productions than in the corpus analyzed by Bybee & Scheibman, and that the lexical combination *don't know* is particularly dominant. Whereas the same caveat mentioned previously concerning the differences between the two corpora may be relevant here, we note that these findings are also consistent with previous research that has pointed to the fact that language learners sometimes tend to over-rely on certain word combinations (see, for example, Fuller [2003]).

- 28 For the second portion of the analysis, we isolated the 410 tokens of *I don't know* and analyzed each to determine if it was used as a discourse marker (and was, thus, phraseological) or whether it was fulfilling another function. Increased use over time was expected on two counts. First, several researchers have hypothesized that phraseological development may receive a boost during a stay abroad (see Ellis *et al.* [2015]). Second, there exists evidence of increased discourse marker use during a stay abroad (Tavakoli [2018]). In the case of the PROLINGSA corpus, evidence of change was slight. We did observe an overall increase in the proportion of *I don't know* tokens used as discourse markers between interviews 1 (47.92%) and 5 (70.08%). However, the percentages of discourse marker use for interviews 2, 3, and 4 are between 49% and 56%. This stark difference between the first four interviews and the final interview leads us to question whether the high rate of discourse marker use in interview 5 may have been influenced by the interview protocol (e.g., many questions in the final interview were prospective, which may have led to more hedging on the part of the speakers). When we observed the relative frequency of *I don't know* as a discourse marker (i.e., its use per 1,000 words), only two participants – Y and M – showed steady increases in *I don't know* tokens functioning as discourse markers over time. However, this trend is modest, and we moreover saw less linear patterns for the three remaining participants (A, N, C). Although our findings stand out from those reported by other authors, this may be because of differences in methodology. For example, we adopted a simple binary distinction in our functional analysis (discourse marker vs. other). This is an obvious simplification of a complex reality. Analyses by, for example, Aijmer [2009] identify numerous discursive functions for *I don't know* (e.g., hedging, floor yielding, speech management). Thus, whereas we saw only slight trends towards greater (numerical) use of *I don't know* as a discourse marker over the course of the stay abroad, it may be that changes occur at the level of individual functions. Moreover, many past studies have tended to focus on numerous discourse markers. Tavakoli, for example, identified all discourse markers and reported that a group of B2-level learners of English increased in the total number used over a four-week stay-abroad period. This

approach masks trends at the level of the lexical combination, and it would not be surprising to find that certain of the strings identified by Tavakoli increased little or not at all in frequency, mirroring the results we report for *I don't know*. Future research, in which we explore the PROLINGSA corpus using Tavakoli's methodology, could allow us to verify whether there is indeed change in the rate of use of discourse markers as a whole. In addition, thanks to the inclusion of five interviews with each PROLINGSA participant, we could go beyond Tavakoli's findings in order to characterize the pace of any potential development, identifying whether or not change occurs gradually over the stay abroad.

- 29 For the final part of our analysis, previous research on native varieties of English inspired us to examine the phonetic realization of *I don't know* tokens. According to Pichler [2007: 184], “[t]he variation between the full and reduced forms of the expression I DON'T KNOW [...] involves functional divergence.” In other words, it has been documented for certain native varieties that the phonetic realization of *I don't know* depends on the discursive or pragmatic function fulfilled by the string. More specifically, Schubotz, Oostdijk and Ernestus [2015: 362] observe that “[i]t appears that the more pragmatic and the less lexical or compositional a given item is, the more it will be reduced.” We thus explored whether the five PROLINGSA learners show this functional divergence in their productions of *I don't know* and whether the phonetic realizations of discourse markers versus other uses may change over the course of their stay in a target-language environment. Focusing only on the highest users of *I don't know*, we observed two very different patterns. The data from A were largely consistent with patterns reported in research on native varieties. This speaker produces *don't* with a more reduced (i.e., shorter) initial consonant and vowel when she uses *I don't know* as a discourse marker. Moreover, this distinction appears to strengthen over time, which may be the result of the stay abroad. On the other hand, Y's phonetic realizations of *I don't know* showed no reliable distinction as a function of phraseological status. Instead, we noted that this speaker showed a clear overall trend towards greater vowel reduction (including complete elision of the vowel, especially in interviews 3 and 5), accompanied simultaneously by increased length of the initial consonant. It thus appears that the stay abroad has influenced Y's production of this string, without leading to a sensitivity to the functional divergence (i.e., greater reduction for phraseological uses of a string) identified by Pichler. Although speculative, there is some evidence in our data to suggest that the trend towards longer initial consonants may reflect a growing sensitivity on Y's part to longer voice onset times in English. More specifically, in our coding of initial consonant length, we noted separately the length of the closure and burst. Interestingly, Y produces few instances of burst before leaving for England, whereas 23.53% of his *I don't know* tokens in interview 3 show burst. Thus, the increasing initial consonant length may reflect an attempt on Y's part to integrate this feature of English into his pronunciation. If our suspicions are correct, we would expect to see an evolution in voice onset time elsewhere in Y's interlanguage system. This possibility remains to be verified.
- 30 Although our analysis of the phonetic realization of *I don't know* is modest, concerning data from three interviews conducted with two participants, we believe that the inclusion of this dimension offers an interesting contribution to discussions of phraseological development. Conceptually speaking, and as mentioned previously, Lin [2010] highlights the fact that spoken language phraseology is often disassociated from its acoustic realization, and she forcefully argues for the need to take the phonetic and

prosodic characteristics of such strings into account in L2 phraseology research. Local [2003: 322] offers a similar argument with respect to linguistic study more generally. He discusses “fine phonetic detail” and suggests that

[m]eaning is much more than lexical meaning. If we want to construct a robust, integrated model of speech perception, speech understanding and phonological representation, we need to entertain richer ideas about the ways in which phonetic detail relates to the construction of meaning.

- 31 Conventional expressions like *I don't know, you see, I mean*, etc., which can be used either as phraseological units fulfilling a particular discursive function or as literal strings, provide an ideal testing ground for investigating the contribution of phonetic detail to the construction of meaning. Although such research exists with respect to native speakers (e.g., Pichler [2009]), to the best of our knowledge, the current study constitutes a first attempt to examine if and how learners appropriate these fine phonetic details with respect to conventional expressions. It is worth highlighting the fact that whereas conventional expressions may be included in teaching materials, often taught as lexical blocks (see Chini [2001]), instruction as to how to pronounce the expressions seems to be far from systematic. Moreover, our own data suggest that even presumably rich and sustained L2 input of the type we might expect during a stay abroad may not be sufficient to lead to the integration of this type of detail into learners' interlanguage. Thus, not only researchers but also language teachers may do well to heed Lin's [2010] call to attend to the acoustic realization of phraseological units. We believe that additional research in this direction should be undertaken. It is also our hope that the current investigation provides food for thought regarding the methodology to be adopted in the study of such details.

---

## BIBLIOGRAPHY

- AIJMER Karin, 2009, “‘So er I just sort I dunno I think it's just because...’ A corpus study of *I don't know* and *dunno* in learners' spoken English”, in JUCKER Andreas H., SCHREIER Daniel & HUNDT Marianne (Eds.), *Corpora: Pragmatics and discourse. Papers from the 29<sup>th</sup> International Conference on English Language Research on Computerized Corpora (ICAME 29)*. Amsterdam: Rodopi, 151-168.
- BARDOVI-HARLIG Kathleen, 2009, “Conventional expressions as a pragmalinguistic resource: Recognition and production of conventional expressions in L2 pragmatics”, *Language Learning*, 59, 755-795.
- BARDOVI-HARLIG Kathleen & BASTOS Maria-Thereza, 2011, “Proficiency, length of stay, and intensity of interaction, and the acquisition of conventional expressions in L2 pragmatics”, *Intercultural Pragmatics*, 8, 347-384.
- BARRON Anne, 2007, “‘‘Ah no honestly we're okay:’’ Learning to upgrade in a study abroad context”, *Intercultural Pragmatics*, 4(2), 129-166.
- BATISTA Roselene & HORST Marlise, 2016, “A new receptive vocabulary size test for French”, *La Revue Canadienne des Langues Vivantes*, 2(2), 211-233.

- BAUMGARTEN Nicole & HOUSE Juliane, 2010, "I think and I don't know in English as lingua franca and native English discourse", *Journal of Pragmatics*, 42, 1184-1200.
- BOERSMA Paul & WEENINK David, 2021. "Praat: doing phonetics by computer [Computer program]", Version 6.1.42, retrieved 15 April 2021 from <http://www.praat.org/>
- BYBEE Joan, 2002, "Phonological evidence for exemplar storage of multiword sequences", *Studies in Second Language Acquisition*, 24, 215-221.
- BYBEE Joan & SCHEIBMAN Joanne, 1999, "The effect of usage on degrees of constituency: The reduction of don't in English", *Linguistics*, 37(4), 575-596.
- CHARKOVA Krassimira D. & HALLIDAY Laura J., 2011, "Second- and foreign-language variation in tense backshifting in indirect reported speech", *Studies in Second Language Acquisition*, 33, 1-32.
- CHINI Danielle, 2001, « Détournement didactique de la locutionnalité : les blocs lexicalisés », in TOLLIS Francis (Ed.), *La Locution et la Périphrase du Lexique à la Grammaire*, Paris: L'Harmattan, 37-53.
- CONKLIN Kathy & SCHMITT Norbert, 2012, "The processing of formulaic language", *Annual Review of Applied Linguistics*, 32, 45-61.
- CROSSLEY Scott, SALSURY Tom & MCNAMARA Danielle, 2010, "The development of polysemy and frequency use in English second language speakers", *Language Learning*, 60(3), 573-605.
- DE COCK Sylvie, 1998, "A recurrent word combination approach to the study of formulae in the speech of native and nonnative speakers of English", *International Journal of Corpus Linguistics*, 3, 59-80.
- DE COCK Sylvie, 2000, "Repetitive phrasal chunkiness and advanced EFL speech and writing", *Corpus linguistics and linguistic theory*, 51-68.
- EDMONDS Amanda & GUDMESTAD Aarnes, 2021, "Collocational development during a stay abroad", *Languages*, 6(1), 12.
- ELLIS Nick C., SIMPSON-VLACH Rita, RÖMER Ute, O'DONNELL Matthew B. & WULFF Stefanie, 2015, "Learner corpora and formulaic language in SLA", in GRANGER Sylviane, GILQUIN Gaëtanelle & MEUNIER Fanny (Eds.), *Cambridge Handbook of Learner Corpus Research*, Cambridge: Cambridge University Press, 357-378.
- FULLER Janet M., 2003, "Discourse marker use across speech contexts: A comparison of native and non-native speaker performance", *Multilingua*, 22, 185-208.
- GEESLIN Kimberly L. & LONG Avizia, 2014, *Sociolinguistics and second language acquisition: Learning to use language in context*, London: Routledge.
- GONZÁLEZ-FERNÁNDEZ Beatriz & SCHMITT Norbert, 2020, "Word knowledge: Exploring the relationships and order of acquisition of vocabulary knowledge components", *Applied Linguistics*, 41(4), 481-505.
- GRANGER Sylviane & BESTGEN Yves, 2014, "The use of collocations by intermediate vs. advanced non-native writers: A Bigram-based study", *International Review of Applied Linguistics and Language Teaching* 52, 229-252.
- GRANGER Sylviane & PAQUOT Magali, 2008, "Disentangling the phraseological web", in GRANGER Sylviane & MEUNIER Fanny (Eds.), *Phraseology: An interdisciplinary perspective*, Amsterdam: Benjamins, 27-50.



- HELLERMAN John & VERGUN Andrea, 2007, "Language which is not taught: The discourse marker use of beginning adult learners of English", *Journal of Pragmatics*, 39, 157-179.
- KECSKES Istvan, 2000, "Conceptual fluency and the use of situation-bound utterances", *Links & Letters*, 7, 145-161.
- KYLE Kristopher & CROSSLEY Scott A., 2015, "Automatically assessing lexical sophistication: Indices, tools, findings, and application", *TESOL Quarterly*, 49(4), 757-786.
- LI Jie & SCHMITT Norbert, 2009, "The acquisition of lexical phrases in academic writing: A longitudinal case study", *Journal of Second Language Writing*, 18, 85-102.
- LIAO Silvie, 2009, "Variation in the use of discourse markers by Chinese teaching assistants in the US", *Journal of Pragmatics*, 41(7), 1313-1328.
- LIN Phoebe M. S., 2010, "The phonology of formulaic sequences: A review", in WOOD D. (Ed.), *Perspectives on formulaic language: Acquisition and communication*, London: Continuum, 174-193.
- LIN Phoebe M. S. & ADOLPHS Svenja, 2009, "Sound evidence: Phraseological units in spoken corpora", in BARFIELD Andy & GYLLSTAD Henrik (Eds.), *Researching collocations in another language: Multiple interpretations*, New York: Palgrave Macmillan, 34-48.
- LOCAL John, 2003, "Variable domains and variable relevance: Interpreting phonetic exponents", *Journal of Phonetics*, 31, 321-339.
- MACWHINNEY Brian, 2000, *The CHILDES project: Tools for analyzing talk* (3<sup>rd</sup> Ed.), New York: Lawrence Erlbaum.
- MEARA Paul & MILTON James, 2003, *X\_Lex. The Swansea Levels Test*, Newbury: Express.
- NATION Paul, 2001, *Learning vocabulary in another language*, Cambridge: Cambridge University Press.
- PAWLEY Andrew K. & SYDER Frances H., 1983, "Two puzzles for linguistic theory: Nativelike selection and nativelike fluency", in RICHARDS Jack C. & SCHMIDT Richard W. (Eds.), *Language and communication*, London: Longman, 191-225.
- PETERS Ann M., 1983, *The units of language acquisition*, New York: Cambridge University Press.
- PICHLER Heike, 2007, "Form-function relations in discourse: The case of I DON'T KNOW", *Newcastle Working Papers in Linguistics*, 13, 174-187.
- PICHLER Heike, 2009, "The functional and social reality of discourse variants in a northern English dialect: I DON'T KNOW and I DON'T THINK compared", *Intercultural Pragmatics*, 6(4), 561-596.
- RAUPACH Manfred, 1984, "Formulae in second language speech production", in DECHERT Hans, MÖHLE Dorothea & RAUPACH Manfred (Eds.), *Second language productions*, Berlin: Gunter Narr Verlag Tübingen, 114-137.
- ROEVER Carsten, 2006, "Validation of a web-based test of ESL pragmalinguistics", *Language Testing*, 23, 229-256.
- SANKOFF Gillian, THIBAUT Pierrette, NAGY Naomi, BLONDEAU Hélène, FONOLLOSA Marie-Odile & GAGNON Lucie, 1997, "Variation in the use of discourse markers in a language contact situation", *Language Variation and Change*, 9, 191-217.
- SCHIEBMAN Joanne, 2000, "I dunno: A usage-based account of the phonological reduction of don't in American English conversation", *Journal of Pragmatics*, 32, 105-124.

- SCHMITT Norbert, 2013, "Formulaic language and collocation", *The Encyclopedia of Applied Linguistics*, Wiley Online Library.
- SCHUBOTZ Louise, OOSTDIJK Nelleke & ERNESTUS Mirjam, 2015, "Y'know vs. you know: What phonetic reduction can tell us about pragmatic function", in LESTRADE Sander, DE SWART Peter & HOGEWEG Lotte (Eds.), *Addenda. Artikelen voor Ad Foolen*, Nijmegen: Radboud University, 361-380.
- SIYANOVA-CHANTURIA Anna & SPINA Stefania, 2020, "Multi-word expressions in second language writing: A large-scale longitudinal learner corpus study", *Language Learning*, 70, 420-463.
- SKEHAN Peter, 1998, *A cognitive approach to language learning*, Oxford: Oxford University Press.
- STRIK Helmer, HULSBOSCH Micha & CUCCHIARINI Catia, 2010, "Analyzing and identifying multiword expressions in spoken language", *Language Resources & Evaluation*, 44, 41-58.
- TAVAKOLI Parvaneh, 2018, "L2 development in an intensive study abroad EAP context", *System*, 72, 62-74.
- WOOD David, 2002, "Formulaic language in acquisition and production: Implications for teaching", *TESL Canada Journal*, 20, 1-15.
- WRAY Alison, 2002, *Formulaic language and the lexicon*, Cambridge: Cambridge University Press.

## NOTES

1. LECLERCQ Pascale, EDMONDS Amanda, SNEED GERMAN Elisa, 2021, *PROLINGSA* [Corpus]. ORTOLANG (Open Resources and TOols for LANGuage) - [www.ortolang.fr](http://www.ortolang.fr), <https://hdl.handle.net/11403/prolingsa>.
2. We note that numerous different terms have been used to refer to what we call conventional expressions, including routine formulae, pragmatic routines, lexical phrases, and situation-bound utterances. Some of the studies we cite have used these different terms.
3. We also searched for all instances of non-contracted *do not*. Only one such token was identified and further excluded from this analysis.
4. We leave the question as to whether *I don't know* participates in other phraseological units to future research.

---

## ABSTRACTS

We explore lexical development during a stay abroad for five francophone learners of English by focusing on how they use a single string, namely *don't*. We adopt a phraseological perspective to explore change over time, opposing uses of *don't* when it appears in the discourse marker *I don't know* to other uses of this string. In addition to exploring the overall distribution of *don't*, we also conduct an acoustic analysis to explore the phonetic realization of *don't* as a function of whether it appears in a phraseological or non-phraseological sequence. Results point to slight changes over time in the use of *I don't know* as a discourse marker and in the phonetic realization of this string. Our results highlight clear instances of inter-individual differences.

Nous étudions le développement lexical au cours d'un séjour en immersion de cinq étudiants francophones de l'anglais. Nous nous focalisons sur les emplois faits d'une séquence fréquente dans notre corpus : *don't*. Pour cette analyse, nous adoptons une perspective phraséologique qui oppose les emplois de *don't* dans le marqueur discursif *I don't know* aux autres emplois de *don't* afin d'étudier l'évolution des emplois à travers le temps. L'analyse de la distribution de *don't* est complétée par une analyse acoustique qui vise à caractériser la réalisation phonétique de cette séquence en fonction de si elle se trouve dans une unité phraséologique ou non. Nos résultats révèlent des changements légers au cours du séjour en immersion, tant au niveau des emplois de *I don't know* en tant que marqueur discursif qu'au niveau de la réalisation phonétique de cette séquence. Par ailleurs, nous relevons des différences inter-individuelles importantes.

## INDEX

**Keywords:** phraseology, discourse markers, phonetic reduction, stay abroad

**Mots-clés:** phraséologie, marqueurs discursifs, réduction phonétique, séjour à l'étranger

## AUTHORS

### AMANDA EDMONDS

Université Côte d'Azur

amanda.edmonds@univ-cotedazur.fr

### ELISA SNEED GERMAN

Université Paul Valéry – Montpellier 3

elisa.sneed-german@univ-montp3.fr

### PASCALE LECLERCQ

Université Paul Valéry – Montpellier 3

pascale.leclercq@univ-montp3.fr