MASSICOTTE-LAFORGE, Sarah; MELANÇON, Andréane; SHI, Rushen. **Prosodic cues to syntatic structures in speech production.** Revista LinguíStica / Revista do Programa de Pós-Graduação em Linguística da Universidade Federal do Rio de Janeiro. Volume 11, número 1, junho de 2015, p. 85-96. ISSN 2238-975X 1. [http://www.letras.ufrj.br/poslinguistica/revistalinguistica] DOI: 10.17074/2238-975X.2015v11n1p85

PROSODIC CUES TO SYNTACTIC STRUCTURES IN SPEECH PRODUCTION

por Sarah Massicotte-Laforge (UQAM), Andréane Melançon (UQAM) e Rushen Shi** (UQAM)*

ABSTRACT

Children's initial syntactic acquisition tasks include finding clausal and phrasal units from continuous speech stream and assigning words to grammatical categories. This study inquires if prosodic cues exist in adult speech and mark syntactic units. Participants were Quebec-French speakers. In Experiment 1 participants read Determiner+Noun and Pronoun+Verb utterances. Determiners and pronouns were French words. Nouns and verbs were pseudo-words (e.g., *mige*, *krale, vare*) counterbalanced in their occurrences in the utterances. Their prosodic properties (duration, pitch, intensity) were measured. Results showed that the two categories did not differ in prosody: noun versus verb productions of these pseudo-words were equivalent. Experiment 2 tested whether larger utterances were produced with prosodic cues supporting different grammatical categories and phrasal groupings. The same pseudo-words were the final words (counterbalanced) in 1) [Determiner+Adjective+Noun] and 2) [[Determiner+Noun]+[Verb]] structures. The second word in both structures was *felli.* Results showed that the last word as nouns versus verbs differed significantly in duration, pitch and intensity. Moreover, the initial consonant of verb productions was longer, with a distinct preceding pause. The second word in (2) exhibited categorical and boundary cues, differing from the second word in (1) in duration, pitch and intensity. We suggest that these acoustic cues may help infants first parse larger utterances and then acquire the syntactic properties of phrases and words based on their distribution.

KEYWORDS: prosody; syntax, input; language acquisition.

*University of Quebec at Montreal, Canada
**shi.rushen@uqam.ca

During early language acquisition infants face various tasks, including finding from continuous speech linguistically relevant units such as words, phrases and sentences, as well as learning the structures of these units appropriate for their native language. Two of the earliest tasks are pertinent for syntactic acquisition, 1) assigning grammatical categories (such as nouns, verbs, etc.) to segmented words, and 2) parsing continuous multi-word speech into syntactic constituents. These tasks are non-trivial for infants. Infant-directed speech is predominantly multi-word utterances rather than single-word utterances (Shi, Morgan, & Allopena, 1998; Van de Weijer, 1998). Furthermore, unlike second language teachers, parents do not label the grammatical categories of words to their infants, nor do they provide overt information about constituents. Infants must rely on the speech input to learn to assign the grammatical categories of individual words and parse running speech into groups corresponding to syntactic constituents. The present study tested the hypothesis that acoustic cues may exist in speech input that mark grammatical categories of lexical items and indicate phrasal groupings.

The existing literature suggests that there are distinct sound patterns for different grammatical categories in natural speech production. The most fundamental distinction of lexical versus functional categories (i.e., content words versus function words) is marked by phonological and acoustical cues in speech (Cutler, 1993; Monaghan, Christiansen, & Chater, 2007; Shi *et al*., 1998), and infants shortly after birth are sensitive to such cues and use them to perceive the fundamental categorical distinction (Shi, Werker, & Morgan, 1999; Shi & Werker, 2001, 2003). The distinction between nouns versus verbs has been a focus in language acquisition research. These two categories are marked by phonological cues. For example, Soreno and Jongman (1990) showed that frequent nouns differ from verbs in English in vowel types. Kelly and colleagues (Cassidy & Kelly, 1991; Kelly, 1992) also reported differences of segment, syllable and stress patterns in English nouns versus verbs. Acoustic measures of infant-directed speech showed that mothers produce prosodic cues supporting the distinction between nouns and verbs in English (Conwell & Morgan, 2012) and in French (Shi & Moisan, 2008).

With respect to cues marking syntactic groupings, Cooper and colleagues (Cooper & Paccia-Cooper, 1980; Cooper & Sorensen, 1981) showed that major syntactic units are marked by cues such as pause, final vowel lengthening and F0 declination. Such cues are exaggerated in infant-directed speech (Bernstein Ratner, 1986). For example, final lengthening at the clausal boundary is more salient in infant-directed than in adult-directed speech. Acoustic markings of phrasal boundaries have also been shown, although they are generally not as strong as cues to clauses. Researchers have inquired how adult listeners resolve ambiguous utterances that have identical phonemic sequences but differ in syntactic constituents. For instance, a recent study in French (Millotte, Wales, & Christophe, 2007) found that speakers produced distinct prosodic cues to ambiguous utterances such as [*les pommes dures*] (i.e., a noun phrase, "hard apples") versus [*les pommes*][*durent* …] (i.e., a noun phrase followed by a main verb, "apples last …"). Furthermore, they found that adults were able to exploit the prosody to identify the grammatical categories of the target words. Both adults and infants perceive such prosodic cues for distinguishing syntactic groupings (Millotte *et al.,* 2007; Nazzi, Kemler Nelson, Jusczyk, & Jusczyk, 2000; Soderstrom, Seidl, Kemler Nelson, & Jusczyk, 2003). More recently, it was shown that as early as two years of age, French children can use phrasal prosody to interpret syntactic structures and grammatical categories of familiar content words (de Carvalho, Dautriche, & Christophe, 2014). Another study (Dautriche, Cristia, Brusini, Yuan, Fisher, &Christohpe, 2013)

investigated the role of prosody for children's interpretation of verb meaning in right-dislocated sentences. They found that 28-month-old French-learning infants used prosodic cues to correctly identify the intended meaning of familiar verbs, although they failed to do so for novel verbs.

The present study aimed at better understanding the nature of acoustic cues to grammatical categorization of words and to syntactic groupings. Previous production studies in the literature typically analyzed real words in speakers' native language. Here we used nonsense content words as production stimuli. The same non-words served as different grammatical categories such as the noun or the verb. This approach not only controlled for any possible acoustic differences inherent to different phonemes, but also eliminated potential effects of lexical frequencies on productions since non-words have the same lexical frequency (i.e., zero occurrence). In Experiment 1 non-words followed determiners in the Determiner+Noun structure (e.g., *ton mige* - your mige), and these non-words also followed subject pronouns in the Pronoun+Verb structure (e.g., *il mige* - he miges). The two kinds of utterances both form one phonological phrase, although they are different syntactically. The question was whether the non-words are produced with prosodic cues distinguishing noun versus verb uses. In Experiment 2 we examined whether longer utterances that are phonemically homophonous but differ in syntactic constituents (e.g., [*ton felli mige*], "your felli mige", as in "your little dog" versus [[*ton felli*] *mige*], "your felli miges", as in "your bunny runs") are marked by corresponding prosodic groupings. Furthermore, we inquired whether different prosodic groupings in Experiment 2, if they exist, affect the acoustic properties of the non-words used as different grammatical categories of words.

EXPERIMENT 1

Methods

Participants and Stimuli

Participants were four native Quebec-French-speaking adults recruited in the Psychology Department of the Université du Québec à Montréal.

We constructed six pseudo-words to be used as production stimuli: *mige, krale, vare, reule, bane, gorte*. The pseudo-words respected the sound patterns of French. They were also equally noun- and verb- like phonologically in French (Monaghan *et al*., 2007). We then created utterances by combining these words with French function words, so that each pseudo-word appeared as a noun or a verb across utterances. The use of the same pseudo-words as nouns and verbs controlled for any other phonetic or phonological factors that may lead to differences in prosody. We chose three determiners (i.e., *le*–the, *ton*–yours, *des*–some) to form noun utterances such as *le mige*. Three pronouns (i.e., *je*–I, *tu*–you, *il*–he) were used to form verb phrases such as *je mige*. To elicit the noun and verb uses of the pseudo-words as naturally as possible, we created filler utterances with the same structures, including 22 familiar French nouns and verbs. Eight were common nouns (*ballon*–ball, *drame*–tragedy, *père*–father, *singe*–monkey, *frère*–brother, *monde*–world, *bec*–spout, *cœur*–heart), which were paired with determiners (e.g., *le ballon*– the ball, *ton ballon*-your ball). Eight others were verbs in singular present tense (*dors*–sleep, *dessine*–draw, *mange*–eat, *goûte*–taste, *rame*–row, *bois*–drink, *trouve*–find, *trace*–trace), which were paired with pronouns (e.g., *je dors* - I sleep, *il dort* – he sleeps). The remaining six filler words were ambi-categorical, as a noun or as a verb (*cri*–cry/to scream, *pêche*–peach/to fish,

porte–door/to carry, *marche*–step/to walk, *travaille*–job/to work, *signe*–wave/to sign). They were paired with determiners and pronouns (e.g., *le cri* – the cry, *je crie* – I screem). These filler utterances were comparable to the target pseudo-word utterances in that both sets involved ambi-categorical words. This manipulation served to induce productions of pseudo-words in a compatible fashion. Final stimuli included 18 noun utterances with pseudo-words, 16 with familiar nouns, 6 with ambi-categorical nouns, 18 verb utterances with pseudo-words, 16 with familiar verbs and 6 with ambi-categorical verbs, for a total 80 phrases (see Table 1).

	Fillers with real French nouns/verbs				
Pseudo-words		**Non-ambiguous**		**Ambi-categorical**	
Noun use	**Verb use**	**Noun use**	**Verb use**	**Noun use**	**Verb use**
le mige	*je mige*	*le ballon*	*je dors*	*la porte*	*je porte*
le krale	*je krale*	*le père*	*je dessine*	*le travail*	*je travaille*
le vare	*je vare*	*le singe*	*je mange*	*ton cri*	*tu cries*
le reule	*je reule*	*le frère*	*je rame*	*ta pêche*	*tu pêches*
le bane	*je bane*	*le bec*	*je trouve*	*des marches*	*il marche*
le gorte	*je gorte*	*le cœur*	*je trace*	*des signes*	*il signe*
ton mige	*tu miges*	*ton drame*	*tu dessines*		
ton krale	*tu krales*	*ton singe*	*tu goûtes*		
ton vare	*tu vares*	*ton frère*	*tu rames*		
ton reule	*tu reules*	*ton monde*	*tu bois*		
ton bane	*tu banes*	*ton cœur*	*tu traces*		
ton gorte	*tu gortes*	*des ballons*	*il dort*		
des miges	*il mige*	*des drames*	*il mange*		
des krales	*il krale*	*des pères*	*il goûte*		
des vares	*il vare*	*des mondes*	*il rame*		
des reules	*il reule*	*des becs*	*il boit*		
des banes	*il bane*				
des gortes	*il gorte*				

Table 1: Materials used in Experiment 1.

Procedure

Participants were comfortably seated in an IAC acoustic chamber facing a microphone connected to a Sound Device 702T recorder. The recorder was outside the booth. Speech recording was made with a sampling frequency of 44.1 kHz, bit rate at 32 bits. Speakers were instructed that they would read some short sentences and phrases, and that some of them included unknown words. Then, they were given a sheet with the six pseudo-words in isolation (i.e., with no co-occurring function word) and were asked to read them aloud. This phase ensured that the speakers had a chance to practice the pseudo-words. Next, they were given three different sheets with the utterances listed. One sheet presented the noun utterances in random order. Another sheet presented the remaining verb utterances in a random order. The third sheet listed all 80 utterances in an alternating noun-verb order, that is, one noun utterance and then one verb utterance, etc.. On each sheet pseudo-word utterances and real word (noun or verb) utterances were intermixed randomly. Participants were asked to read each sheet

two times, from top to bottom and then from bottom to top, before moving on to the next sheet. They were instructed not to look at the next sheet before finishing one sheet. Half of the participants started with the noun sheet, followed by the verb sheet. The other half of the participants started with the verb sheet and then the noun sheet. Both groups read the alternating sheet last. Each participant was recorded individually and the whole session lasted approximately 20 minutes.

Prosodic Analyses

Prosodic measures were conducted for the tokens of four of the pseudo-words: *mige, reule, vare, krale*. Using Praat (version 5.1.05) sound processing software (Boersma & Weenink, 2009), we first segmented the pseudo-words from onset to offset in both noun and verb sentences. There were 24 noun tokens and 24 verb tokens (i.e., 4 pseudo words x 3 function word contexts x 2 recordings) per participant, for a total of 96 tokens in each grammatical category. For each token, word duration, word mean pitch, and word mean intensity were measured. In addition, the vowel onset and offset for each word were marked. We then also measured the duration, mean pitch and mean intensity for the vowel of each target word.

Results and Discussion

Independent samples *t*-tests were conducted on the acoustical measures to compare all the noun uses versus the verb uses of the pseudo-word tokens. The results showed no significant difference in any of the comparisons. At the word level, durations for the targets used as nouns (mean = .635 s; SD = .091 s) and as verbs (mean = .617 s; SD = .098 s) did not differ, $t(183) = 1.312$, $p = .191$. Targets used as nouns were not different from their use as verbs in mean pitch (Nouns: mean = 214.105 Hz; SD = 30.101 Hz; Verbs: mean = 216.211 Hz; SD = 22.237 Hz), $t(182) = -.542$, $p = .588$. The values of mean intensity of targets in noun utterances (mean = 58.382 dB; SD = 9.512 dB) and in verb utterances (mean = 58.494 dB; SD = 8.2 dB) were not different, $t(183) = -.086$, $p = .931$.

Vowel analyses also yielded no acoustical differences for noun versus verb uses. Vowel durations for noun targets (mean = .299 s; SD = .073 s) and for verb targets (mean = .291 s; SD = .074 s) were similar, $t(183) = .694$, $p=.488$. Vowel mean pitch was not different for noun targets (mean = 212.176 Hz; SD = 29.821 Hz) and verb targets (mean = 214.663 Hz; SD = 24.641 Hz), $t(183) = -.620$, $p = .536$. Finally, vowel mean intensity was also not different for noun uses (mean = 60.303 dB; SD = 9.423 dB) versus verb uses (mean = 60.306 dB; SD = 7.915 dB), $t(183) = -.002$, $p = .998$. All analyses were two-tailed. The results are summarized in Table 2.

Acoustic measures	**Average values for noun uses**	**Average values for verb uses**	***p* (unpaired *t*-test, 2-tailed)**
Total word duration (s)	.635 (.091)	.617 (.098)	.191
Word mean pitch (Hz)	214.105 (30.101)	216.211 (22.237)	.588
Word mean intensity (dB)	58.382 (9.512)	58.494 (8.2)	.931
Vowel duration (s)	.299 (.073)	.291 (.074)	.488
Vowel mean pitch (Hz)	212.176 (29.821)	214.663 (24.641)	.536
Vowel mean intensity (dB)	60.303 (9.423)	60.306 (7.915)	.998

Table2. Acoustic measures (and standard deviations) across multiple exemplars of pseudo-words.

These results suggest that noun versus verb productions in our experiment were not different prosodically. Our utterances were short two-word combinations consisting of a function word followed by a content word (i.e., Determiner+Noun; Pronoun+Verb). Both noun utterances and verb utterances formed one phonological phrase, with the target words falling in the same position. This might be a reason why we found no prosodic differences between noun versus verb uses in Experiment 1. Natural speech productions, however, often contain longer utterances and may contain more than one prosodic phrase. Nouns and verbs can occupy distinct positions in prosodic phrases. In processing longer utterances listeners not only need to categorize speech materials to grammatical categories, but also have to segment the speech into word groups corresponding to syntactic constituents (e.g., noun phrases). In Experiment 2 we examined whether the longer utterances are produced as prosodic groups corresponding to syntactic constituents, and whether grammatical categories such as nouns and verbs are also prosodically marked. To test this question, we created pairs of utterances, each pair containing the same phonemic sequences but are distinct in grammatical categories as well as syntactic constituents. Two syntactic structures were used: 1) [Determiner+Adjective+Noun], and 2) [[Determiner+Noun]+[Verb]]. Pseudo-words were also used for content words in Experiment 2 to control for phonetic/phonological and lexical frequency factors.

EXPERIMENT 2

Methods

Participants and Stimuli

One new speaker, a female native Quebec-French-speaking adult from the Psychology Department of the Université du Québec à Montréal, served as the participant for Experiment 2. We used a subset of pseudo-words from Experiment 1 as our stimuli: *mige, krale*. These pseudo-words were the last word serving either as a noun or a verb in the two syntactic structures mentioned above: Structure 1) [Determiner+Adjective+Noun], i.e., a noun phrase, and Structure 2) [[Determiner+Noun]+[Verb]], i.e., a sentence consisting of a subject noun phrase and a verb phrase. We chose determiners in French (i.e., *un*–a, *ton*–your, *des*–some) as the first element for both syntactic structures. The pseudo-word *felli* was also created as the second word, i.e., an adjective in Structure 1 and a noun in Structure 2. Thus, the three-word sequences were intended to represent the two distinct structures, yielding phonemic homophones. For example, [*des+fellis+miges*] being Structure 1 and [[*des+fellis*]+[*migent*]] being Structure 2 were phonemically homophonous despite the spelling differences in French. In total, there were six ambiguous pairs: [*un+felli+mige*] - [[*un+felli*]+[*mige*]]; [*ton+felli+mige*] - [[*ton+felli*]+[*mige*]]; [*des+fellis+miges*] - [[*des+fellis*]+[*migent*]]; [*un+felli+krale*] - [[*un+felli*]+[*krale*]]; [*ton+felli+krale*] - [[*ton+felli*]+[*krale*]]; [*des+fellis+krales*] - [[*des+fellis*]+[*kralent*]]. We also created filler sentences including elicitation questions, each of which was followed by an answer with real French words (see Table 3). For the utterances of Structure 1, the question was "*Qu'est-ce-que tu vois?*" (What do you see?), and the answer was "*Ton petit chat.*" (Your little cat.). The same question was then used to elicit the target utterance answer (e.g., *Qu'est-ce-que tu vois? Ton felli mige.*). Similar question-answer sentence pairs were created for eliciting Structure 2. For example, an elicitation question with real French words "*Qui dort?*" (Who sleeps?) was followed by the filler answer "*Ton cheval dort. Il dort très fort*." (Your horse sleeps. He sleeps deeply.). Comparable questions and answers were used

for the target pseudo-word for Structure 2, e.g., "*Qui mige? Ton felli mige. Il mige très fort.*" (Who miges? Your felli miges. He miges deeply.). The use of questions and real word answers ensured that participants would produce the two distinct syntactic structures with the pseudo-words. We predicted that the prosodic cues in production should exist and distinguish the two structures. In particular, prosodic cues were expected to mark the syntactic phrasal boundary between the second and third words in Structure 2. In addition, we predicted that homophonous pseudo-words used as different grammatical categories may exhibit distinct prosodic properties.

Structure 1 (noun phrase)			**Structure 2 (noun phrase + verb)**		
Elicitation question	**Filler response**	**Target response**	**Elicitation question**	**Filler response**	**Target response**
Qu'est-ce-que tu vois?	*Ton petit chat.*	*Ton felli mige.*	*Qui dort?*	*Ton cheval dort. Il dort très fort.*	*Ton felli mige. Il mige très fort.*
Qu'est-ce-que tu vois?	*Des petits chats.*	*Des fellis miges.*	*Qui mige?*		*Des fellis migent.* *Ils migent très fort.*
Qu'est-ce-que tu vois?	*Un petit chat.*	*Un felli mige.*	*Qui mige?*		*Un felli mige. Il mige très fort.*
Qu'est-ce-que tu vois?	*Ton petit chat.*	*Ton felli krale.*	*Qui dort?*	*Ton cheval dort. Il dort très fort.*	*Ton felli krale. Il krale très fort.*
Qu'est-ce-que tu vois?	*Des petits chats.*	*Des fellis krales.*	*Qui mige?*		*Des fellis kralent. Ils kralent très fort.*
Qu'est-ce-que tu vois?	*Un petit chat.*	*Un felli krale.*	*Qui mige?*		*Un felli krale. Il krale très fort.*

Table 3. Materials used in Experiment 2.

Procedure

Recording apparatus was identical to Experiment 1. The participant was instructed to read the utterances presented on paper. One sheet presented the question-answer pairs for Structure 1 in a quasi-random order. The second sheet contained the question-answer pairs for Structure 2 in a quasi-random order. A question-answer pair with a real word answer was always listed before each question-answer pair involving pseudo-words. The participant was recorded while reading the materials multiple times, yielding a total of 45 target token utterances. The stimuli productions included 21 Structure 1 utterances with the pseudo-words and 24 Structure 2 utterances with the pseudo-words. Specifically, for the noun uses in Structure 1, there were 12 *mige* tokens (i.e., 4 tokens for each *mige/miges* noun phrase) and 9 *krale* tokens (i.e., 3 tokens for each *krale/krales* noun phrases). For the verb uses in Structure 2, there were 12 *mige* tokens (i.e., 4 tokens for each *mige/migent* sentence) and 12 *krale* tokens (i.e., 4 tokens for each *krale/kralent* sentence). There were 21 *felli* tokens used as an adjective in Structure 1, and there were 24 *felli* tokens used as a noun in Structure 2.

Acoustic Analyses

Prosodic measures were conducted on the tokens of the pseudo-words *felli, mige, krale*. As in Experiment 1, we used the analysis software Praat (version 5.1.05). We first segmented out the utterances of the intended target structures from the recording, for a total of 21 Structure 1 productions and 24 Structure 2 productions. For each token, word duration, word mean pitch, and word mean intensity were measured. Tokens intended as the noun (i.e., the last word) in Structure 1 utterances were statistically compared to those as a verb in Structure 2 utterances for each acoustic measure. In addition, we conducted the same acoustic measures for *felli,* comparing its tokens indented as an adjective in structure 1 and as a noun in structure 2. These comparisons of *felli* also served to show whether there were acoustical cues at the end of the second word marking the phrasal boundary in Structure 2 (i.e., [[Determiner+Noun]+[Verb]]) as opposed to the phrase-internal position in Structure 1. Additional measures were conducted to examine whether there was a boundary between the second and third word in Structure 2 but not in Structure 1. These included the pause duration (the lack of a pause would have the value of 0) at the end of the second word, the *m* duration of the word *mige,* as well as the mean pitch, mean intensity and duration of the last syllable *li* in *felli*. We predicted that the last word as nouns versus verbs should differ in duration, pitch and intensity. The word *felli* should also differ in these measures in Structure 1 versus Structure 2. We also predicted that the initial consonant in verb productions in Structure 2 should be longer, with a more distinct preceding pause, in comparison to the same consonant in Structure 1. Furthermore, the syllable immediately preceding the verb in Structure 2, i.e., *li* from *felli,* should exhibit boundary cues (in measures of duration, mean pitch and mean intensity of the syllable) distinct from the non-boundary *li* in Structure 1.

Results

Independent samples *t*-tests were conducted on the prosodic measures to compare all the noun uses in Structure 1 versus the verb uses in Structure 2, i.e., the third word *mige* or *krale*. The durations for the pseudo-word targets used as nouns (mean = .619 s; SD = .084 s) and as verbs (mean = .712 s; SD = .126 s) differed significantly, $t(43) = -2.855$, $p = .007$. The mean pitch of the noun use was significantly higher than that of the verb use (Nouns: mean = 239.907 Hz; SD = 12.408 Hz; Verbs: mean = 203.818 Hz; SD = 42.491 Hz), $t(43) = 3.750$, $p = .001$. The values measured for the mean intensity were also different for the noun use (mean = 64.391 dB; SD = 2.048 dB) versus for the verb use (mean = 61.878 dB; SD = 3.207 dB), $t(43) = 3.082$, $p = .004$.

The duration for the pseudo-words *felli* used as nouns in Structure 2 (mean = .596 s; SD = .099 s) was significantly longer than its use as an adjective in Structure 1: (mean = .411 s; SD = .038 s), $t(43) = 8.050$, $p = .000$. The mean pitch of *felli* in Structure 1 (mean = 227.114 Hz; SD = 19.657 Hz) differed significantly from *felli* in Structure 2 (mean = 214.035 Hz; SD = 16.950 Hz), $t(43) = 2.397$, $p = .021$. Mean intensity of *felli* production in Structure 1 (Structure 1: mean = 64.082 dB; SD = 1.534 dB) was marginally different from that in Structure 2 (mean = 63.271 dB; SD = 1.312 dB), $t(43) = 1.911$, $p = .063$.

The *m* in *mige* differed significantly in duration in Structure 1 utterances as opposed to Structure 2 utterances: (Structure 1: mean = .111 s; SD = .010 s; Structure 2: mean = .145 s; SD = .038 s), $t(22) = -3.033$, $p = .006$. Thus, *m* was longer in Structure 2, demonstrating acoustic strengthening at the onset of the new unit following a phrasal boundary. Moreover, the pause duration at the phrasal boundary following the second word in Structure 2 was salient (mean = .097 s; SD = .083 s), significantly longer

than the measure between the second and third words (i.e., phrase-internal) in Structure 1 (mean = .023 s; SD = .029 s), $t(43) = 3.884$, $p = .000$. In the latter case a small pause only appeared before the *k* segment in *krale*, since a brief silence closure was acoustically obligatory for the production of the voiceless stop consonant regardless of syntactic environment. In contrast, the extended pause duration in Structure 2 showed the influence of the syntactic phrasal boundary.

The *li* of *felli* was another measure related to the distinction of the two structures. The syllable duration of *li* was significantly lengthened at the phrasal boundary position (i.e., Structure 2: mean = .308 s; SD = .062 s) in comparison to the phrase-internal position (i.e., in Structure 1: mean = .188 s; SD = .016 s), $t(43) = 8{,}620$, $p = .000$. The mean pitch of *li* in Structure 1 (mean = 233.121 Hz; SD = 17.851 Hz) and in Structure 2 (mean = 222.013 Hz; SD = 29.818 Hz) did not differ significantly, $t(43) = 1.488$, $p = .144$. On the other hand, there was a significant difference for the mean intensity of *li* in Structure 1 (mean = 64.302 dB; SD = 2.391 dB) versus in Structure 2 (mean = 62.394 dB; SD = 2.645 dB), $t(43) = 2.523$, $p = .015$. All analyses were two-tailed and are summarized in Table 4.

Acoustic measures	**Average values in Structure 1**	**Average values in Structure 2**	***p* (unpaired *t*-test, 2-tailed)**
Duration of the third word (s)	.619 (.084)	.712 (.126)	.007
Mean pitch of the third word (Hz)	239.907 (12.408)	203.818 (42.491)	.001
Mean intensity of the third word (dB)	64.391 (2.048)	61.878 (3.207)	.004
Duration of *felli* (s)	.411 (.038)	.596 (.099)	.000
Mean pitch of *felli* (Hz)	227.114 (19.657)	214.035 (16.950)	.021
Mean intensity of *felli* (dB)	64.082 (1.534)	63.271 (1.312)	.063
Consonant duration of *m* (s)	.111 (.010)	.145 (.038)	.006
Pause duration (s)	.023 (.029)	.097 (.083)	.000
Duration of *li* (s)	.188 (.016)	.308 (.062)	.000
Mean pitch of *li* (Hz)	233.121 (17.851)	222.013 (29.818)	.144
Mean intensity of *li* (dB)	64.302 (2.391)	62.394 (2.645)	.015

Table 4. Acoustic measures (and standard deviations) across multiple exemplars in Experiment 2.

GENERAL DISCUSSION

The present study inquired whether there are acoustic cues in speech that mark syntactic structures. The existence of such cues would be particularly beneficial to infants, potentially enabling them to break into the syntax at the initial stage of acquisition. Our results showed that salient acoustic cues are produced in speech supporting the parsing of continuous speech into prosodic groups. The cues are present as long as utterances are sufficiently long. In Experiment 2 the boundary between the subject noun phrase and the main verb in Structure 2 was indicated by multiple cues, including phrasal final lengthening, increased pause, phrasal onset strengthening, etc.. The same words (i.e., the second and third words) in Structure 1 exhibited phrasal internal acoustic properties, with no prosodic break. These results are consistent with the production data of Millotte *et al*., (2007), who used real ambiguous content words in French. In our study all content words were non-words, eliminating

any possible effect of inherent production differences related to varying lexical frequencies. The acoustical differences were strictly caused by the syntactic constituent structures.

In addition to acoustic cues to syntactic groupings, Experiment 2 also showed that words of different grammatical categories were different in their acoustic properties. Thus, homophonous words used as nouns versus verbs differed in production, so did adjective-noun ambi-categorical words. On the other hand, in Experiment 1, which used two-word short utterances, no acoustic difference was found for ambi-categorical words. The short utterances in the two syntactic structures did not differ in their prosodic structure, both forming one phonological phrase according to the idea of prosodic phonology (Nespor & Vogel, 1986; Selkirk, 1982). It appears that whether or not grammatical categories of content words are acoustically distinguished in production is tied to prosodic groupings. There may not be independent acoustic cues to distinct grammatical categories for different content word categories. It should be noted that these results differ from those of Conwell and Morgan (2012) in English and Shi and Moisan (2008) in French, both of which showed acoustic cues to nouns versus verbs. The non-words serving as both nouns and verbs in Shi and Moisan were bi-syllabic and were embedded in longer, rhythmically controlled sentence. Thus, the stimuli conditions of that study and the present study differed significantly and seem to have had an impact on the results. The most important difference between the present study and those of Conwell and Morgan and Shi and Moisan is that the latter two analyzed parental speech to infants whereas the present study involved adult-directed speech. It is possible that infant-directed speech contains more salient acoustic cues to grammatical categories of words than do adult-directed speech.

It is striking that the non-infant-directed speech in the present study yielded many acoustic cues to syntactic groupings (Experiment 2). Bernstein Rather (1986) reported that prosodic cues to clause boundaries are more salient in infant-directed speech than in adult-directed speech. Our study showed that smaller constituents such as noun phrases are also robustly marked in adult-directed speech.

In conclusion, our results suggest that speech production contains acoustic cues to syntax, especially to constituent groupings. Such cues may affect the online syntactic processing in adults. They may particularly impact early acquisition, as they can bootstrap infants into very initial syntactic analysis. Indeed, infants during the first year of life have been shown to parse clauses and phrases using prosodic cues (e.g., Nazzi *et al*., 2000; Soderstrom *et al*., 2003). Given that infants this young have either no or little vocabulary knowledge, it is plausible that acoustic cues are the primary means that infants rely on for initial syntactic acquisition.

ACKNOWLEDGMENTS

This study was supported by funding from NSERC and SSHRC to the third author.

REFERENCES

Bernstein Ratner, N. (1986). Durational cues which mark clause boundaries in mother-child speech. *Journal of Phonetics*, v. 14, n.2, p. 303-309.

Boersma, P., & Weenink, D. Praat: Doing phonetics by computer (version 5.1.05) [computer program]. Available at: <www.praat.org>

Cassidy, K., & Kelly, M. (1991). Phonological information for grammatical category assignments. *Journal of Memory and Language*. v. 30, p. 348-369.

de Carvalho, A., Dautriche, I., & Christophe, A. (2014). Phrasal prosody constrains online syntactic analysis in two-year-old children. In: *Boston University Conference on Language Development*, 39. USA.

Conwell, E., & Morgan, J. L. (2012). Is it a noun or is it a verb? Resolving the ambicategoricality problem. *Language Learning Development*, v. 8, n. 2, p. 87-112.

Cooper, W. E., & Paccia-Cooper, J. (1980). *Syntax and speech*. Cambridge: Harvard University Press.

Cooper, W. E., & Sorensen, J. M. (1981). *Fundamental frequency in sentence production.* New York: Springer-Verlag.

Cutler, A. (1993). Phonological cues to open- and closed-class words in the processing of spoken sentences. *Journal of Psycholinguistic Research*, v. 22, p. 109-131.

Dautriche, I., Cristia, A., Brusini, P., Yuan, S., Fisher, C., & Christophe, A. (2014). Toddlers default to canonical surface-to-meaning mapping when learning verbs. *Child Development*, v. 85, p. 1168-1180.

Kelly, M. H. (1992). Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. *Psychological Review*, v. 99, p. 349-364.

Millotte, S., Wales, R. & Christophe, A. (2007). Phrasal prosody disambiguates syntax. *Langage and Cognitive Processes*, v. 22*,* n. 6, p. 898-909.

Monaghan, P., Christiansen M. H., & Chater, N. (2007). The phonological-distributional coherence hypothesis: Cross-linguistic evidence in language acquisition. *Cognitive Psychology*, v. 55, p. 259–305.

Nazzi, T., Kemler Nelson, D. G., Jusczyk, P. W., & Jusczyk, A. (2000). M. Six-month-olds' detection of clauses in continuous speech: Effects of prosodic well-formedness. Infancy, v. 1, n. 1, p. 123-147.

Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. The Netherlands: Foris, Dordrecht.

Selkirk, E. (1982). *The syntax of words.* Cambridge: MIT Press.

Shi, R., & Moisan, A. (2008). Prosodic cues to noun and verb categories in infant-directed speech. In *Boston University Conference on Language Development*, 32. USA.

Shi, R., Morgan, J. L., & Allopenna, P. (1998). Phonological and acoustic bases for earliest grammatical category assignment: A cross-linguistic perspective. *Journal of Child Language*, v. 25, p. 169-201.

Shi, R., & Werker, J. F. (2003). Basis of preference for lexical words in six-month-old infants. *Developmental Science*, v. 6, n. 5, p. 484-488.

Shi, R., & Werker, J. F. (2001). Six-month-old infants' preference for lexical words. *Psychological Science*, v. 12, n. 1, p. 70-75.

Shi, R., Werker, J. L., & Morgan, J. L. (1999). Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition*, v. 72, p. B11-21.

Soderstrom, M., Seidl, A., Kemler Nelson, D. G., & Jusczyk, P. W. (2003). The prosodic bootstrapping of phrases: Evidence from prelinguistic infants. *Journal of Memory and Language*, v. 49, p. 249-267.

Soreno, J. A., & Jongman, A. (1990). Phonological and form class relations in the lexicon. *Journal of Psycholinguistic Research*, v. 19, p. 387–404.

Van de Weijer, J. (1998). *Language input for word discovery*. Thesis (doctorate Degree).

Recebido em: 28/06/2015
Aceito em: 03/07/2015