Teaching listening in L2:
A successful training method using the word-spotting task

Delvaux Véronique[1,2], Huet Kathy[2], Calomme Mélanie[2], Piccaluga Myriam[2], Harmegnies Bernard[2],

[1]FNRS, [2]Institut de Recherche en Sciences et Technologies du Langage, UMONS, Belgium
delvaux@umons.ac.be

ABSTRACT

In this study, we assessed the efficiency of a learning sequence based on both implicit (using the word-spotting task) and explicit teaching of word boundary detection in an EFL classroom, using two cues: lexical stress and phonotactic constraints. Results showed that: (i) students from the experimental group achieved significantly more progress from pre-test to post-test than students from the control group; (ii) low-skilled listeners most benefited from the learning sequence; (iii) both explicit instruction (and related exercises) targeting segmentation subskills and practice using the word-spotting task were beneficial to the EFL learners.

Keywords: L2, listening, English as a Foreign Language, word-spotting task, classroom.

1. INTRODUCTION

Listening in a foreign language is a complex task, which ranges from perception to comprehension and requires the interaction between top-down and bottom-up cognitive processes partly mediated by attention and memory mechanisms.

As stated by Flowerdew and Miller [1], for years "listening has been treated as the Cinderella of the four macro-skills: speaking, listening, reading, and writing." Indeed, listening comprehension skills are typically assessed rather than taught in language classrooms [2,3]. L2 learners generally perform better in reading than in listening comprehension [4,5], so that some students may develop anxiety and/or loss of motivation for listening tasks [6], a phenomenon which is well-known of foreign language teachers, although not easily overcome.

In the last two decades, a growing interest for listening in L2 research has mainly focussed on (meta-)cognitive listening strategies (for a review, see [7]), so that a strong relationship between differences in strategy use and L2 listening proficiency is now established [8-13]. However, the efficiency of improving the listening skills of L2 learners − in particular, of low-proficiency learners − by explicitly teaching them listening strategies has been recently called into question, in the context of the "extensive listening" method [14,15].

Alternatively, a couple of researchers have considered specific training targeting the speech signal decoding processes as a potential tool to improve nonnative listening skills, making use of the "word-spotting task" [16,17]. When performing this task, listeners must detect real words which are embedded in non-word context. The word-spotting task is typically used in psycholinguistic research in order to study the structural factors influencing the segmentation of continuous speech in native and nonnative listeners [e.g.17,18,19].

Recently, Cutler and Shanley have shown that the task meets the basic requirements to be used as a training procedure in L2 listening [17]. Indeed, while primarily focussing on segmentation processes (vs. further syntactic and semantic processing), the word-spotting task provides a good balance between ecological validity from the listener's point of view and control of the speech material from the experimenter/teacher's point of view. However, to our knowledge, there is no report in the literature of the implementation of the word-spotting task as a training procedure in an L2 classroom, as is the case in the present paper.

In this study, we assessed the efficiency of a learning sequence based on both implicit (using the word-spotting task) and explicit teaching of word boundary detection in an EFL (English as a Foreign Language) classroom, using two cues that have been proved to be useful to native English speakers [19,20]: (i) lexical stress, and (ii) phonotactic constraints. Namely, teenage French EFL learners were taught to exploit the fact that a majority of English words exhibit word-initial lexical stress (i), and that many three-consonant clusters are illegal word-medially, thus signalling the presence of a word boundary (ii). Moreover, we documented the attitudes and self-reported practices of the participants in relation with listening comprehension before intervention, as well as their opinion on the usefulness of the learning sequence after intervention, and confronted these informations with the learners' performances in the word-spotting tasks.

2. MATERIAL AND METHODS

2.1. Stimuli

The stimuli consist in 3 lists (one for the pre- and post-test; two for the learning sequence) of 90 trisyllabic strings made of real words and/or non-words. Real words were selected based on the active vocabulary shared by the participants. Each list comprises 3 repetitions of: (i) 5 trisyllabic non-words (e.g. [trəpɪnʌk] (ii) 5 trisyllabic words (e.g. [jestərdeɪ]; (iii) 8 disyllabic words, e.g. [brʌðər], each repetition being preceded by a different non-word monosyllabic context (e.g. [hjustbrʌðər, trænkbrʌðər, tʃɪəldbrʌðər]); (iv) 12 monosyllabic words, each preceded by a different non-word disyllabic context (e.g. [mɪnəltfɪʃ, kɑbɑnsfɪʃ, rətɪŋfɪʃ]).

Therefore, in this corpus all strings are trisyllabic (but there are more monosyllabic than trisyllabic *words*, as in the English lexicon), and each syllable may or may not coincide with a word onset. Word onsets are signalled by two concurrent cues: word-initial lexical stress and the presence of a consonant cluster which would be illegal word-medially, and is made of a legal two-consonant coda cluster (/sp/, /sk/, /st/, /kt/, /ld/, /nk/, /tʃ/, /nd/, /nt/, /lt/, /ns/, /ng/) at the offset of the nonsense context, followed by the consonantal onset of the real word.

In order to include speaker variability in the training speech material (which has been shown to improve perceptual training in L2: [24]), the corpus was recorded by four female speakers of British English, who were instructed to avoid pausing between nonsense syllables and target words, but to make sure that the real word carried the main stress (as in English adjective+noun combinations).

2.2. Participants

Participants were 46 Belgian French highschool students aged 13-14 who had been engaged in an EFL course for a year and a half, at a rate of four hours a week. They composed three classes, from which class C (13 students, 5 male, 8 female) was randomly selected as the experimental group.

2.3. Design

The study followed a pre-test/post-test design.

2.3.1 Pre-test and post-test

The pre-test and post-test were exactly similar. They were administered to all 46 participants at an interval of 4 weeks. They consisted in a word-spotting task based on the first of the three 90 strings lists described above (voice#1 only). In order to adjust the methodology to the everyday classroom conditions, the stimuli were played twice to the whole class using a CD player and (good quality)

external loudspeakers. Students were instructed to write the words they detected on a sheet of paper (spelling errors were not taken into account for the computation of the word detection scores).

2.3.2. Treatment

The treatment was administered to the experimental group only. It consisted in a learning sequence made of 9 one-hour sessions (Table 1), administered twice or thrice a week during 4 weeks, in place of the usual EFL course. In these sessions, there was a progression from: (i) individual exercises targeting specific subskills enabling efficient segmentation (syllabification, stress detection) followed by a joint correction and a shared structuring of some basic theoretical knowledge (on English-specific phonotactic constraints and rules of lexical stress); to (ii) practice of segmentation based on the word-spotting task, using the last two lists of stimuli described above. In both types of activities, additional voices pronouncing the speech materials were gradually introduced.

Table 1. Summary of the 9 sessions composing the learning sequence.

Session#	Main task	Speech material	Voices
1	syllabification	30 words	1
2	syllabification, stress detection	30 words	1
3	word recognition, stress detection	10 sentences	1
4	word recognition, stress detection	10 sentences	3,4
5	word spotting	45 strings	3
6	word spotting	45 strings	2,3
7	word recognition, stress detection	10 sentences	2,3,4
8	word spotting	45 strings	3,4
9	word spotting	45 strings	2,3,4

2.3.3. Questionnaires

Two customized questionnaires were filled out by the participants. The 'initial questionnaire' was filled out by all 46 students before the pre-test. It was made of 23 questions regrouped in three sections ('before listening, 'while listening', 'after listening') collecting information on the students' attitudes (motivation, anxiety, etc.) towards L2 listening comprehension tasks and the strategies they report using in such tasks. The 'final questionnaire' was filled out by the experimental group only, after the post-test. It was made of 5 questions dealing with the students' opinion on whether, and how, their listening skills might have improved due to the treatment. For both questionnaires, participants responded using a four-level Likert-type scale.

2.4. Data processing

Data were collected (in written form), preprocessed and digitized by the third author. Raw scores from the 90-item- (pre&post-test) and 45- item- (treatment) word-spotting tasks were transformed into correct responses rates (%) (below: 'scores').

3. RESULTS

3.1. Treatment effect: pre-test to post-test evolution

In order to assess the treatment effect, a repeated-measure ANOVA[1] was first carried out, with Score as the dependent variable, Time (pre-test vs. post-test) as a within-subject factor and Class (A vs. B vs. C) as a between-subject factor. The analysis showed that both Time ($F_{(1,43)}$=64,185; $p<0,001$) and Class ($F_{(2,43)}$=10,855; $p<0,001$) had a significant effect on scores, and that there was a significant interaction between Time and Class ($F_{(2,43)}$=23,941; $p<0,001$). Fig.1 illustrates the interaction effect by displaying the scores achieved in the pre-test vs. the post-test across the 3 classes of EFL students.

Fig.1. Proportion of correct responses in the word-spotting task (pre-test vs. post-test) across class.
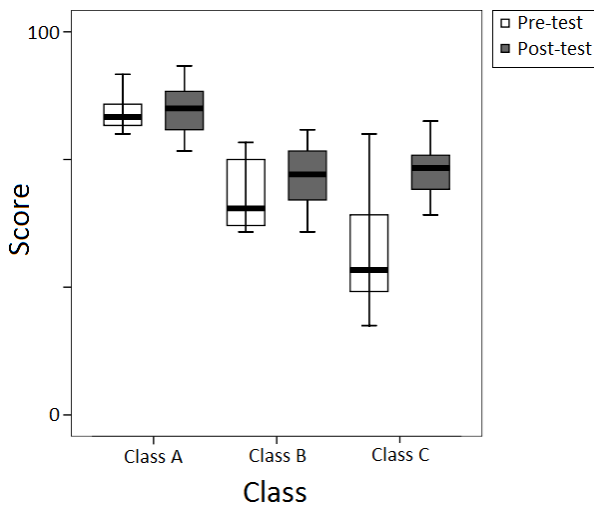
It is apparent from Fig.1 that the original sampling of participants was quite unbalanced, in that the three classes of EFL students exhibited contrasted performances in the pre-test[2].

As illustrated in Fig.1, from pre-test to post-test students from both class B and class C improve their performance on average, the more so for students of class C, whose initial performances (in the pre-test) were significantly poorer. Thus, a complementary analysis was conducted. Absolute and relative gains were computed:

$$(1)\ \text{Absolute gain} = \text{Score}_{(Post-test)} - \text{Score}_{(Pre-test)}$$

$$(2)\ \text{Relative Gain} = \frac{\text{Absolute gain}}{100 - \text{Score}_{(Pre-test)}}$$

An ANOVA was carried out with Relative Gain as the dependent variable, and Class (A, B, C) as independent variable, revealing a significant effect of Class on relative gains ($F_{(2,43)}$ =10,794; $p<0,001$). Post Hoc tests showed that there was no significant difference in terms of Relative Gain between Class A (Mean: 0,07) and Class B (Mean: 0,13), which form the control group, whereas Relative Gain was significantly higher for Class C (Mean: 0,36), the experimental group. Thus, overall results showed that only students from the experimental group achieved significant progress, even controlling for the fact that the margin for improvement was larger for them than for control groups.

Note, however, that within-class variability was especially high in the pre-test − and notably reduced in the post-test − in class C (Fig.1). Fig.2 displays absolute gains as a function of pre-test scores for Class C. The significant, negative correlation coefficient (r = -0,705, p= 0,007) indicates that the students from class C who most benefited from the treatment were those who performed the poorest in the pre-test.

Fig.2. Absolute gains as a function of Pre-test scores (Class C).
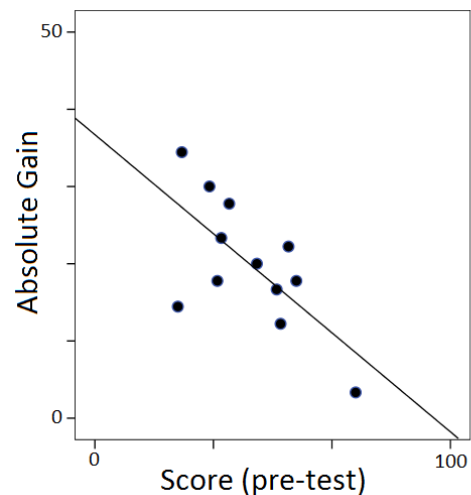
3.2. Within-treatment evolution

In order to explore how the treatment helped the students to improve their segmentation skills, we compared their scores in the successive word-spotting tasks: in the pre-test, during the treatment, and in the post-test (Fig.3). Fig.3 suggests that one major improvement occurred between the pre-test and the first word-spotting task of the learning sequence (session#5). From session#5 to session#9, performances remained approximately at the same level on average, although additional voices were

progressively introduced (cf. Table 1). When reverting to one voice in the post-test, variability in the performances substantially decreased, which is an indicator of improvement at the class level.

Fig.3. Scores to the successive word-spotting tasks (pre-test/treatment/post-test) (experimental group).
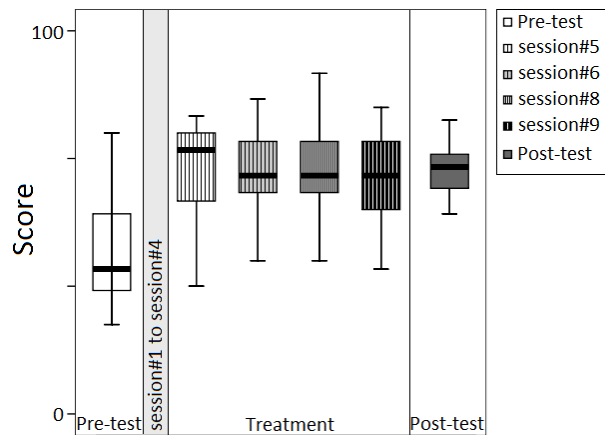
3. Questionnaires

The initial questionnaire was built to collect information about attitudes and self-reported practices related with listening comprehension from a pool of Belgian EFL learners aged 13-14. By lack of space, these results will not be detailed here. In the following, we focus on the relation between responses to the questionnaires and individual performances in word-spotting tasks.

Two stepwise linear regressions were performed, with the responses to, respectively, the 23 questions of the initial questionnaire/the 5 questions of the final questionnaire as independent variables, and the scores (pre-test)/relative gains as dependent variables. The associated output models revealed that: (i) the students who performed best at the pre-test were those who were the most optimistic beforehand ("Before listening, I am very focused, I tell myself I'm going to succeed") and who reported making notes of the keywords in English while listening; (ii) the students who felt like they had been given helpful tools during the treatment achieved more progress between pre- and post-test.

4. DISCUSSION

In this paper, we presented a learning sequence targeting 'bottom-up', speech signal decoding, processes (vs. listening strategies) and, using a pre-test/post-test design, we showed that it could be successfully implemented in the context of an EFL classroom (i.e. over a relatively short period of time, using no student-specific activities and little feedback). Specifically, all participants improved

from pre-test to post-test, but students who benefited from the treatment achieved significantly more progress, even controlling for the fact that, by happenstance (cf. Note 1), their margin for improvement was larger.

Building on pioneer work from Cutler and Shanley [17] and Al-Jasser [16], we devised an original methodology which used the word-spotting task as a performance assessment procedure as well as (part of) the training method. Since lexical stress and phonotactic constraints were extensively shown to be exploited by native and nonnative listeners in listening tasks (e.g. [16-22]), we combined both cues in our stimuli. Similarly, the learning sequence alternated individual exercises and theory building, implicit and explicit learning, activities targeting specific segmentation subskills and general practice using the word-spotting task, while several voices were progressively added in the speech materials. Results suggest that this *combination* of didactic means and phonetic cues was efficient (Fig.3), but the specific contribution of each modality, and the underlying processes at work, certainly deserve further investigation.

From an SLA perspective, two specific results merit particular attention. First, within the experimental group, the poorest listeners (as assessed by the pre-test) were those who made the most progress in the post-test. This is promising for potential applications in instructional contexts because it has been claimed that lower proficiency EFL learners little benefit from explicit teaching of top-down learning strategies [14]. Second, there were meaningful relations between the performances in the word-spotting task and the students' answers to the questionnaires. Students who were (self-reportedly) focused on the listening task and approached it with a positive attitude performed better, which resonates with previous results showing that, in contrast, when listeners are anxious, their ability to concentrate falters, and comprehension declines [23]. The fact that actual improvement from pre- to post-test was related with the impression that helpful tools were provided during the treatment is an asset of the method, because it may sustain the learners' metacognitive skills and motivation.

To conclude, the learning sequence we devised allowed significant improvement of the French EFL learners' segmentation skills, as assessed by their performances in a word-spotting task. Further work will have to assess the time robustness of this improvement, its transferability to connected speech, and its relations with other (syntactic, semantic) processes involved in L2 listening comprehension.

5. ACKNOWLEDGMENTS

6. REFERENCES

[1] Flowerdew, J., Miller, L. 2005. *Second Language Listening: Theory and Practice*. New York: Cambridge University Press.

[2] Sheerin, S. 1987. Listening comprehension: teaching or testing? *ELT Journal* 41(2), 126-31.

[3] Field, J. 2008. Guest editor's introduction. Emergent and divergent: A view of second language listening research. *System* 36(1), 2-9.

[4] Lund, R. J. 1991. A taxonomy for teaching second language listening. *Foreign Language Annals* 23(2), 105–115.

[5] van Zeeland, H. 2013. L2 vocabulary knowledge in and out of context: is it the same for reading and listening? *Australian Review of Applied Linguistics* 36(1), 52-70.

[6] Xu, F. 2011. Anxiety in EFL Listening Comprehension. *Theory and Practice in Language Studies* 1(12), 1709-1717.

[7] Berne, J. E. 2004. Listening Comprehension Strategies: A Review of the Literature. *Foreign Language Annals* 37, 521–531.

[8] Vandergrift, L. 1997. The comprehension strategies of second language (French) listeners: A descriptive study. *Foreign Language Annals* 30(3), 387-409.

[9] Vandergrift, L. 2007. Recent developments in second and foreign language listening comprehension research. *Language Teaching* 40(3), 191–210.

[10] Goh, C. 2002. Exploring listening comprehension tactics and their interaction patterns. *System* 30(2), 185-206.

[11] Chao, Y-G., Chien, L-Y. 2005. College English majors' listening strategies and difficulties while taking TOEFL. *Selected papers from the 14th International Symposium on English Teaching* Taipei, 292-301.

[12] Chang, A. 2009. EFL Listeners' Task-based Strategies and Their Relationship with Listening Performance. *TESL-EJ* 13(2), 1-28.

[13] Lynch, T. 2009. *Teaching Second Language Listening*. Oxford: Oxford University Press.

[14] Renandya, W., Farrell, T. 2011. Teacher, the tape is too fast! Extensive listening in ELT. *ELT Journal* 65(1), 52-59.

[15] Blyth, A. 2012. Extensive Listening vs. Listening Strategies: Response to Seigel. *ELT Journal* 66(2), 236-239.

[16] Al-jasser, F. 2008. The effect of teaching English phonotactics on the lexical segmentation of English as a foreign language. *System* 36(1), 94-106.

[17] Cutler, A., Shanley, J. 2010. Validation of a training method for L2 continuous-speech segmentation. *Proceedings of Interspeech 2010* Makuhari, 1844-1847.

[18] Weber, A., Cutler, A. 2006. First-language phonotactics in second-language listening. *Journal of the Acoustical Society of America* 119(1), 597-607.

[19] Lugt, A. van der. 2001. The use of sequential probabilities in the segmentation of speech. *Perception and Psychophysics* 63, 811-823.

[20] Cutler, A., Norris, D. G. 1988. The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance* 14, 113-121.

[21] Cutler, A., Wales, R., Cooper, N., Janssen, J. 2007. Dutch listeners' use of suprasegmental cues to English stress. In: J. Trouvain, & W. J. Barry (eds.), *Proceedings of the 16th International Congress of Phonetics Sciences*. Dudweiler: Pirrot, 1913-1916.

[22] Schwab, S., Llisterri, J. 2011. Are French speakers able to learn to perceive lexical stress contrasts? *Proceeding 17th ICPhS* Hong Kong, China, 1774-1777.

[23] Bloomfield, A., Wayland, S. C., Rhoades, E., Blodget, A., Linck, J., Ross, S. 2010. What makes listening difficult? Factors affecting second language listening comprehension. *University of Maryland Center for Advanced Study of Language.* http://www.dliflc.edu/file.ashx?path=archive/documents/CASL_study_FINAL_Lit_Rev.pdf

[24] Lively, S.E., Logan, J.S., Pisoni, D.B. 1993. Training Japanese listeners to identify English /r/ and /l/: II. The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America* 94, 1242-1255.

[25] Calomme, M. 2014. *Comment améliorer la segmentation lors de la compréhension orale en anglais chez des élèves francophones du premier degré de l'enseignement secondaire ?* Unpublished master's thesis, University of Mons, Belgium.

---

[1] This analysis and subsequent ANOVAs in 3.1 were confirmed using non-parametric Kruskall-Wallis tests and Mann-Whitney pairwise comparisons.

[2] Note that this across-class variability in the pre-test was unexpected since information collected prior to the experiment ensured that EFL grades were similar in the three classes (Class A: Mean=66%, SD=11% ; Class B: Mean=61%, SD=11%; Class C: Mean=66%, SD=18%). A one-way ANOVA yielded no effect of Class on grades: $(F_{(2,42)}=0,799; p=0,457$ n.s.). After debriefing with the teachers, it appeared that Class A was indeed known as a "strong" class and Class C as a "weak" class, and that their similar overall grades in EFL may have resulted from the teachers' tendency to rank the students of a single class against each other, rather than against a fixed set of learning goals.