

Une archive Internet des sons de la langue roumaine

Suite à un travail de plusieurs années d'une équipe mixte, les étudiants et les chercheurs ont à leur disposition un site web à libre accès, pour apprendre la prononciation de la langue roumaine, pour faire des études comparatives entre la langue roumaine et d'autres langues, surtout latines, pour développer des systèmes de parole synthétique, ou pour des myriades d'autres applications en linguistique, phonétique, technologie de la langue, sociolinguistique ou médecine. L'adresse du site trilingue (roumain, et en grande partie en anglais et en français) est http://www.etc.tuiasi.ro/sibm/romanian_spoken_language/index.htm.

En outre d'aide à l'enseignement et l'apprentissage de la phonation de la langue roumaine dans d'autres pays, l'archive fournit, aux chercheurs ainsi qu'aux étudiants, une base de données phonétiques de la langue roumaine, structurée selon des critères précis, mais aussi enregistrée, documentée et annotée en conformité avec une méthodologie bien définie. Les fichiers sont en formats 16 bits et 24 bits .wav et .ogg. Le site web contient plus de 2000 enregistrements de syllabes, mots et propositions, avec diverses tonalités, prononcées avec ou sans charge émotionnelle. La base de données contient des enregistrements des voix professionnelles (prononciations «élaborées»), mais aussi des voix normales («la voix des personnes communes»), de la région du Nord-Est de la Roumanie, sans accent dialectal marqué.

Mais le site ne contient pas seulement une archive, il représente même plus qu'un corpus, car il inclut des études statistiques vastes visant les sons de la langue roumaine, et même des enregistrements effectués par des personnes ayant diverses pathologies (voire la section «Gnatosonie et Gnatophonie»).

Le volet «émotions» contient des pages avec enregistrements des voix émotive, donc prononcées avec une certaine émotion – joie, furie, tristesse – ou sans émotion, ton neutre. Nous avons utilisé ces pages pour extraire les paramètres caractéristiques de chaque émotion dans la langue roumaine et pour les comparer avec les résultats obtenus pour d'autres langues. Une section spécialisée est représentée par l'étude réalisée sur les aspects phonétiques en liaison avec les aspects sémantiques; le cas des constructions avec sujet double dans la langue roumaine est traité de manière exhaustive. L'existence du

sujet double est un aspect assez controversé, entre autre parce que cette construction n'existe pas dans les langues européennes modernes. Nous croyons avoir pu démontrer, en corroborant les différences phonétiques entre appositions et expressions soupçonnées d'être sujet double, que l'hypothèse du sujet double semble véridique.

Les enregistrements peuvent servir aux utilisateurs pour étudier les différences ou les similitudes entre la langue roumaine et d'autres langues latines.

Le site offre aussi une collection d'instruments d'analyse du signal vocal (les fichiers exécutables et aussi des lignes de programmes) conçus par notre collective. Les instruments existant sont: extracteur du spectre du signal (pour voir les composantes fréquentielles du signal), filtres digitaux, détecteur du ligne prosodique (la trace du fréquence fondamentale) par plusieurs méthodes, bloc de décision prosodique et un extracteur du formants. On rappelle que les formants sont utilisés souvent comme paramètres dans la reconnaissance du locuteur ou des phonèmes. Les traces des formants peuvent aider à corriger les prononciations déficitaires, en logopédie. Toute personne peut enregistrer sa voix et utiliser les outils accessibles sur le site pour l'étudier à son gré.

Le site a été créé par la coopération de l'Institut d'Informatique Théorique de l'Académie Roumaine, avec l'Université Technique "Gheorghe Asachi" de Iași - Centre pour l'Excellence dans la Recherche "CERFS" (directeur Prof. Horia-Nicolai Teodorescu) et avec le Groupe pour le Traitement de la Parole, de la Faculté d'Informatique, Université "Al. I. Cuza" Iași.

1. Teodorescu H.N., Tandabăț D., Feraru M., Zbancioc M., Luca R.(2006), *A corpus of the sounds in the Romanian spoken language for language-related education*. Chapitre six de Carlos Periñán Pascual (Editor), "Revisiting Language Learning Resources", Cambridge Scholars Publishing (CSP), UK, ISBN 1-84718-156-2; ISBN 13: 9781847181565, pp. 73-89, 2007
2. Horia-Nicolai Teodorescu, Monica Feraru, *Analyzing emotions in spoken Romanian*, Proceedings Romanian Academy, Série A, Volume 8, No. 2/2007, pp. 161-168
3. H.N. Teodorescu, Diana Trandabăț, *The Prosody of the Double-Subject Sentences in Romanian* Revue Roum. Linguistique (to appear 2009)
4. Horia-Nicolai Teodorescu, Monica Feraru, *A study on Speech with Manifest Emotions*, 10th International Conference on Text, Speech and Dialogue, TSD 2007,

Pilsen, Czech Republic, 3 – 7 Septembre, 2007, Lecture Notes in Computer Science, Springer Verlag, vol. 4629/2007, pp. 254-262, ISBN 978-3-540-74627-0

5. Horia-Nicolai Teodorescu, Diana Trandabăț, *Appositions Versus Double Subject Sentences - What Information the Speech Analysis Brings to a Grammar Debate*, 10th International Conference on Text, Speech and Dialogue, TSD 2007, Pilsen, Czech Republic, 3 – 7 Septembre, 2007, Lecture Notes in Computer Science, Springer Verlag, vol. 4629/2007, pp. 286-293, ISBN 978-3-540-74627-0