Sandrine Brognaux — Aspirante FNRS – Cental UCL
Sophie Roekhaut — Cental UCL

Train&Align : Outil d'alignement phonétique de corpus oraux

Avec l'émergence de la linguistique de corpus est apparue la nécessité d'annoter, à divers niveaux linguistiques, des corpus de taille conséquente. Ainsi, l'analyse prosodique ou phonétique de données orales requiert généralement un alignement temporel du son et de sa transcription phonétique.
Plusieurs logiciels permettent d'encoder des couches d'annotation synchronisées avec le son. Praat[1] est l'un des plus répandus. Le son et sa transcription phonétique peuvent y être alignés manuellement. Ceci offre un résultat précis mais requiert un temps considérable. Plusieurs outils ont donc été développés afin de proposer un alignement automatique de corpus oraux (EasyAlign[2], SPPAS[3], P2FA[4], SPRAAK[5], etc.). Ils sont très similaires en terme de fonctionnement et produisent un alignement lisible dans Praat[1] qui doit généralement être vérifié manuellement.
Ces systèmes se basent tous sur une modélisation du langage par modèles de Markov cachés (HMM), implémentés par des outils tels que HTK[6]. L'entraînement sur un corpus manuellement aligné ou non permet d'extraire un modèle qui en décrit la réalisation des différents phonèmes. L'utilisation de ce modèle permet ensuite d'aligner un nouveau corpus. La qualité de l'alignement dépend du modèle ainsi que de l'adéquation entre le corpus d'entraînement et celui à aligner. Ainsi, un modèle entraîné en partie sur des données manuellement alignées permettra généralement un alignement plus précis.

Les systèmes précités fournissent des modèles préalablement entraînés ainsi que les méthodes qui permettent d'aligner un nouvel enregistrement sur base de ces modèles. Leur faiblesse majeure est de ne pas permettre l'entraînement de nouveaux modèles ou la modification des modèles existants.
Ce choix a plusieurs conséquences néfastes :

- Les modèles n'existant que pour un nombre limité de langues, il n'est pas possible d'aligner des corpus dans toutes les langues.
- Les modèles ont été entraînés sur un certain type de parole. Ils ne fournissent donc pas toujours un alignement concluant pour une parole de type différent.
- Un avantage des modèles HMM est qu'ils peuvent être adaptés au locuteur ou au type de parole à aligner, ce qui améliore la qualité de l'alignement. Ceci n'est pas possible dans ces logiciels car ils n'offrent pas la possibilité d'adapter les modèles existants.

L'outil que nous présentons permet également d'aligner un son et sa transcription phonétique, dans un format compatible avec Praat[1]. Il se base également sur une modélisation par HMM. Cependant, notre outil permet également à l'utilisateur d'accéder à la phase d'entraînement. Il peut ainsi créer ses propres modèles et les adapter en fonction du corpus à aligner. Ceci apporte plusieurs avantages :

- Le système ne se base pas sur des modèles pré-entraînés mais s'entraîne directement sur le corpus à aligner. Il peut ainsi **s'adapter à toutes les langues et tous les types de parole**, pourvu qu'un corpus non aligné de taille suffisante soit fourni.
- L'alignement automatique ne nécessite pas d'alignement manuel partiel préalable. Cependant, si un alignement de quelques minutes existe ou peut être réalisé, il est possible de l'utiliser afin d'améliorer l'entraînement du modèle.
- Une méthode d'adaptation est proposée. Celle-ci peut considérablement améliorer le résultat dans le cas d'un corpus comprenant plusieurs locuteurs ou plusieurs types de parole.

Enfin, un mode d'évaluation est prévu si une partie alignée existe. Cette évaluation permet alors d'optimiser le choix des paramètres d'entraînement avant d'aligner l'intégralité du corpus.
Cet outil devrait être distribué prochainement.

[1] Boersma, P. & D. Weenink. 2009. Praat : doing phonetics by computer, http://www.praat.org
[2] Goldman, J.-Ph., 2011. EasyAlign : an automatic phonetic alignment tool under Praat. INTERSPEECH 2011, Florence, Italy, August 27-31, 3233-3236.
[3] Bigi, B. & D. Hirst. 2012. Speech Phonetization Alignment and Syllabification (SPPAS) : a tool for the automatic analysis of speech prosody, *Speech Prosody,* Shangai, May 2012.
[4] Yuan, J. & M. Liberman. 2008. Speaker Identification on the Scotus Corpus. *Proceedings of Acoustics, 2008,* 5687-5690.
[5] Demuynck, K., Roelens, J., Van Compernolle, D. & P. Wambacq. 2008. SPRAAK : an open source « Speech Recognition and Automatic Annotation Kit », *Interspeech 2008*, p. 495.
[6] Young, S., Kershaw, D., Odell, J., Ollason, D., Valtchev, V. & P. Woodland, The HTK Book (for HTK Version 3). Cambridge university, 1995.