Syllabation graphémique automatique à l'aide d'un dictionnaire phonétique aligné

20 mars 2012

Sophie Roekhaut[1,2] Sandrine Brognaux[1,3,4] Richard Beaufort[1,2]
(1) Centre de traitement automatique du langage (CENTAL)
(2) Institut Langage et Communication
(3) Institute of Information and Communication Technologies, Electronics and Applied Mathematics
(4) Aspirante FNRS
Université catholique de Louvain, Louvain-la-Neuve, Belgique
{sophie.roekhaut,sandrine.brognaux,richard.beaufort}@uclouvain.be

Résumé. La syllabation graphémique d'une phrase consiste à segmenter cette phrase en séquences de lettres correspondant strictement aux syllabes phonétiques qui la constituent. La syllabe graphémique se distingue de la syllabe orthographique classique, qui ne respecte pas toujours les frontières syllabiques au profit de règles de césure des mots. L'algorithme proposé repose sur un dictionnaire phonétique aligné, où chaque phonème de la transcription phonétique est aligné sur la première lettre du graphème correspondant. Il exploite également la frontière de mots et la frontière de groupes rythmiques pour limiter l'extension de certaines syllabes. Nos tests ont été réalisés en français, sur mots isolés et mots en contexte. L'article présente l'algorithme en lui-même, mais également la méthode d'alignement automatique et de correction semi-automatique permettant d'obtenir le dictionnaire aligné.

Abstract. Automatic Graphemic Syllabification using an Aligned Pronouncing Dictionary The graphemic syllabification of a sentence consists in splitting it into graphemes, each of which strictly corresponds to a phonetic syllable of the sentence. Graphemic syllabification differs from written syllabification (hyphenation), which sometimes does not keep to phonetic syllable boundaries because of etymological or morphological principles. Our algorithm relies on an aligned pronouncing dictionary, in which each phoneme is aligned with the first letter of the grapheme it corresponds to. It also exploits word and rhythm group boundaries to avoid the expansion of some syllables. Our tests were carried out in French, on both isolated words and words in context. The paper presents the algorithm itself, but also the automatic alignment algorithm and the semi-automatic correction method which, together, provided us with the aligned dictionary.

Mots-clés : Syllabation graphémique, graphèmes, phonèmes, alignement.

Keywords: Graphemic syllabification, graphemes, phonemes, alignment.

1 Introduction

Trois notions différentes mais apparentées font référence à la syllabe : la syllabe *phonétique*, la syllabe *graphémique* et la syllabe *orthographique*, illustrées en table 1.
Une *syllabe phonétique* (ligne 1) est un regroupement de phonèmes qui se prononcent en une seule émission. Elle comporte toujours un noyau, ou *nucleus*. En français, le noyau est toujours une voyelle (V) alors que certaines langues, comme l'anglais, possèdent également des consonnes dites *vocalisées* qui jouent le rôle de noyau. Le noyau peut être précédé par une consonne (C) ou un groupe consonantique et éventuellement une semi-voyelle (S). L'ensemble constitue l'attaque de la syllabe, ou *onset*, marquée par un positionnement des organes phonatoires. On est en phase explosive, l'énergie augmente. Le noyau est éventuellement suivi d'une semi-voyelle et d'une ou de plusieurs consonnes. L'ensemble constitue la queue de la syllabe, ou *coda*, marquée par un relâchement des organes phonatoires. On est en phase implosive, l'énergie diminue. Il existe donc des syllabes de structures et de longueurs différentes.
La *syllabation graphémique* (ligne 2) est une transposition fidèle de la syllabation phonétique dans l'orthographe du mot. La syllabe graphémique respecte donc toujours les frontières de la syllabe phonétique.
La *syllabation orthographique* (ligne 3) applique les règles de césure qui doivent être respectées à l'écrit. Les règles varient d'une langue à l'autre. Pour le français par exemple, elles sont décrites dans (Grevisse-Goosse, 2008, pp.35–37). Du fait de ces règles, la syllabe orthographique ne correspond pas toujours à la syllabe phonétique. Parfois, elle regroupe les syllabes (*bi - blio - thèque*), parfois, elle en déplace les frontières (*bon - ne - ment*).

mot	*bibliothécaire*	*bonnement*	*action*	*descendrais*
(1) *S.P.*	b i - b l i - ɔ - t e - k ɛ ʁ	b ɔ - n ə - m ɑ̃	a k - s j ɔ̃	d ɛ - s ɑ̃ - d ʁ ɛ
	C V - C C V - V - C V - C V C	C V - C V - C V	V C - C S V	C V - C V - C C V
(2) *S.G.*	bi - bli - o - thé - caire	bo - nne - ment	ac - tion	de - scen - drais
(3) *S.O.*	bi - blio - thé - caire	bon - ne - ment	ac - tion	des - cen - drais

TABLE 1 – Comparaison des syllabations phonétique (S.P.), graphémique (S.G.) et orthographique (S.O.). Le tiret (-) marque les frontières de syllabes

Si la syllabe orthographique ne respecte pas strictement la syllabe phonétique, dans l'acception courante, elle est pourtant *la* syllabe écrite à laquelle nous faisons naturellement référence. La syllabe graphémique, en fait, est rarement représentée.

Pourtant, la syllabe graphémique semble occuper une place de choix dans l'apprentissage de la langue. Plusieurs chercheurs estiment en effet qu'elle est l'une des unités de base à partir desquelles se construit progressivement la reconnaissance des mots de la langue (Lecocq, 1991, p.34) : dans le processus d'apprentissage de la lecture, les jeunes enfants ont tendance à lire syllabe par syllabe, parfois en les séparant de pauses assez longues (Leroy-Boussion, 1971, p.155), avant de passer, progressivement, à une lecture lexicale qui démontre la compréhension.
Sur cette base, nous faisons l'hypothèse que les applications d'Apprentissage et d'Enseignement des Langues Assistés par Ordinateur (ALAO/ELAO) pourraient bénéficier d'un outil de détection des syllabes graphémiques d'un texte. Ce serait certainement le cas de notre plateforme PlatON (Beaufort & Roekhaut, 2011), proposant en ligne des exercices de dictée complètement automatisés. Lors de la correction de l'exercice, PlatON affiche en rouge la séquence de lettres erronées et propose un diagnostic linguistique. Dans le cas d'une erreur phonétique concernant une syllabe phonétique entière, l'affichage

pourrait s'étendre à la totalité de la syllabe graphémique correspondante, et le diagnostic linguistique être adapté en conséquence.
On peut également supposer qu'une application qui proposerait une lecture automatique d'un texte, avec mise en surbrillance de la syllabe à prononcer, pourrait s'avérer une aide précieuse pour les personnes souffrant d'un déficit de lecture.
C'est sur la base de ce constat que nous avons développé l'algorithme de syllabation graphémique présenté dans cet article. Cet algorithme s'intègre dans l'étape de conversion phonétique du module de traitement automatique du langage (TALN) du système de synthèse de la parole eLite.

La suite de cet article s'articule comme suit. La section 2 présente les algorithmes de syllabation phonétique et orthographique existants. La section 3 est consacrée à notre algorithme de syllabation graphémique : après avoir brièvement présenté le système de synthèse, nous passons en revue les modules et données du système qui ont dû être adaptés afin d'obtenir cette syllabation. La section 4 évalue ensuite les résultats obtenus, et la section 5 conclut et propose quelques perspectives.

2 Etat de l'art

Il existe plusieurs outils libres de découpage en syllabes à partir de la transcription phonétique des mots pour le français(Bigi *et al.*, 2010; Pallier, 2004; Goldman, 2008; Adda-Decker *et al.*, 2005). Ces systèmes de syllabation reposent sur une série de règles basées sur la répartition des phonèmes en différentes classes : voyelles, semi-voyelles, occlusives, liquides, constrictives, nasales. Les algorithmes appliquent les mêmes règles sur des mots isolés et sur des mots en contexte. La prononciation des mots en contexte donne lieu à un certain nombre de phénomènes en frontières de mots, tels que les liaisons, les élisions ou les chutes de schwas, qui provoquent l'apparition de groupes consonantiques complexes, absents des transcriptions phonétiques des mots isolés. Cette réorganisation syllabique due a l'enchaînement des mots en contexte est appelée *re-syllabation*. Bigi *et al.* (2010) comparent les performances des différents algorithmes de syllabation phonétique du français à la syllabation réalisée par deux annotateurs humains à partir de transcriptions phonétiques de dialogues spontanés. Les divergences entre les syllabations produites par ces systèmes automatiques et celles proposées par les annotateurs humains sont principalement liées (1) à l'apparition de groupes consonantiques complexes lorsque les mots sont en contexte pour lesquels les systèmes et les annotateurs appliquent des règles différentes et (2) au fait que les frontières de mots ne sont pas prises en compte pour déterminer les frontières de syllabes alors qu'il semble que ces frontières constituent parfois des frontières de syllabe.
Le seul outil libre qui propose une syllabation orthographique des mots est Lexique 3 (New, 2006). Cet outil n'a cependant pas été présenté et évalué en détail. Les auteurs précisent qu'il s'agit d'un champ expérimental et que l'algorithme pour syllaber les mots présente des différences par rapport à celui utilisé pour la syllabation phonétique. En effet, l'algorithme de Lexique 3 tente de respecter les règles de césure des mots à l'écrit. Si nous souhaitons comme Lexique 3 établir des frontières de syllabes dans le mot graphique, nous souhaitons découper le mot en syllabes graphémiques plutôt qu'en syllabes orthographiques.

3 Algorithme

3.1 eLite : vue d'ensemble

eLite, prononcé [i l a j t], est un système de synthèse de la parole à partir du texte développé à Multitel ASBL [1] de 2001 à 2008, et maintenu et amélioré au CENTAL depuis. Etant un système complet de synthèse de la parole à partir du texte, eLite comporte un module de traitement du langage naturel (TALN) et un module de traitement du signal (TS). Cette section se concentre sur les éléments et les caractéristiques du module TALN nécessaires à la compréhension de l'algorithme de syllabation graphémique présenté. Le lecteur intéressé trouvera une description du système dans (Beaufort & Ruelle, 2006), l'algorithme d'analyse syntaxique dans (Beaufort *et al.*, 2002) et les algorithmes du TS dans (Colotte & Beaufort, 2004; Bozkurt *et al.*, 2004).

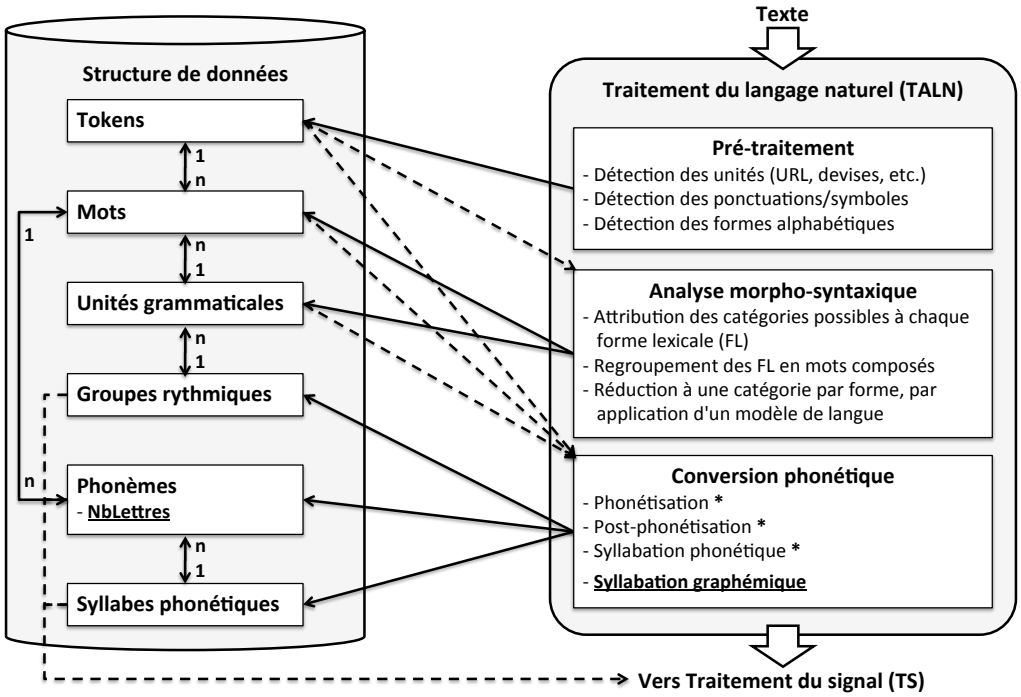

FIGURE 1 – eLite. Architecture TALN et structure de données. La syllabation graphémique a nécessité l'ajout des éléments soulignés, et l'adaptation de ceux suivis d'une astérisque

La figure 1 présente l'architecture du TALN d'eLite et la structure de données correspondante, remplie par le TALN et exploitée par le TS. L'architecture du TALN est somme toute assez classique en synthèse : pré-traitement, analyse morpho-syntaxique et conversion phonétique [2]. Le pré-traitement segmente le texte en *Tokens*, soit en séquences de caractères formant un tout, comme les URL, les numéros de téléphone ou les devises, et bien sûr les séquences alphabétiques et les ponctuations. Ce module est le seul qui travaillle au niveau du texte. Sur la base des ponctuations fortes, les deux modules suivants travaillent au niveau de la phrase.

L'analyse morpho-syntaxique désambiguïse les séquences alphabétiques. Elle les segmente en formes lexi-

1. Centre de recherche belge situé à Mons, Hainaut, Belgique.
2. eLite ne produit pas de prosodie acoustique, ce type d'information n'étant pas nécessaire au TS intégré.

cales, les *Mots*, et leur attribue une catégorie syntaxique, l'*Unité grammaticale*, qui correspond à plusieurs *Mots* dans le cas de formes composées (noms ou verbes).
Avant l'intégration de la syllabation graphémique, la conversion phonétique comportait trois étapes : la phonétisation, qui produit la séquence de *Phonèmes* de chaque mot de la phrase indépendamment du contexte, la post-phonétisation, qui modifie au besoin la transcription phonétique d'un mot en fonction du contexte, et la syllabation phonétique, qui construit d'abord les *Groupes rythmiques* de la phrase, puis les utilise pour rassembler les phonèmes en *Syllabes phonétiques*.
La syllabation graphémique est une quatrième étape, qui ne crée aucune couche dans la structure de données, mais nécessite une information supplémentaire dans la couche *Phonèmes* : *NbLettres*, le nombre de lettres auxquelles un phonème correspond. Outre cette légère modification de la structure de données, la syllabation graphémique a aussi et même surtout nécessité des adaptations importantes des trois étapes qui la précède.

3.2 Phonétisation

Principe. Ce module a pour tâche de produire la transcription phonétique d'un mot indépendamment de son contexte.
Notre phonétisation repose sur un arbre de décision de type ID3. Pour les détails du fonctionnement d'un arbre ID3, nous renvoyons le lecteur à (Pagel *et al.*, 1998). Nous ne nous intéressons ici qu'à ce qui influence notre algorithme de syllabation.
Un arbre de décision s'entraîne sur un corpus, en l'occurrence un dictionnaire phonétique de triplets {forme graphique, catégorie, transcription phonétique}. Le dictionnaire que nous avons utilisé pour le français est BRULEX (Content & Radeau, 1990), qui contient 282 402 mots phonétisés.
L'arbre n'apprend pas à phonétiser des mots, mais des lettres. A l'utilisation, l'arbre produit donc la phonétisation d'un mot *lettre par lettre*. Or, quelle que soit la langue, un mot compte bien souvent plus de lettres qu'il ne génère de phonèmes. C'est le cas de « quelqu'un » → [k ɛ l k œ̃], qui compte neuf lettres avec l'apotrophe, mais 5 phonèmes, ou de « pain » → [p ɛ̃], qui compte 4 lettres, mais 2 phonèmes. A l'inverse, il arrive qu'une lettre émette plusieurs phonèmes, comme « axiale » → [a k s j a l], dont le « x » génère [k s]. Pour permettre à l'arbre de toujours produire un et un seul phonème par lettre, le dictionnaire phonétique doit subir deux aménagements. D'une part, lorsqu'une suite de lettres génère un seul phonème, par exemple « ain » → [ɛ̃], on considère que la première lettre émet le phonème et que les autres émettent un silence, représenté par un souligné '_' : « ain » → [ɛ̃ _ _]. D'autre part, lorsqu'une lettre émet plusieurs phonèmes, comme le « x » → [k s] de « axiale », les phonèmes sont rassemblés en une seule séquence appelée *pseudo-phonème* : « x » → [k+s]. Ces deux modifications de la transcription produisent un dictionnaire phonétique *aligné*, où toute lettre émet un et un seul phonème :

âne	NVERB	a n _	pain	NVERB	p ɛ̃ _ _
axiale	NVERB	a k+s j a l _	quelqu'un	NVERB	k _ ɛ l k _ _ œ̃ _
coeur	NVERB	k œ _ _ ʁ	revendre	VERB	ʁ ə v ɑ̃ _ d ʁ _
hasard	NVERB	_ a z a ʁ _	vil	NVERB	v i l

A la fin de l'entraînement de l'arbre, une procédure de test soumet à l'arbre chaque forme graphique du dictionnaire d'entraînement. Toutes les formes dont l'arbre ne regénère pas correctement la phonétisation sont alors ajoutées à un dictionnaire d'exceptions, contenant également de nombreux noms propres et emprunts. Le taux de regénération correcte de l'arbre est de 98,96%. L'arbre permet donc de compresser

efficacement le dictionnaire phonétique. Il présente également un autre avantage : il permet de produire la phonétisation de mots non rencontrés lors de l'entraînement.
Sur cette base, le module de phonétisation procède comme suit. Un mot à phonétiser est d'abord recherché dans le dictionnaire d'exceptions, qui retourne sa phonétisation si le mot y figure. Dans le cas contraire, le mot est phonétisé au vol par l'arbre de décision.

Adaptation. Notre méthode de syllabation graphémique part du constat suivant : puisqu'il procède par lettre, l'arbre produit toujours une phonétisation alignée, dont les silences sont supprimés et les pseudo-phonèmes, segmentés, avant que la phonétisation ne soit sauvegardée dans la structure de données. Or, les silences et les pseudo-phonèmes permettent de calculer une correspondance exacte entre phonèmes et graphèmes [3] d'un mot, information précieuse et nécessaire à la syllabation graphémique. Conserver les silences et pseudo-phonèmes dans la couche *Phonèmes* de la structure de données n'était cependant pas souhaitable : cela aurait eu de lourdes conséquences sur la totalité des modules suivants. A la place, nous avons ajouté dans la couche *Phonèmes* l'information *NbLettres* qui permet de connaître, pour chaque phonème, le nombre de lettres auxquelles il correspond. En voici quelques exemples :

pain	p ɛ̃ _ _	:	[p] : 1, [ɛ̃] : 3
quelqu'un	k _ ɛ l k _ _ œ̃ _	:	[k] : 2, [ɛ] : 1, [l] : 1, [k] : 3, [œ̃] : 2
vil	v i l	:	[v] : 1, [i] : 1, [l] : 1
axiale	a k+s j a l _	:	[a] : 1, [k] : 1, [s] : 0, [j] : 1, [a] : 1, [l] : 2

Le cas du « x » de « axiale » montre que lorsqu'une lettre émet plusieurs phonèmes, seul le premier phonème ([k]) a un *NbLettres* valant 1. Le second par contre ([s]) vaut 0, ce qui permet d'éviter un décalage malheureux dans l'association phonèmes–graphèmes.
Pour être applicable, ce principe a cependant nécessité deux modifications. D'une part, le dictionnaire d'exceptions n'était pas aligné. En effet, lors de la procédure de test, tous les mots ajoutés dans le dictionnaire étaient nettoyés de leurs silences et pseudo-phonèmes. Ce dictionnaire a donc dû être regénéré. D'autre part, les premiers tests du système complet ont rapidement mis en évidence un fait de taille : de nombreux mots du dictionnaire phonétique utilisé étaient mal alignés. En voici quelques exemples :

bandeau	NVERB	b ɑ̃ _ d _ o _	:	[o] aligné sur « *au* » et non sur « *eau* »
coeur	NVERB	k _ œ _ ʁ	:	[œ] aligné sur « *eu* » et non sur « *oeu* »
quelqu'un	NVERB	k _ ɛ l k œ̃ _ _ _	:	[o] aligné sur « *u'un* » et non sur « *un* »

Initialement, ces erreurs d'alignements n'étaient pas gênantes, étant donné que les silences et pseudo-phonèmes étaient supprimés de la phonétisation. Dans une perspective de syllabation graphémique, il en allait tout autrement. Nous avons donc réaligné le dictionnaire phonétique. La procédure mise en place, détaillée ci-dessous, comporte un alignement automatique suivi d'une correction semi-automatique. Une comparaison de l'alignement corrigé avec l'alignement original a permis de constater que 30 711 entrées sur 281 402 étaient mal alignées.

Algorithme initial : récursion. L'algorithme initial, proposé par Pagel *et al.* (1998), est récursif. Au cours d'un cycle, il apprend un modèle d'association lettre–phonème à partir d'un alignement donné, et produit un nouvel alignement en utilisant le nouveau modèle appris. Le cycle apprentissage–alignement s'interrompt lorsque deux alignements successifs sont identiques.

3. Un graphème est une séquence de lettres à laquelle correspond un seul (pseudo-)phonème.

Pour initialiser le système, un premier alignement est produit en dehors du cycle d'itérations. Cet alignement est simplement obtenu en ajoutant à la fin de toute transcription trop courte le nombre de silences nécessaires pour obtenir une transcription de même longueur :

âne	→	a n _	hasard	→	a z a ʁ _ _
ânesse	→	a n ɛ s _ _	quelqu'un	→	k ɛ l k ɔ̃ _ _ _ _
coeur	→	k ɔ ʁ _ _	vil	→	v i l

Cet alignement permet d'apprendre le tout premier modèle d'association lettre–phonème, qui servira au cours du premier cycle d'itérations.
Chaque couple {forme graphique, transcription phonétique} est aligné par programmation dynamique, au travers d'une DTW (Myers & Rabiner, 1981) qui exploite les probabilités du modèle d'association :

$$ali(W, P) = \arg\max \prod_{i,j} p(W_i, P_j) \tag{1}$$

où W est la forme graphique, P est la transcription phonétique, W_i est la i^e lettre de W, P_j est le j^e phonème de P, et $p(W_i, P_j)$ est la probabilité, provenant du modèle d'association, que la lettre i corresponde au phonème j.
L'algorithme converge au bout de 5 itérations et produit le dictionnaire aligné que nous avions au départ, soit celui contenant 30 711 alignements erronés. Dans la suite de ce document, cette version de l'alignement est appelée *DicoAli0*.

Extension de l'algorithme : bootstrapping. Une rapide analyse de notre dictionnaire phonétique a mis en évidence un fait de taille : rares sont les cas où forme graphique et transcription phonétique sont de même longueur. Comme le montre la table 2, la transcription phonétique est bien souvent plus courte que la forme graphique, avec un pic marqué pour les transcriptions plus courtes de 2 ou 3 phonèmes et un total cumulé de seulement 17,07% pour les transcriptions présentant au maximum 1 phonème de moins. Sur cette base, nous avons décidé de travailler par *bootstrapping*, en appliquant le principe suivant : nous avons progressivement augmenté la taille du dictionnaire à aligner, en commençant par les 7 598 couples {forme graphique, transcription phonétique} de même longueur et en ajoutant, à chaque *bootstrap*, les couples présentant une différence de longueur supplémentaire. Dans la suite de cette section, un dictionnaire de couples {forme graphique, transcription phonétique} présentant de 0 à n différences de longueur est appelé « dictionnaire n ». Le dictionnaire 0 étant par nature déjà aligné, seul le modèle d'association a dû être appris à ce niveau. Ensuite, l'alignement récursif de tout dictionnaire $n > 0$ a été initialisé à l'aide du dernier modèle d'association calculé sur le dictionnaire $n - 1$. L'alignement récursif du dictionnaire 1 a ainsi été initialisé à partir de l'unique modèle 0, celui du dictionnaire 2, à partir du dernier modèle 1 et ainsi de suite, jusqu'au dictionnaire 4. Le dictionnaire complet a ensuite été aligné sur le dernier modèle 4. Le dernier alignement produit à ce stade est appelé *DicoAli1* dans la suite de ce document.
La comparaison de *DicoAli1* avec l'alignement corrigé a montré l'intérêt de la méthode. Grâce au *bootstrapping*, seules 6 832 transcriptions présentent encore au moins 1 erreur d'alignement, ce qui correspond à 97,57% d'alignement correct au mot.
Tous les dictionnaires $n > 0$ ont nécessité 4 itérations pour que l'alignement converge, soit 20 itérations au total. On peut donc supposer que l'alignement récursif, directement appliqué au dictionnaire complet, convergeait trop vite du fait du peu de couples de même longueur.

Correction de l'alignement automatique. Afin de vérifier et de corriger l'alignement automatique obtenu, nous avons défini et appliqué la procédure itérative suivante :

Différence	*Occurrences*	*%*	*Cumul*	*% cumulé*
0 (même longueur)	7 598	2,70	7 598	2,70
- 1	40 435	14,36	**48 033**	**17,07**
- 2	**78 163**	**27,77**	126 196	44,84
- 3	**76 533**	**27,19**	202 729	72,04
- 4	46 882	16,66	249 611	88,70
- 5 et plus	31 791	11,29	281 402	100,00

TABLE 2 – Différences de longueur entre formes graphiques et transcriptions phonétiques

1. A partir du dictionnaire aligné, nous générons automatiquement une liste de correspondances, où chaque phonème est associé à la liste de graphèmes sur lesquels il a été aligné. Ainsi, nous observons les correspondances suivantes :
   [ø] → {*eu*, *u*, *ue*,*e*, *œu*, *heu*, *i*, *œ*, *hu*, *o*, *oo*, *l*, *a*, *uc*, *k*, *ea*}.
   [b] → {*b*, *be*, *bes*, *bb*, *bent*, *bh*, *bs*, *ba*, *bbe*, *bo*, *b'*}.
2. Les correspondances sont ensuite vérifiées manuellement, à la recherche d'associations suspectes. Dans le cas de l'exemple précédent, on peut notamment s'étonner des associations :
   [ø] → {*i*, *hu*} et [b] → {*bo*}.
   Si des associations douteuses sont trouvées, on passe en (3). Sinon, la correction est terminée.
3. L'association douteuse est recherchée dans le dictionnaire aligné et corrigée au besoin. Ici, on remarque que [ø] associé à *hu* est correct (dans le mot *hum*) mais que l'association à *i* est due à la présence de mots anglais (comme *flirt*) qu'il faudra supprimer du dictionnaire. On retrouve [b] associé à *bo* dans le mot *boeuf* dont l'alignement doit donc être corrigé. Ceci provoquera l'apparition du nouveau graphème *oeu* dans la liste des associations au phonème [ø].
4. Lorsque toutes les associations douteuses ont été traitées, retour en (1).

Le dictionnaire aligné et corrigé, obtenu à la fin de cette procédure, est appelé *DicoAli2*.

Un dictionnaire de schwas optionnels. BRULEX propose des phonétisations sans schwas optionnels. Pour les besoins d'un synthétiseur de parole multi-styles (Roekhaut *et al.*, 2010), nous avons produit à partir de *DicoAli2*, une variante avec schwas optionnels, appelée *DicoSchwa* par la suite. Dans le cadre de cet article, *DicoSchwa* a été employé pour comparer notre algorithme de syllabation à la syllabation orthographique proposée par Lexique 3.
Notre système de phonétisation manipule donc trois types de dictionnaires :

Non aligné (*DicoPho*) : `revendre VERB ʁ ə v ɑ̃ d ʁ`
Alignés (*DicoAli{0,1,2}*) : `revendre VERB ʁ ə v ɑ̃ _ d ʁ _`
Aligné avec schwas (*DicoSchwa*) : `revendre VERB ʁ ə v ɑ̃ _ d ʁ ə`

3.3 Post-phonétisation

Principe. L'objectif de ce module est de gérer les phénomènes phonétiques qui se produisent en frontière de mots. En français, il s'agit d'une part de la liaison, et d'autre part de la lubrification du discours par insertion ou suppression de schwas. Le système utilise des règles de réécriture de la forme

« $\alpha \rightarrow \beta$: γ _ δ », signifiant que α se réécrit β lorsqu'il est entouré par γ à gauche et δ à droite. Voici une règle concrète de notre système :

```
e|es|ent  → ə  :  _ VERB; <C><A>* PREP
```

Cette règle se lit comme suit : on prononce (→ ə) les finales « e », « es » et « ent » (e|es|ent) d'un verbe (_ VERB) suivi (;) par une préposition dont la première lettre est une consonne (<C><A>* PREP). Il s'agit donc d'une règle d'insertion du schwa, qui s'appliquera par exemple à « *ils rest<u>ent</u> pour l'aider* », présenté comme ceci au système :

```
restent VERB; pour PREP
```

Adaptation. Pour la syllabation graphémique, les sorties des règles (β) ont dû être alignées sur les entrées (α). Dans le cas de notre exemple, la règle initiale, composées de 3 entrées, a dû être scindée en trois règles différentes. Voici le résultat obtenu :

```
e    → ə     :  _ VERB; <C><A>* PREP
es   → ə _   :  _ VERB; <C><A>* PREP
ent  → ə _ _ :  _ VERB; <C><A>* PREP
```

La présence des silences (_) permet au système de connaître la longueur du graphème à stocker dans *NbLettres* au moment de l'insertion du schwa dans la couche *Phonèmes* de la structure de données.

3.4 Syllabation phonétique

Principe. L'algorithme de syllabation phonétique que nous proposons (voir figure 2) tâche de respecter au mieux la définition de la syllabe posée en début d'introduction (voir section 1).
Dans son principe, il est assez proche des algorithmes présentés dans l'état de l'art : un test articulatoire est réalisé sur le phonème courant afin de décider s'il appartient ou non à la syllabe courante. Son originalité tient à trois caractéristiques. Premièrement, il est appliqué *après* la post-phonétisation et tient donc compte de l'ensemble des phénomènes articulatoires, sans nécessiter de re-syllabation. Deuxièmement, il tient compte de la frontière de mots (lignes 10 et 11), qui impose une frontière syllabique dans certains cas articulatoires précis. Troisièmement, il tient également compte de la frontière de *groupes rythmiques* (lignes 03 et 07), avec laquelle il fait *toujours* coïncider une frontière de syllabe.
Le groupe rythmique est une notion propre à eLite. Il s'agit d'un groupe de souffle, susceptible de porter un accent tonique et d'être suivi d'une pause. C'est la présence potentielle d'une pause entre deux groupes rythmiques qui impose de faire coïncider la frontière de groupe rythmique à une frontière syllabique. L'exemple ci-dessous illustre la différence de syllabation obtenue par le même algorithme, s'il tient ou ne tient pas compte des groupes rythmiques :

	« *dans les parcs c'est un peu limité* »
(1) Sans groupes rythmiques :	d ɑ̃ l e p a ʁ k s e œ p ø l i m i t e
(2) Avec groupes rythmiques :	d ɑ̃ l e p a ʁ k ‖ s e œ p ø l i m i t e

Dans la syllabation (1), le [k] de *parcs* ouvre la syllabe [k s e] et est explosif, alors que dans la syllabation (2), qui exploite les frontières de groupes rythmiques, le [k] ferme la syllabe [p a ʁ k] et devient implosif. A l'écoute, la syllabation (2) améliore nettement le naturel de la parole (Beaufort, 2008). On constate en outre que des annotateurs humains produisent également la syllabation (2), au contraire de LPL-syllabeur (Bigi *et al.*, 2010) qui, ne disposant pas des frontières de groupes, produit la syllabation (1).

Construction des groupes rythmiques. Si la notion de *groupe rythmique* est originale, sa méthode de construction est par contre bien connue. Le groupe rythmique est constitué d'une ou de plusieurs *Unités grammaticales*, regroupées selon l'algorithme *chinks 'n chunks* proposé par Liberman & Church (1991). Cet algorithme repose sur une répartition des catégories syntaxiques en deux classes : celle des *chinks* ou *mots-fonctions* (déterminant, préposition...) et celle des *chunks* ou *mots-contenus* (nom, verbe...). Sur cette base, le rassemblement des catégories en groupes respecte l'expression régulière suivante : « (*chinks** *chunks**)* ». En somme, dès qu'un *chunk* est suivi par un *chink*, une frontière de groupe est posée.

Adaptation. Pour la syllabation graphémique, un cas particulier a dû être traité : celui de deux phonèmes produits par une seule et même lettre. L'algorithme initial les séparait parfois entre deux syllabes ; c'était le cas du « x » de « axial », syllabé [a k] – [s j a l]. L'adaptation de l'algorithme les réunit toujours dans la seconde syllabe : [a] – [k s j a l]. Cela étant, il nous semble que cette nouvelle syllabation phonétique, guidée par la syllabation graphémique, correspond peut-être mieux aux phénomènes articulatoires (implosion, explosion) réellement présents dans ce cas.

3.5 Syllabation graphémique

A ce stade, la structure de données est remplie. Pour reconstituer les syllabes graphémiques, nous exploitons les relations (voir figure 1) existant entre les différentes couches de la structure : un *Mot* connaît ses *Phonèmes* ; chaque *Phonème* connaît sa *Syllabe*, son *Mot* et la longueur du graphème (*NbLettres*) auquel il correspond ; chaque *Syllabe* connaît ses *Phonèmes*.

```
01.  ajout_ok := TRUE
02.  If syll_cour != EMPTY Then
03.      If pho_cour == DEBUT_GR_RYTHM
04.      Or If pho_cour == VOC And pos_cour == CODA
05.      Or If pho_cour == (SEMI-VOC And INTERVOC) Or SILENCE Then
06.          ajout_ok := FALSE
07.      Else If pho_cour == (CONS And CODA) And pho_suiv != DEBUT_GR_RYTHM Then
08.                  If group_pho_suiv == (LIQUID ? SEMI-VOC ? VOC)
09.                  Or If pho_prec == CONS And MEME_ARTICULATION
10.                 Or If pho_cour == (CONSTRICTIVE And DEBUT_MOT)
11.                 Or If pho_cour == (PLOSIVE And ALVEODENTAL And DEBUT_MOT) And
                            group_pho_suiv == (POSTALVEOLAR VOC)
12.                 Or If pho_cour == LIQUID And pho_suiv == VOC Then
13.                     ajout_ok := FALSE
14.  Return ajout_ok
```

FIGURE 2 – Syllabation phonétique. Test d'appartenance à la syllabe courante

Nous rappelons qu'une syllabe graphémique doit strictement correspondre à une syllabe phonétique. Sur cette base, la syllabe graphémique est simplement la somme des lettres attribuées aux phonèmes de la syllabe phonétique correspondante.
La figure 3 montre pour les mots « *bandeau* » et « *mois* » comment les syllabes graphémiques sont construites à partir des diverses informations de la structure. Dans la figure, les rectangles gris matérialisent les syllabes graphémiques, et les cercles blancs, les graphèmes, correspondant à l'information *NbLettres* stockée dans les *Phonèmes*.

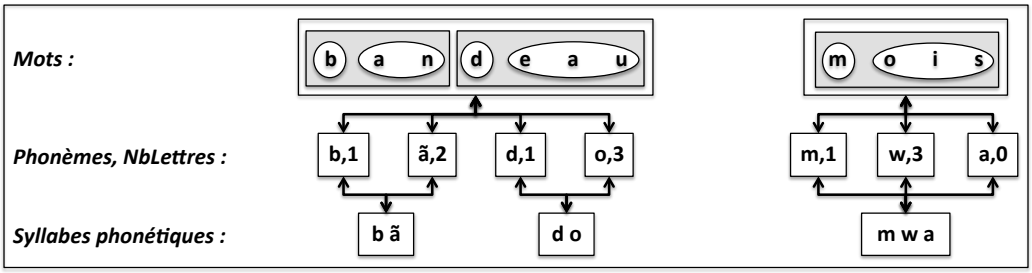

FIGURE 3 – Syllabation graphémique, guidée par les relations entre couches et *NbLettres*

4 Evaluation

Evaluation sur mots isolés. Le corpus de test contient 1 470 mots français, de longueur et de structures syllabiques différentes. Les résultats sont présentés en table 3. Nous avons commencé par comparer deux versions de la syllabation graphémique, entraînées sur des dictionnaires alignés différents : SG_1 sur *DicoAli1*, et SG_2 sur *DicoAli2*. Nous rappelons que la seule différence entre ces deux dictionnaires alignés est la correction semi-automatique dont a bénéficié *DicoAli2*. Les syllabations graphémiques ont été validées par deux experts, qui n'ont accepté la syllabation d'un mot que lorsque la totalité des syllabes proposées étaient valides. On constate que la correction semi-automatique dont a bénéficié *DicoAli2* augmente d'environ 4% le taux de syllabation correcte de SG_2 par rapport à SG_1. Les seules erreurs produites par SG_2 (11 formes, soit 0,7%) résultent d'une mauvaise phonétisation du mot. Notons que les experts ont au préalable évalué la phonétisation produite par l'arbre appris sur *DicoAli2*. Le taux de phonétisation correcte est de 98,77%, soit 1 452 mots correctement phonétisés. Les erreurs de phonétisation se produisent sur des formes qui n'appartiennent pas au dictionnaire sur lequel l'arbre a été entraîné (13 cas) ou dont la transcription est erronée dans le dictionnaire (2 cas). Enfin, certaines erreurs de phonétisation sont liées à une mauvaise analyse grammaticale des mots (3 cas). On peut observer que les erreurs de phonétisation n'entraînent pas toujours d'erreurs de syllabation graphémique. Par exemple, le mot 'merguez' phonétisé erronément [m ɛ ʁ g e], est syllabé correctement : 'mer-guez'.
Nous avons ensuite comparé SG_3, une syllabation graphémique entraînée sur *DicoSchwa*, à la syllabation orthographique proposée par Lexique 3. Le choix de *DicoSchwa* se justifie par le fait que les règles de césure sont plus proches d'une phonétisation où les schwas optionnels sont prononcés. Etant donné que les syllabations orthographique et graphémique reposent sur des règles différentes, les découpages en syllabes des deux systèmes ont été validés indépendamment. La syllabation obtenue avec SG_3 est proche de celle obtenue par SG_2, mais présente quelques erreurs supplémentaires car certains schwas finaux n'ont pas

été générés par le système (0,4% des formes)[4]. Lexique 3 obtient un taux de syllabation orthographique correct de 95,99%. Etant donné qu'aucune évaluation de cette syllabation orthographique n' a été réalisée par ses auteurs, nous en proposons ici une petite analyse. Outre les erreurs liées à l'absence complète de forme syllabée et celles liées à une phonétisation erronée, certains mots présentent des syllabes sans noyau vocalique (*an–té–ch–rist*) ou ne contenant que des lettres non prononcées (*poing–s*). Aucune de ces erreurs n'est commise par notre système.

Evaluation sur mots en contexte. Le corpus de test contient 100 phrases de 10 à 15 mots, choisies aléatoirement dans un corpus de presse francophone belge. Ces phrases totalisent 1 242 mots et 1 912 syllabes. Le modèle testé est SG_2, et l'objectif est d'évaluer la réorganisation syllabique qui s'opère lorsque les mots sont en contexte. Au niveau phonétique et selon la définition que nous avons donnée de la syllabe en section 1, l'algorithme de syllabation phonétique tâche de respecter les principes suivants :
1) Un phonème de liaison appartient toujours à la première syllabe du mot qui a entraîné la liaison : « les oiseaux » → [l e - z w a - z o].
2) La frontière rythmique marque toujours une frontière syllabique :
« il part ‖ en avion » → [i l - p a ʁ - ɑ̃ - n a - v j ɔ̃].
3) Lorsqu'une frontière de mot ne correspond pas à une frontière rythmique, la règle générale est de construire les syllabes comme si les phonèmes appartenaient au même mot :
« bonne idée » → [b ɔ - n i - d e].
4) Un cas, cependant, fait exception à la règle (3) : si le premier mot se termine par une plosive ou une constrictive, il y a frontière syllabique, quelle que soit la consonne qui commence le second mot. En voici un exemple : « avec lui » → [a - v ɛ k - l ɥ i]. Le [k] est ici en position implosive, et donc en coda. Dans une élocution rapide, on constate d'ailleurs que si la phase d'occlusion du [k] est maintenue, la phase d'explosion, elle, est presque inaudible. On comparera avec le même groupe phonétique dans « j'acclame » → [ʒ a - k l a m] où le [k] est clairement en position explosive, et ne risque pas de s'affaiblir dans une élocution rapide.

Nous avons validé notre syllabation graphémique au niveau de la phrase complète, des mots et des syllabes, en tenant compte des règles énoncées ci-dessus. Les résultats sont présentés en table 4. Les erreurs relevées sont le fait :
1) d'une erreur de phonétisation ou de post-phonétisation : « presque trop » → [p ʁ ɛ s k - t ʁo] où il manque un schwa de lubrification entre [k] et [t] ;
2) d'un mauvais usage des groupes rythmiques : « prendre ‖ en main » → [p ʁ ɑ̃ - d ʁ - ɑ̃ - m ɛ̃] où le rattachement de [d ʁ] à [ɑ̃] est impossible à cause du groupe rythmique. Cette syllabe *sans noyau* auraît dû être rattachée à la précédente : [p ʁ ɑ̃ d ʁ] ;
3) d'une liaison qui aurait dû être évitée, et qui modifie la syllabation phonétique : « vous et votre véhicule » → [v u - z e - v ɔ - t ʁ ə - v e - i - k y l].

	Modèle	Corrects	%
Eval. 1	SG_1	1 399	95,17%
	SG_2	1 459	**99,25**%
Eval. 2	SG_3	1 451	**98,71**%
	Lexique 3	1 411	95,99%

TABLE 3 – Evaluation sur 1 470 mots isolés

	Total	Corrects	%
Phrases	100	87	87,00
Mots	1242	1226	98,71
Syllabes	1912	1892	98,95

TABLE 4 – Evaluation sur mots en contexte

4. Il s'agit du schwa en finale de syllabe, précédé du phonème [j] (« *réveilleront* », phonétisé [ʁ e v ɛ j ʁ ɔ̃] est syllabé « *ré - veille - ront* ») ou [ɲ] (« *témoignerais* », phonétisé [t e m w a ɲ ʁ ɛ] est syllabé « *té - moi - gnerais* »).

5 Conclusion et perspectives

Nous avons présenté un algorithme de syllabation graphémique original qui repose sur le lien strict existant entre les graphèmes et les phonèmes du français. Les performances de cet algorithme dépendent d'une part de la qualité de l'algorithme de syllabation phonétique, et d'autre part de l'utilisation d'un dictionnaire du français qui respecte fidèlement les associations graphèmes-phonèmes de la langue. Le dictionnaire est obtenu grâce à une méthode d'alignement originale qui apprend les associations graphèmes-phonèmes en augmentant progressivement la taille du dictionnaire à aligner. L'alignement du dictionnaire est amélioré grâce à une correction semi-automatique des erreurs récurrentes. Ce dictionnaire est d'ailleurs une ressource riche d'informations qui peut être le point de départ d'études linguistiques sur les associations graphèmes-phonèmes de la langue. Les méthodes d'alignement automatique et de correction semi-automatique peuvent quant à elles être facilement appliquées à d'autres dictionnaires phonétiques dans d'autres langues.
L'algorithme a été conçu pour syllaber tant les mots isolés que les mots en contexte, grâce à l'utilisation des frontières de mots et des groupes rythmiques pour limiter l'étendue de certaines syllabes, là où le lecteur humain ferait vraisemblablement de même. La validation sur mots isolés a montré que les seules erreurs commises sont dues à des erreurs de phonétisation en amont. La validation sur mots en contexte a montré l'intérêt des groupes rythmiques et de l'utilisation de la frontière de mots pour poser certaines frontières de syllabes, mais a également mis au jour certaines erreurs de phonétisation, et surtout certains cas où l'algorithme exploite mal le groupe rythmique et produit des syllabes sans noyau vocalique. Ceci devra être corrigé.
Nous avons l'intention d'intégrer la syllabation graphémique dans notre système de correction automatique de dictées. La syllabe graphémique nous semble une unité pertinente à mettre en évidence lors de la correction de la dictée. L'analyse des erreurs d'associations graphèmes-phonèmes peut également être utile pour détecter les erreurs typiques chez un apprenant d'une langue étrangère qui serait influencé par les associations graphèmes-phonèmes de sa langue maternelle.
Notons enfin que l'algorithme de détection des syllabes graphémiques pourrait aider à découper automatiquement les sous-titres apparaissant dans les clips de karaoké : la mise en évidence de la syllabe à chanter semble en effet fort proche d'un découpage en syllabes graphémiques.

Références

ADDA-DECKER M., BOULA DE MAREUIL P., ADDA G. & LAMEL L. (2005). Investigating syllabic structures and their variation in spontaneous french. *Speech Communication*, **46**(2), 119–139.

BEAUFORT R. (2008). *Application des Machines à Etats Finis en Synthèse de la Parole. Sélection d'unités non uniformes et Correction orthographique*. PhD thesis, FUNDP, Namur.

BEAUFORT R., DUTOIT T. & PAGEL V. (2002). Analyse syntaxique du français. Pondération par trigrammes lissés et classes d'ambiguïtés lexicales. In *Proc. JEP*, p. 133–136.

BEAUFORT R. & ROEKHAUT S. (2011). Automation of dictation exercises. A working combination of CALL and NLP. *Computational Linguistics in the Netherlands Journal*, **1**, 1–20.

BEAUFORT R. & RUELLE A. (2006). eLite : Systéme de synthése de la parole á orientation linguistique. In *Proc. JEP*, p. 509–512.

BIGI B., MEUNIER C., BERTRAND R., NESTERENKO I. *et al.* (2010). Annotation automatique en syllabes d'un dialogue oral spontané. In *Acte des XVIIIémes Journées d'Etudes sur la Parole*, Mons, Belgique.

BOZKURT B., DUTOIT T., PRUDON R., D'ALESSANDRO C. & PAGEL V. (2004). Chapter 1 : Reducing discontinuities at synthesis time for corpus-based speech synthesis. In S. NARAYANAN & A. ALWAN, Eds., *Text To Speech Synthesis : New Paradigms and Advances*. Prentice Hall PTR.

COLOTTE V. & BEAUFORT R. (2004). Synthèse vocale par sélection linguistiquement orientée d'unités non-uniformes : LiONS. In *Proc. JEP'04*.

CONTENT, A. MOUSTY P. & RADEAU M. (1990). Brulex. une base de données lexicales informatisée pour le français écrit et parlé. *L'année psychologique*, **90**(4), 551–566.

GOLDMAN J. (2008). Easyalign : a semi-automatic phonetic alignment tool under praat. *Computer script]. Online : http ://latlcui. unige. ch/phonetique/easyalign/, accessed on October*, **1**, 2008.

GREVISSE-GOOSSE A. (2008). *Le Bon Usage*. Bruxelles : De Boeck-Duculot.

LECOCQ P. (1991). *Apprentissage de la lecture et dyslexie*, volume 190. Mardaga Editions.

LEROY-BOUSSION A. (1971). Maturité mentale et apprentissage de la lecture. *Enfance*, **24**(3), 153–208.

LIBERMAN M. & CHURCH K. (1991). Text Analysis and Word Pronunciation in Text-to-Speech Synthesis. In S. FURUI & M. SONDHI, Eds., *Advances in Speech Signal Processing*, p. 791–831. Dekker.

MYERS C. & RABINER L. (1981). A comparative study of several Dynamic Time-Warping algorithms for connected word recognition. *The Bell System Technical Journal*, **60**(7), 1389–1409.

NEW B. (2006). Lexique 3 : une nouvelle base de données lexicales. In *Verbum ex machina. Actes de la 13e conférence sur le Traitement automatique des langues naturelles, Cahiers du Cental*, volume 2, p. 892–900.

PAGEL V., LENZO K. & BLACK A. (1998). Letter-to-sound rules for accented lexicon compression. In *Proc. ICSLP'98*, p. 252–255.

PALLIER C. (2004). Syllabation des représentations phonétiques de brulex et de lexique. Manuscrit disponible : `http://www.pallier.org/ressources/syllabif/syllabation.pdf`.

ROEKHAUT S., GOLDMAN J.-P. & SIMON A.-C. (2010). A Model for Varying Speaking Style in TTS systems. In *Proceedings of Speech Prosody 2010*, Chicago, Illinois, USA.