UNIVERSITÉ DE GENÈVE

Archive ouverte UNIGE
https://archive-ouverte.unige.ch

Chapitre d'actes 1996

Published version Open Access

This is the published version of the publication, made available in accordance with the publisher's policy.

Comparaison des structures syllabiques en français et en anglais

Goldman, Jean-Philippe; Content, Alain; Frauenfelder, Ulrich Hans

How to cite

GOLDMAN, Jean-Philippe, CONTENT, Alain, FRAUENFELDER, Ulrich Hans. Comparaison des structures syllabiques en français et en anglais. In: XXIèmes Journées d'Etude sur la Parole. Avignon (France). [s.l.] : [s.n.], 1996. p. 119–122.

This publication URL: https://archive-ouverte.unige.ch/unige:82786

© This document is protected by copyright. Please refer to copyright holder(s) for terms of use.

COMPARAISON DES STRUCTURES SYLLABIQUES EN FRANÇAIS ET EN ANGLAIS

Jean-Philippe Goldman[1], Alain Content[1,2], Uli H.Frauenfelder[1]
goldman,content,uli@fapse.unige.ch
[1]Laboratoire de Psycholinguistique, Université de Genève, Suisse
tél: ++41.22.705.97.41 fax: ++41.22.300.14.82
[2]Laboratoire de Psychologie Expérimentale, Université libre de Bruxelles, Belgique

ABSTRACT

This study present a quantitative comparison between the phonological properties of French and English using the BRULEX and CELEX lexical databases, respectively. These lexical statistics give a description of the syllabic structure of these two languages. These analyses revealed some interesting differences in the size of the two respective syllabic inventories, in syllabic complexity and finally in the frequency distribution of the syllable types and tokens. The processing consequences of these differences are discussed.

1. INTRODUCTION

Dès les années 70, la difficulté à identifier des indices phonétiques invariants dans le signal de parole a conduit les psycholinguistes à envisager un rôle possible de la syllabe comme unité privilégiée d'accès au lexique. Cette hypothèse a été initialement renforcée par des travaux (Mehler et al., 1981) montrant la sensibilité des locuteurs francophones à la structure syllabique. Plus récemment, des comparaisons inter-langues ont conduit à l'hypothèse que l'importance de l'unité syllabique pourrait différer selon les propriétés phonétiques, phonologiques, prosodiques et distributionnelles des langues (Sebastiàn-Gallès et al., 1992). En particulier, les différences observées entre des locuteurs francophones et anglophones ont été expliquées en référence aux différences dans la complexité et le degré d'ambiguïté des structures syllabiques des deux langues (Cutler et al., 1986).

Cependant, les différences structurales entre les deux langues n'ont jamais été analysées en détail. Par ailleurs, les études sur la syllabation du français (Aubergé et al., 1988; Laporte 1988; Dell 1995) montrent des désaccords et des ambiguïtés non négligeables. C'est probablement un des facteurs qui expliquent qu'il n'existe que peu de données quantitatives sur l'inventaire et la distribution des syllabes du français (Wioland, 1995). Le travail présenté constitue une première tentative de préciser les différences structurales entre le français et l'anglais, par des analyses quantitatives sur des lexiques informatisés.

2. METHODE

De manière à pouvoir prendre en compte la fréquence d'usage des mots, nous avons sélectionné deux bases de données qui contiennent cette information, BRULEX (Content et al., 1990), pour le français (noté Fr), et CELEX (Celex, 1993), pour l'anglais (An).

BRULEX est basé sur le corpus du Micro-Robert et contient environ 35000 entrées. Après suppression des formes fléchies, des mots composés, des onomatopées, abréviations, et des emprunts, le corpus restant comporte 24494 entrées. L'information sur la fréquence d'usage est basée sur les données du Trésor de la Langue française. BRULEX ne fournit pas l'information sur la structure syllabique des représentations. Celle-ci a été générée sur base de l'analyse des groupes consonantiques intervocaliques dans le lexique BDLEX (Pérennou et al.).

La réduction du lexique anglais s'est faite en éliminant les mots composés, les abréviations et les mots empruntés, ramenant le corpus de 52447 à 40911 entrées. Un élément souvent considéré comme un facteur d'ambiguïté pour la syllabation de l'anglais est la notion d'ambisyllabicité, proposée notamment par Kahn (1980), selon laquelle une consonne intervocalique peut, dans certains contextes phonologiques, appartenir à la fois à la syllabe précédente et à la syllabe suivante. Pour évaluer les conséquences de ce facteur, nous avons introduit pour l'anglais une représentation re-syllabifiée (notée AnAmb).

Les conditions d'application de l'ambisyllabicité ne font pas l'objet d'un accord unanime. Nous avons ici opté pour l'analyse la plus conservatrice, c'est-à-dire celle qui donne lieu au moins de changements, sur base de l'algorithme proposé par Gussenhoven (1986). Il propose d'attacher à toute syllabe sans coda, la première consonne de l'attaque de la suivante à la condition que celle-ci comporte une voyelle réduite (par exemple, le mot 'villa', qui est codé phonétiquement vI.l@,

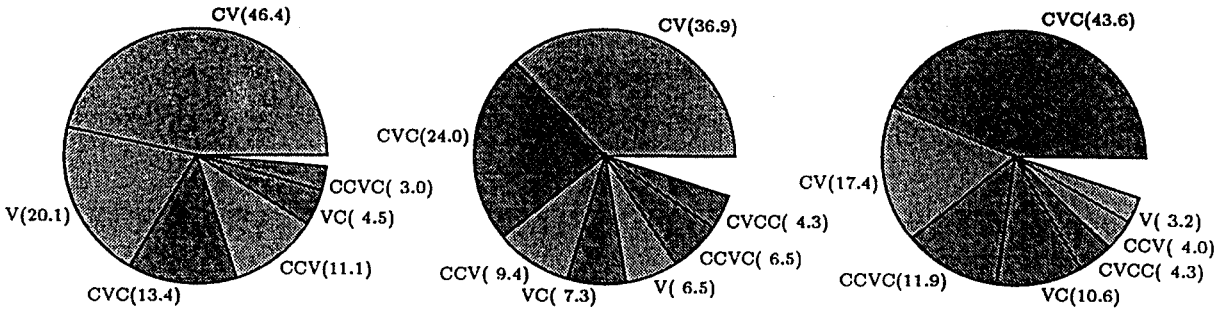

Figure 1: Répartition des structures syllabiques dans Fr., An., AnAmb. (pondérée par la fréq. d'utilisation)

devient vIl.l@, la consonne /l/ étant notée à la fois dans la première et la seconde syllabe).

Les analyses présentées portent sur la comparaison des structures et inventaires syllabiques des deux langues. Il faut remarquer que malgré la sélection effectuée, une grande différence persiste dans la taille des corpus.

3. RESULTATS

3.1. Inventaire phonémique et syllabique

Un premier indice éventuel de différences de complexité concerne le nombre absolu de syllabes et de structures syllabiques différentes attestées dans chacune des deux langues. La première analyse présentée concerne la taille de l'inventaire des segments et des syllabes. Les résultats principaux apparaissent dans la table 1.

Table 1: Inventaire phonémique et syllabique

	Fr	An	AnAmb
#mots	24.494	40.911	40.911
phonèmes	34	44	44
voyelles	15	23	23
consonnes	19	21	21
syllabes	3246	6781	8090
structures	21	22	22

Une différence importante apparaît en ce qui concerne l'inventaire des syllabes des deux langues. Le nombre de syllabes anglaises est plus de deux fois supérieur au répertoire syllabique français, et cette différence est encore renforcée par l'introduction de l'ambisyllabicité. Par contre, le nombre de structures syllabiques, définies en fonction de la classification des segments en consonnes (C) et voyelles (V) ne diffère pratiquement pas. En outre, le nombre de structures n'est pas modifié par la prise en compte de l'ambisyllabicité pour l'anglais.

Cette première analyse offre donc des indications contradictoires : la taille de l'inventaire syllabique confirme la plus grande complexité de l'anglais, mais il apparaît que la diversité des syllabes en anglais n'est pas liée à une plus grande variation de leur structure. L'analyse complète des relations entre ces deux observations dépasse le cadre du présent article, mais on peut d'ores et déjà signaler que la taille de l'inventaire phonémique (44 pour l'anglais vs. 34 pour le français) semble un facteur pertinent.

3.2. Fréquence des structures syllabiques

Une seconde analyse porte sur la fréquence des différentes structures syllabiques. On pourrait en effet faire l'hypothèse que malgré l'équivalence globale du nombre de structures, il existe des différences entre les deux langues en termes du nombre de structures syllabiques prédominantes. La figure 1 indique les proportions de syllabes des différents types dans chacun des corpus. Dans cette analyse, la fréquence d'usage des mots a été prise en compte, de telle sorte que la fréquence de chaque syllabe est pondérée par la fréquence d'usage des mots dans lesquels elle apparaît. Pour faciliter la lecture, les structures de fréquence inférieure à 3% ont été omises. Les résultats indiquent qu'un nombre très limité de structures syllabiques (6 ou 7) prédominent, et ils suggèrent effectivement des différences entre les deux langues: alors que 91.0% des syllabes attestées dans les mots français se répartissent entre 4 structures (CV, V, CVC et CCV), les quatre structures dominantes en anglais ne couvrent que 76.6% à 83.5%(pour An et AnAmb, respectivement). En outre, comme on peut le voir sur la figure 1, la proportion de syllabes non-couvertes par les 6 (Fr) ou 7(An) structures les plus fréquentes est plus importante en anglais.

La figure 1 montre également la nette prédominance des syllabes ouvertes (gris clair) par rapport aux syllabes fermées (gris foncé) en français. L'importance de la classe des syllabes ouvertes est fortement atténuée en anglais et la prédominance des syllabes fermées est encore renforcée lorsque l'ambisyllabicité est prise en compte (Cf. table 2). Il est intéressant de noter que, pour le français, nos observations s'avèrent très

proches de celles rapportées par Wioland, malgré les différences de procédure et de corpus. L'analyse réalisée par Wioland porte en effet sur un corpus de français parlé, transcrit et syllabé par l'auteur et il obtient un taux de 80.4% de syllabes ouvertes.

Table 2: Proportion de syllabes ouvertes

	Fr	An	AnAmb
Syll. ouv.	79.1%	54.1%	26.6%
Syll. fermées	20.9%	45.9%	73.4%
Ambiguïté	26.9%	39.6%	15.8%

On sait qu'en français, certaines voyelles apparaissent préférentiellement en syllabe ouverte et d'autres en syllabe fermée. Nous avons tenté d'estimer l'influence de ces contraintes phonotactiques en calculant la proportion de syllabes ouvertes et fermées pour chaque voyelle séparément. Le taux d'ambiguïté est la moyenne sur l'ensemble des voyelles de la plus petite des deux valeurs. Ainsi, un taux de 0 indiquerait que la voyelle n'apparaît que dans une classe de structure syllabique, tandis qu'un taux de 50% correspondrait au cas où les deux classes sont équiprobables. En moyenne, il apparaît que le taux d'ambiguïté est nettement plus élevé en anglais qu'en français. Par contre, cette différence est complètement éliminée si l'ambisyllabicité est prise en compte, du fait de la prédominance des syllabes fermées.

La différence dans le taux d'ambiguïté constitue donc un indice complémentaire en faveur de l'hypothèse que la structure syllabique peut être plus saillante en français. Contrairement à une suggestion assez répandue, il apparaît dans cette analyse que la prise en compte de l'ambisyllabicité pourrait simplifier, et non rendre plus ardue la segmentation syllabique, dans la mesure où, dans la plupart des cas, la voyelle attire la consonne subséquente.

3.3. Longueur des mots et des syllabes

Un indice potentiel de la complexité des syllabes est fourni par leur longueur en termes du nombre de phonèmes. La figure 2 (panneau gauche) montre la distribution des deux lexiques selon le nombre de phonèmes. Les longueurs moyennes sont similaires (6.61 et 6.74, resp. pour le français et l'anglais) et les distributions sont largement superposables. Mais, cette équivalence se réalise par des biais différents dans les deux langues. Comme l'indique le panneau médian, le mode du nombre de syllabes est de 3 pour le français et 2 pour l'anglais (moyenne resp.:2.83 et 2.62 syl./mot), et cette tendance est contrebalancée par la variation du nombre de phonèmes par syllabe (panneau droit, respectivement 2.33, 2.57 et 2.96 pour Fr, An et AnAmb en moyenne). Les syllabes comportant plus de trois phonèmes représentent une plus grande proportion pour le lexique anglais que pour le français. A l'inverse les syllabes formées d'un ou deux phonèmes sont en plus grand pourcentage en français. Cette relation d'échange correspond au principe général proposé par Menzerath (1954) qui stipule que le nombre de phonèmes par syllabe diminue à mesure qu'augmente le nombre de syllabes par mot. Nous avons donc examiné si la différence de longueur des syllabes entre les deux langues était observée indépendamment de la longueur des mots en syllabes. Les résultats (figure 3) montrent que si on considère la syllabation fournie par CELEX, les différences de longueur des syllabes s'estompent et disparaissent pour les mots longs par rapport au français. Par contre, pour la version ambisyllabifiée, la différence se maintient indépendamment de la longueur des mots.

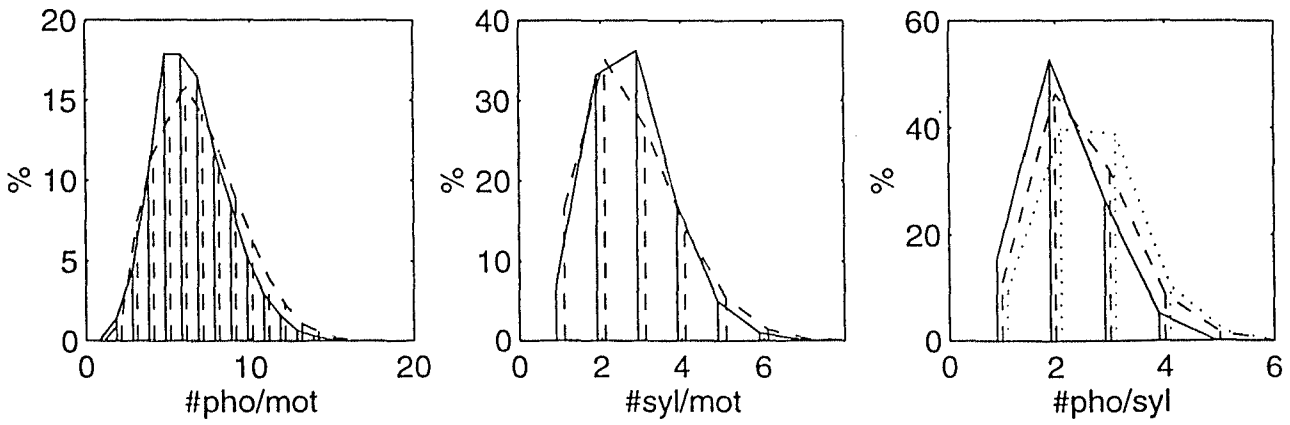

Figure 2: Distribution des mots et des syllabes selon la longueur (Fr: continu, An: tireté, AnAmb: pointillé)

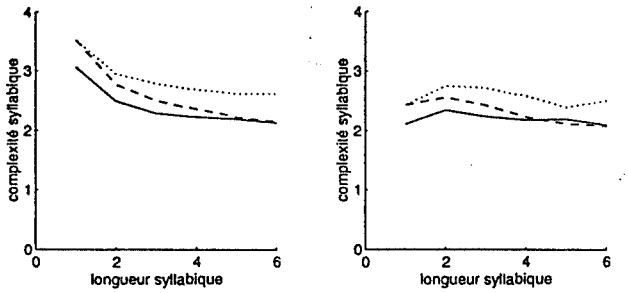

Figure 3: Longueur syllabique en fonction du nombre de syllabes par mot, sans (à gauche) et avec (à droite) pondération par la fréquence d'utilisation (Fr: continu, An: tireté, AnAmb: pointillé)

4. CONCLUSION

La comparaison quantitative des structures syllabiques en français et en anglais confirme l'existence de différences inter-langues. Nos analyses indiquent qu'en anglais, le nombre de syllabes attestées est plus grand, les structures syllabiques sont plus variées, les syllabes sont en moyenne plus longues. Toutefois, mis à part les variations de taille de l'inventaire, et la dominance des syllabes ouvertes en français, les différences observées ne sont pas extrêmement marquées. De plus,la prise en compte de l'ambisyllabicité pouvait dans le cas de l'ambiguïté, réduire l'écart entre les deux langues.

Le résultat le plus clair concerne la différence de taille des inventaires syllabiques des deux langues. La dimension du "lexique syllabique" pourrait effectivement constituer le point de départ d'une explication des différences dans le traitement de la chaîne parlée. Mais pour établir la nature de la relation entre les caractéristiques des deux langues et leur traitement, des analyses complémentaires plus fines selon la fréquence d'usage des mots seraient nécessaires, ainsi que l'examen de l'occupation de l'espace syllabique théorique des deux langues

En outre, une interprétation pertinente de ces données comparatives dépend aussi d'un modèle théorique du traitement plus explicite et de données expérimentales plus riches.

REMERCIEMENTS

Cette recherche a été financée par le F.N.R.S. 11-39553.93. Nous remercions F. Dell et P. Encrevé pour leurs commentaires et suggestions sur ces travaux.

5. BIBLIOGRAPHIE

Mehler J., Dommergues J.Y., Frauenfelder U., et Segui J. (1981) *The syllable role's in speech segmentation.* Journal of Verbal Learning and Verbal Behavior, 20:298-305.

Sebastiàn-Gallès N., Dupoux E., Segui J. et Mehler J.(1992) *Contrastive syllabic effect in catalan and spanish.*Journal of Memory and Language, 31:18-32.

Cutler A., Mehler J., Norris D. et Segui J. (1986) *The syllabe's differing role in the segmentation of french and english.* Journal of Memory and Language, 385-400.

Aubergé V., Boé L.-J., Lefèvre J.-P. (1988) *Lexiques et groupes consonantiques.*XVIIèmes J.E.P., 55-60

Laporte E. (1992) *Phonetic syllables in french: combinatorics, structure and formal definitions.* Acta Linguistica Hungarica, 41:175-189.

Dell F.*Consonant clusters and phonological syllables in french.* In press.

Wioland F. (1985) *Les structures syllabiques du français.* Slatkine.

Content A., Mousty P. et Radeau M. (1990) *Brulex. une base de données lexicales informatisée pour le français écrit et parlé.* Année Psycho., 90:551-556,

Celex (1993) *Centre for Lexical Information MPI for Psycholinguistics*, Nijmegen

Pérennou G. et de Calmès M. *BDLEX. Base de donnéees lexicales du français écrit et parlé.* Labo. CERFIA, Toulouse.

Kahn D. (1980) *Syllable-based generalizations in English phonology.* Garland, New-York.

Gussenhoven C. (1986) *English plosive allophones and ambisyllabicity* GRAMMA 10.2:119-141.

Menzerath (1954) *Die architektonik des deutschen wortschatzes.* Phonetische Studien,3