

Étude sur la distribution des syllabes en français

Cheryl Goodenough-Trepagnier et Robert M. Frankston

Numéro 7, 1978

URI : <https://id.erudit.org/iderudit/800052ar>

DOI : <https://doi.org/10.7202/800052ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Les Presses de l'Université du Québec

ISSN

0315-4025 (imprimé)

1920-1346 (numérique)

[Découvrir la revue](#)

Citer cet article

Goodenough-Trepagnier, C. & Frankston, R. M. (1978). Étude sur la distribution des syllabes en français. *Cahier de linguistique*, (7), 43–70.
<https://doi.org/10.7202/800052ar>

ÉTUDE SUR LA DISTRIBUTION DES SYLLABES EN FRANÇAIS

Les travaux dont nous allons décrire les résultats ont été entrepris dans le but de mettre au point un système de communication qui devrait servir aux malades paralytiques cérébraux incapables de parler à la suite d'une infirmité des organes articulatoires. Ce but a déterminé l'orientation de notre travail et a imposé certaines contraintes qu'il convient d'expliquer avant d'aborder l'analyse des résultats eux-mêmes.

L'infirmité motrice cérébrale recouvre une variété de troubles allant d'un déficit minime jusqu'à l'incapacité la plus globale, mais ils proviennent tous de dommages cérébraux survenus avant, durant ou peu après la naissance et ne sont pas de nature progressive. De plus, la proportion des troubles de la motricité est plus grande que les troubles sensoriels, en particulier des déficits auditifs et visuels. Le déficit moteur peut toucher différentes parties du corps plutôt que d'autres, par exemple l'hémicorps, les membres inférieurs ou supérieurs. Parmi la population d'infirmes moteurs cérébraux, il en existe une petite minorité qui, malgré le haut degré de compétence linguistique dont leur compréhension apparente témoigne, ne peut pas produire de parole articulée, à cause d'une atteinte à la motricité des organes de la parole. Si, de plus, comme c'est généralement le cas, la motricité des membres supérieurs est aussi défectueuse, la personne se trouve démunie de tout moyen d'exprimer ses besoins, ses sentiments, ses opinions et ses idées. Elle se trouve restreinte à un ensemble appauvri de signaux, se limitant souvent à des expressions faciales, des signes oui-non, que seul l'entourage peut comprendre.

Il existe actuellement deux approches à la compensation de ce déficit : l'une est fondée sur l'exploitation de l'alphabet, l'autre sur celle d'un ensemble de mots ou de symboles pictographiques.

L'alphabet présente l'avantage d'un nombre très petit de symboles, et de permettre l'accès à toute la langue. L'écriture manuscrite étant en dehors des capacités d'un IMC (infirmes moteurs cérébraux), il lui reste la possibilité de dactylographier, ou de désigner les lettres de l'alphabet inscrites sur une planche. A part les problèmes moteurs inhérents à l'usage de la machine à écrire qu'il ne convient pas de traiter dans cet article, il y a ceux qui résultent de la nature de l'alphabet et du système orthographique du français, de sa redondance, qui rendent complexe son utilisation. L'inventaire des symboles a beau être petit, le nombre de gestes à faire pour produire un mot est très grand à raison d'au moins un geste par lettre. C'est un désavantage de taille qui ralentit énormément la communication jusqu'à la rendre pratiquement impossible.

Dans la communication par mots écrits, ou par symboles pictographiques, le principe de base d'une telle approche est que chaque symbole représente un mot ou un concept. Pour des référents plutôt concrets, le symbole pictographique peut avoir une certaine ressemblance avec l'objet, par exemple, un petit bonhomme pour représenter "homme". Pour des mots plus abstraits, le symbole devient forcément arbitraire, bien qu'il puisse conserver un certain aspect mnémorique. Le système actuellement utilisé¹ fournit une traduction sous chaque pictogramme. L'utilisateur désigne les symboles nécessaires à la transmission de son message et peut recourir à l'alphabet orthographique ou, au besoin, à des suites de symboles pour les mots qui ne figurent pas dans l'ensemble présenté sur sa planche. L'avantage de ce système est que, pour chacun

1. C.K. Bliss *Semantography-Blissymbolics*, 1965.

des deux cents ou cinq cents mots qui figurent sur le tableau, un seul geste par mot suffit. Cependant on perd cette efficacité dès qu'il s'agit de représenter un mot qui ne se trouve pas sur le tableau. L'utilisateur se voit obligé de construire, au moyen d'un ensemble limité de concepts sémantiques, la définition du mot qu'il veut communiquer. Ce recours à la périphrase, tout en augmentant le nombre de gestes par mots, a pour effet de diminuer la précision et la valeur communicative du message².

La présente tentative cherche à réunir les avantages principaux des deux approches décrites ci-dessus : inventaire minimal, mais également, nombre minimal de gestes par mot. La syllabe, plutôt que l'orthographe ou le lexique, nous a paru offrir le point de départ le plus avantageux. Nous nous sommes tournés vers une unité phonologique, plutôt qu'orthographique afin d'économiser le nombre de symboles. Le français offre souvent de nombreuses façons d'écrire ce qui se prononce de la même façon : par exemple pour /e/ on trouve *e, er, ai, aie, ez, ef*. De plus, le nombre de lettres est souvent disproportionné par rapport aux sons qu'une suite de lettres représente : *eau*, qui se prononce /o/, *maillot* qui se prononce /majo/. Un système fondé sur des suites de symboles orthographiques serait nécessairement moins économique, du point de vue de son inventaire et du nombre de gestes par mot, qu'un système fondé sur des suites de symboles de nature phonologique.

La combinaison d'un alphabet à base phonologique avec un ensemble d'unités phonologiques plus grandes que le phonème, à la fois pas trop nombreuses et suffisamment fréquentes pour mériter d'être incluses

-
2. Quelques exemples relevés auprès d'enfants utilisant le Bliss :
semblable-contre-contre-petit semblable-autobus : "gros camions"
une chose avec un oeil pour faire un beau carré avec un visage :
"caméra"
petit oiseau me fait mal : "un moustique m'a piqué".

dans un tableau de communication, nous a semblé offrir la solution idéale. Nous avons donc décidé d'étudier les possibilités de la syllabe en ce sens. Notre analyse a porté sur la structure syllabique du français parce qu'elle est relativement simple et facilement accessible à l'intuition linguistique du locuteur natif³.

Deux choses étaient évidentes au début de cette recherche : l'inventaire total des syllabes en français dépassait de plusieurs milliers le nombre de deux à trois cents qui nous paraissait le plus pratique pour constituer un tableau de communication ; de plus, il faudrait certainement inclure dans ce tableau un alphabet, phonologique cette fois, pour compenser l'absence inévitable d'un très grand nombre de syllabes possibles du français, ce qui allait diminuer encore le nombre total des syllabes qu'on pourrait inclure. La possibilité de construire un syllabaire répondant à ces contraintes dépendait donc de la nature de la distribution des syllabes dans la langue. Il faudrait qu'un petit nombre de syllabes soit très fréquent et que la proportion des occurrences représentées par les syllabes de moins en moins fréquentes diminue très rapidement. Autrement dit, il faudrait que cette distribution soit sujette à une espèce de loi de Zipf⁴, bien que Zipf lui-même n'ait pas étudié cette aspect du langage⁵.

-
3. Par comparaison, la facilité d'accès à la syllabe du point de vue de l'intuition linguistique ne fonctionne pas pour l'anglais.
 4. ZIPF, G. (1935), *The Psycho-biology of Language*, Boston : Houghton Mifflin.
 5. Il existe une étude de la distribution des syllabes en anglais, faite sur un corpus de langage écrit : G. Dewey, *The Relative Frequency of English Speech Sounds*. Cambridge, Mass., Harvard University Press, 1923.

Puisque nous cherchions à construire une "prothèse" linguistique, c'est-à-dire un dispositif qui permettrait de suppléer aux carences du langage parlé, nous avons décidé de baser notre première étude des fréquences des syllabes sur le langage parlé. La liste de mots établie par G. Gougenheim⁶ nous semblait très utile à cette fin. Cette liste de 1 063 mots est composée à partir d'un ensemble d'interviews menées en France, principalement à Paris, pendant les années cinquante, auprès de personnes d'âge, de sexe et de niveau social différents. La liste représente un total de 285 404 occurrences, et n'inclut que les mots qui ont eu une fréquence minimale de vingt occurrences. Une locutrice native du français de Montréal⁷ a transcrit la liste en phonèmes du français standard et a divisé les mots en syllabes, d'après les règles de syllabation décrites dans le Tableau II. Les conventions utilisées par Gougenheim ont été adoptées, sauf que *l'* a été assimilé à *le* ou à *la* selon le cas, et évidemment le fait de transcrire en phonèmes a fait confondre les distinctions entre mots phonétiquement identiques mais grammaticalement différents, distinctions que Gougenheim avait respectées (exemple, *que* comme conjonction et comme relatif). Nous avons remanié la liste initiale pour fournir une liste de fréquences des syllabes.

Le tableau I résume les résultats de cette première enquête sur la distribution des syllabes. En ce qui concerne le but pratique de notre entreprise, le résultat principal s'est avéré très positif : 188 syllabes différentes suffisaient pour représenter 90% du corpus de 285 404 occurrences. En fait la moitié du corpus se trouvait repré-

-
6. G. Gougenheim, P. Rivenc, R. Michea, A. Sauvageot. *L'Elaboration du Français Fondamental*. Paris, Didier, 1964.
 7. Je voudrais remercier Mlle Francine Drolet, qui a fait la transcription et la syllabation des deux premiers corpus, et de la moitié du troisième, et M. Daniel Gagné, qui a fait l'autre moitié du troisième corpus.

sentée par les vingt-cinq premières syllabes. Sans avoir procédé à une analyse statistique à ce moment, nous pouvions voir, de façon impressionniste, que la distribution semblait bien être celle que nous espérions.

Malgré le résultat si encourageant de cette étude, nous n'avions quand même pas d'indication empirique qui permettraient de se fier à cet inventaire de syllabes. En pratiquant la même analyse sur un autre corpus, devions-nous nous attendre à retrouver une forte proportion de ces mêmes syllabes, surtout celles qui apparaissaient en premier dans l'ordre de fréquence, et qui étaient alors les syllabes cruciales pour notre travail. De plus, il nous semblait intéressant de prendre un corpus non-tronqué, contrairement au premier, et donc d'inclure même des mots de fréquence 1. Nous avons donc pris un corpus tout à fait différent, un conte pris dans un livre d'enfants⁸. Le texte a été tapé à l'aide de la même transcription phonologique que la précédente, sans tenir compte des liaisons (s' devenant se, par exemple), et les mots ont été divisés en syllabes. Nous avons fait faire la liste de fréquences des mots d'abord, puis la fréquence des syllabes, toujours présentée de trois façons. Nous avons cependant inclus une différence systématique avec le corpus Gougenheim : tous les verbes ont été transcrits dans leurs formes conjuguées. Nous avons eu dès le début l'intention d'encourager chez nos utilisateurs IMC une forme de communication agrammaticale, où on emploierait, pour économiser les gestes, des formes non conjuguées de verbes. Cependant, nous voulions voir si ce traitement différent des verbes et nous semblait-il, favorable à l'augmentation du nombre de syllabes fermées au détriment des syllabes ouvertes, allait changer de façon importante la constitution du "noyau" de syllabes fréquentes.

8. L. Bellocq *Conte de mes bêtes au vent*, Belgique, Casterman, 1962.

Le Tableau I résume les résultats de cette deuxième étude. Il y avait 1 076 mots différents dans le texte, un nombre à peu près égal à celui de la liste de Gougenheim. Dans ce deuxième corpus il y avait 704 syllabes différentes, ce qui donnait un total d'occurrences de mots de 3 998. Contrairement à la liste de fréquences de Gougenheim, où les mots les plus rares apparaissaient au moins 20 fois, il y avait évidemment dans le conte pour enfants un grand nombre de mots d'une seule occurrence. Et dans ce corpus non présélectionné en fonction de la fréquence des mots, l'emploi des syllabes fréquentes était moins économique : il fallait 304 syllabes pour rendre compte de 90% des occurrences.

Il nous semblait raisonnable de croire que le corpus de Gougenheim, basé sur du langage parlé, auprès de personnes différentes, sur des sujets différents, devait offrir une représentation plus fidèle du langage que le conte pour enfants, fait par une seule personne, sur un sujet précis, et, évidemment, en langage écrit. Si, en effet, le "Français fondamental" de Gougenheim était fondamental, on devrait pouvoir s'attendre à retrouver les syllabes très fréquentes de Gougenheim parmi les syllabes les plus fréquentes du conte pour enfants. Ce qui voudrait dire que le noyau de syllabes fréquentes qu'on recherchait était réel et non pas illusoire.

Le Tableau III résume la comparaison des deux listes de fréquences des syllabes, le Gougenheim et le conte. Dans les 188 syllabes de Gougenheim qui rendent compte de 90% du corpus, les 100 premières, sauf une, se trouvent aussi dans le conte, exception faite de deux syllabes qui se trouvent dans la première liste et non pas dans la seconde à cause du traitement différent accordé aux verbes : /ɛtr/ et /avwar/. De plus, toutes les autres syllabes de la première centaine à part trois (/prɑ̃ndr/, infinitif ou deuxième syllabe d'infinitif ; /kwa/, langage écrit versus langage parlé, et /otr/) figurent dans les 90% du corpus-conte. Dans les vingt-cinq suivantes, il y a deux absences :

/bẽ/, qu'on ne retrouve pas dans le langage écrit du conte, et /krwar/, encore un infinitif. Les omissions vont en croissant, et dans la soixantaine de syllabes suivantes, il y en a six.

Cette comparaison confirmait notre impression que nous convergions vers un ensemble restreint de syllabes fréquentes, si fréquentes que la construction d'un syllabaire serait possible. Cependant nous restions insatisfaits de nos deux premiers corpus. Le Gougenheim était de beaucoup préférable au conte, mais il s'agissait d'un corpus restreint qui, du fait qu'il était tronqué, à un niveau assez haut, de vingt occurrences par mot, risquait de ne pas être représentatif du français parlé au Québec.

Nous avons donc choisi d'analyser le corpus Sankoff-Cedergren⁹. Ce corpus est composé d'interviews menées à Montréal auprès de 120 personnes qui constituent un échantillon représentatif de la population montréalaise. Il comportait au total à peu près un million et demi d'occurrences de mots, quelque vingt mille mots différents. Nous avons pris les 9 866 premiers mots, incluant des mots d'une fréquence minimale de deux. Cette liste représentait à peu près 1 032 000 mots en termes d'occurrences. La liste présentait les verbes sous leurs formes conjuguées, et c'est de cette façon qu'ils ont été transcrits pour le remaniement du corpus en syllabes. L'ambiguïté des verbes de la première conjugaison (exemple, MANGE, qui pouvait être /mãž ou mãže/ a été résolu en faveur des syllabes ouvertes, donc /mãže/. Les mots anglais non assimilés au français québécois, les noms étrangers, ainsi que les chiffres, ont été omis, (les chiffres sont représentés indépendamment sur le tableau de communication).

9. D. Sankoff et H. Cedergren. Corpus du français montréalais. Département d'Anthropologie, Université de Montréal, document inédit.

Le Tableau I présente les résultats du réaménagement de cette liste. Il y avait 2479 syllabes différentes dans 10 000 mots. Mais pour rendre compte de 90% du total des occurrences (1 032 000 mots), il suffisait d'avoir les 269 premières syllabes de la liste. Pour un nombre de mots différents dix fois plus grand que le Gougenheim, et pour un nombre d'occurrences à peu près quatre fois plus grand, la proportion de syllabes qu'il faut ajouter aux 188 premières est de 40% seulement.

Il fallait se demander, comme nous l'avions fait pour le Gougenheim et le conte, si c'était les mêmes syllabes qu'on retrouvait dans le nouveau corpus et surtout, si les syllabes les plus fréquentes du Gougenheim se retrouvaient parmi les syllabes les plus fréquentes du corpus Sankoff-Cedergren. Le Tableau IV résume la comparaison entre les deux corpus.

Mises à part les omissions dues aux traitements différents des deux listes de mots (à savoir la fréquence disproportionnée des infinitifs dans le Gougenheim, et la présence des mots désignant les chiffres, qu'on a omis dans le Sankoff-Cedergren), il n'y a qu'une syllabe dans les 150 premières du Gougenheim qui manque au groupe de syllabes fréquentes du Sankoff-Cedergren, la syllabe /da/. Elle se trouve en fait plus bas dans la deuxième liste, avec un rang de 406. Entre 151 et 188, il y a 7 omissions, /sjø, dam, kε, vɔtr, ʒys, plə, so/. Des différences culturelles sont peut-être à l'origine de ces variations : par exemple, l'emploi moins fréquent en conversation québécoise de "monsieur" et "madame", source majeure des deux premières syllabes de cette liste. Il faut ajouter aussi que ces syllabes se trouvent en effet dans la liste Sankoff-Cedergren, mais en dehors du "noyau" des 269 qui rend compte de 90% du corpus. Le Tableau V présente la liste des syllabes les plus fréquentes du français parlé à Montréal.

Ces résultats étant connus pour le français, nous nous demandons naturellement si cette démarche peut être adaptée à d'autres langues. Nous sommes en train de mener une enquête pilote sur les intuitions des locuteurs natifs de l'anglais sur la syllabation.

Cheryl Goodenough-Trepagnier
Université du Québec à Montréal

Robert M. Frankston
Massachusetts Institute of Technology

TABLEAU I

Caractéristique des 3 corpus et résumé des résultats.

	1	2	3	4	5	6
A.	1 063	285 404	684	375 078	1,31	25 39 62 101 188
B.	1 076	3 998	704	5 290	1,32	43 67 107 173 304
C.	9 866	1 032 876	2 479	1 256 369	1,22	33 51 80 135 269

- A. Corpus Gougenheim : langage parlé, corpus tronqué, fréquence minimale de 20 occurrences par mot, verbes à l'infinitif.
- B. Corpus conte : langage écrit, corpus non-tronqué, verbes conjugués.
- C. Corpus Sankoff-Cedergren : langage parlé, corpus tronqué, fréquence minimale de 2 occurrences par mot, verbes conjugués, verbes ambigus de la 1ère conjugaison interprétés comme syllabes ouvertes, mots étrangers et chiffres omis.

1. Nombre de mots différents.
2. Nombre total d'occurrences de mots.
3. Nombre de syllabes différentes.
4. Nombre total d'occurrences de syllabes.
5. Nombre moyen de syllabes par mot.
6. Nombre de syllabes nécessaires pour rendre compte de 50 à 90% des occurrences : 50 60 70 80 90.

TABLEAU II
Règles de syllabation

- (1) VCV devient V CV
- (2) VCCV devient VC CV
- exception :
- VC $\left\{ \begin{matrix} 1 \\ r \end{matrix} \right\}$ V devient V C $\left\{ \begin{matrix} 1 \\ r \end{matrix} \right\}$ V
- (3) VCCC(C)V devient VC CC(C)V
- exception :
- VCsC(C)V devient VCs C(C)V

TABLEAU III
Comparaison du Gougenheim et du conte

Syllabes figurant dans les 188 premières du Gougenheim (rendant compte de 90% des occurrences) et

ne figurant qu'après le niveau 304 (=90%) du conte	ne figurant pas du tout dans le conte	rang dans le Gougenheim
	etr*	1-25
	lwar*	26-50
prãdr*		51-75
kwa	da	76-100
otr		
tir*	bě**	101-125
	krwar*	
pã	trwa	126-150
do	frã	
eg	vě	
jær	ke	151-175
ba		
mwě		
katr		
tro		
sěk		
fam	plø	176-188
	tetr	

* Traitement différent du verbe dans les deux listes.

** Langage parlé du Gougenheim, versus langage écrit du conte.

TABLEAU IV

Comparaison du Gougenheim et du Sankoff-Cedergren

Syllabes figurant dans les 188 premières du Gougenheim (rendant compte de 90% des occurrences) et

ne figurant qu'après
le niveau 269 (=90%)
du Sankoff-Cedergrenne figurant pas
du tout dans le
Sankoff-Cedergrenrang dans le
Gougenheim

		1-25
lwar*		26-50
		51-75
dø*		76-100
da		
bě**		101-125
metr*		
krwar*		
trwa*	vě*	126-150
sjø	katr*	151-175
dam		
ke		
netr*		
votr		
ȝys		176-188
plə		
so		

* Traitement différent du verbe et des chiffres.

** Différence de représentation dans les deux listes, 'ben' ayant été représenté comme 'bien' dans le Sankoff-Cedergren.

TABLEAU V

Les syllabes les plus fréquentes du français parlé à Montréal

Compilation de données

Nombre total de mots : 1 032 876 Nombre total de syllabes : 1 237 600.

Explication des titres de colonnes :

Nombre : Le rang de chaque syllabe.

Syllabe : Les syllabes en transcription phonémique.

Fréquence
absolue : Nombre total des occurrences de cette syllabe.

Nombre de mots : Nombre de mots différents dans lesquels on retrouve cette syllabe.

Longueur
moyenne : La longueur moyenne des mots où se trouve cette syllabe, en termes de nombre de syllabes.

Début % des fois que cette syllabe constitue la première ou la seule syllabe du mot
Milieu : (=début), la syllabe finale d'un mot de 2 syllabes ou plus (=fin); ou une
Fin syllabe intérieure (=milieu).

% du total
des syllabes : Cette syllabe à elle seule rend compte de cette proportion du total des occurrences.

% cumulatif : Cette syllabe plus les syllabes qui la précèdent, rendent compte de cette proportion du total des occurrences.

NOMBRE	SYLLABE	FRÉQUENCE ABSOLUE	NOMBRE DE MOTS	LONGUEUR MOYENNE	DÉBUT	MILIEU	FIN	% DU TOTAL DES SYLLABES	% CUMULATIF
1	a	70292	631	1.58	99.62	0.37	0.00	5.68	5.68
2	ʒə	37331	28	1.00	99.66	0.31	0.03	3.02	8.70
3	e	33107	307	1.71	99.57	0.36	0.07	2.68	11.37
4	il	30623	6	1.00	99.99	0.00	0.01	2.47	13.85
5	la	30423	68	1.07	96.49	0.64	2.86	2.46	16.30
6	ɛ	29829	86	1.17	99.98	0.00	0.02	2.41	18.71
7	də	27218	93	1.13	97.54	2.45	0.00	2.20	20.91
8	ã	25785	333	1.40	99.21	0.68	0.11	2.08	23.00
9	sa	25566	4	1.00	99.98	0.00	0.02	2.07	25.06
10	pa	22874	16	1.03	99.51	0.22	0.27	1.85	26.91
11	kə	20044	12	1.08	92.28	0.29	7.42	1.62	28.53
12	õ	19560	9	1.01	99.83	0.00	0.17	1.58	30.11
13	pu i	19274	11	1.02	97.72	0.05	2.22	1.56	31.67
14	le	17116	77	1.22	81.59	0.27	18.13	1.38	33.05
15	lə	16819	44	1.04	98.16	1.78	0.05	1.36	34.41
16	mɛ	14807	67	1.26	86.11	0.34	13.55	1.20	35.61
17	bjẽ	13590	8	1.01	99.07	0.00	0.92	1.10	36.71
18	dɛ	13361	86	1.16	93.46	1.89	4.64	1.08	37.78
19	tɛ	13210	130	2.04	2.15	1.03	96.81	1.07	38.85
20	i	12180	87	1.35	99.41	0.33	0.26	0.98	39.84
21	tə	11515	404	2.71	8.23	7.80	83.97	0.93	40.77

NOMBRE	SYLLABE	FRÉQUENCE ABSOLUE	NOMBRE DE MOTS	LONGUEUR MOYENNE	DÉBUT	MILIEU	FIN	% DU TOTAL DES SYLLABES	% CUMULATIF
22	œ	11507	3	1.00	99.76	0.00	0.23	0.93	41.70
23	se	11048	157	1.51	68.61	5.00	26.39	0.89	42.59
24	mã	10659	306	2.75	6.02	17.91	76.06	0.86	43.45
25	o	10558	102	1.78	98.84	1.08	0.07	0.85	44.30
26	wi	10451	5	1.00	99.93	0.04	0.03	0.84	45.15
27	sə	10450	57	1.36	83.61	2.07	14.32	0.84	45.99
28	ve	9985	80	1.92	13.13	1.18	85.69	0.81	46.80
29	tu	9808	22	1.28	91.11	0.21	8.68	0.79	47.59
30	dã	9520	43	1.20	89.34	0.34	10.32	0.77	48.36
31	si	9278	175	1.88	43.37	11.77	44.86	0.75	49.11
32	di	8858	159	1.86	87.22	4.93	7.84	0.72	49.83
33	ki	8538	18	1.01	99.45	0.55	0.00	0.69	50.52
34	ty	7822	71	1.24	86.17	7.78	6.05	0.63	51.15
35	yn	7349	3	1.00	99.74	0.00	0.26	0.59	51.74
36	par	7314	72	1.79	95.39	2.79	1.82	0.59	52.33
37	se	7278	202	2.31	8.60	2.88	88.51	0.59	52.92
38	fe	7139	45	1.13	93.20	5.07	1.72	0.58	53.50
39	pur	6810	8	1.07	99.94	0.06	0.00	0.55	54.05
40	mwa	6687	34	1.10	97.35	1.17	1.48	0.54	54.59
41	le	6477	129	2.28	1.05	13.32	85.62	0.52	55.11
42	ply	6474	6	1.14	99.87	0.00	0.12	0.52	55.64

NOMBRE	SYLLABE	FRÉQUENCE ABSOLUE	NOMBRE DE MOTS	LONGUEUR MOYENNE	DÉBUT	MILIEU	FIN	% DU TOTAL DES SYLLABES	% CUMULATIF	g
43	nõ	6383	17	1.06	96.96	2.60	0.44	0.52	56.15	
44	de	6156	385	3.18	44.43	23.68	31.88	0.50	56.65	
45	el	6023	11	1.11	92.34	3.72	3.93	0.49	57.13	
46	ma	5938	151	1.90	90.40	9.56	0.03	0.48	57.61	
47	tã	5828	98	1.70	61.58	11.54	26.87	0.47	58.09	
48	vu	5799	46	1.37	92.65	0.05	7.29	0.47	58.55	
49	otr	5793	7	1.43	56.74	0.00	43.26	0.47	59.02	
50	ne	5655	171	2.92	5.69	8.70	85.60	0.46	59.48	
51	mõ	5642	53	1.67	92.75	0.85	6.40	0.46	59.94	
52	pø	5622	6	1.31	100.00	0.00	0.00	0.45	60.39	
53	dy	5207	65	1.42	77.87	4.42	17.71	0.42	60.81	
54	ẽ	5119	148	2.86	99.86	0.04	0.10	0.41	61.22	
55	kã	5035	35	1.14	97.85	1.51	0.63	0.41	61.63	
56	kõm	5004	4	1.03	100.00	0.00	0.00	0.40	62.03	
57	mæ	4892	38	1.22	92.17	7.83	0.00	0.40	62.43	
58	sõ	4778	30	1.22	78.61	0.35	21.03	0.39	62.82	
59	pars	4756	1	1.00	100.00	0.00	0.00	0.38	63.20	
60	zur	4708	16	2.21	22.49	36.11	41.40	0.38	63.58	
61	kõ	4629	196	2.49	92.09	7.90	0.00	0.37	63.95	
62	ve	4439	86	2.35	4.14	2.52	93.33	0.36	64.31	
63	mi	4353	157	3.09	26.85	46.20	26.94	0.35	64.67	
64	pa	4337	163	2.30	77.49	18.74	3.76	0.35	65.02	

NOMBRE	SYLLABE	FRÉQUENCE ABSOLUE	NOMBRE DE MOTS	LONGUEUR MOYENNE	DÉBUT	MILIEU	FIN	% DU TOTAL DES SYLLABES	% CUMULATIF
65	fer	4312	33	1.43	63.59	4.61	31.79	0.35	65.36
66	ku	4251	105	2.11	37.42	17.26	45.31	0.34	65.71
67	nu	4224	16	1.49	95.24	0.00	4.76	0.34	66.05
68	re	4180	180	2.39	10.21	13.37	76.41	0.34	66.39
69	ti	4143	206	2.41	3.38	27.92	68.69	0.33	66.72
70	re	4091	285	2.91	32.71	30.68	36.61	0.33	67.05
71	mɛ:m	3987	10	1.05	94.48	0.00	5.52	0.32	67.37
72	va	3667	9	1.01	99.92	0.00	0.08	0.30	67.67
73	u	3623	33	1.30	99.56	0.44	0.00	0.29	67.96
74	pã	3603	58	2.12	81.54	5.97	12.49	0.29	68.25
75	syr	3598	24	1.24	99.03	0.00	0.97	0.29	68.55
76	vek	3499	4	2.00	0.00	0.00	100.00	0.28	68.83
77	dir	3476	2	1.00	99.71	0.00	0.29	0.28	69.11
78	me	3464	139	2.52	22.05	19.77	58.17	0.28	69.39
79	kõ	3438	275	2.62	92.79	6.16	1.05	0.28	69.67
80	ɛ:tr	3299	4	1.54	45.32	0.00	54.68	0.27	69.93
81	pre	3269	31	1.89	32.76	15.42	51.82	0.26	70.20
82	rə	3173	192	2.82	97.70	2.30	0.00	0.26	70.45
83	ri	3107	166	2.87	6.63	60.09	33.28	0.25	70.70
84	ka	3013	183	2.96	73.88	25.95	0.16	0.24	70.95
85	sjõ	2952	239	3.36	0.00	0.23	99.76	0.24	71.19
86	rã	2913	96	2.32	42.12	11.36	46.51	0.24	71.42

NOMBRE	SYLLABE	FRÉQUENCE ABSOLUE	NOMBRE DE MOTS	LONGUEUR MOYENNE	DÉBUT	MILIEU	FIN	% DU TOTAL DES SYLLABES	% CUMULATIF
87	sqi	2906	12	1.05	99.79	0.20	0.00	0.23	71.66
88	ʒe	2861	133	2.63	14.01	3.25	82.73	0.23	71.89
89	je	2816	99	2.86	0.00	8.41	91.58	0.23	72.12
90	bo	2768	27	1.77	99.60	0.18	0.22	0.22	72.34
91	sa	2762	56	1.88	95.36	3.44	1.19	0.22	72.56
92	per	2695	76	1.76	85.64	6.90	7.46	0.22	72.78
93	vwar	2679	10	1.73	29.15	0.00	70.85	0.22	73.00
94	zõ	2673	23	2.04	0.00	3.10	96.89	0.22	73.21
95	pə	2610	23	2.07	92.91	7.09	0.00	0.21	73.42
96	ʃe	2523	84	1.73	45.22	0.12	54.66	0.20	73.63
97	y	2504	60	1.95	88.22	11.78	0.00	0.20	73.83
98	ʃoz	2437	5	1.01	98.56	0.00	1.43	0.20	74.03
99	fa	2401	60	2.17	99.46	0.54	0.00	0.19	74.22
100	vã	2389	64	2.16	11.93	7.32	80.74	0.19	74.41
101	lqi	2360	8	1.13	91.06	2.88	6.06	0.19	74.60
102	ne	2353	101	2.47	1.78	20.27	77.94	0.19	74.79
103	tra	2346	63	2.87	94.33	5.67	0.00	0.19	74.98
104	vi	2295	93	2.54	47.27	43.92	8.80	0.19	75.17
105	ze	2238	28	2.03	0.13	0.00	99.87	0.18	75.35
106	vø	2216	9	1.10	91.20	0.00	8.80	0.18	75.53
107	vwa	2152	52	1.84	85.41	7.53	7.06	0.17	75.70
108	rir	2106	10	1.05	94.54	0.00	5.46	0.17	75.87

NOMBRE	SYLLABE	FRÉQUENCE ABSOLUE	NOMBRE DE MOTS	LONGUEUR MOYENNE	DÉBUT	MILIEU	FIN	% DU TOTAL DES SYLLABES	% CUMULATIF
109	ter	2102	72	4.00	11.13	65.17	23.69	0.17	76.04
110	fwa	2084	10	1.11	94.62	0.09	5.28	0.17	76.21
111	li	2032	161	3.16	26.18	66.44	7.38	0.16	76.38
112	la	1987	104	2.60	28.94	0.25	70.81	0.16	76.54
113	sã	1971	97	1.83	60.73	9.28	29.98	0.16	76.70
114	zã	1946	35	1.50	67.57	1.28	31.14	0.16	76.85
115	va	1945	57	2.91	14.03	85.96	0.00	0.16	77.01
116	o	1940	63	2.41	99.12	0.88	0.00	0.16	77.17
117	kel	1899	18	1.82	96.63	0.00	3.37	0.15	77.32
118	lœr	1887	16	1.11	88.39	0.00	11.60	0.15	77.47
119	tut	1873	3	1.11	88.52	0.00	11.48	0.15	77.62
120	rjẽ	1872	3	1.01	99.36	0.37	0.27	0.15	77.78
121	bõ	1835	15	1.13	94.38	0.22	5.39	0.15	77.92
122	ra	1831	134	2.73	62.91	31.13	5.95	0.15	78.07
123	mal	1801	14	1.19	92.06	0.83	7.10	0.15	78.22
124	pro	1771	90	2.71	98.19	1.81	0.00	0.14	78.36
125	dui	1766	7	2.96	0.00	0.17	99.83	0.14	78.50
126	tə	1764	45	2.11	58.39	41.61	0.00	0.14	78.65
127	frã	1756	27	2.26	90.26	9.56	0.17	0.14	78.79
128	ni	1725	113	3.15	16.52	47.88	35.59	0.14	78.93
129	və	1717	47	2.39	64.47	35.53	0.00	0.14	79.07
130	lœr	1706	3	2.00	2.64	0.17	97.19	0.14	79.20

NOMBRE	SYLLABE	FRÉQUENCE ABSOLUE	NOMBRE DE MOTS	LONGUEUR MOYENNE	DÉBUT	MILIEU	FIN	% DU TOTAL DES SYLLABES	% CUMULATIF
131	ver	1647	57	2.79	20.89	40.68	38.43	0.13	79.34
132	sẽ	1610	112	2.59	91.36	0.56	8.07	0.13	79.47
133	fə	1605	16	2.00	99.69	0.31	0.00	0.13	79.60
134	tru	1598	17	1.99	97.18	2.81	0.00	0.13	79.73
135	fã	1597	25	2.05	1.75	3.25	94.99	0.13	79.85
136	res	1582	38	2.23	97.03	1.01	1.96	0.13	79.98
137	ø	1581	13	1.76	97.97	0.25	1.77	0.13	80.11
138	su	1575	34	2.00	94.67	0.00	5.33	0.13	80.24
139	fo	1558	16	1.15	97.30	0.00	2.69	0.13	80.36
140	mõd	1526	2	1.00	100.00	0.00	0.00	0.12	80.49
141	dɔ	1513	57	2.17	90.02	9.98	0.00	0.12	80.61
142	mɛr	1494	27	1.41	73.63	4.82	21.55	0.12	80.73
143	gar	1476	32	2.49	52.03	47.56	0.41	0.12	80.85
144	ʒa	1476	9	2.01	99.39	0.61	0.00	0.12	80.97
145	ke	1475	65	2.36	45.90	1.15	52.95	0.12	81.09
146	tre	1471	42	1.26	84.02	1.97	14.00	0.12	81.21
147	fi	1465	92	2.77	53.99	39.72	6.28	0.12	81.32
148	ʒɛn	1449	2	1.00	100.00	0.00	0.00	0.12	81.44
149	pe	1403	104	2.73	11.19	35.35	53.45	0.11	81.55
150	kɔr	1397	19	2.02	4.15	2.15	93.70	0.11	81.67
151	prə	1367	28	2.22	85.88	14.12	0.00	0.11	81.78
152	pri	1362	43	2.01	49.85	22.17	27.97	0.11	81.89

NOMBRE	SYLLABE	FRÉQUENCE ABSOLUE	NOMBRE DE MOTS	LONGUEUR MOYENNE	DÉBUT	MILIEU	FIN	% DU TOTAL DES SYLLABES	% CUMULATIF
153	ba	1360	40	1.46	74.41	0.59	25.00	0.11	82.00
154	set	1358	11	1.01	98.75	0.00	1.25	0.11	82.11
155	tro	1349	13	1.13	92.51	2.67	4.82	0.11	82.22
156	ga	1347	16	1.16	98.81	0.00	1.19	0.11	82.33
157	vjě	1340	23	1.53	51.19	1.34	47.46	0.11	82.43
158	ǰã	1324	62	2.01	90.10	3.77	6.12	0.11	82.54
159	na	1322	87	3.64	23.52	73.45	3.02	0.11	82.65
160	ta	1297	110	2.82	36.31	62.22	1.46	0.10	82.75
161	to	1284	50	2.32	2.33	16.04	81.62	0.10	82.86
162	te	1278	16	1.83	92.64	0.63	6.73	0.10	82.96
163	kol	1272	10	1.99	1.73	0.94	97.33	0.10	83.06
164	al	1257	21	2.93	8.19	0.63	91.17	0.10	83.16
165	trwa	1251	2	1.00	100.00	0.00	0.00	0.10	83.26
166	ser	1251	44	2.21	83.69	7.83	8.47	0.10	83.37
167	ny	1251	43	2.42	3.12	5.44	91.45	0.10	83.47
168	ry	1236	13	1.07	96.36	0.00	3.64	0.10	83.57
169	kwa	1213	9	1.56	60.76	0.16	39.07	0.10	83.66
170	vre	1195	15	1.46	83.35	0.25	16.40	0.10	83.76
171	pu	1179	34	2.07	95.76	4.07	0.17	0.10	83.86
172	sor	1158	20	1.96	85.40	0.69	13.90	0.09	83.95
173	ær	1153	7	1.07	96.70	0.00	3.29	0.09	84.04
174	tre	1149	29	2.17	0.52	0.17	99.30	0.09	84.14

NOMBRE	SYLLABE	FRÉQUENCE ABSOLUE	NOMBRE DE MOTS	LONGUEUR MOYENNE	DÉBUT	MILIEU	FIN	% DU TOTAL DES SYLLABES	% CUMULATIF	Σ
175	eg	1122	32	2.53	100.00	0.00	0.00	0.09	84.23	
176	fɪj	1103	6	1.00	99.82	0.00	0.18	0.09	84.32	
177	grã	1047	20	1.62	98.57	0.28	1.14	0.08	84.40	
178	mo	1044	31	1.55	88.02	5.75	6.23	0.08	84.48	
179	võ	1039	6	1.06	93.45	0.00	6.54	0.08	84.57	
180	mwě	1036	2	1.01	100.00	0.00	0.00	0.08	84.65	
181	je	1034	51	2.51	0.00	0.00	100.00	0.08	84.74	
182	prã	1026	27	1.60	51.46	7.11	41.42	0.08	84.82	
183	lõ	1020	31	1.86	72.55	3.14	24.31	0.08	84.90	
184	bl	1007	72	3.09	12.91	81.43	5.66	0.08	84.98	
185	eks	998	63	2.96	100.00	0.00	0.00	0.08	85.06	
186	mɔ	996	65	2.49	84.74	15.26	0.00	0.08	85.14	
187	vy	980	14	1.14	89.59	0.00	10.41	0.08	85.22	
188	ko	972	27	2.12	91.36	2.06	6.58	0.08	85.30	
189	sær	955	13	1.43	74.13	0.00	25.86	0.08	85.38	
190	gle	953	4	2.03	0.00	0.21	99.79	0.08	85.46	
191	m	920	1	1.00	100.00	0.00	0.00	0.07	85.53	
192	reɟ	917	7	2.00	0.00	0.00	100.00	0.07	85.60	
193	tě	908	29	2.10	1.76	3.08	95.15	0.07	85.68	
194	rje	906	25	2.02	0.00	0.88	99.11	0.07	85.75	
195	tɪt	900	2	2.00	0.00	0.00	100.00	0.07	85.82	

NOMBRE	SYLLABE	FRÉQUENCE ABSOLUE	NOMBRE DE MOTS	LONGUEUR MOYENNE	DÉBUT	MILIEU	FIN	% DU TOTAL DES SYLLABES	% CUMULATIF
196	be	881	51	2.34	38.02	23.83	38.14	0.07	85.89
197	pɔr	881	41	2.62	22.13	61.41	16.46	0.07	85.97
198	ga	875	76	2.88	45.37	54.63	0.00	0.07	86.04
199	vɪl	870	23	1.30	87.93	0.00	12.07	0.07	86.11
200	mʃø	865	2	1.00	99.65	0.00	0.34	0.07	86.18
201	sy	859	44	2.63	40.86	12.22	46.91	0.07	86.25
202	sæl	834	4	1.67	100.00	0.00	0.00	0.07	86.31
203	pɛl	811	16	1.99	3.33	0.74	95.93	0.07	86.38
204	ky	807	66	2.92	14.25	66.91	18.83	0.07	86.44
205	ze	807	104	3.07	2.73	2.35	94.92	0.07	86.51
206	nɔtr	801	1	1.00	100.00	0.00	0.00	0.06	86.57
207	mɛn	800	19	2.04	2.50	0.50	97.00	0.06	86.64
208	fe	796	47	2.93	9.42	76.88	13.69	0.06	86.70
209	pɛ	794	54	2.11	68.39	12.47	19.14	0.06	86.77
210	ba	790	106	2.68	70.25	26.58	3.16	0.06	86.83
211	kur	783	14	1.13	90.04	0.00	9.96	0.06	86.89
212	no	778	32	1.61	67.09	6.81	26.09	0.06	86.96
213	ak	772	37	2.54	96.89	3.11	0.00	0.06	87.02
214	nə	767	8	1.15	95.83	4.17	0.00	0.06	87.08
215	tʃe	766	28	2.07	0.00	2.87	97.13	0.06	87.14
216	ʒyst	764	2	1.00	100.00	0.00	0.00	0.06	87.21

NOMBRE	SYLLABE	FRÉQUENCE ABSOLUE	NOMBRE DE MOTS	LONGUEUR MOYENNE	DÉBUT	MILIEU	FIN	% DU TOTAL DES SYLLABES	% CUMULATIF	∞
217	mar	762	34	2.18	88.19	6.30	5.51	0.06	87.27	
218	ka	760	33	1.59	83.81	0.79	15.39	0.06	87.33	
219	swa	745	26	1.50	80.67	1.74	17.58	0.06	87.39	
220	pɔ	741	50	2.77	88.80	11.20	0.00	0.06	87.45	
221	lã	726	39	2.43	51.79	29.61	18.59	0.06	87.51	
222	ly	724	31	2.75	20.86	44.20	34.94	0.06	87.57	
223	ʃə	719	29	2.45	57.44	42.56	0.00	0.06	87.62	
224	par	714	3	1.02	97.90	0.00	2.10	0.06	87.68	
225	tõ	713	30	1.72	59.89	0.56	39.55	0.06	87.74	
226	ar	707	39	2.12	99.15	0.85	0.00	0.06	87.80	
227	prãdr	702	4	1.31	68.80	0.00	31.19	0.06	87.85	
228	sěk	696	1	1.00	0.00	0.00	0.00	0.06	87.91	
229	fě	695	6	1.46	56.26	0.00	43.74	0.06	87.97	
230	młj	694	2	2.00	0.00	0.00	100.00	0.06	88.02	
231	swar	676	7	1.13	90.83	0.00	9.17	0.05	88.08	
232	fɔr	666	44	1.84	78.38	9.16	12.46	0.05	88.13	
233	fõ	663	24	1.22	90.95	6.48	2.56	0.05	88.18	
234	plas	660	10	1.07	98.18	0.45	1.36	0.05	88.24	
235	zãpl	660	2	2.00	0.00	0.00	100.00	0.05	88.29	
236	vr	659	27	2.73	42.03	7.13	50.83	0.05	88.34	
237	frer	652	7	1.09	91.10	0.00	8.89	0.05	88.40	

NOMBRE	SYLLABE	FRÉQUENCE ABSOLUE	NOMBRE DE MOTS	LONGUEUR MOYENNE	DÉBUT	MILIEU	FIN	% DU TOTAL DES SYLLABES	% CUMULATIF
238	nɪr	651	17	2.38	0.00	0.00	100.00	0.05	88.45
239	mwa	650	1	1.00	100.00	0.00	0.00	0.05	88.50
240	pɪ	645	57	2.48	35.66	57.36	6.98	0.05	88.55
241	sãbɪ	642	6	1.42	57.63	0.00	42.37	0.05	88.60
242	zjɔ̃	641	14	3.80	0.00	0.00	100.00	0.05	88.66
243	po	634	58	2.47	33.12	37.38	29.50	0.05	88.71
244	tur	628	30	2.07	51.11	20.06	28.82	0.05	88.76
245	kar	619	23	1.97	96.77	3.23	0.00	0.05	88.81
246	sɥɪt	613	3	1.47	52.04	0.00	47.96	0.05	88.86
247	ʃer	612	15	1.64	95.75	4.25	0.00	0.05	88.91
248	zɪ	608	67	3.01	0.66	68.59	30.76	0.05	88.96
249	twa	606	26	1.46	77.06	7.59	15.34	0.05	89.01
250	a	584	5	1.13	99.31	0.00	0.68	0.05	89.05
251	fam	583	2	1.00	100.00	0.00	0.00	0.05	89.10
252	krwa	575	19	1.20	96.87	2.61	0.52	0.05	89.15
253	tɪr	574	7	2.01	0.00	0.00	100.00	0.05	89.19
254	dɪs	568	50	2.57	64.26	0.53	35.21	0.05	89.24
255	parɪ	557	4	1.00	100.00	0.00	0.00	0.05	89.28
256	sɔn	557	11	2.00	2.15	1.07	96.77	0.05	89.33
257	trə	555	25	3.06	0.00	100.00	0.00	0.04	89.37
258	nã	555	19	2.10	0.00	1.98	98.02	0.04	89.42

NOMBRE	SYLLABE	FRÉQUENCE ABSOLUE	NOMBRE DE MOTS	LONGUEUR MOYENNE	DÉBUT	MILIEU	FIN	% DU TOTAL DES SYLLABES	% CUMULATIF
259	vo	554	21	2.01	21.48	7.76	70.76	0.04	89.46
260	bɔr	552	10	1.55	49.45	0.36	50.18	0.04	89.51
261	mjer	549	6	2.16	0.00	13.30	86.70	0.04	89.55
262	bɛk	539	3	2.02	0.37	0.00	99.63	0.04	89.60
263	vjø	539	2	1.03	100.00	0.00	0.00	0.04	89.64
264	gro	538	6	1.05	99.44	0.56	0.00	0.04	89.68
265	dök	535	1	1.00	100.00	0.00	0.00	0.04	89.73
266	bɛl	533	21	1.41	90.80	0.00	9.19	0.04	89.77
267	jær	532	13	2.05	0.00	0.00	100.00	0.04	89.81
268	bɔn	531	4	1.03	100.00	0.00	0.00	0.04	89.85
269	ma	526	15	2.21	74.90	6.84	18.25	0.04	89.90
270	sll	524	14	2.72	3.43	11.64	84.92	0.04	89.94
271	bll	519	22	3.09	0.00	93.83	6.16	0.04	89.98
272	ʒwal	518	2	1.00	100.00	0.00	0.00	0.04	90.02
273	sɔ	517	44	2.67	72.73	27.27	0.00	0.04	90.07
274	tyr	506	37	2.52	0.00	0.59	99.40	0.04	90.11
275	aʒ	506	2	1.00	100.00	0.00	0.00	0.04	90.15
276	but	501	1	1.00	100.00	0.00	0.00	0.04	90.19
277	vɔ	500	36	3.09	49.00	51.00	0.00	0.04	90.23
278	so	494	27	2.95	28.95	49.59	21.46	0.04	90.27
279	mě	492	14	1.96	33.74	0.00	66.26	0.04	90.31
280	by	488	23	2.32	46.11	21.51	32.38	0.04	90.35