

3 Listes de 10 phrases françaises phonétiquement équilibrées

par Matthew LENNIG*

Résumé

On présente ici 3 listes de 10 phrases françaises phonétiquement équilibrées. La valeur du χ^2 des distributions phonémiques entre chaque liste de phrase et un corpus de référence (français de théâtre) est inférieure à 4,0.

PHONETICALLY BALANCED FRENCH SENTENCES

Abstract

Three phonetically balanced lists of ten sentences in French are proposed. The ten sentences are determined such that the distribution of phonemes in each list is kept as close as possible from that of a reference corpus (spoken French used in Theater). When the difference of distributions between any of the list and the reference is measured by χ^2 distance, the value of χ^2 was less than 4.0.

Il existe des phrases phonétiquement équilibrées pour l'anglais [1] mais pas pour le français. Ces listes de dix phrases chacune ont des fréquences relatives par chaque phonème qui reflètent celles de la langue. Les listes de phrases phonétiquement équilibrées sont utiles pour les essais de systèmes de communication où il est important d'avoir un corpus de longueur réduite qui soit phonologiquement représentatif de la langue entière. Par exemple, si l'on veut évaluer l'intelligibilité ou la qualité de la parole codée ou synthétisée sur un texte réduit, une liste de phrases phonétiquement équilibrée est utile.

Cet article offre trois listes de phrases phonétiquement équilibrées. Chaque liste consiste en dix phrases. Les listes sont phonétiquement équilibrées dans le sens expliqué ci-dessous.

Ce que l'on appelle ici équilibre phonétique est plus précisément un équilibre phonémique que phonétique. L'analyse phonémique standard utilisée ici emploie 36 phonèmes. Cette analyse correspond à celle du *Petit Robert* à l'exclusion de 3 phonèmes étrangers, [h], [ŋ] et [x], et du phonème abstrait « h-aspiré ». On ne distingue pas de longueur vocalique. Par contre, on distingue bien les paires [a] ~ [ɑ], [e] ~ [ɛ] et [ɛ̃] ~ [œ̃]. Toutes les liaisons usuelles de la prononciation soignée sont transcrites. Le taux d'occurrence de chaque phonème dans une liste est calculé par ordinateur à partir d'une transcription phonémique. Ces taux sont ensuite comparés aux fréquences relatives des phonèmes du français calculées sur un grand corpus de théâtre français du XX^e siècle [2].

La valeur du χ^2 est calculée pour mesurer la différence de distribution phonémique entre une liste proposée et le corpus de référence. Si le χ^2 total entre les fréquences de la liste et celles du corpus de référence est inférieur à 4,0 on accepte la liste comme étant « phonétiquement équilibrée ». Si χ^2 est supérieur ou égal à 4,0, on modifie les phrases de la liste en essayant de changer des mots pour faire baisser le χ^2 tout en

gardant la grammaticalité et le naturel des phrases. On exige aussi que chaque phonème apparaisse au moins une fois dans chaque liste.

Ce processus permet d'obtenir des listes qui sont semblables au corpus de référence pour les statistiques de premier ordre. On n'a pas pu rendre les listes semblables au corpus de référence pour les statistiques d'ordres plus élevés sans allonger considérablement les listes. Néanmoins, les caractéristiques de second ordre sont probablement importantes pour certaines utilisations du corpus d'essai.

Le choix du seuil d'acceptation d'une liste de χ^2 total = 4,0 était pragmatique : la manipulation des phrases a montré qu'il est difficile, bien que possible de faire baisser le χ^2 total d'une liste en dessous de 4,0. On voit dans les Tableaux 1, 2 et 3 que les fréquences relatives d'occurrence des phonèmes des listes et du corpus de référence sont très proches.

Liste N° 1

1. Est-ce que le conducteur arrête l'auto ?
/ɛs kə lə kɔ̃dyktœr arɛt loto/
2. C'est toujours comme ça depuis dix ans, tu sais.
/sɛ tuzur kɔ̃m sa dɔ̃pmi di zɑ̃ ty-sɛ/
3. Ce cheval ne peut pas marcher au pas.
/sə ʃəval nɔ̃ pə pa marʃe o pa/
4. La bière est moins forte que le rhum.
/la bjɛr ɛ mwɛ̃ fɔ̃rt kə lə rɔ̃m/
5. Ici il fait toujours très froid en hiver.
/isi il fɛ tuzur trɛ frwa ɑ̃ nivɛr/
6. J'aime Sylvie quand elle est mignonne.
/zɛm silvi kɑ̃ tɛl ɛ mi ʒɔ̃n/
7. Diane ne reviendra pas avant lundi.
/djan nɔ̃ rɔ̃vjɛ̃dra pa zavɑ̃ lœ̃di/
8. Aimez-vous de dessin ?
/eme vu sɔ̃ desɛ̃/
9. J'ai déjà lu la réponse qu'il m'a envoyée par la poste.
/ze deza ly la repɔ̃s kil ma ɑ̃vwaje par la pɔ̃st/
10. Mes gants sont usés.
/mɛ ɡɑ̃ sɔ̃ tyzɛ/

* Recherches Bell-Northern, 3 Place du Commerce - Ile des Sœurs, Québec, Canada H3E 1H6

Tableau I
Statistiques de fréquences relatives pour la liste 1

Pho- nème	Nombre d'occurrences par phrase										Total pour les 10 phrases	Fréquence (%)		χ^2
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10		Pour les 10 phrases	Pour le grand corpus	
r	2	1	1	3	4	0	2	0	2	0	15	7.04	7.55	0.03
a	1	1	2	1	1	0	3	0	6	0	15	7.04	7.14	0.00
ə	2	1	3	2	0	0	2	1	0	0	11	5.16	5.59	0.03
l	2	0	1	2	1	2	1	0	4	0	13	6.10	5.43	0.08
s	1	3	1	0	1	0	2	2	1		12	5.63	5.35	0.02
ɛ	2	2	0	2	3	3	0	0	2		13	6.10	5.25	0.14
e	0	0	1	0	0	0	0	3	4	1	9	4.23	5.12	0.16
t	3	2	0	1	2	1	0	0	1	1	11	5.16	5.05	0.00
i	0	2	0	0	4	3	1	0	1	0	11	5.16	4.93	0.01
m	0	1	1	2	0	2	0	1	1	1	9	4.23	4.09	0.00
d	1	2	0	0	0	0	3	11	0		8	3.76	3.99	0.01
k	3	1	0	1	0	1	0	0	1	0	7	3.29	3.74	0.06
p	0	1	3	0	0	0	1	0	3	0	8	3.76	3.68	0.00
v	0	0	1	0	1	1	2	1	1	0	7	3.29	3.21	0.00
n	0	0	1	0	1	1	2	0	0	0	5	2.35	2.89	0.10
ɹ	0	1	0	0	1	1	1	0	1	1	6	2.82	2.89	0.00
u	0	2	0	0	2	0	0	1	0	0	5	2.35	2.86	0.09
z	0	1	0	0	1	1	0	0	2	0	5	2.35	2.08	0.03
y	1	1	0	0	0	0	0	0	1	1	4	1.88	2.04	0.01
ʊ	0	1	0	2	0	1	0	1	0	0	5	2.35	1.94	0.09
o	1	0	0	0	0	0	0	0	1	1	3	1.41	1.85	0.11
j	0	0	0	1	0	0	2	0	1	0	4	1.88	1.81	0.00
w	0	0	0	1	1	0	0	0	1	0	3	1.41	1.50	0.01
f	0	0	0	1	2	0	0	0	0	0	3	1.41	1.31	0.01
b	0	0	0	1	0	0	0	0	0	0	1	0.47	1.20	0.44
a	0	0	2	0	0	0	1	0	0	0	3	1.41	1.19	0.04
o	2	0	1	0	0	0	0	0	0	0	3	1.41	1.10	0.09
ɛ	0	0	0	1	0	0	1	1	0	0	3	1.41	1.09	0.09
z	0	1	0	0	0	0	1	0	0	1	3	1.41	0.73	0.63
ɸ	0	0	1	0	0	0	0	0	0	0	1	0.47	0.71	0.08
ʃ	0	0	2	0	0	0	0	0	0	0	2	0.94	0.62	0.16
ɣ	0	1	0	0	0	0	0	0	0	0	1	0.47	0.54	0.01
ʁ	1	0	0	0	0	0	0	0	0	0	1	0.47	0.51	0.00
g	0	0	0	0	0	0	0	0	1		1	0.47	0.46	0.00
ʁ	0	0	0	0	0	0	1	0	0	0	1	0.47	0.43	0.00
ɲ	0	0	0	0	0	1	0	0	0	0	1	0.47	0.12	1.02

Nombre total de phonèmes = 213

$\chi^2 = 3.58$

Liste N° 2

1. C'et en forgeant qu'on devient forgeron.
/sɛ tã fãrzã kõ dð vjẽ fãrzðrõ/
2. Je voudrais des frites s'il vous plaît.
/zð vudrɛ dɛ frit sil vu plãt/
3. Quant à moi, je n'ai pas envie d'y aller.
/kã ta mwa zð ne pa zãvi di ale/
4. C'est une fille tout à fait à la page.
/sɛ tyn fij tu ta fɛ ta la paz/
5. Tu mets combien d'œufs pour préparer une quiche ?
/ty mɛ kõbjẽ dɤ pur prepare yn kif/
6. Marie-Hélène roule en moto tandis que Céline se déplace en métro.
/mari elɛn rul ã moto tãdi kðselinsð deplas ã metro/
7. Voulez-vous lui passer le beurre ?
/vule vu ly pase lðbœr/
8. Moi, ce que je veux c'est la tranquillité pour chacun.
/mwa sðkðzðvɤ sɛ trãkilitɛ pur fakõ/
9. Il s'occupe de ses oignons !
/il sɔkyp dð sɛ zɤnõ/
10. La perte de ma mère m'a jeté hors de mes gonds.
/la pɛrt dð ma mɛr ma zðteɔr dð mɛ gõd/

Tableau 2
Statistiques de fréquences relatives pour la liste 2

Pho- nème	Nombre d'occurrences par phrase										Total	Fréquence (%)		χ²
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	Pour les 10 phrases	Pour les 10 phrases	Pour le grand corpus	
r	3	2	0	0	3	3	1	2	0	3	17	6.94	7.55	0.05
a	0	0	3	4	1	2	0	3	0	3	16	6.53	7.14	0.05
ð	2	1	1	0	0	2	1	3	1	3	14	5.71	5.59	0.00
l	0	2	1	1	0	4	3	2	1	1	15	6.12	5.43	0.09
s	1	1	0	1	0	3	1	2	2	0	11	4.49	5.35	0.14
ɛ	1	3	0	2	1	1	0	1	1	3	13	5.31	5.25	0.00
e	0	0	2	0	2	4	2	1	0	1	12	4.90	5.12	0.01
t	1	1	1	4	1	3	0	2	0	2	15	6.12	5.05	0.23
i	0	2	2	1	1	3	1	2	1	0	13	5.31	4.93	0.03
m	0	0	1	0	1	3	0	1	0	4	10	4.08	4.09	0.00
d	1	2	1	0	1	2	0	0	1	2	10	4.08	3.99	0.00
k	1	0	1	0	2	1	0	3	1	0	9	3.67	3.74	0.00
p	0	1	1	1	3	1	1	1	1	1	11	4.49	3.68	0.18
v	1	2	1	0	0	0	2	1	0	0	7	2.86	3.21	0.04
n	0	0	1	1	1	2	0	0	0	0	5	2.04	2.89	0.25
ã	2	0	2	0	0	3	0	1	0	0	8	3.27	2.89	0.05
u	0	2	0	1	1	1	2	1	0	0	8	3.27	2.86	0.06
z	2	1	1	1	0	0	0	1	0	1	7	2.86	2.08	0.29
y	0	0	0	1	2	0	0	0	1	0	4	1.63	2.04	0.08
ɔ	2	0	0	0	0	0	0	0	2	1	5	2.04	1.94	0.01
o	2	0	0	0	1	0	0	0	1	1	5	2.04	1.85	0.02
j	1	0	0	1	1	0	0	0	0	0	3	1.22	1.81	0.19
w	0	0	1	0	0	0	0	1	0	0	2	0.82	1.50	0.31
f	2	1	0	2	0	0	0	0	0	0	5	2.04	1.31	0.41
b	0	0	0	0	1	0	1	0	0	0	2	0.82	1.20	0.12
a	0	0	1	0	0	0	1	0	0	0	2	0.82	1.19	0.12
o	0	0	0	0	0	3	0	0	0	0	3	1.22	1.10	0.01
ẽ	1	0	0	0	1	0	0	0	0	0	2	0.82	1.09	0.07
z	0	0	1	0	0	0	0	0	1	0	2	0.82	0.73	0.01
ɤ	0	0	0	0	1	0	0	1	0	0	2	0.82	0.71	0.02
ʃ	0	0	0	0	1	0	0	1	0	0	2	0.82	0.62	0.06
ɣ	0	0	0	0	0	0	1	0	0	0	1	0.41	0.54	0.03
œ	0	0	0	0	0	0	1	0	0	0	1	0.41	0.51	0.02
g	0	0	0	0	0	0	0	0	0	1	1	0.41	0.46	0.01
œ	0	0	0	0	0	0	0	1	0	0	1	0.41	0.43	0.00
η	0	0	0	0	0	0	0	0	1	0	1	0.41	0.12	0.69

Nombre total de phonèmes = 245.

χ²

= 3.64

Liste N° 3

1. Le ronronnement du moteur a réveillé les gens.
/lθ rθrɔnmã dy mɔtœr a reveje lɛ zã/
2. La salade se mange à la vinaigrette.
/la salad sθ mãz a la vinɛgrɛt
3. Il y a un chien qui aboie.
/il i a œ fʃjɛ ki abwa/
4. Lorsque Jean a écrit cet article, sa femme était tellement fâchée qu'elle l'a quitté.
/lɔrskθ zã a ekri sɛt artiklθ sa fam etɛ tɛlmã fafɛkɛl la kite/
5. Par considération pour vos voisins, veuillez à ce que le moins de bruit possible émane de vos fêtes.
/par kɔsiderasjɔ pur vo vwazɛ veje a sθ kθ lθ mwɛ dθbryi posiblθeman dθvo fɛt/

6. Nous nous sommes baignés à Saint-Jean.
/nu nu sɔm beŋ a sɛ zã/
7. On ne peut pas chanter la chanson sans musique.
/ɔ nθpθ pa ʃãtɛ ʃãsɔ sã mysik/
8. Le mauve te va très bien.
/lθ mov tθ va trɛ bjɛ/
9. Des jardins ont été aménagés pour les gens qui voudraient faire pousser leurs propres légumes et fruits.
/dɛ zardɛ ɔ tete amenaze pur lɛ zã ki vudrɛ fɛr puse lœr proprθ legym ɛ frʃi/
10. Voici une liste de services qui peut être utile.
/vwasi yn list dθ sɛrvɛs ki pθ tɛtr ytil/

Tableau 3
Statistiques de fréquences relatives pour la liste 3

Pho- nème	Nombre d'occurrences par phrase										Total pour les 10 phrases	Fréquence (%)		χ²
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10		Pour les 10 phrases	Pour le grand corpus	
r	4	1	0	3	4	0	0	1	8	2	23	7.17	7.55	0.02
a	1	5	3	5	4	1	1	1	3	1	25	7.79	7.14	0.06
θ	1	1	0	2	6	0	1	2	1	1	15	4.67	5.59	0.15
l	2	3	1	5	2	0	1	1	3	2	20	6.23	5.43	0.12
s	0	2	0	3	4	2	2	0	1	4	18	5.61	5.35	0.01
ɛ	1	2	0	4	1	0	0	1	5	2	16	4.98	5.25	0.01
e	3	0	0	4	4	2	1	0	6	0	20	6.23	5.12	0.24
t	1	1	0	5	1	0	1	2	2	4	17	5.30	5.05	0.01
i	0	1	3	3	3	0	1	0	2	5	18	5.61	4.93	0.09
m	2	1	0	2	2	1	1	1	2	0	12	3.74	4.09	0.03
d	1	1	0	0	3	0	0	0	3	1	9	2.80	3.99	0.35
k	0	0	1	5	2	0	1	0	1	1	11	3.43	3.74	0.03
p	0	0	0	0	3	0	2	0	4	1	10	3.12	3.68	0.09
v	1	1	0	0	4	0	0	2	1	2	11	3.43	3.21	0.01
n	1	1	0	0	2	2	1	0	1	1	9	2.80	2.89	0.00
zã	2	1	0	2	0	1	3	0	1	0	10	3.12	2.89	0.02
u	0	0	0	0	1	2	0	0	3	0	6	1.87	2.86	0.34
z	1	1	0	1	0	1	0	0	3	0	7	2.18	2.08	0.00
y	1	0	0	0	0	0	1	0	1	2	5	1.56	2.04	0.11
ɔ	2	0	0	1	1	1	0	0	1	0	6	1.87	1.94	0.00
ɑ	1	0	0	0	2	0	2	0	1	0	6	1.87	1.85	0.00
j	1	0	1	0	2	0	0	1	0	0	5	1.56	1.81	0.04
w	0	0	1	0	2	0	0	0	0	1	4	1.25	1.50	0.04
f	0	0	0	2	1	0	0	0	2	0	5	1.56	1.31	0.05
b	0	0	1	0	2	1	0	1	0	0	5	1.56	1.20	0.11
a	0	0	0	1	1	0	1	0	0	0	3	0.93	1.19	0.05
o	0	0	0	0	2	0	0	1	0	0	3	0.93	1.10	0.02
ɛ̃	0	0	1	0	2	1	0	1	1	0	6	1.87	1.09	0.56
z	0	0	0	0	1	0	1	0	0	0	2	0.62	0.73	0.02
φ	0	0	0	0	0	0	1	0	0	1	2	0.62	0.71	0.01
ʃ	0	0	1	1	0	0	2	0	0	0	4	1.25	0.62	0.63
y	0	0	0	0	1	0	0	0	1	0	2	0.62	0.54	0.01
œ	1	0	0	0	0	0	0	0	1	0	2	0.62	0.51	0.03
g	0	1	0	0	0	0	0	0	0	1	2	0.62	0.46	0.06
œ̃	0	0	1	0	0	0	0	0	0	0	1	0.31	0.43	0.03
η	0	0	0	0	0	1	0	0	0	0	1	0.31	0.12	0.31

Nombre total de phonèmes = 321 χ² = 3.68

BIBLIOGRAPHIE

[1] IEEE Recommended Practice for Speech Quality Measurements, IEEE Transaction on Audio and Electroacoustics, Vol. AU-17, N° 3, September 1969.

[2] SZKLARCZYK, Lillian, 1961, Essai sur la structure phonologique du français. Thèse de doctorat inédite, University of Pennsylvania.

(Manuscrit reçu le 26 décembre 1980, revu le 28 janvier 1981)