

La loi de Menzerath appliquée à un ensemble de textes

Marc Hug

Professeur émérite, Université Marc Bloch, 23, rue Descartes, 67084 Strasbourg Cedex

ABSTRACT : The so-called "Menzerath law" is well known by German linguists dealing with quantitative linguistics, while it is almost ignored by French linguists. It says that in any linguistic construct, the more constituents a given unit is made of, the smaller they are. Here this "law" is confronted with observations made on a corpus of texts taken from the newspaper Le Monde. It appears that globally articles (texts) that have long sentences use longer words than articles having shorter sentences - but inside most of these articles, long sentences use shorter words than short sentences. Similar measures made with an intermediate level between "sentence" and "word", or made on phonetic transcriptions of the texts, do not change the conclusions. This shows that a research on the Menzerath law should always provide precise information on the exact nature of the units that are being counted and the way they have been obtained.

KEYWORDS : Menzerath law, sentence, clause, word, punctuation, linear regression, linear correlation.

RESUME : La loi dite de Menzerath est bien connue des linguistes allemands faisant de la linguistique quantitative, mais presque totalement ignorée de leurs collègues français. Elle dit que "plus une unité linguistique d'un niveau donné comporte de constituants du niveau inférieur, plus ces constituants sont petits". Cette "loi" est ici confrontée aux observations faites sur un ensemble de textes tirés du quotidien Le Monde. On constate que globalement les articles qui ont les phrases les plus longues font aussi usage des mots en moyenne plus longs que les articles dont les phrases sont plus courtes - mais qu'à l'inverse, à l'intérieur de la plupart de ces articles, les phrases les plus longues ont des mots en moyenne plus courts que les phrases courtes. Des mesures analogues faites en prenant en compte un niveau intermédiaire entre la phrase et le mot, ou en opérant sur les transcriptions phonétiques des textes, ne conduisent pas à des conclusions différentes. Il apparaît que si l'on travaille sur la loi de Menzerath, il est indispensable de dire très précisément quelles unités on a comptées et de quelle façon précise on a obtenu les données sur lesquelles on calcule.

MOTS CLES : Loi de Menzerath, phrase, proposition, syntagme, mot, ponctuation, régression linéaire, corrélation linéaire.

La loi de Menzerath

Die Architektonik des deutschen Wortschatzes

mot

syntagme proposition phrase

Störfaktoren

Teilsatz

prescrites

Je crois / que tu devrais acheter les lunettes / que l'oculiste t'a

lexique

Le corpus étudié

Le Monde

Définition de la phrase et du mot

Monsieur

etc.

M. ...

aujourd'hui *quelqu'un*

Moyennes

Il arrive.

du texte

Corrélations entre textes

r		u
R		u'
r	R	

r		u
R		u'

Corrélation interne aux textes

Monde

Prise en compte d'un niveau de "syntagmes"

r
r
r

u
u
u

bémol

or

Traitement de textes oraux

temps *fini*

r
R

r
R

Tableau 1.		
Texte	Mots/Phrase	Phon/Mot
T3	15,3529	3,3978
T7	19,9388	3,4626
T26	20,8182	3,5735
T29	15,9149	3,3035
T36	21,6111	3,5874
T48	36,5833	3,9408
T53	12,1695	3,3593
T73	32,0476	3,7712
T80	28,0000	3,6752
T98	22,5532	3,6981
Moy.	19,5022	3,5488

Tableau 2		
Texte	Mots/Phrase	Lettres/Mot
T3	15,3529	4,6558
T7	19,8571	4,7595
T26	19,8788	4,7515
T29	15,9362	4,4446
T36	19,8333	4,7759
T48	36,0000	5,1181
T53	12,1695	4,5515
T73	31,6667	5,1278
T80	27,2188	4,8347
T98	21,9574	4,9331
Moy.	19,12329	4,77185

Les droites de régression de G. Altmann

X

Y

$$y' = ax' + b$$

$$\bar{y}_i = \sum_j \frac{y_{ij}}{n_i x_i}$$

$$\sum_i n_i Z_i^2 - \frac{\left(\sum_i n_i Z_i\right)^2}{n}$$

$$\sum_i n_i Z_i Y_i - \frac{\sum_i n_i Z_i \sum_i n_i Y_i}{n}$$

$$\sum_i n_i Y_i^2 - \frac{\left(\sum_i n_i Y_i\right)^2}{n}$$

$$\bar{Y} = \frac{1}{n} \sum_i n_i Y_i \quad \bar{Z} = \frac{1}{n} \sum_i n_i Z_i$$

$$b = \frac{M_{ZY}}{M_{ZZ}} \quad \bar{Y} - b\bar{Z} \quad a \quad e$$

b

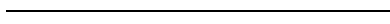
y

b

b

b

b



y

b

b

"Loi" de Menzerath et "loi" d'Arens

Conclusions provisoires

b

non

eh', Brunst, stracks, Splint

an, bei, zu, um, von, und

der, die, das, den, dem

orale

Sprechtakte

Sprechakte

Références bibliographiques

*Das Menzerathsche Gesetz in
informationsverarbeitenden Systemen*

Allgemeine Sprachwissenschaft, Sprachtypologie und Textlinguistik

Verborgene Ordnung

Die Architektonik des deutschen Wortschatzes

Initiation aux méthodes de la statistique linguistique

*Trierer Kolloquium zur Quantitativen Linguistik
of Quantitative Linguistics*

*4.
Journal*

Grundzüge der Phonetik

*Glottometrika
Linguistics*

Quantitative