Glottometrics 24, 2012, 88-94

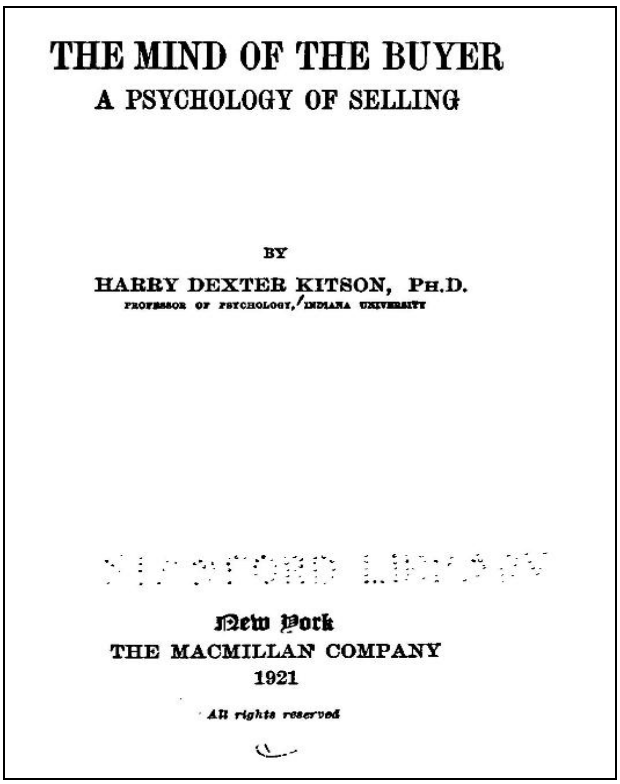
History of Quantitative Linguistics

Since a historiography of quantitative linguistics does not exist as yet, we shall present in this column short statements on researchers, ideas and findings of the past – usually forgotten – in order to establish a tradition and to complete our knowledge of history. Contributions are welcome and should be sent to Peter Grzybek, peter.grzybek@uni-graz.at.

Harry Dexter Kitson (1886-1959)

Peter Grzybek

Harry Dexter Kitson, born in 1886 in Mishawaka, Indiana, taught applied psychology at Teachers' College, Columbia University. He was a charter member of the American Psychological Association and a pioneer in the field of vocational guidance. His main field of professional interest throughout his life (see references below), and it would definitely be incorrect to rank him among the precursors of quantitative linguistics. Yet, some ideas and analyses represented in his 1921 book *The Mind of the Buyer. A Psychology of Selling* illustrate the need of his time for a solid basis in linguostatistics and quantitative linguistics, and therefore deserve mention in an historical flashback.



THE MIND OF THE BUYER
A PSYCHOLOGY OF SELLING

BY
HARRY DEXTER KITSON, PH.D.
PROFESSOR OF PSYCHOLOGY, INDIANA UNIVERSITY

New York
THE MACMILLAN COMPANY
1921

All rights reserved

Kitson's booklet was meant to be a guide in advertisement strategies for salesmen, in his words "every one who is engaged in influencing men to buy" (p. v). For Kitson, such a work must necessarily be based on theoretical psychology and deal with profound psychological questions, particularly mental processes such as attention, interest, desire and confidence (ibd., v).

With this orientation and phrasing, Kitson's booklet was a typical child of its time. After all, the booklet was published in the very same year when the famous *AIDA* formula was first used as an acronym by C.P. Russell (1921) to refer to the relevant components (or steps, as which they were considered at that time) of successful advertising: "attract **A**ttention, maintain **I**nterest, create **D**esire and get **A**ction". Such components were usually traced back to psychological theories of that time (usually related to some kind of association psychology).

Yet, although Russell is considered to have casted this concept into a concise verbal form – i.e., the *AIDA* formula –, he is not responsible for having developed the general idea and concept behind it. It is commonly held that it is American advertising and sales pioneer Elias St. Elmo Lewis who should be credited for having established the term and approach in 1903: postulating at least three principles to which a successful advertisement should conform, for Lewis, the "mission of an advertisement" was "to attract a reader […]; then to interest him, […]; then to convince him […]." The first published instance of the general concept seems to be a 1904 article by Frank Hutchinson Dukesmith, according to whom the four most important steps were attention, interest, desire, and conviction. Later important references are Ralph Starr Butler's (1911) *Advertising, Selling, and Credits. Part II: Selling and Buying*, with a whole chapter on "Principles of Salesmanship" (p. 410ff.) focusing on attention, interest, desire, action. Butler, in turn, refers to Arthur F. Sheldon, founder of the Sheldon School of Scientific Salesmanship, and his 1911 book *The art of selling, for business colleges*, containing similar ideas.

In this respect, Kitson's approach is not genuinely innovative. What makes him differ from preceding approaches, however, is his definition and treatment of what he termed the "collective buyer". According to him, persons who are served by a given selling medium constitute a collectivity. For Kitson, such a public is not a simple arithmetic summation of individual minds, nor is a some kind of super-mind transcending its components (ibd., 54). For him, newspapers and magazines offer good evidence of the existence of the collective mind: "psychologically speaking, the readers of a sales medium constitute an entity, a public, which is not a loose aggregation of isolated and individual minds but an organic union, coalesced into one collective mind" (ibd., 55). Furthermore, each public is unique, and readers of different newspapers differ from each other, what does of course not exclude the possibility that a given individual may belong to more than one public.

In trying to develop measurements of such publics which, in Kitson's terms, are "buying publics" (ibd., 56), Kitson suggests to study a number of relationships, mainly geographical, economic, sociological, and psychological. In his effort to establish some kind of "yard sticks" for the psychological, or mental, dimension (including ideas, feelings, motives), Kitson suggested, among others, the analysis of linguistic criteria of different journals and newspapers. Admitting that the kind of measures he suggested are still very fragmentary (ibd., 63), he suggested to concentrate on word length and sentence length, which he considered to be indicators of psychological differences between periodicals.

With regard to word length, Kitson first chose the *Chicago Evening Post* and the *Chicago American* for his analyses. From the editorial, news and feature columns of six parallel issues of each of these two papers, ca. 5000 words were taken in consecutive order and tabulated according to the number of syllables they contained. Likewise, two magazines were analyzed, the *Century* and the *American* magazines. Unfortunately, Kitson did not give

the complete results, concentrating on words with more than two syllables only. The data are represented in Table 1.

Table 1

	Word Length			
	> 2	> 3	> 4	> 5
Chicago Evening Post	13,20	4,60	1,20	0,00
Chicago American	7,70	2,70	0,70	0,00
Century	13,50	4,30	1,00	0,20
American	9,90	2,70	0,60	0,10

In Kitson's interpretation, the results show that the number of words with more than two syllables in the *Post* is greater than that in the *American* by ca. 71%, a ratio approximately holding for all the polysyllabic words. The results of the magazine analyses are quite similar to the ones from the newspaper analysis; here, the number of words with more than two syllables in the *Century* is greater than the corresponding number in the *American* by ca. 36%. Kitson therefore concludes that the two journals and the two magazines clearly differ in their profiles.

A re-analysis of his data shows that his interpretation, albeit correct, is not unproblematic: first and foremost, because the dominating amount of one- and two-syllable words have been totally omitted from the analyses – after all, their percentage is > 90% in all cases, ranging from 90.8% to 95% across the four samples. But even concentrating on the word length frequencies of words with more than two syllables shows that Kitson's conclusions are far from being self-evident. Table 2 offers Kitson's data in re-ordered form, presenting them in non-cumulative form.

Table 2

	Word Length			
	3	4	5	> 5
Chicago Evening Post	8,60	3,40	1,20	0,00
Chicago American	5,00	2,00	0,70	0,00
Century	9,20	3,30	0,80	0,20
American	7,20	2,10	0,50	0,10

Comparing word length of both the two journals (*Chicago Evening Post*, *Chicago America*n) and the two magazines (*Century*, *American*) by way of the non-parametric Mann-Whitney *U*-test yields non-significant differences in both cases ($p = 0.73$ and $p \approx 1$, respectively), after weighting word length by the percentages given; the same holds true for the two journals' data and the two magazines's data taken together in combination ($p = 0.84$). Also a Kruskal-Wallis test for differences between all four groups shows the differences to be non-significant ($p = 0.98$); quite logically, post-hoc comparisons of averages yield no homogeneous subgroups. It seems reasonable, therefore, to conclude that, in contrast to Kitson's observations, there are no significant differences across the four journalistic samples with regard to word length.

In case of his sentence length analyses, based on parallel issues and columns of the same four journals and magazines, Kitson provides a better data basis which, as a con-

sequence, allows for more reliable re-analyses. A total of 8000 sentences were analyzed for sentence length, measured in the number of words per sentence. Here, all data are presented, pooled in intervals per 10, in a similar fashion as the word length data in Table 1; the data are accordingly reproduced in Table 3.

Table 3

	Sentence Length (in intervals per 10)										
	1-10	>10	>20	>30	>40	>50	>60	>70	>80	>90	>100
Chicago Evening Post	16,9	83,1	49	22,3	8,5	2,7	0,8	0,2			
Chicago American	23,1	76,9	43,4	20,6	10,3	2,3	1,8	0,6	0,3	0,2	0,2
Century	22,8	77,2	45,4	24,4	10,6	5,5	2,4	0,9	0,4	0,2	
American	30,5	69,5	33,5	14,5	5,2	1,8	0,7	0,3	0,1	0,1	0,1

Comparing sentence length for the two journals, the *Chicago Evening Post* and the *Chicago American*, Kitson states that the results show a greater number of "long" sentences (considering sentences with > 20 words as long) in the *Post*; he likewise finds the *Century* to favor long sentences as compared to the *American*.

Attempting to re-analyze the data, it seems reasonable to re-order them without cumulation, in analogy to the word length data presented in Table 2. The corresponding sentence length data are presented in Table 4.

Table 4

	Sentence Length (in intervals per 10)										
	1-10	>10	>20	>30	>40	>50	>60	>70	>80	>90	>100
Chicago Evening Post	16,9	34,1	26,7	13,8	5,8	1,9	0,6	0,2			
Chicago American	23,1	33,5	22,8	10,3	8	0,5	1,2	0,3	0,1		0,2
Century	22,8	31,8	21	13,8	5,1	3,1	1,5	0,5	0,2	0,2	
American	30,5	36	19	9,3	3,4	1,1	0,4	0,2			0,1

Again, a re-analysis is not unproblematic, since not the raw data are given, but the pooled data in intervals per 10. Nevertheless, after weighting the sentence length categories with the percentages given allows a test for differences, in analogy to the above word length analyses, first between the two journals and the two magazines, then between all four samples. Whereas a Mann-Whitney *U*-test yields no significant differences between the two journals ($p = 0.31$), it shows the differences between the two magazines to be significant ($p = 0.03$). As to a comparison between all four groups, a Kruskal-Wallis test shows the differences to be significant ($p = 0.03$), but a post-hoc comparison of means yields no homogeneous subgroups.

This seemingly contradictory result might well be due to the fact that all four samples follow a common frequency distribution model, though with different weights for the individual length classes, what should result in different parameter values for the given model. In order to test this assumption, it would be necessary to have the original data at hand, what is not the case. Nevertheless, by way of some approximation, one might try to reconstruct original sample sizes given the fact that on the whole, 8000 sentences were

analyzed, based on four approximately equal sample sizes. The results of reconstructing the corresponding absolute frequencies are represented in Table 5.

Table 5

	Sentence Length (in intervals per 10)										
	1	2	3	4	5	6	7	8	9	10	11
	1-10	>10	>20	>30	>40	>50	>60	>70	>80	>90	>100
Chicago Evening Post	338	682	534	276	116	38	12	4	0	0	0
Chicago American	462	670	456	206	160	10	24	6	2	0	4
Century	456	636	420	276	102	62	30	10	4	4	0
American	610	720	380	186	68	22	8	4	0	0	2

In trying to find a theoretical frequency distribution as an adequate model for these data, it turns out that the negative binomial distribution defined as

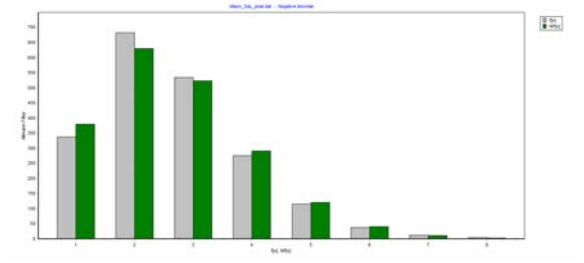
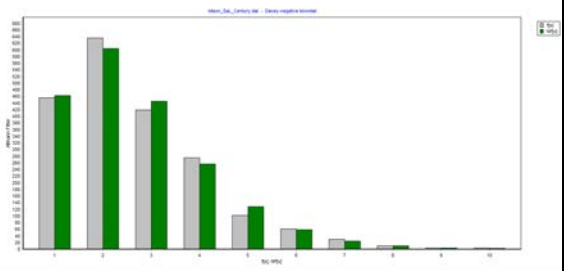
$$(1) \qquad P_x = \binom{k+x-1}{x} p^k q^x$$

is an excellent model for the three of the data sets (*Chicago Evening Post*, *Century*, *American*), whereas the *Chicago American* data can be fitted by the mixed negative binomial distribution defined as

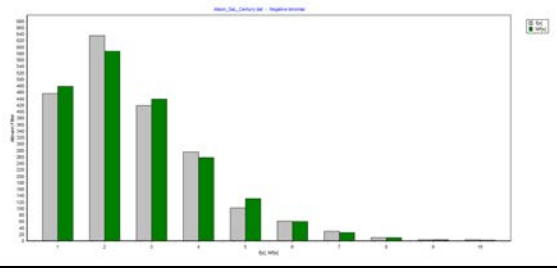
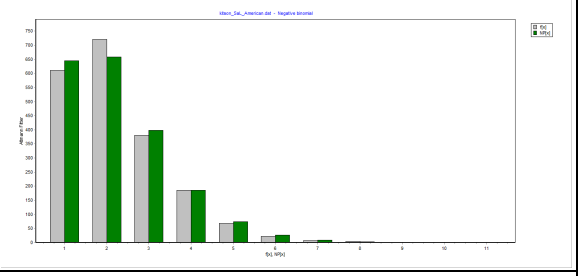
$$(2) \qquad P_x = \alpha \binom{k_1+x-1}{x} p_1^{k_1} q_1^{x} + (1-\alpha) \binom{k_2+x-1}{x} p_2^{k_2} q_2^{x},$$

both of course in one-displaced form. Taking into account that we are concerned with mixed data, in all cases, the need for a mixed model seems to be reasonable – quite evidently, with $\alpha = 0$ or $\alpha = 1$, the mixed model (2) has the ordinary model (1) as a special case..

Figures 1-4 show the fitting results, with parameter values for k and p given below the graphs, as well as the discrepancy coefficients $C = X^2/N$, with $C < 0.02$ indicating a good, $C < 0.01$ a very good fitting result.[1]

Chicago Evening Post	*Chicago American*
	
$k = 826.59$, $p = 0.998$, $C = 0.0051$	$k = 3.36$, $p = 0.70$, $\alpha = 0.23$, $C = 0.0060$

[1] In case of the *Chicago Evening Post* data, with parameter $k \to \infty$ and $1\text{-}p = q \to 0$, the negative binomial distribution converges against the Poisson distribution, yielding an equally good fit with $a = 1.66$ and $C = 0.0051$. – The Century data can also be modeled by the Mixed Poisson distribution: with $a = 3.31$, $b = 1.30$, and $\alpha = 0.19$ the result is almost identical, with $C = 0.0054$.

Century	*American*
	
k = 4.59, p = 0.73, C = 0.0077	k = 5.41, p = 0.81, C = 0.0056

Figures 1-4. Fitting the distribution of sentence length by the negative binomial distribution

The sentence length data thus indeed follow one and the same model, albeit with some "local" modifications.

As a result, one can say that Kitson has indeed raised interesting and important questions which, in one way or another, would today be treated between the fields of applied and quantitative linguistics. Whereas earlier word length and sentence length studies had mainly treated them in terms of individual author characteristics, on the basis of literary texts – in order to determine authorship, for example, or literary development –, Kitson, not referring to the works of Sherman, Mendenhall and others, extended the field of interest to "everyday" journalistic texts, asking for recipient specific and, in this sense, pragmatic differences. Almost simultaneously with and subsequent to his work, the influential discipline of text difficulty and readability research would become increasingly important: after the first readability formula suggested by Lively and Pressey in 1923, this line of research faced a first highlight in Rudolf Flesch's works (e.g., Flesch 1948), very much later leading to, among others, systematic analyses of different journalistic sources (e.g., Amstad 1978). And although Kitson did not create a readability formula, he is considered to have shown how the principles work (cf. DuBay 2004: 13), since in almost all readability studies, word and sentence length have always played a crucial role, though not as separate, but specifically related factors, more often combined with further linguistic levels and units.

In this context, it may be important to emphasize that Kitson explicitly stated that from these findings he would not reason that superiority in long words or sentences proves conclusively a corresponding intellectual superiority. Admitting that "long words and long sentences are not an absolute criterion of erudition or short of ignorance" (ibd., 63), he nevertheless admits "that in the long run, the chances favor a greater number of long words being associated with more enlightened people" (ibd., 63). Interestingly enough, Kitson (ibd., 63) refers to a relation between vocabulary richness (in terms of the size of an individual's subjective lexicon) on the one hand, and word length on the other: "Measurements made by various vocabulary tests have shown that there are more words in the vocabulary of the more enlightened; hence we might expect a greater number of long words there."

References

Amstad, Toni (1978). *Wie verständlich sind unsere Zeitungen?* Diss., Univ. Zürich.

Butler, Ralph Starr (1911). *Advertising, Selling, and Credits. Part II: Selling and Buying.* New York: Alexander Hamilton Institute.

DuBay, William H. (2004). *The Principles of Readability*. Costa Mesa, CA: Impact Information. [http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.91.4042]

Dukesmith, Frank Hutchinson (1904). Three Natural Fields of Salesmanship. *Salesmanship 2(1), January 1904, p. 14.*

Flesch, Rudolf (1948). A New Readability Yardstick. In: *Journal of Applied Psychology* 32(3), 1948; 221-233.

Kitson, Harry Dexter (1916). *The scientific Study of the College student***.** Phil. Diss, Chicago. [1st reprint: Princeton, NJ, 1917]

Kitson, Harry Dexter (1916). *How to use your mind***.** *A psychology of study; being a manual for the use of students and teachers in the administration of supervised study***.** Philadelphia**:** Lippincott Co.

Kitson, Harry Dexter (1920). *Manual for the study of the psychology of advertising and selling***.** Philadelphia, London: J.B. Lippincott Co.

Kitson, Harry Dexter (1921). *The Mind of the Buyer. A Psychology of Selling*. New York: Macmillan.
[http://archive.org/stream/mindbuyerapsych00kitsgoog#page/n76/mode/2up]

Kitson, Harry Dexter (1929). *How to find the right vocation.* **N**ew York: Harper & Brothers Publishers**.**

Kitson, Harry Dexter (1933). *Finding a job during the depression*. New York City: The Robert C. Cook Co.

Kitson, Harry Dexter (1947). How to find the right vocation**.** New York: Harper & Brothers Publishers.

Kitson, Harry Dexter; Newton, Juna Barnes (1950). *Helping people find jobs: how to operate a placement office.* New York: Harper & Brothers Publishers.

Kitson, Harry Dexter (1954). *I find my vocation.* New York: McGraw-Hill.

Lewis, Elias St. Elmo (1903). Catch-Line and Argument. In: *The Book-Keeper, Vol. 15, February*; 124.

Lively, Bertha A.; Pressey, Sidney L. (1923). A method for measuring the 'vocabulary burden' of textbooks. *Educational administration and supervision*, 9; 389–398.

Russell, C.P. (1921). How to Write a Sales-Making Letter. *Printers' Ink, June 2.*

Sheldon, Arthur Frederick (1911). *The art of selling, for business colleges.* Published/ Created: Libertyville, Ill., The Sheldon University Press.