

FRANCO MELETTI S. J.

PROPORTION BETWEEN NUMBER OF WORDS
AND NUMBER OF FORMS IN THE WORKS
OF THE "INDEX THOMISTICUS"

1. THE PROBLEM

At the conclusion of his long work of elaboration of the *Index Thomisticus* (the concordance of all the works of St. Thomas Aquinas), Father ROBERTO BUSA s.j. proposed the following problem to me: For each of the 179 elaborated works, we know the exact number of words and the exact number of graphic forms. Obviously, from this data it also results that the ratio between the number of forms and the number of words decreases with an increasing number of words. Does a formula exist that will compare two works of different length so that one can affirm that one work has comparatively more forms than another? For instance, the work with the code number 017-QDI contains 10,238 words and 1,471 forms. Does this work have comparatively more or less forms than work 012-QDV which has 114,991 words and 9,016 forms?

Father Busa asked me for the pure quantitative data. He also told me that he would not interpret or evaluate this data in terms of "lexical wealth" as he doubts that anything can be concluded in this regard on the basis of this data alone. Indeed, he believes that lexical wealth must be measured on the basis of a much wider range of parameters.

2. THE ANSWER

Following the process reported below, a relation between the number of words (W) and the number of graphic forms (F) has been found: it results that this relationship can be expressed with a straight line on a bi-logarithmic diagram:

$$(1) \quad \log F = \log A + B \cdot \log(W/10,000)$$

where $\log A$ and B are two constants which characterize the linear function. Therefore the function $F = f(W)$ is expressed by a generalized parabola:

$$(2) \quad F = A \cdot (W/10,000)^B$$

The values of A and B are the following for the set of the works of the entire *Index Thomisticus Corpus* and for the various groups of works according to the subdivisions indicated below:

TABLE 1.

		A	B
<i>Index Thomisticus Corpus</i>	(n. 179)	2197	0.685
<i>Thomistic Works</i>	(n. 118)	2163	0.660
<i>Parathomistic Works</i>	(n. 61)	2301	0.777

TABLE 2. *Thomistic Works*

		A	B
<i>Proper Works</i>	(n. 49)	2114	0.628
<i>Commentaries</i>	(n. 27)	2015	0.702
<i>Stenographic Reports</i>	(n. 24)	2604	0.655
<i>Works of Dubious Authent.</i>	(n. 18)	1901	0.663

3. GRAPHIC AND CALCULUS PROCEDURE

To establish first of all whether a relationship exists between the values of W and F in the works of the *Index Thomisticus Corpus*, the points corresponding to couples W, F are traced on a bi-logarithmic diagram, that is with scales $\log W$ and $\log F$. The choice of the logarithmic scales was suggested directly by the distribution of the values of W and F arranged in increasing order. These values on a logarithmic scale would result well distributed along the entire W and F interval.

From the diagram $\log W - \log F$ it immediately appeared clear that a straight line could well approximate the distribution of the points all over the W interval, from minimal to maximal values (see figure 1).

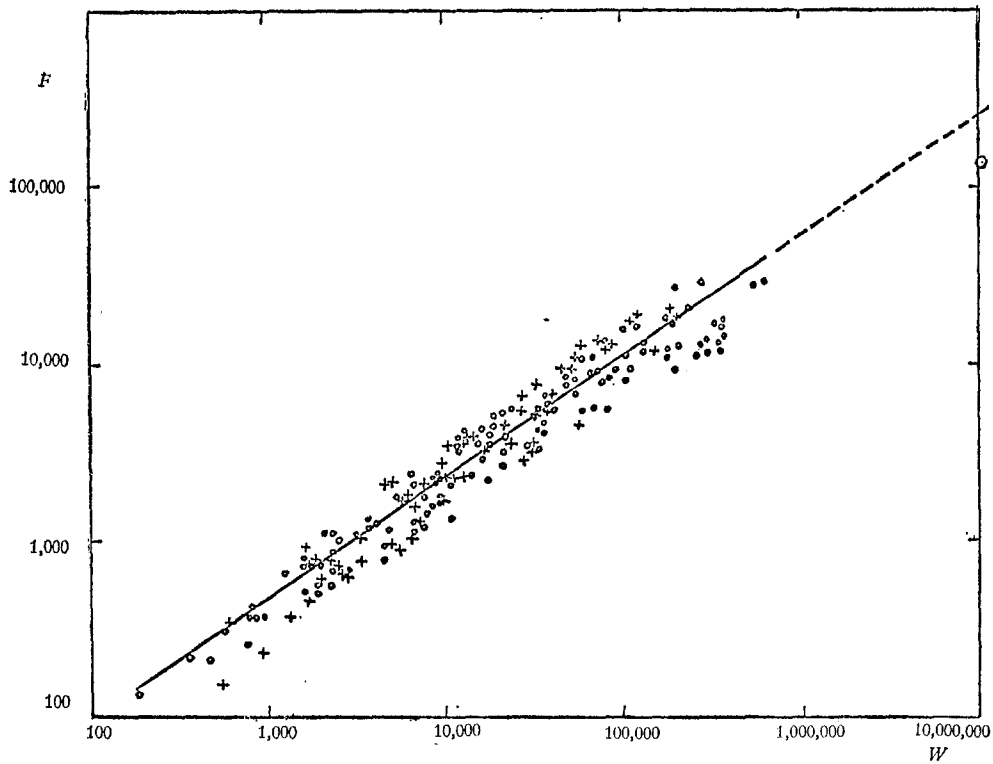


Fig. 1. *Words-Forms Diagram*
 ○ *Thomistic Works* (n. 118)
 + *Parathomistic Works* (n. 61)
 ○ *Whole Thomistic Corpus*

Therefore the relationship between $\log W$ and $\log F$ is linear as expressed in (1) and the function $F = f(W)$ is as expressed in (2).

The expression (2) has been preferred to the expression $F = A \cdot W^B$, because in (2) the constant A has a significant value, that is the number of forms corresponding to 10,000 words, mean value in the interval of $\log W$. It is evident that the expressions (1) and (2) are valid only within the intervals of the respective works. For lower values of W the function $F = f(W)$ obviously links with the straight line $F = W$, which is valid for very short texts.

The constants $\log A$ and B of (1) were computed by least-squares polynomial approximation on the logarithmic diagram.

The complete results of the computations are reported in Tables 3

and 4, for the entire *Index Thomisticus Corpus*; for two major divisions of the Corpus: *Thomistic Works* and *Parathomistic Works* (works of other authors included in the Corpus); for four divisions of Thomistic works: *Proper Works* (OPERA PROPRIA), *Commentaries* (COMMENTARIA: Aristotle's, Bible's, etc.), *Stenographic reports* (REPORTATIONES), works of *dubious authenticity* (OPERA DUBIAE AUTHENTICITATIS).

In Tables 3 and 4 the following values are reported:

- the constants $\log A$, A , B ,
- the $\log F$ RMS deviation, that is, the RMS deviation between the true values of $\log F$ and the mean values of $\log F$ computed by (1),
- the *max. upper deviation*, that is the ratio $(F_{true}/F_{mean})_{max}$, for values $F_{true} > F_{mean}$,
- the *max. lower deviation*, that is the ratio $(F_{true}/F_{mean})_{min}$, for values $F_{true} < F_{mean}$.

TABLE 3.

	$\log A$	A	B	$\log F$ RMS deviation	<i>max. upper deviation</i>	<i>max. lower deviation</i>
<i>Index Thomisticus Corpus</i>	3.3417	2197	0.685	0.126 = $\log 1.34$	1.92	1/2.01
<i>Thomistic Works</i>	3.3351	2163	0.660	0.116 = $\log 1.30$	1.83	1/1.73
<i>Parathomistic Works</i>	3.3620	2301	0.777	0.126 = $\log 1.34$	1.75	1/1.94

TABLE 4. *Thomistic Works*

	$\log A$	A	B	$\log F$ RMS deviation	<i>max. upper deviation</i>	<i>max. lower deviation</i>
<i>Proper Works</i>	3.3250	2114	0.628	0.093 = $\log 1.24$	1.60	1/1.64
<i>Commentaries</i>	3.3042	2015	0.702	0.154 = $\log 1.43$	1.73	1/1.83
<i>Stenographic Reports</i>	3.4156	2604	0.655	0.045 = $\log 1.11$	1.13	1/1.45
<i>Works of Dubious Authenticity</i>	3.2790	1901	0.663	0.121 = $\log 1.32$	1.80	1/1.34

4. NOTES, RESERVATIONS AND COMMENTS

The values of the constants A and B with reference to four divisions of 118 Thomistic works (Tables 2 and 4) present not significant differences, which are included in the random variations. On the other hand, the constants A and B of the Thomistic works compared with those of the Parathomistic works (Table 1 and 3) present significant differences, using the Student-Fisher's test (see Figure 2).

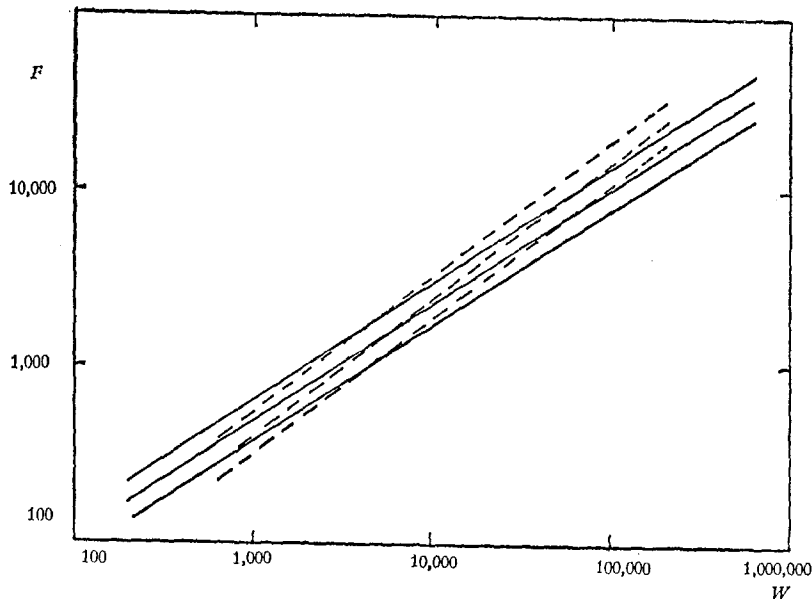


Fig. 2. *Words-Forms Mean Diagram with limits of RMS error.*
 ———— Thomistic Works
 - - - - - Parathomistic Works

In Table 4 one can notice that among the Thomistic works the Stenographic reports (REPORTATIONES), which are distinct from the truly dictated works (these are among the proper works), present a RMS deviation of F which is about half of that of the other groups of Thomistic works; is this a significant fact?

The straight-line behaviour of the function $F = f(W)$ on the logarithmic diagram is seen along the *whole* interval of W , from the

minimum of about $W = 200$ up to a maximum of about 600,000 words, without showing a tendency to flatten with regard to the W axis. However the sum of the words of the entire *Thomistic Corpus* ($W = 10,600,085$) presents a number of graphic forms $F = 132,125$ which is lower than the mean (extrapolated) value computed by the expression (1); but this point does not much exceed the double of the RMS deviation (see figure 1).

5. LIST OF THE WORKS OF THE CORPUS OF THE INDEX THOMISTICUS

Serial number, abbreviation, title, number of words, number of graphic forms are reported for every work.

A)	OPERA THOMISTICA	(<i>Thomistic works</i>)	W : from 183 to 597821	
	1 - OPERA PROPRIA (<i>Proper works</i>)		W : from 351 to 597821	
001	1SN	In I Sententiarum	268965	12266
002	2SN	In II Sententiarum	296309	16610
003	3SN	In III Sententiarum	334431	17504
004	4SN	In IV Sententiarum	597821	28109
005	SCG	Summa contra Gentiles	325820	18406
006	ADL	Autographi deleta	70763	7543
007	ST1	Summae Theologiae Prima Pars	364483	16906
008	ST2	Summae Theologiae Prima Secundae	358954	19603
009	ST3	Summae Theologiae Secunda Secundae	513890	28013
010	ST4	Summae Theologiae Tertia Pars	336106	21557
011	QDV	Qu. Disp. de Veritate I-XXII A. 11 Co.	287123	13864
012	QDV	Qu. Disp. de Veritate XXII A. 11 RA - Finis	114991	9016
013	QDP	Qu. Disp. de Potentia	184562	10860
014	QDM	Qu. Disp. de Malo	180464	12319
015	QDS	Qu. Disp. de Spirit. Creaturis	35279	4002
016	QDA	Qu. Disp. de Anima	59920	5190
017	QDI	Qu. Disp. de Unione Verbi	10238	1471
018	QDW	Qu. Disp. de Virtutibus	80073	7475
019	QDL	Quodlibeta I-XI	134153	13209
020	OCE	Contra Errores Graecorum	16602	2947
021	OCG	De Rationibus Fidei	9446	2558
022	OSF	De Forma Absolutionis	3698	1227
023	OTS	De Substantiis Separatis	21615	3701
024	O2D	Super Decretalem	7420	1965
025	OCI	Contra Impugnantes etc.	66708	10375
026	OAP	De Perfectione Spiritualis Vitae	23626	5224
027	OCR	Contra Doctrinam Retrahentium etc.	19402	5136
028	OCA	De Unitate Intellectus	14452	2407
029	OEE	De Ente et Essentia	7318	1255
030	OPN	De Principiis Naturae	4335	761
031	OPC	De Motu Cordis	1729	586
032	OPX	De Mixtione Elementorum	961	376

033 OPO	De Occultis Operibus Naturae	1917	702
034 OCM	De Aeternitate Mundi	2188	692
035 OTR	De Sortibus	5203	1863
036 OTE	De Judiciis Astrorum	351	231
037 ORI	De Regimine Judaeorum	1660	788
038 ORP	De Regimine Principum	13504	4069
039 OTT	Compendium Theologiae	69896	8371
040 OSS	De Articulis Fidei etc.	6478	2105
041 OPW	De Emptione et Venditione	546	309
042 OS1	Responsio ad Lect. Venetum de art. 30	2147	813
043 OS2	Responsio ad Lect. Venetum de art. 36	3877	1254
044 OS3	Responsio ad Lect. Vercell. de art. 42	3719	1339
045 OS4	Responsio ad Lect. Vercell. de art. 108	8077	1575
046 OS5	Responsio ad Lect. Bisuntinum	773	422
047 OS6	Responsio ad Bernardum	754	388
048 OTC	Principium Biblicum	2496	1015
049 OTD	Breve Principium	1271	656

 2 - COMMENTARIA (*Commentaries*)

W: from 6596 to 269016

050 CCM	In Libros de Caelo et Mundo	108691	7622
051 CGC	In Libros de Gener. et Corrupt.	21979	2677
052 CPY	In Libros Physicorum	197747	8955
053 CTC	Sententia Libri Ethicorum	199955	12836
054 CTE	Tabula Libri Ethicorum	31905	4951
055 CPO	Sententia Libri Politicorum	56192	6751
056 CPA	In Libros Posteriorum Analyticorum	86684	5538
057 CME	In Libros Metereologicorum	34673	4562
058 CPE	In Libros Perihermeneias	31750	3276
059 CAN	In Libros de Anima II et III	66800	5620
060 CSS	In Libros de Sensu et Sensato	33118	4291
061 CMR	In Libros de Memoria et Reminiscentia	10629	2044
062 CMP	In Libros Metaphysicorum	260960	11624
063 CDC	In Librum de Causis	29665	3435
064 CBT	In Librum Boethii de Trinitate	41502	5483
065 CBH	In Librum Boethii de Hebdomadibus	6596	1336
066 CDN	In Dionysii de Divinis Nominibus	96689	9245
067 CIO	In Job	115239	16212
068 CIS	In Isaiam	100331	15232
069 CPH	In Hieremiam	58473	10921
070 CTH	In Threnos Hieremiae	12147	3782
071 CMT	Catena Aurea in Matthaicum	269016	29390
072 CMC	Catena Aurea in Marcum	78115	14510
073 CLC	Catena Aurea in Lucam	191638	27765
074 CJO	Catena Aurea in Johannem	175419	19643
075 CRO	Super Ep. ad Romanos	131357	13499
076 C1C	Super I ad Cor. I-VII L. 2	38215	6383

 3 - REPORTATIONES (*Stenographic Reports*)

W: from 1668 to 320019

077 RQD	Quodlibetum XII	8300	2194
078 RAN	In Aristotelis de Anima L. I	23798	3164
079 RSV	In Symbolum Apostolorum	15582	3674

080	RSU	In Orationem Dominicam	8254	2304
081	RST	In Salutationem Angelicam	1668	709
082	RAC	De Duobus Praeceptis Charitatis	16437	4061
083	RSR	Sermones	35371	6733
084	RPS	In Psalmos	193568	18876
085	REM	Super Evangelium Matthaei	233176	20597
086	REI	Super Evangelium Johannis	320019	22144
087	RIL	Reportationes Ineditae Leoninae	34363	5818
088	R1C	Super I ad Cor. XI-XVI	54098	7896
089	R2C	Super II ad Cor.	65409	8542
090	RGL	Super ad Galatas	48078	7173
091	REP	Super ad Eph.	48331	7350
092	RPL	Super ad Philip.	18068	3941
093	RCL	Super ad Coloss.	19190	4170
094	R1T	Super ad Thess. I	12341	3148
095	R2T	Super ad Thess. II	8456	2346
096	RT1	Super ad Thim. I	27755	5589
097	RT2	Super ad Thim. II	16816	3981
098	RTT	Super ad Titum	12484	3306
099	RPM	Super ad Philemonem	3082	1182
100	RHE	Super ad Hebraeos	100965	11787

4 - DUBIAE AUTHENTICITATIS (*Works of dubious authenticity*)

W: from 183 to 17509

101	DTX	De Secreto	795	344
102	DTI	De Natura Verbi Intellectus	2912	689
103	DP5	De Instantibus	4622	992
104	DPM	De Nat. Materiae et Dimens. Int.	9030	1536
105	DP9	De Principio Individuationis	1511	511
106	DPA	De Natura Accidentis	1925	553
107	DPG	De Natura Generis	17509	2310
108	DP4	De Quatuor Oppositis	6842	1206
109	DP3	De Fallaciis	9820	1621
110	DPP	De Propositionibus Modalibus	753	255
111	DAS	Epistola de Modo Studendi	183	142
112	DSG	Officium de Festo Corporis Christi	2144	1123
113	DSR	Sermones	6249	2503
114	DSP	Piae Preces	2050	1139
115	DP2	De Concordantiis Suiipsius	7290	1793
116	DOA	Utrum in Creaturis sit Ordo Agendi	457	217
117	DIA	Qu. de Immortalitate Animae	4883	1199
118	DLV	Utrum Al. de Libro Vitae Deleri possit	2302	593

B) OPERA PARATHOMISTICA (*Works of other authors included in the I. T. Corpus*)

W: from 555 to 191110

Author: Haimo Altissiodorensis O..S.B. † 850 c.				
119	XCC	In Canticum Exp. I - Salomon Inspiratus	26958	6468
Author: Richardus de Wedinghausen O. Praem. † 1180 c.				
120	XEM	Expositio Missae	3087	1073
Author: Adamus de Buckfield † 1285				
121	XSV	In de Somno et Vigilia	6955	1385

122 XSM	In de Somniis	4437	999
123 XDS	In de Divinatione per Somnum	2396	718
	Author: Aegidius de Lessinia O.P. † 1278 c.		
124 XTU	De Usuris in Communi	21951	3713
	Author: Aegidius de Roma O.S.A. † 1316		
125 XC2	In Canticum Exp. II - In Principio	31238	5601
	Author: Albertus Magnus O.P. † 1280		
126 XDP	De Pulchro et Bono	9154	1604
127 XPZ	De Fato	5759	1832
	Author: Aldobrandinus de Cavalcantibus O.P. † 1279		
128 XAW	Sermones pro Dominicis etc.	86141	12653
	Author: Aldobrandinus de Toscanella O.P. † 1292		
129 XSU	Insertum in Opusc. in Pater	565	325
	Author: Guillelmus Peraldus O.P. † 1271 c.		
130 XRE	De Eruditione Principum	75552	14301
	Author: Guillelmus Wheatley † post 1317		
131 XBC	In Boethii de Consol. Philos.	116573	17726
132 XBD	In Boethii de Scholarium Disc.	56040	13710
	Author: Helvicus Theutonicus O.P. † 1300		
133 XAO	De Dilectione Dei et Proximi	44831	8935
	Author: Hugo de S. Caro O.P. † 1263		
134 X1A	Super Apocalypsim I - Vidit Jacob	182105	21074
	Author: Jacobus de Benevento O.P. † 1260/70		
135 XTA	De Adventu etc. Antichristi	12527	3417
136 XTP	De Praeambulis ad Judicium etc.	16420	4420
	Author: Johannis Michaelis O.F.M. † 1320 c.		
137 XDA	In Daniele	55605	10720
138 XMB	In Libros Machabaeorum	31242	7065
	Author: Leonardus Pistoriensis O.P. † 1280 c.		
139 XTZ	De Praescientia et Praedestinatione	9204	2225
	Author: Nicolaus de Cotrone † 1260 c.		
140 XFT	Liber de Fide Trinitatis	27730	5013
	Author: Nicolaus de Gorran O.P. † 1295		
141 XEC	In 7 Epistolas Canonicas	89395	13333
	Author: Petrus de Alvernia † 1304		
142 XCM	In de Caelo et Mundo Continuatio	29267	3091
143 XPO	In Politicorum Continuatio	150392	11332
	Author: Petrus de Scala O.P. † 1295		
144 XLM	Inserta in Lectura super Matthaicum	14066	3953
	Author: Petrus Johannis Olivi O.F.M. † 1298		
145 XGN	Postilla in Librum Geneseos	128398	19714
	Author: Petrus de Tarantasia † 1295 (Nic. de Gorran)		
146 X1C	Postilla in I ad Corinthios	23711	4865
	Author: Ptolomaeus de Lucca O.P. † 1237		
147 XRP	De Regimine Principum Continuatio	44545	9268
	Author: Thomas de Sutton O.P. † 1315 c.		
148 XPF	De Pluralitate Formarum	15078	2453
	Author: Matthaeus de Cracovia † 1410		
149 XSC	De Modo Confitendi et Pur. Consc.	11922	3658
	Author: Vincentius de Castronovo O.P. † 1506		
150 XQP	Insertum in Qu. Disp. de Potentia	11962	2089
151 XQW	Insertum in Qu. Disp. de Virtutibus	1951	609

Author: Card. Caietanus O.P. † 1533		
152 XPE	In Perihermeneias Continuatio	30203 3380
Unknown Authors:		
153 XGC	In de Gener. et Corrupt. Continuatio	31602 3462
154 XME	In Meteorologicorum Continuatio	39760 5333
155 XPT	De Tempore	5240 931
156 XP6	De Natura Loci	2880 679
157 XPI	De Intellectu e Intelligibili	1620 497
158 XP8	De Potentiis Animae	6300 1584
159 XPS	De Natura Syllogismorum	2863 689
160 XPL	De Totius Logicae Aristotelicae Summa	55779 4500
161 XPV	De Inventione Medii	555 151
162 XPU	De Universalibus I	3272 794
163 XU2	De Universalibus II	2688 653
164 XSA	De Vener. Sacramento Altaris	37820 7806
165 XSE	De Sacramento Eucharistiae	7285 2150
166 XTH	De Humanitate J.C.D.N.	44809 8917
167 XTM	De Divinis Moribus	4850 2037
168 XTB	De Beatitudine	13743 3847
169 XSO	De Officio Sacerdotis	9248 2755
170 XAV	De Vitiis et Virtutibus	4666 2086
171 X2A	Expositio II super Apocalypsim - Vox Domini	191110 19079
172 XPB	De Motoribus Corporum Caelestium	2283 743
173 XSR	Sermo: Intravit Rex	1583 707
174 XSX	Officium de Festo S. Augustini	1581 960
175 XPD	De Demonstratione	891 248
176 XNB	Qu. de Natura Beatitudinis	1371 476
177 XUN	De Universalibus	6456 1062
178 XAM	De Arte Musica	1866 762
179 XM2	Ars Musyce	5024 1766