

12-11-1981

24

Textus quem R. Busa adduxerat ad Colloquium romanum de vocum frequentiis (apud « Lessico Intellettuale Europeo », 1981) hic denuo editur quippe cuius dactiloscRIPTa recensio fuerat omnibus Gallaratum invitatis praemissa ut continuationis ratio ac veluti thema.

ROBERTO BUSA S. I.

DE VOCUM FREQUENTIIS IN USUM LEXICOLOGIAE
PRIUSQUAM STYLISTICAE INDUCENDIS

I

1. Saepe, praesertim ab iis qui in ephemeridibus scribunt, numeri seu frequentiae vocum conclamantur probativa in dirimendis quaestionibus de authenticitate aut chronologia operum. Sed periti communiter timent ne facile invalidetur conclusio quae examen frequentiarum direxerit ad statim invenienda vel proprietates styli characteristicas et distinctivas personae auctoris, vel etiam insuper modos seu modulos quibus illae characteristicae ita in eodem auctore evolvantur ut exinde chronologia ipsius operum statui possit.

2. Stylistica enim quae dicitur disciplina, profundior est et subtilior et forte etiam acerbior, quam multi animo advertant. Ne certa quidem adhuc videtur possibilitas reperiendi in stylo paucas singillatas particularitates, quae possint, ut hodie dicimus, formalizari seu formulizari in usum automati, ad hoc ut idem probent quod pellis sub extremis digitis delineata.

Si quando vero tale parvum quid in uno textu forte detegeretur, adhuc inquirendum superesset utrum necne in omni et quolibet auctore, scilicet in omnis hominis sermone, aliqua singillatio characteristicae expectari debeat.

Si possibiles sunt sosiae in figura corporis, cur non etiam in stylo?

Nullus bonarum artium criticus quid unquam conclusit ex numero lapidum aut quantitate colorum aut numero notarum musicalium ... Haec enim se tenent ex parte elementorum materialium, dum proprietates in forma et formalibus consistunt. Forma scilicet et aspectus formales id sunt quo multa diversaque materialia elementa in unum quid, novum, totum, integrum, pulchrum, simul assumuntur et constructurantur.

Quis autem denegaret validitatem omnibus illis inquisitionibus stylisticis quae antehac innitebantur in formalibus et globalibus appretiationibus et degustationibus, potius quam in quantitativis illis computationibus, quae automata hodie nobis consentiunt?

3. Ipse sentio examen frequentiarum seu quantitatum prius dirigendum esse ad finalitates lexicologicas; secundo atque exinde ad inquisitiones stylisticas generales; quae tertio tandem momento applicandae erunt quaestionibus particularibus de authenticitate aut chronologia.

Finem autem lexicologicum voco delineationem et descriptionem – ac veluti italice « carta geografica » – integram, completam et systematicam, totius systematis verbalis prostantis in textu de quo agitur.

Haec descriptio (A) comportat prius quidem computationes elementorum materialium seu vocum, sed (B) postea statim cogitur ascendere ad examen architecturae seu structurarum aut stylematum etc. quibus elementa illa progressive componuntur in tota concentrica: propositiones - peryodos (sensu italico) - paragraphos - ratiocinia - capitula etc. ac tandem totum opus simul.

4. Frequentias ipse aliquas ex meis 179 scriptis scholasticis medievalibus iam computaveram: plures quidem de (A) – non omnes quia sine fine ... – et aliquas etiam de (B). Sectionis enim Primae mei Indicis Thomistici octo volumina continent quinque tabulas distributionis, duo volumina triginta et octo tabulas lexi-

cologicas. Inde hic excerpsi summarias quasdam mearum frequentiarum notitias.

In « Revue » (LASLA, Liège) 1976, n. 3, pp. 1-45, meus articulus « The quantities of the Latin vocabulary documented in the Index Thomisticus » iam praebebat plures frequentias et numeros ex iis quae ex IT emergerant. Ibi tamen non adhuc continebantur quantitates operis « Tabula Ethicorum » (054 CTE), quod posteriore tempore inclusum est. Hoc opus constat 31.095 vocibus; hae adunantur in 4.965 formas atque 2.048 lemmata; harum, lemmata 22 et formae 177 sunt simpliciter novae additiones, lemmata 250 et formae 1.118 sunt novae variationes graphicae. Sed idem articulus plura praebet quae in tabulis IT iam non adsunt, ut sunt praecipue quantitates plurium operum, ut poetarum et scriptorum, quae comparantur cum numeris vocum et formarum singulorum IT operum.

II

1.1. Opera Omnia S. Thomae (= ST) divisa a me sunt in 118 scripta, quae linguistice ordinavi. Simul continent 8.767.855 voces.

His addidi 61 scripta ex aliis Medii Aevi auctoribus (= AA), a s. IX ad XVI, quae simul continent 1.864.125 voces.

Haec 179 scripta simul ergo continent 10.631.980 voces.

1.2.	<i>continent voces</i>	<i>scripta ST</i>	<i>scripta AA</i>
	1-999	9	3
	1.000-9.999	36	26
	10.000-99.999	45	27
	100.000-499.999	26	5
	513.890	1	—
	597.821	1	—

1.3. Harum vocum agnitae sunt occurrere:

in citationibus ad litteram

ST 1.527.803 AA 321.966

in citationibus ad sensum, ubicumque scilicet sermones duorum forte miscentur (praeter commentaria, i.e. omnia opera 050-076 et quaedam ex 077-100, quae ex sua natura iam intelliguntur esse citationes ad sensum ex integro)

ST 518.461 AA 13.121

in ubicationibus relatis vel ad alios vel ad alias partes proprii operis

ST 690.287 AA 77.955

in sermone ipsi auctori proprio

ST 6.031.304 AA 1.451.083

1.4. Ex his emerunt 133.752 formae vocum graphice diversae. Multiplicatis quibusdam formis ratione compositionis aut homographiae, hae factae sunt 147.088.

Harum, 126.488 sunt « plenae » (fp); 20.600 sunt « vacuae » (fv) diversa ratione.

Formas plenas et vacuas ego hic non eas intelligo quas quidam « mots pleins » et « mots vides » appellant, sed eas voco plenas quibus in IT una saltem occurrentia agnoscitur, vacuas vero quibus nulla. Hoc diversis evenit de causis. Unum exemplum sufficiat: forma « caro », tres habens posibles valores, occurrit 1.684 vicibus, quibus singillatim perlectis atque iuxta significationem dispertitis, hoc emerit quod ut *caro-carnis* occurrit in ST 1.456 in AA 224; ut *carus-a-um* occurrit in ST 2 in AA 2; ut *caro-carere* numquam occurrit. Inde prima et secunda vocantur « plenae » sed tertia remanet « vacua ».

Ex ipsa enim rerum natura apparuit haec categoria agnoscenda esse in systemate verbali, cui locus non est in textu operum.

1.5. Hae formae adunantur in 20.173 lemmata « compressissima », quibus annectuntur 4.840 lemmata variantia (usque ad 8 variationes eiusdem).

Tres classes lemmatum emerunt: A) voces communes; B) nomina propria eorumque derivata proxima; C) voces speciales, quas, infra, n. 3 declarat.

	<i>lemmata</i>	<i>formae</i>	<i>fp</i>	<i>ST</i>	<i>AA</i>
A)	17.436	134.510	117.221	8.157.237	1.709.546
B)	2.710	7.278	5.775	231.878	61.415
C)	27	5.300	3.492	378.740	93.164

Nominibus propriis alia adderet qui examini subiiceret voces incompletas seu abruptas lemmatis 96.500, voces brevias lemmatis 96.550 et non agnitas lemmatis 96.075.

Forcellinius 1940 recenset 92.052 lemmata, quorum 27.392 in Onomastico.

Attentis repetitionibus lemmatum graphice eorundem atque disiectis syntagmatibus, lemmata graphice diversa habet 84.607.

2.1. LEMMATA 1786 integre advocantur seu repraesentantur a formis vacuis (sunt 4.557). Haec probabilius occurrunt sensu vocis sibi homographae, sed non adhuc certo constat utrum necne suo etiam sensu occurrant.

2.2. Lemmata classis A:

<i>sunt</i>	<i>L</i>	<i>fp</i>	<i>ST</i>	<i>AA</i>	<i>in Forcellini L.</i>
nomina	11.366	46.417	3.205.264	724.754	72.883
verba	3.796	70.350	2.060.649	419.810	9.837
invariab.	362	441	2.481.895	506.106	4.285
inter quae					
sunt praepos.	45	57	866.793	177.609	
ps-lemmata	10	13	409.429	58.876	5.047

Tabula 15 classificat et dividit iuxta singulos typos morphologicos seu flexionis omnia nostra lemmata. De lemmatibus autem linguae latinae in universali, vide meum: « The morphological categories of the Forcellini's Latin lemmas » in « Lexicologie een

bundel opstellen voor F. De Tollenaere [...] », Groningen, Wolters-Noordhoff 1977 pp. 79-106.

Tabula 26 in IT numerat circiter 1.100 typos morphologicos formarum classis A repertos in nostris textibus. Horum plures, iuxta tabulam 12, sunt homographi aliis flexionibus eiusdem paradigmatis.

Appendices tabulae 26 eosdem typos ordinant prius iuxta numerum formarum et dein iuxta numerum vocum, quae eos verificant. Verbi gratia, numerosissimus in formis est nom. sing. fem. lae decl. sed frequentissimae in vocibus sunt primo loco formae invariabiles, dein 3 pers. sing. ind. praes. act. coniug. irreg. (v.g. dicit, est) et tandem nom. sing. fem. lae declinationis.

Tabula 16 praebet 275 illa lemmata quae plures typos verificant intra vel verba vel nomina (numquam intra utraque): fp 1.884 ST 86.247 AA 22.323.

2.3. Inter nostra dantur lemmata quae Forcellinius non exponit ut lemmata, ei propterea ut « nova », forte sed non adhuc certo, adiungenda:

classis A	2.657	fp	6.313	ST	35.150	AA	7.797
classis B	445	fp	688	ST	4.105	AA	1.321

2.4. De lemmatibus classis A quae non sint hapax, Tabula 17 praebet diversam habitudinem inter numerum formarum et occurrentiarum.

Maximum numerum fp habet *facio*: 146.

Minimus (12) est numerus lemmatum habentium fp 90-99.

Maximus numerus lemmatum (1.314) est habentium 3 fp.

Sed maximam occurrentiam (ST 1.664.753) habent 836 lemmata habentia 1 fp non hapax: *obvie*, ratione invariabilium.

3. TYPI FORMARUM

	<i>fp</i>	<i>ST</i>	<i>AA</i>
3.1. lacunae	1	24	1
figurae	1	17	21
voces abruptae	477	877	—
disiectae partes vocum	22	37	21
voces breviatae	801	92.006	23.744
voces non agnitae	24	1	24
3.2. numeri vocabulo dicti	545	180.385	47.863
numeri numero dicti	851	100.163	20.075
symbola logicae	26	5	26
symbola musicae	68	—	182
reliqua symbola	160	4.664	902
3.3. voces aliarum linguarum	516	561	305
3.4. voces abnormes	73	34	43
formae-lemma	48	—	48
suff. cum enclitica	1.524	4.085	1.914
non stab. compositae	140	9.653	587

3.5. In Tabula 25 nota: v.g. maximus (58.492) est numerus formarum habentium 8-11 litteras; maximam vero occurrentiam (3.394.695) habent 4.376 formae habentes 3-5 litteras.

3.6. Voces-claves sunt:

in concordantia prima:	ST 2.660.442	AA 655.204
in altera:	ST 5.962.980	AA 1.166.900
in nulla (sed in indicibus)	ST 144.433	AA 42.021

Ex 117.221 *fp* classis A, iuxta Tabulas 18-20:

102 recipiunt contextum breviorum in conc. prima:

ST 85.644 AA 12.978

2.204 invariables aut grammaticales assignantur conc. alteri:

ST 4.885.525 AA 925.893

4. POLIMORPHIA-POLISEMIA

4.1. Conspectus polimorphiae formarum:

7.746	in Tabula 3 sunt variationes graphicae additae 6.625 formis (max 8 eidem)
1.180	in Tabula 24 sunt variationes formales (max 4 eiusdem) 564 formarum
140	in Tabula 21 sunt formae non stabiliter compositae
1.524	in Tabulae 2-3 lemmate 97.000 univoce suffixatae ab enclitica
48	in Tabula 22 sunt formae-lemma
73	in Tabula 23 sunt formae abnormes

4.2. Conspectus homographiae (intra polisemiam) vocum scil. pertinentium ad plura nostra lemmata:

(in Tabula 6) 10.810 formae graphice diversae multiplicatae sunt (max usque ad 7) in:

dv	fp	2.145	fv	1.036	ST	618.919	AA	127.396		
nd	fp	8.806	fv	543	ST	2.054.084	AA	387.869		
ns	fp	—	fv	10.363	ST	— —	AA	— —		
		<hr/>								
		f	22.893							

(in Tabula 7) 41 lemmata integre homographa non divisa
f 758 ST 322.506 AA 68.184

(ex Tabula 1) 10 pseudolemmata classis A seu 13 formae non divisae
f 13 ST 409.429 AA 58.876

(in Tabula 10) formae quae possent etiam esse nomina propria
f 12.532 ST 1.508.691 AA 351.608

(in Tabula 11) formae quarum terminatio posset etiam esse enclitica
f 1.838 ST 70.228 AA 15.066

<hr/>										
omnes	f	38.034	ST	4.983.857	AA	1.008.999				
%		25,85		56,84		54,12				

Hic terminationes homographae minime connumerantur.

4.3. Tabula 12 praebet omnes terminationes regulares quae *intra eandem coniugationem vel declinationem* habent plures valores (non ergo, hic, illas etiam quae in diversis occurrunt).

Recensui 2.750 possibiles terminationes regulares latinas, numerum forte augendum.

Harum 1.268 sunt verborum ex themate supini: duplicandae propterea, si quando suffixum thematicum ab eis excludatur. Non multae terminationes stricte verbales, sed plurimae nominales et participiales sunt « homographae » intra eundem typum flexionis.

En quae supersunt univocae:

in decl. 1-2: am, arum, orium, os
 3: em, um, ium
 4: ui, um, ua, uum
 5: em, e, erum

5. FREQUENTIA OCCURRENDI

5.1. Frequentiam absolutam complexivam vocum classis A in ST praebant Tabulae 33-34. En quae supra 100.000:

<i>lemmata</i>		<i>formae</i>	
466.781	sum	295.593	et
295.593	et	250.863	est
262.293	quam-quod	249.459	in
249.459	in	233.626	quod
189.544	non	189.544	non
184.800	qui	157.103	ad
172.544	dico	107.012	sed
157.103	ad		
121.628	hic		
107.012	sed		

Lemmata 128 (communia seu cl. A) quorum frequentia excedit 9.999, simul occurrunt 5.285.963 vicibus seu 60,28 %.

Formae 880 communes, quarum frequentia excedit 999, simul occurrunt 5.872.138 vicibus seu 66,97 %.

Inter has formas dantur 85 quarum frequentia excedit 9.999: hae simul occurrunt 3.675.575 vicibus seu 41,92 %.

Tabulae autem 35-36 praebent comparationem frequentiae absolutae dispersitae per series operum. Inde ordo vocum iuxta frequentiam decrescentem absolutam in omnibus operibus ST simul sumptis, constat variari intra ipsius opera singula aut series partiales operum. Ad exemplum vide septem frequentissimas formas:

ST	001-004	in	et	est	quod	non	ad	sed
ST	005-006	est	in	et	quod	non	autem	ad
ST	007-010	est	quod	et	in	ad	non	sed
ST	011-019	in	est	quod	et	non	ad	sed
ST	020-049	et	in	quod	est	non	ad	ab
ST	050-066	et	quod	est	in	non	ad	autem
ST	067-076	et	in	non	est	quod	ad	autem
ST	077-100	et	in	est	quod	non	ad	quia
ST	101-118	est	et	in	non	quod	ad	esse
<hr/>								
ST	001-118	et	est	in	quod	non	ad	sed
<hr/>								
AA	119-179	et	in	est	quod	non	ad	de

5.2. Tabulae 27-28 praebent voces classis A quae sunt hapax ST: formae 29.637, seu 30,22 % (fere dividia pars in citationibus); non hapax vero sunt 68.430.

Harum, 2.229 sunt unicae formae sui lemmatis, quare in ST lemmata hapax sunt 2.229, seu 16,9 %; non hapax autem sunt 11.044.

Distribuuntur hoc modo: sunt hapax legomena:

<i>in opp. ST</i>	<i>lemmata</i>	<i>formae</i>
001-004	216	2.954
005-006	46	567
007-010	254	3.462
011-019	112	1.462
020-049	85	1.272
050-066	439	3.627
067-076	753	11.436 (in cit. ad litt. 9.180)
077-100	295	4.618
101-118	29	329

5.3. Formae classis A quae occurrunt solum in AA, ut ex tabulis 29-30, sunt 18.306, lemmata autem 2.261.

5.4. Tabulae 29-30 ostendunt ubi voces classis A prima vice occurrant in ordine operum, atque utrum necne ibi tantum occurrant, vocibus quae sunt hapax hic non computatis.

En specimen numericum: occurrunt prima vice

<i>in opp. ST</i>	<i>habentibus voces</i>	<i>lemmata</i>	<i>formae</i>
001-049/A	3.809.392	8.407	45.258
001-049/B	1.066.101	1.154	9.440
050-066	1.315.536	619	3.675
067-076	1.169.941	760	8.693
077-100	1.325.609	99	1.310
101-118	81.277	5	54

5.5. Tabulae tandem 37-38 ostendunt in quot ST operibus occurrant voces classis A non hapax.

Verbi gratia, in plus quam 99 operibus occurrunt formae 137, lemmata 140: operum ST sexdecim constant minus quam 2.000 vocibus. Summaria ratione, frequentia occurrendi, sed non numerus formarum, respondet decremento numeri operum in quae distribuuntur.

6. Hucusque de vocibus singulis. De syntagmatibus, seu locutionibus, et de trinomiis, quaedam notabilia etiam comprobata sunt, alibi declaranda. Unum ad interim sufficiat: ea numero sunt valde minora quam quis forte expectaret: scilicet catenae vocum identicae parum repetuntur.

6.1. In Concordantia Prima ST:

514	binomia occurrunt, simul sumpta	103.578
61	trinomia occurrunt, simul sumpta	9.462
9	catenae 4 vocum occurrunt, simul sumpta	2.021
1	catena 5 vocum occurrit	19
<hr/>		<hr/>
585	syntagmata occurrunt, simul sumpta	115.080

6.2. In Concordantia Altera ST:

70.430 formae sunt voces-claves in 5.962.981 trinomiis.

Horum, 3. 360.425, seu 56,35 %, sunt diversa, constant scilicet vocibus quarum una saltem est diversa.

Horum, 2.758.653 occurrunt semel;

545.984 occurrunt plus quam 99 sed minus quam 1.000;

117 occurrunt plus quam 1.000: haec constant solum vocibus grammaticalibus aut dicendi.

Ex eis 49 sunt trinomia, quorum maximum est 4.851 (ad primum ergo); 68 sunt binomia (quia contigua interpunctioni forti), quorum suprema sunt 15.556 (et ideo) atque postea 6.870 (et sic).

Si deus dabit, plura de his omnibus in futuro tempore vitae.