LingAeg 16 (2008), 105-134

On Consonant Frequency in Egyptian and Other Languages[1]

Carsten Peust, Konstanz

Abstract

Counts of consonant frequency in running texts from a sample of 50 languages as well as from two diachronic varieties of Egyptian (Middle Egyptian and Bohairic Coptic) are provided. Based on the 50 language sample, 14 typological generalizations on phoneme frequency are proposed. It is then tested whether current reconstructions of the Egyptian sound system agree with these generalizations. Frequency data can thus provide additional, hitherto unexploited evidence for determining the sound values of Egyptian.

It can be shown that ‹*p*›, ‹*t*›, ‹*k*› of earlier Egyptian were probably plain voiceless stops, whereas the distribution of ‹*b*›, ‹*d*›, ‹*g*› does not favour their interpretation as voiced stops, but rather as possibly some kind of emphatics. It is also shown that ‹*ꜣ*› is likely to have been a liquid, that the behaviour of ‹*z*› would better agree with some kind of affricate than with /z/, that ‹*š*› was probably not originally a /ʃ/, and that Rössler's reconstruction of ‹*ḫ*› as /γ/, which has become widely accepted, is implausible.

Statistical approaches to language have received a somewhat marginal attention in linguistics, although proponents of statistical linguistics have ascribe them a crucial role in assuming that all linguistic laws are essentially of a stochastic nature.[2] Grammar books usually enumerate the phonemes of a language without giving any information on frequency. One problem of this way of presentation is the fact that there is no clear-cut boundary between phonemes that exist and phonemes that do not. Most or possibly all languages possess so-called "marginal" phonemes, which can originate when speakers, possibly bilingual speakers, retain phonemes of (originally) foreign words such as in *genre* /ʒãrə/ or *loch* /lɔx̱/ in English or in *Garage* /gara:ʒə/, *Teint* /tɛ̃/ or *Thread* /θɹɛd/ in German. Quantitative data would give a more realistic characterization of the status of a phoneme in a language than a binary decision about its "existence" can ever do.

Also for non-marginal phonemes, frequency counts may allow for interesting conclusions. The first major researcher in statistical linguistics, G.K. Zipf, posited that articulatorily "simple" phonemes are more frequent than "complex" phonemes: "The accent or degree of conspiciousness of any word, syllable, or sound is inversely proportionate to the relative frequency of that word, syllable, or sound, among its fellow words, syllables, or sounds in the stream of spoken language" (Zipf 1932: 1) / "it appears plausible to believe that the magnitude of complexity of a phoneme bears an inverse (not necessarily proportionate) relationship to its relative frequence of occurrence" (Zipf 1935: 73). Although Zipf's notion of "conspiciousness" or "complexity"

---

1 My thanks go to Wolfgang Schenkel (Tübingen), whose electronic Coffin Text data were essential for me in preparing this article, as well as to Dörte Borchers (Berlin) and to Sónia Frota (Lisboa) who provided me with data on Sunwar and Portuguese.

2 E.g. Altmann *et al.* (2002: 6): "Heute nehmen wir an, daß hinter allen Spracherscheinungen stochastische Mechanismen stehen".

is too subjective to render his idea fully satisfying (cf. Altmann & Lehfeldt [1980: 113-119] and Berger [1987: 5] for criticism), there is evidently some truth in it: Sounds intuitively perceived as "simple", such as *m*, *n*, *t*, *k*, tend to be more frequent in the world's languages, and are also more commonly found in grammatical affixes, than ejective uvular affricates, aspirated labiovelar stops or voiced lateral clicks.

The question what the frequency distribution of phonemes can contribute to our knowledge about historical sound change has not been raised very often. Among the few examples are Zvelebil (1972, who takes the low frequencies of voiced stops in modern Dravidian languages as an indication that they were absent from Proto-Dravidian) and Colless (1992, who uses frequency counts as an argument for determining the sound values of the Byblos script). With respect to Egyptian, Kammerzell (2005: 184-193) employs frequency data in order to explain why certain elementary graphemes are likely to have changed their sound value during the Old Kingdom. On the use of statistical linguistics in general in the field of Egyptology see also Lepper (2006) (with further references).

There are several studies on phoneme or grapheme frequency of single, for the most part European languages. Only a few studies compare data from more than one language (Zipf 1932: 2; Zipf 1935: 68-79; Krámský 1959; Gamkrelidze 1978: 40). I have decided to collect data anew for 50 languages, together with two diachronic varieties of Egyptian (Coffin Texts from the early Middle Kingdom, ca. 2000 BC; translation of the New Testament into Bohairic Coptic, 1st millennium AD). The following remarks have to be made:

- I count frequencies in running texts, not in dictionaries (in other words: token counts, not type counts).[3]
- This is a study of consonant frequency only. Vowels are ignored.
- Not only vowels, but also the "weak" consonants /j/, /w/ and /ʔ/ are ignored. Their status on the borderline between consonants and vowels is doubtful in several languages, and these sounds are represented only inconsistently in Egyptian writing.
- Long (double) consonants are counted as a single token. Consonant length, if it existed, is not noted in the Egyptian writing system. However, a few of the authors whose data I reproduce below have counted double consonants as two tokens (Bengali, Gujarati).
- I have preferred languages whose orthographies can be translated into a phonemic representation by applying a limited number of mapping rules. For some languages

3 I prefer text counts because of their better statistical behaviour: It can be assumed that the figures will more and more converge towards a limit as the amount of text increases. A similar convergence is not assured for dictionary counts, since the use of larger dictionaries may lead to the inclusion of more and more rare and atypical lexical material, whose phoneme distributions may differ from those of the core vocabulary. Another crucial difference is the treatment of grammatical affixes, which are considered in text counts as opted for here but are ignored in counts on a dictionary base. I believe that it is preferable for grammatical affixes to contribute to the quantitative description of a language.

A third possibility, which is likely to yield results somewhere intermediate between text and dictionary counts, is to count entries from a dictionary restricted to the most basic (and frequent) vocabulary, as done by Sokarno (2002) for a dialect of Nile Nubian.

with a large discrepancy between graphemes and phonemes, I relied on other authors who performed counts on phonemicized texts.

- The relative frequency of each consonant phoneme is given in ‰ so that the sum in each row of the charts amounts to 1000.
- The figures should not be taken too literally: Modifications of the phonemic analysis will often be possible which would influence the numbers to some degree. Some of my text corpora are contaminated by foreign names, abbreviations etc., which, despite some effort, I have not always been able to remove completely. In order not to suggest undue accuracy, I round all values to ‰ without decimals.
- As far as available from my source data, I provide separate numbers for word-initial (*initial*) and non-word-initial position (*non-init.*). I do not attempt to make more subtle distinctions such as between syllable-initial and syllable-final consonants, which would often be impossible to decide for Egyptian.
- Some more or less arbitrary decisions could not be avoided. One dubious issue is the definition of word boundaries, for which I have tended to simply follow the orthographic conventions (e.g. I took the definite article as a part of the noun in Arabic, Hebrew and Coptic, but not in European languages).

I know that there are several weak points in what will follow. Although I attempted to analyse languages of different parts of the world and of different genetic affiliations, the language sample is certainly not unbiased. While European languages are represented well, I was not able to include data from any Papuan, Australian or Northern American language. The phonemic analysis is often non-trivial, and certain ad-hoc decisions had to be made without further discussion. Last but not least, it is a gross oversimplification to identify phonemes across different languages, since the systematic position of the same sound may be different in different languages: A phoneme /*t*/ of English, which contrasts with sounds such as /*d*/ or /*þ*/, is not the same as a phoneme /*t*/ of Chinese, which contrasts with neither /*d*/ nor /*þ*/ but instead with /t^h/.

Nevertheless, I believe that generalizations as attempted here are not complete nonsense. Recurrent patterns seem to emerge from the data despite the imperfections that still exist. The patterns are already so robust as to allow for first tentative conclusions about the Egyptian sound system from a quantitative point of view. I hope that others will continue this line of research and will be able to improve on the methodology, which may make any conclusions more certain than they can be now.

Frequency data from a 50 language sample

In the following charts, up to three data rows are presented for each language. They provide counts for any position, word-initial position, and non-word-initial position, respectively. Sources and further details are given in an appendix at the end of this paper. I was only able to include languages for which either statistical data have been published or a sufficient amount of texts in electronic form was easily available. In doing this, I have attempted to achieve some typological diversity in my sample. Languages whose consonant inventory is extremely small or extremely large have been ignored.

	p	*b*	*f*	*v*	*m*	*t*	*d*	*þ*	*ð*	*n*	*tʲ*	*dʲ*	*ɲ*	*ts*	*dz*	*s*	*z*
Albanian	62	21	15	29	69	167	45	6	17	115	18	11	9	11	1	72	13
initial	134	32	22	35	79	132	59	9	31	98	36	13	23	9	0	64	11
non-init.	29	15	11	26	64	184	38	5	10	123	9	10	2	12	1	76	14
Bah.Indon.	52	47	3	1	83	99	84			128			20			80	1
initial	99	102	2	0	137	90	166			21			1			141	2
non-init.	36	28	3	1	64	102	55			165			26			58	0
Bambara	2	84	45	0	85	64	57			202			16			67	1
initial	1	111	73	0	90	101	93			100			19			88	0
non-init.	2	56	18	0	81	27	20			306			13			46	1
Basque	16	61	3		23	116	63			168			1	27		79	
initial	29	208	7		45	16	129			73			0	–		133	
non-init.	12	20	2		17	144	45			195			1	35		64	
Breton	32	43	11	47	44	98	100			125			1			79	58
initial	69	102	19	56	60	47	165			36			0			91	27
non-init.	18	20	7	43	37	119	74			161			1			73	70
Czech	56	33	13	70	66	85	49			82	15	9	36	25		84	35
initial	120	47	29	88	61	68	49			92	9	7	14	15		143	62
Dholuo	22	27	3		133	57	76	10	10	169			21			26	
initial	29	33	3		204	59	30	6	8	157			20			12	
non-init.	17	24	3		89	56	104	13	11	177			21			35	
English	31	31	32	36	50	132	58	8	43	120						74	38
initial	45	77	66	12	69	92	46	13	122	50						92	0
non-init.	24	11	18	46	41	149	63	6	9	149						67	55
French	72	21	30	46	64	99	78			54			2			118	40
initial	127	26	43	51	71	49	142			35			0			145	0
non-init.	30	17	21	42	58	137	27			68			3			97	71
German	13	29	37	36	47	141	73			168						73	35
initial	20	55	86	96	61	56	215			54						0	84
Greek(mod)	82	4	23	20	62	153	13	22	36	119						182	9
initial	147	7	15	19	90	245	3	24	66	46						134	8
non-init.	60	3	26	20	52	122	16	21	25	144						198	9
Guaraní	134	36	3	87	57	90	23		9	33			29			44	
initial	160	48	7	24	40	82	24		7	19			40			22	
non-init.	125	32	2	108	63	92	22		9	38			26			51	
Hungarian	19	31	17	38	69	141	36			98	0	26	12	5	0	33	52
initial	31	41	65	84	151	94	20			62	0	9	8	7	0	69	3
non-init.	16	28	6	27	49	152	40			106	0	30	13	4	0	25	64
Latin	49	27	20		96	157	48			114					0	147	
initial	125	10	72		89	105	76			62					0	131	
non-init.	29	31	6		98	170	41			128					0	151	
Portuguese	65	23	23	37	70	107	90			50			13			76	20
initial	131	25	49	38	105	74	124			101			–			101	2
non-init.	27	22	7	36	50	126	70			21			21			62	30
Sami	16	44	6	50	68	116	115	0	9	93		16	4	7	1	108	
initial	4	100	14	66	129	3	191	–	–	21		–	3	16	–	140	
non-init.	21	20	2	43	43	163	83	0	12	123		22	5	3	1	95	
Swahili	32	52	21	12	146	75	28	1	3	137			17			43	27
initial	26	43	5	17	223	28	5	2	3	190			18			42	21
non-init.	34	55	26	10	118	93	36	1	3	117			16			43	29
Tagalog	55	42	1		77	88	30			199						111	
initial	71	41	2		126	40	33			242						219	
non-init.	47	43	0		53	111	29			177						59	
Tok pisin	90	45	2	9	139	90	17	1	0	105						99	0
initial	72	121	3	0	139	86	31	1	0	116						100	0
non-init.	98	11	1	13	139	92	11	0	0	99						98	0

	tś	*ś*	*č*	*dʒ*	*ʃ*	*ʒ*	*k*	*g*	*x*	*γ*	*ŋ*	*h*	*r*	*rr/ř*	*l*	*ɫ*	*ʎ*	∑
Alban.			3	1	36	2	69	15				13	136	5	32	6		1000
init.			*6*	*0*	*36*	*2*	*98*	*8*				*12*	*21*	*6*	*21*	*1*		*1000*
n-in.			*2*	*1*	*35*	*2*	*55*	*18*				*14*	*190*	*5*	*37*	*9*		*1000*
B.Ind.			9	18	1		97	23	1		55	49	91		62			1000
init.			*10*	*24*	*1*		*108*	*11*	*1*		*0*	*31*	*23*		*28*			*1000*
n-in.			*8*	*16*	*1*		*94*	*27*	*1*		*74*	*55*	*114*		*74*			*1000*
Bamb.			15	22	1		137	50			10	3	62		76			1000
init.			*26*	*32*	*2*		*208*	*8*			*4*	*4*	*6*		*34*			*1000*
n-in.			*5*	*11*	*1*		*65*	*93*			*15*	*1*	*120*		*118*			*1000*
Basq.	4	40	5		2	8	99	56				37	109	26	56		0	1000
init.	*–*	*34*	*7*		*2*	*32*	*37*	*92*				*108*	*–*	*2*	*45*		*0*	*1000*
n-in.	*5*	*41*	*5*		*2*	*1*	*116*	*46*				*18*	*139*	*32*	*59*		*0*	*1000*
Breton					5	6	52	49	7	9		6	153		75		1	1000
init.					*4*	*7*	*80*	*89*	*5*	*23*		*22*	*50*		*47*		*0*	*1000*
n-in.					*5*	*6*	*40*	*33*	*8*	*3*		*0*	*194*		*86*		*1*	*1000*
Czech			17		22	15	72	7	22			23	65	22	78			1000
init.			*13*		*8*	*16*	*54*	*11*	*10*			*27*	*27*	*7*	*23*			*1000*
Dhol.			46	37			124	71			39	8	79		43			1000
init.			*27*	*70*			*177*	*90*			*25*	*6*	*29*		*16*			*1000*
n-in.			*57*	*16*			*91*	*59*			*48*	*9*	*110*		*60*			*1000*
Engl.			9	7	14	1	47	17			17	50	125		63			1000
init.			*6*	*9*	*18*	*0*	*63*	*38*			*–*	*97*	*43*		*41*			*1000*
n-in.			*10*	*6*	*12*	*1*	*39*	*7*			*25*	*30*	*160*		*72*			*1000*
French					9	34	79	10			0		137		108			1000
init.					*10*	*58*	*112*	*7*			*–*		*20*		*102*			*1000*
n-in.					*7*	*15*	*53*	*13*			*1*		*229*		*112*			*1000*
Germ.					25	0	28	35	47		11	18	119		64			1000
init.					*51*	*0*	*49*	*66*	*–*		*–*	*55*	*21*		*30*			*1000*
Gr.(m)							87	4	26	16			92		51			1000
init.							*131*	*2*	*27*	*9*			*10*		*17*			*1000*
n-in.							*72*	*4*	*26*	*18*			*120*		*62*			*1000*
Guar.				38	25		92	5		44	3	101	132		16			1000
init.				*44*	*27*		*107*	*0*		*38*	*1*	*218*	*83*		*10*			*1000*
n-in.				*36*	*24*		*87*	*7*		*46*	*3*	*62*	*149*		*18*			*1000*
Hung.			6	0	73	1	93	42				30	74		106			1000
init.			*17*	*0*	*29*	*1*	*130*	*10*				*95*	*30*		*45*			*1000*
n-in.			*4*	*0*	*84*	*1*	*84*	*49*				*15*	*84*		*120*			*1000*
Latin							117	25				17	129		55			1000
init.							*160*	*21*				*49*	*48*		*52*			*1000*
n-in.							*106*	*26*				*8*	*150*		*56*			*1000*
Portg.					113	15	90	18					122	10	50		10	1000
init.					*6*	*20*	*168*	*10*						*17*	*26*		*4*	*1000*
n-in.					*175*	*12*	*44*	*23*					*192*	*5*	*65*		*13*	*1000*
Sami			18	5	14		33	70			3	71	42		91		1	1000
init.			*27*	*–*	*8*		*2*	*137*			*–*	*18*	*24*		*98*		*–*	*1000*
n-in.			*14*	*7*	*16*		*46*	*42*			*4*	*93*	*50*		*88*		*2*	*1000*
Swah.			9	26	12		173	29		0	22	36	21		78			1000
init.			*11*	*19*	*4*		*197*	*5*		*1*	*4*	*95*	*6*		*33*			*1000*
n-in.			*8*	*29*	*15*		*164*	*38*		*0*	*28*	*14*	*27*		*95*			*1000*
Tagal.							80	55			131	33	24		75			1000
init.							*123*	*17*			*12*	*36*	*4*		*35*			*1000*
n-in.							*60*	*73*			*189*	*31*	*34*		*94*			*1000*
Tok p.			0	1	2		82	29			60	12	37		181			1000
init.			*1*	*1*	*1*		*98*	*32*			*–*	*31*	*17*		*150*			*1000*
n-in.			*0*	*0*	*2*		*74*	*27*			*88*	*3*	*47*		*196*			*1000*

	p	*b*	*ṗ*	*f*	*v*	*m*	*t*	*d*	*ṭ*	*n*	*ts*	*dz*	*tṣ*	*s*	*z*	*č*	*dʒ*
Georg.(anc.)	10	21	4		48	122	80	102	13	83	22	8	17	125	6	6	3
initial	*6*	*6*	*12*		*45*	*223*	*56*	*170*	*4*	*16*	*12*	*11*	*24*	*60*	*13*	*19*	*3*
non-init.	*12*	*26*	*2*		*49*	*86*	*88*	*77*	*16*	*107*	*25*	*7*	*15*	*148*	*4*	*1*	*2*
Hebrew(anc.)	23	80				122	87	34	8	69	9	11	19				
initial	*22*	*130*				*114*	*45*	*12*	*5*	*30*	*6*	*13*	*11*				
non-init.	*24*	*64*				*125*	*100*	*41*	*10*	*80*	*10*	*10*	*22*				
Hebrew(mod.)	10	38		19	60	125	99	41		78	23			24	17	0	0
initial	*15*	*105*		*0*	*83*	*91*	*28*	*15*		*37*	*12*			*12*	*17*	*0*	*0*
non-init.	*8*	*14*		*26*	*52*	*137*	*124*	*50*		*93*	*27*			*28*	*17*	*0*	*0*
Ingush	6	37	0	3	59	36	43	80	11	78	43	0	5	44	28	35	0
initial	*11*	*70*	*1*	*10*	*89*	*66*	*27*	*133*	*27*	*25*	*73*	*0*	*14*	*70*	*6*	*15*	*0*
non-init.	*4*	*19*	*0*	*0*	*43*	*21*	*51*	*51*	*2*	*107*	*27*	*0*	*0*	*30*	*40*	*46*	*0*
Japanese	6	21		7	0	53	121	52		188	22	36		123			
Maltese	31	39		36	12	81	144	61		115	15	0		63	16	9	16
initial	*54*	*57*		*75*	*10*	*120*	*143*	*69*		*32*	*0*	*0*		*56*	*10*	*5*	*19*
non-init.	*21*	*31*		*20*	*13*	*64*	*144*	*58*		*151*	*21*	*0*		*67*	*18*	*10*	*14*
Ossetic	9	40	0	37	12	80	102	90	0	106	34	24	3	85	37	6	8
initial	*11*	*96*	*0*	*92*	*1*	*67*	*34*	*71*	*0*	*68*	*52*	*22*	*7*	*101*	*47*	*16*	*0*
non-init.	*8*	*18*	*0*	*16*	*16*	*85*	*128*	*97*	*0*	*120*	*27*	*25*	*1*	*79*	*34*	*2*	*11*
Persian	22	85		19	17	61	56	120		128				49	50	16	15
Turkish	10	63		7	16	82	50	100		141				52	31	30	18
initial	*12*	*230*		*10*	*42*	*25*	*40*	*136*		*27*				*96*	*11*	*56*	*11*
non-init.	*10*	*15*		*6*	*9*	*99*	*53*	*90*		*175*				*39*	*37*	*23*	*20*

	p	*b*	*ṗ*	*ḅ*	*f*	*v*	*m*	*t*	*d*	*ṭ*	*ḍ*	*þ*	*ð*	*ð̣/ẓ*	*n*	*ts*	*s*	*z*	*ṣ*	*č*
Arabic		52			40		122	60	27	6	8	6	21	4	159		27	7	9	
initial		*94*			*115*		*160*	*62*	*10*	*4*	*3*	*10*	*13*	*4*	*28*		*24*	*3*	*8*	
non-init.		*42*			*22*		*113*	*59*	*31*	*6*	*9*	*6*	*22*	*4*	*189*		*28*	*8*	*9*	
Hausa		56		3	22		92	50	75		17				179		96	23	6	29
initial		*72*		*2*	*24*		*105*	*53*	*121*		*21*				*62*		*122*	*42*	*9*	*45*
non-init.		*43*		*5*	*21*		*81*	*48*	*37*		*15*				*273*		*76*	*7*	*4*	*16*
Kabyle		31			22		81	72	99	10	12			4	174	22	83	12	3	2
initial		*24*			*14*		*59*	*147*	*182*	*3*	*1*			*2*	*155*	*35*	*103*	*9*	*2*	*1*
non-init.		*33*			*24*		*89*	*46*	*71*	*12*	*16*			*4*	*180*	*17*	*76*	*13*	*3*	*3*
Oromo	2	65	2		53	0	97	106	41	3	26				163		77	1		14
initial	*5*	*99*	*0*		*72*	*0*	*79*	*69*	*57*	*4*	*45*				*34*		*55*	*2*		*1*
non-init.	*1*	*52*	*2*		*46*	*1*	*105*	*121*	*34*	*3*	*18*				*213*		*86*	*1*		*19*

	p	*b*	*pʰ*	*bʰ*	*ɓ*	*f*	*v*	*m*	*t*	*d*	*tʰ*	*dʰ*	*ts*	*dz*	*tsʰ*	*s*	*z*	*n*	*ṭ*	*ḍ*
Bengali	28	70	3	13				57	62	45	8	7				51		111	25	1
Gujarati	48	22	6	12			72	75	44	26	24	18				48	0	123	16	15
initial	*91*	*44*	*6*	*15*			*56*	*67*	*73*	*27*	*21*	*8*				*92*	*0*	*58*	*5*	*3*
non-init.	*26*	*11*	*6*	*10*			*80*	*79*	*29*	*26*	*25*	*23*				*26*	*0*	*155*	*21*	*22*
Sunwar	64	63	5	2	2		2	167	51	47	14	4	31	23	9	49		131	1	1
initial	*103*	*92*	*11*	*5*	*5*		*2*	*153*	*40*	*78*	*24*	*8*	*56*	*35*	*13*	*64*		*51*	*0*	*0*
non-init.	*42*	*46*	*2*	*1*	*0*		*2*	*175*	*57*	*29*	*9*	*1*	*17*	*16*	*7*	*41*		*178*	*1*	*1*
Telugu	51	11		6		3	59	80	69	59		8	17	6		52		155	44	44
initial	*98*	*28*		*15*		*8*	*140*	*110*	*72*	*29*		*2*	*36*	*7*		*65*		*88*	*7*	*12*
non-init.	*38*	*6*		*3*		*1*	*38*	*72*	*68*	*67*		*10*	*12*	*5*		*49*		*173*	*53*	*52*

	č̣	ʃ	ʒ	ś	k	g	ḳ	x	γ	q	q̇	ħ	ʕ	h	r	l	Σ
Georg.(a.)	2	20	1		21	40	15	22	14	4	24			12	98	59	1000
init.	*4*	*57*	*3*		*11*	*55*	*19*	*26*	*11*	*8*	*20*			*34*	*68*	*2*	*1000*
n-in.	*1*	*7*	*1*		*24*	*34*	*14*	*21*	*15*	*3*	*26*			*3*	*108*	*80*	*1000*
Hebrew(a.)		59		17	58	10				23		35	55	87	88	105	1000
init.		*50*		*8*	*72*	*11*				*17*		*23*	*84*	*184*	*18*	*145*	*1000*
n-in.		*62*		*19*	*54*	*10*				*25*		*39*	*46*	*57*	*109*	*93*	*1000*
Hebrew(m.)		59	0		47	21		56						95	84	105	1000
init.		*74*	*0*		*79*	*21*		*23*						*249*	*24*	*116*	*1000*
n-in.		*53*	*0*		*36*	*21*		*68*						*41*	*106*	*101*	*1000*
Ingush	1	56	9		20	46	5	64	22	27	12	42	27	7	74	79	1000
init.	*4*	*43*	*5*		*22*	*16*	*15*	*57*	*18*	*28*	*10*	*67*	*27*	*17*	*2*	*32*	*1000*
n-in.	*0*	*63*	*11*		*19*	*63*	*0*	*68*	*24*	*26*	*13*	*29*	*27*	*2*	*112*	*103*	*1000*
Japanese					158	47								49	118		1000
Maltese		23			58	10						29			91	152	1000
init.		*19*			*86*	*11*						*30*			*21*	*182*	*1000*
n-in.		*24*			*46*	*9*						*29*			*120*	*139*	*1000*
Ossetic	0	1			40	42	7	51	16	13					108	49	1000
init.	*0*	*0*			*112*	*13*	*18*	*67*	*2*	*28*					*48*	*27*	*1000*
n-in.	*0*	*1*			*13*	*52*	*3*	*45*	*21*	*7*					*131*	*58*	*1000*
Persian		50	3		43	39		24		4				49	138	12	1000
Turkish		33	0		80	28			12					18	127	100	1000
init.		*28*	*0*		*108*	*100*			*–*					*51*	*10*	*7*	*1000*
n-in.		*34*	*0*		*72*	*8*			*15*					*8*	*161*	*127*	*1000*

	dʒ	č̣	ʃ	ʒ	ɲ	yʾ	k	g	ḳ	x	γ	q	ħ	ʕ	h	r	r̃/ṛ	l	ɫ	Σ
Ar.	14		9				49			11	6	32	19	43	68	54		140	9	1000
init.	*19*		*17*				*59*			*19*	*8*	*60*	*22*	*82*	*24*	*38*		*115*	*–*	*1000*
n-init.	*13*		*8*				*46*			*10*	*5*	*26*	*18*	*33*	*78*	*58*		*146*	*11*	*1000*
Ha.	15		24			3	134	37	12						23	36	41	26		1000
init.	*17*		*39*			*4*	*134*	*49*	*18*						*31*	*18*	*3*	*9*		*1000*
n-in.	*14*		*12*			*1*	*133*	*27*	*8*						*17*	*51*	*72*	*40*		*1000*
Kb.	4		17	3			37	30		11	39	19	16	21	7	67	30	75		1000
init.	*1*		*10*	*3*			*32*	*21*		*6*	*51*	*10*	*6*	*4*	*3*	*5*	*23*	*86*		*1000*
n-in.	*5*		*19*	*3*			*38*	*33*		*12*	*35*	*22*	*20*	*27*	*8*	*88*	*32*	*71*		*1000*
Or.	31	5	5		4		62	42	18						33	97		54		1000
init.	*78*	*11*	*7*		*1*		*128*	*79*	*43*						*91*	*13*		*28*		*1000*
n-in.	*12*	*2*	*4*		*5*		*35*	*27*	*8*						*10*	*130*		*64*		*1000*

	ṭʰ	ḍʰ	ṇ	č	dʒ	čʰ	dʒʰ	ʃ	k	g	kʰ	gʰ	ŋ	h	r	ṛ	l	ḷ	Σ
Bengali	4	0		36	44	12	0		105	29	33	4	2	16	124	25	84		1000
Gujarati	4	1	24	15	38	23	2	25	73	25	15	4		37	114		39	13	1000
init.	*2*	*1*	*–*	*16*	*69*	*60*	*2*	*23*	*99*	*23*	*7*	*6*		*63*	*39*		*23*	*–*	*1000*
n-in.	*4*	*1*	*37*	*15*	*23*	*5*	*2*	*25*	*59*	*26*	*20*	*2*		*23*	*152*		*47*	*20*	*1000*
Sunwar	1	1	1					46	67	21	12		27	23	63		68	3	1000
init.	*4*	*2*	*0*					*12*	*50*	*38*	*25*		*34*	*35*	*20*		*40*	*–*	*1000*
n-in.	*0*	*0*	*1*					*67*	*77*	*11*	*3*		*23*	*16*	*88*		*85*	*5*	*1000*
Telugu			21	26	3			10	83	32				8	79		67	7	1000
init.			*–*	*56*	*6*			*1*	*113*	*25*				*7*	*37*		*35*	*–*	*1000*
n-in.			*26*	*18*	*2*			*13*	*76*	*33*				*9*	*90*		*76*	*8*	*1000*

	p	b	pʰ	f	v	m	t	d	tʰ	ð	n	ts	dz	tsʰ	s	z	č
Armenian	18	17	4		89	72	54	24	16		170	12	12	50	61	12	4
initial	*39*	*41*	*13*		*133*	*116*	*47*	*39*	*25*		*114*	*12*	*26*	*3*	*32*	*9*	*8*
non-init.	*12*	*11*	*2*		*77*	*60*	*56*	*19*	*14*		*186*	*12*	*8*	*64*	*70*	*13*	*3*
Danish	24		23	31	39	59	164		36	56	141				140		
initial	*24*		*37*	*56*	*71*	*58*	*219*		*47*	*2*	*71*				*191*		
non-init.	*24*		*12*	*12*	*14*	*61*	*123*		*27*	*96*	*193*				*101*		
Greek(anc.)	70	10	22			67	113	64	34		160		4		146		
initial	*125*	*18*	*31*			*92*	*121*	*141*	*34*		*32*		*6*		*36*		
non-init.	*49*	*7*	*18*			*58*	*110*	*36*	*34*		*207*		*4*		*186*		
Manchu	99		2	36		86	67		41		116	1		0	61		57
initial	*189*		*5*	*21*		*54*	*121*		*78*		*13*	*2*		*1*	*92*		*103*
non-init.	*55*		*0*	*44*		*102*	*40*		*23*		*168*	*1*		*0*	*46*		*34*
Maori	49			28		85	263				88						
initial	*46*			*46*		*125*	*307*				*44*						
non-init.	*53*			*9*		*42*	*214*				*137*						
Thai	49	17	48	6		88	68	40	50		176				48		31
initial	*54*	*27*	*78*	*11*		*80*	*44*	*60*	*78*		*89*				*75*		*50*
non-init.	*44*	*7*	*16*	*1*		*95*	*93*	*20*	*21*		*263*				*20*		*11*

	p	b	pʰ	ṗ	f	m	kp	gb	t	d	tʰ	ṭ	s	ṣ	n	č	dʒ	čʰ	č̣
Aymara	64		5	2		67			120		14	14	94		123	26		5	7
initial	*72*		*17*	*6*		*126*			*63*		*15*	*9*	*135*		*45*	*30*		*9*	*19*
non-init.	*63*		*2*	*1*		*54*			*132*		*14*	*15*	*85*		*140*	*26*		*4*	*4*
Korean	41		8	2		66			69		11	25	72	13	177	75		19	7
initial	*76*		*16*	*3*		*85*			*94*		*16*	*19*	*125*	*5*	*69*	*133*		*41*	*1*
non-init.	*33*		*6*	*2*		*61*			*63*		*10*	*27*	*59*	*15*	*203*	*60*		*14*	*8*
Wolof	12	73			23	97			68	78			47		150	32	31		
initial	*5*	*102*			*27*	*97*			*57*	*108*			*57*		*121*	*61*	*48*		
non-init.	*19*	*44*			*18*	*96*			*78*	*49*			*36*		*178*	*4*	*15*		
Yoruba	69				36	63	38	45	103	55			49		117		52		
initial	*73*				*51*	*74*	*41*	*34*	*125*	*40*			*61*		*175*		*34*		
non-init.	*66*				*24*	*54*	*35*	*55*	*84*	*67*			*39*		*65*		*67*		

	p	b	pʰ	f	v	m	t	d	tʰ	þ	n	ts	tsʰ	s	z	tʂ	dʐ	tʂʰ
Chinese	38		9	17		35	98		41		163	42	22	37		39		19
initial	*51*		*12*	*23*		*47*	*133*		*55*		*28*	*57*	*30*	*51*		*53*		*26*
non-init.	*–*		*–*	*–*		*–*	*–*		*–*		*540*	*–*	*–*	*–*		*–*		*–*
Fagauvea	36	7		67	21	83	86	120		1	153			28		1	3	
initial	*47*	*11*		*85*	*13*	*110*	*70*	*176*		*1*	*72*			*25*		*2*	*5*	
non-init.	*24*	*3*		*46*	*30*	*52*	*104*	*55*		*0*	*245*			*31*		*1*	*1*	
Vietnamese	12	33		15	44	55	84	67	39		116			8	13	28		
initial	*0*	*49*		*22*	*65*	*46*	*72*	*100*	*59*		*41*			*12*	*19*	*42*		
non-init.	*35*	*–*		*–*	*–*	*74*	*110*	*–*	*–*		*265*			*–*	*–*	*–*		

	p	b	pʰ	m	t	d	tʰ	n	ṭ	ḍ	ṭʰ	ts	tsʰ	s	z	tʂ	dʐ	tʂʰ
Burushaski	16	73	5	130	52	31	5	171	26	5	0	21	5	67	5	5	5	0

	dʒ	ʧ^h	ʃ	ʒ	ɲ	k	g	k^h	x	ɣ	ŋ	h	r	rr	l	Σ
Armenian	8	19	10	5		43	18	34	10	19		30	149	10	28	1000
init.	*2*	*36*	*12*	*14*		*66*	*44*	*35*	*19*	*1*		*91*	*0*	*0*	*24*	*1000*
n-in.	*10*	*15*	*10*	*2*		*37*	*10*	*34*	*8*	*24*		*13*	*192*	*13*	*29*	*1000*
Danish			3			59		28		11	16	33	73		67	1000
init.			*0*			*53*		*38*		*2*	*1*	*62*	*15*		*53*	*1000*
n-in.			*4*			*64*		*20*		*17*	*27*	*11*	*117*		*77*	*1000*
Greek(anc.)						68	28	24				45	87		58	1000
init.						*118*	*35*	*23*				*168*	–		*19*	*1000*
n-in.						*49*	*26*	*25*				–	*119*		*72*	*1000*
Manchu		35	9		31	72		44	89		34		66		54	1000
init.		*32*	*13*		*58*	*108*		*23*	*76*		–		–		*9*	*1000*
n-in.		*36*	*6*		*17*	*54*		*55*	*95*		*51*		*98*		*76*	*1000*
Maori						185					64	106	131			1000
init.						*223*					*51*	*83*	*75*			*1000*
n-in.						*142*					*78*	*132*	*192*			*1000*
Thai		19				80		45			87	23	66		61	1000
init.		*31*				*76*		*78*			*7*	*37*	*49*		*76*	*1000*
n-in.		*7*				*83*		*11*			*169*	*8*	*83*		*45*	*1000*

	ʃ	ɲ	k	g	k^h	ḳ	x	ŋ	q	q^h	q̇	χ	h	r	l	ʎ	Σ
Aymara		21	118		12	5	55		35	9	5	79		80	26	13	1000
init.		*3*	*85*		*16*	*8*	*215*		*52*	*19*	*11*	*1*		*4*	*31*	*6*	*1000*
n-in.		*25*	*125*		*11*	*5*	*21*		*31*	*7*	*3*	*95*		*96*	*25*	*15*	*1000*
Korean			127		8	13		71					62	133			1000
init.			*189*		*4*	*7*		–					*105*	*11*			*1000*
n-in.			*112*		*10*	*14*		*88*					*51*	*163*			*1000*
Wolof		39	71	66			38	23	2					51	101		1000
init.		*57*	*79*	*37*			*24*	*31*	–					*12*	*76*		*1000*
n-in.		*21*	*63*	*94*			*51*	*15*	*4*					*89*	*124*		*1000*
Yoruba	34		94	19									12	104	109		1000
init.	*38*		*102*	*12*									*5*	*36*	*101*		*1000*
n-in.	*31*		*88*	*26*									*19*	*164*	*116*		*1000*

	ʂ	ʐ	tɕ	dʑ	tɕ^h	ɕ	ɲ	k	g	k^h	x	ɣ	ŋ	h	r	l	Σ
Chinese	73		43		17	18		41		16			126	33	28	47	1000
init.	*99*		*58*		*22*	*24*		*56*		*22*			*13*	*45*	*32*	*63*	*1000*
n-in.	–		–		–	–		–		–			*443*	–	*17*	–	*1000*
Fagauvea			5	4			4	55	123				19	54		131	1000
init.			*1*	*1*			*3*	*34*	*203*				*3*	*69*		*70*	*1000*
n-in.			*10*	*8*			*4*	*79*	*30*				*38*	*35*		*202*	*1000*
Vietnam.	22	16	36				27	139			18	5	141	34	10	39	1000
init.	*34*	*23*	*53*				*41*	*132*			*27*	*7*	*30*	*51*	*16*	*59*	*1000*
n-in.	–	–	–				–	*153*			–	–	*363*	–	–	–	*1000*

	ʂ	tɕ	dʑ	tɕ^h	ɕ	ẏ	k	g	k^h	x	ɣ	ŋ	q	h	r	l	Σ
Burushaski	10	21	16	0	26	0	52	21	10	5	10	10	10	26	67	93	1000

Egyptian frequency data

The following charts present phoneme counts for two diachronic varieties of Egyptian: Middle Egyptian and Bohairic Coptic. I provide not only relative (‰) but also absolute numbers here.

Middle Egyptian. Source: Coffin Texts (CT) volumes 1-7, which Wolfgang Schenkel (Tübingen) generously put at my disposal in electronic form. I remove the 3911 occurrences of *wsjr* "Osiris", the second most frequent noun in the Coffin Texts after *nṯr* "god", since its high frequency is obviously special to this text sort, as well as the 1983 instances of the noun "head" including derivations, whose reading (*tp*? *dp*? *ṯp*? *ḏp*?) is controversial. This leaves a corpus of 328175 text words. The sound values are actually those of late Old Egyptian rather than of Middle Egyptian proper since the sound changes ‹*z*› > ‹*s*›, ‹*ṯ*› > ‹*t*›, ‹*ḏ*› > ‹*d*› are not yet reflected in my data. The reader may compare these numbers with the counts of consonants in roots rather than in running texts from Peust (1999: 295f.).

	‹*p*›	‹*b*›	‹*f*›	‹*m*›	‹*t*›	‹*d*›	‹*s*›	‹*z*›	‹*n*›	‹*ṯ*›	‹*ḏ*›
total	46	30	40	92	116	18	65	13	157	24	28
initial	*78*	*19*	*3*	*147*	*32*	*14*	*84*	*22*	*182*	*19*	*35*
non-init.	*27*	*36*	*62*	*61*	*164*	*21*	*54*	*9*	*143*	*27*	*25*
absolute	32250	20856	28493	65215	81811	12953	46073	9374	110809	16791	20030
abs. in.	*20028*	*4744*	*773*	*37732*	*8139*	*3512*	*21656*	*5531*	*46631*	*4757*	*9006*
abs. n-in.	*12222*	*16112*	*27720*	*27483*	*73672*	*9441*	*24417*	*3843*	*64178*	*12034*	*11024*

	‹*š*›	‹*k*›	‹*g*›	‹*q*›	‹*ḫ*›	‹*ẖ*›	‹*h*›	‹*ḥ*›	‹*ꜥ*›	‹*ꜣ*›	‹*r*›	∑
total	16	35	5	7	35	7	6	49	37	65	108	1000
initial	*27*	*10*	*9*	*8*	*47*	*13*	*12*	*93*	*44*	*22*	*80*	*1000*
n-in.	*9*	*49*	*3*	*6*	*29*	*3*	*2*	*24*	*33*	*89*	*123*	*1000*
abs.	11087	24709	3578	4781	25014	4784	4028	34667	26172	45770	75935	705180
a. in.	*7004*	*2624*	*2378*	*2056*	*11925*	*3284*	*3175*	*23792*	*11404*	*5601*	*20593*	*256345*
a. n.i.	*4083*	*22085*	*1200*	*2725*	*13089*	*1500*	*853*	*10875*	*14768*	*40169*	*55342*	*448835*

Bohairic Coptic. Source: New Testament from http://www.biblical-data.org/coptic/Bohairic_NT.pdf (the 950 occurrences of *iēsous* omitted, which leaves a corpus of 122412 text words). I closely follow the spellings, e.g. in the distinction between p^h and *p*+*h* or in the rendering of ⲅ and ⲇ, without much interpretation. ⲝ and ⲯ are analyzed as ⲕ+ⲥ and ⲡ+ⲥ. Word separation as in the text edition.

	ⲡ *p*	ⲫ p^h	ϥ *f*	ⲃ *β*	ⲙ *m*	ⲧ *t*	ⲑ t^h	ⲇ *d*	ⲥ *s*	ⲍ *dz*	ⲛ *n*
total	67	23	52	30	86	127	22	14	71	1	205
initial	*108*	*37*	*8*	*3*	*129*	*57*	*11*	*37*	*20*	*0*	*364*
non-init.	*54*	*18*	*65*	*39*	*73*	*148*	*25*	*7*	*86*	*2*	*156*
absolute	22197	7539	17185	10019	28725	42285	7292	4602	23559	429	68220
abs. in.	*8475*	*2918*	*588*	*214*	*10097*	*4489*	*896*	*2911*	*1604*	*39*	*28534*
abs. n-in.	*13722*	*4621*	*16597*	*9805*	*18628*	*37796*	*6396*	*1691*	*21955*	*390*	*39686*

	ϫ *č*	ϭ $č^h$	ϣ *š*	ⲕ *k*	ⲭ k^h	ⲅ *g/ŋ*	ϧ *x*	ϩ *h*	ⲣ *r*	ⲗ *l*	∑
total	44	7	32	25	9	8	23	62	66	27	1000
initial	*54*	*2*	*18*	*11*	*3*	*14*	*59*	*57*	*4*	*1*	*1000*
non-init.	*41*	*9*	*36*	*30*	*11*	*6*	*12*	*63*	*85*	*35*	*1000*
absolute	14626	2352	10527	8431	3050	2504	7745	20633	21934	9122	332976
abs. in.	*4205*	*136*	*1420*	*900*	*253*	*1102*	*4633*	*4489*	*341*	*83*	*78327*
abs. n-in.	*10421*	*2216*	*9107*	*7531*	*2797*	*1402*	*3112*	*16144*	*21593*	*9039*	*254649*

I am now proposing a number of generalizations based on the frequency data of the 50 languages sample, discussing for each item how the Egyptian data fit into this picture.

1 The most frequent consonant

n is the most frequent consonant in about half of the languages within my sample and therefore a good candidate for being the world's most frequent consonant phoneme.[4] There are languages in which /*n*/ is almost twice as frequent as the next common consonant (Kabyle, Telugu, Thai). But in the top position we also find *t* (Albanian, Czech, Danish, English, Hungarian, Latin, Maori, Sami) and *r* (Bengali, Breton, French, Ossetic, Persian, Portuguese), more rarely *d* (Ingush), *k* (Swahili), *l* (Maltese, Tok Pisin), *m* (ancient and modern Hebrew, Sunwar), *ŋ* (Vietnamese), *p* (Guaraní), or *s* (Georgian, modern Greek). It needs to be remarked, of course, that not all these rankings can be taken for granted, because some rank orders might still be expected to change if the counts were performed on greater text corpora, particularly in languages where my numbers for the most frequent and the second most frequent phoneme are very close.[5]

Egyptian agrees well with this observation since the most frequent consonant is clearly *n* both in the Coffin Texts and in the Bohairic New Testament. This confirms the accepted reading of 𓈖 as /*n*/,[6] a result which is, however, not likely to arouse much interest, since this sound value belongs to the least controversial among all Egyptian phonemes.

It can be observed that some languages in which /*n*/ is relatively rare either have phonemic nasal vowels (French, Guaraní, Portuguese), or are known to have had them at an earlier period (Czech). The obvious explanation is that the nasal vowels here came about through the loss of an earlier consonantal /*n*/. As to judge from the quantitative point of view, Egyptian does not appear to have experienced such a process during the traceable stages of its history.

2 Labials frequent initially

Labials (sounds such as *p*, *b*, *f*, *v*, *m*) tend to be more frequent in word initial position than in non-initial position. This generalization is particularly clear when we do not consider the labials individually but sum them all up into a single category. It then appears that in 44 languages from my sample, the percentage of initial labials exceeds that of non-initial labials. One of the best examples is Basque, where the four labials

4 The globally most frequent phoneme is probably the vowel /*a*/.

5 It would be possible to compute significance figures which would state how probably my identification of the most frequent consonant is correct for each language. I have not attempted to do this here.

6 In hieratic handwriting, this sign is merely a horizontal stroke, thus having the simplest shape among all phonograms. It is reasonable for a writing system to reserve the simplest shapes for the most frequent functions; this strategy was deliberately observed in the design of the Morse alphabet, whose simplest signs are • = *e* and — = *t*, the two most frequent letters of English.

that are used in the language (*p*, *b*, *f*, *m*) account for 28.9% of all consonants in initial position, but only for 5.1% of all consonants in non-initial position. On the other hand, there are only two languages (Guaraní, Kabyle) in which labials taken as a group are rarer in initial than in non-initial position.

When we consider all the labial consonants separately, it is still true for 16 languages that each of their labials is more frequent in initial than in non-initial position (Albanian, Arabic, Armenian, Aymara, Basque, Breton, Chinese, French, German, ancient Greek, Hungarian, Ingush, Korean, Portuguese, Telugu, Yoruba). The other languages display one or few exceptions, often in lesser used consonants. One such language is Maltese where our generalization holds true for *p*, *b*, *f*, and *m*, with only the relatively rare *v* being disfavoured word-initially.

It can occasionally be observed that labials which fail to obey our generalization, i.e. which are particularly frequent in non-initial position, experienced an increase of frequency in this position by sound changes during the more recent history, so that there may have been, so to say, not enough time for the language to return to a typologically more normal state. One example is Modern Greek, where /f/ and (to a lesser degree) /v/ are relatively infrequent initially. The gain of non-initial labial fricatives is here largely due to the Byzantine development of these sounds from the second element of the diphthongs αυ, ευ, and ηυ.

Egyptian does agree with our generalization when the labials are taken as a group. In the Coffin Texts, all the labials taken together account for 24.7% of the consonants in initial position, but only for 18.6% in non-initial position. On the individual level, however, there are striking exceptions for both *f* and *b* (= Coptic *β*), two sounds which appear to be strongly disfavoured word-initially. This leads to the hypothesis that *f* and *b* may have developed or at least gained in frequency as a result of sound changes that affected non-initial elements during a not too remote period in the prehistory of Egyptian. It might, for example, be stipulated that *f* developed from *p* in certain weak positions (intervocally or syllable-finally), similarly to the *begadkefat* rule of Hebrew. One possible trace of *f* being a spirantic alternant of *p* might be the suffix pronoun 3rd pers. sg. masc. *=f*, if this is etymologically related to the masc. demonstrative base *p-* (thus proposed e.g. by Vycichl 1953: 386f.; rejected, in my view not convincingly, by Takács 1999ff., II: 550f.).[7]

3 Sonorants rare initially

Sonorants tend to be rarer in initial position than in non-initial position, which kind of counters the behaviour of the labial group of consonants. This concerns sounds such as *r*, *l*, *n*, *ɲ*, *ŋ*, but not *m*, which is instead better included into the previous generalization about labials. This generalization appears to be particularly valid for *r*, which is rare in word-initial position in almost every language, despite the fact that the symbol *r* serves to cover relatively different kinds of articulation. The only exceptions in my

---

7 Francis Breyer (Berlin) is currently preparing a new assessment of this etymology.

corpus are Chinese and Vietnamese.[8] In languages which possess two varieties of *r*, the generalization holds true either for both of them (Armenian, Basque, Czech, Hausa, Kabyle), or at least for the more frequent of them (Albanian, Portuguese).

There are quite a number of languages, both inside and outside my corpus, in which word initial *r-* is not only rare but absent (some of these, however, tolerating initial *r-* in loan words). Examples are the Agaw languages (Ethiopia), Armenian, Basque, Brazilian Portuguese, Dyirbal (Australia), ancient Greek (with only initial ῥ /hr-/), Greenlandic, Hittite, Hurritic, Manchu, Mongolian, Nama (Khoisan), Nile Nubian, most or all Omotic languages, Tamil, and Turkish.[9]

As for *l*, the same generalization that this sound is infrequent initially is rather obvious but with some more exceptions (again Chinese and Vietnamese, and additionally Aymara, ancient and modern Hebrew, Kabyle, Maltese, Sami, Thai).

The generalization is again rather strong for (plain) *n*, exceptions being known only from Czech, Portuguese, Swahili, Tagalog, Tok Pisin, and Yoruba, for some of which explanations are available (relatively recent dissolution of non-initial *-n* into nasal vowels in Portuguese; complications in the phonological analysis of Yoruba, for which see the comment in the language entry below).

A strong generalization can be made for the velar nasal *ŋ*. Nearly all languages in my sample, and probably the majority of languages of the world, disfavour *ŋ* in word-initial position. Languages are common which possess *ŋ* as a phoneme but completely lack it as a word initial (e.g. English, German, Korean, Manchu, Sami). A large sample presented in Anderson (2005) records 234 languages that have *ŋ* as a phoneme, out of which 88 do not use it at all in initial position. There are only two languages in my sample in which initial *ŋ-* is more frequent than non-initial *ŋ*, namely Sunwar and Wolof. Other such exceptional languages are not abundant but probably exist (good candidates might be worth to be looked for in Australia and Siberia, although exact counts for such languages are not available to me; possibly also Old Nubian if the conventional reading of the letter ⲅ, frequent word-initially, as /ŋ/ is correct).

Looking at Coptic first, the generalization on sonorants holds very well for *r* and *l*, but not for *n*. This has to do with the fact that ⲛ is so frequent as a preposition and attributive morpheme, which show up as initials in my counting. In Egyptian, among the consonants for which a sound value /r/ can be envisaged at all, it is ‹ꜣ› which shows the property of initial rarity most clearly, though not quite as clearly as ⲣ does in Coptic. The effect is also visible, but more weakly, for ‹r›.[10] In sum, I feel con-

8 In both these languages, the sound conventionally represented by the symbol *r* is not a very prototypical *r*.

9 The absence of initial *r-* in several languages has often been noted (e.g. Restle & Vennemann 2001: 1315), less so the nearly universal rarity of initial *r-*.

10 One may ask why the effect of initial rarity is so much stronger for Coptic ⲣ than for its Egyptian etymological predecessor ‹r›. It may be noted that among the five most frequent words with initial *r-* in the Coffin Texts, three (*r* "to", *rḫ* "to know", *rḏi̯* "to give") lost their *r-* for reasons still not well understood, possibly by irregular sound change, whereas two (*rꜥw* "sun", *rn* "name") preserved their *r-* but are nouns typically preceded in Coptic by a proclitic article so that I do not then count /r/ as initial.

fident that the reconstruction of ‹ꜣ› as /r/ – or at least as some kind of liquid –, though still being disputed[11], is supported by these figures.

None of the Egyptian consonant symbols has ever been reconstructed as /ŋ/[12], and the counts actually do not suggest such an interpretation for any of them.

4 *r* more frequent than *l*

r generally tends to be more frequent than *l*.[13] Since, as we saw in the preceding section, there exists a particularly strong restriction against word-initial *r*-, this generalization becomes better when we consider these phonemes in non-initial position only.

The generalization that non-initial *r* is more frequent than non-initial *l* holds true for 33 languages of my sample as against 9 exceptions (Arabic, Czech, Hungarian, Maltese, Sami, Swahili, Tagalog, Tok Pisin, Wolof; also Vietnamese where, however, both sounds are only used as initials). In two languages, there is an *r* but no *l* (Japanese, Maori), whereas Fagauvea has an *l* but no *r*; Korean has one liquid phoneme whose realization alternates between *r* and *l*.

The prediction is clearly confirmed for Coptic where *r* is more than twice as frequent as *l*.[14] In the Coffin Texts, however, ‹r› /ˀl/ is more frequent than ‹ꜣ› /ˀr/ (I am following here the phonological interpretation by Rössler 1971: 311-319 and Schenkel 1990: 34-36, 53) both generally and non-initially. This observation would favour, but does not suffice to prove, a reverse reconstruction of ‹r› as /r/ and of ‹ꜣ› as /l/, as it was supported in particular by C.T. Hodge (e.g. Hodge 1991: 383 and 1997: 375).

5 Stops compared by manner of articulation

It seems to be a very robust rule, with no exception within my sample, that plain voiceless stops are always more frequent than aspirate voiceless stops, if both classes

---

11 Osing (1997) sticks to the traditional interpretation of ‹ꜣ› as /ʔ/. Takács (1999ff., I: 50-78 and 273-275) acknowledges Afro-Asiatic comparisons with /r/, /l/, and /ʔ/, but considers the evidence for a liquid to be stronger.

12 But Kammerzell (2005: 177-180) assumes the existence in earlier Egyptian of a spoken phoneme /ŋ/ for which no elementary grapheme would have existed.
Note also the paradoxical observation by Maddieson (1984: 11) that a phoneme /ŋ/ is more likely to occur in a language with a small phoneme inventory than in a language with a large phoneme inventory.

13 This statement seems to become false when not the frequencies of /r/ vs. /l/ in individual languages are compared, but the existence of these sounds in phoneme inventories is counted: "(...) some 81.4% of languages have one or more lateral segments, whereas 76.0% have one or more *r*-sounds." (Maddieson 1984: 73). Languages with two liquids normally have one lateral and one *r*-sound, less frequently two laterals, very rarely two *r*-sounds (Maddieson 1984: 84). Note that this conflicts with the reconstruction of the liquid system of earlier Egyptian by Loprieno (1995: 15, 31-33).

14 With the exception of the Fayyumic dialect in which most instances of ⲣ are replaced by ⲗ.

contrast in a language (Armenian, Aymara, Bengali, Burushaski, Chinese, Danish, ancient Greek, Gujarati, Korean, Manchu, Sunwar, Thai, Vietnamese).[15]

A weaker generalization can be made concerning the voice distinction: Voiceless stops tend to be more frequent than voiced stops. This is true for all the three major stop positions (labial, dental, velar) in Albanian, Armenian, Bahasa Indonesia, Czech, English (a borderline case), French, ancient and modern Greek, Guaraní, Gujarati, Latin, Portuguese, Sunwar, Tagalog, Telugu, Thai, and Tok Pisin. In a formalized notation (where > means "more frequent than", < "less frequent than"), which I want to introduce at this point, this situation can be expressed as $p > b$, $t > d$, $k > g$.

It has been argued (e.g. Maddieson 2005) that front voiced stops are easier to pronounce than back voiced stops because the exhalation process, which needs to continue during the whole articulation of a voiced stop, in combination with the closure which prevents air from leaving the mouth, leads to an increasing air pressure in the oral cavity which is easier to tolerate as the cavity is large (i.e. the closure is in front). We may therefore expect that exceptions to our generalization will most often concern fronted places of articulation. This is indeed the case, since the most frequent exceptions to our generalization are the following:

$p < b$, $t > d$, $k > g$: Arabic, Bambara, Basque, Bengali, Burushaski, ancient and modern Hebrew, Hungarian, Japanese, Maltese, Oromo, Swahili, Vietnamese, Yoruba; $p < b$, $t < d$, $k > g$: Breton, Dholuo, Hausa, Kabyle, Persian, Turkish, Wolof.

Other types of exceptions occur more rarely, namely $p < b$, $t > d$, $k < g$ only in German, Ossetic, Sami; $p > b$, $t < d$, $k < g$ only in Fagauvea. Finally, two languages (Georgian, Ingush) completely reverse our generalization in that voiced stops are more frequent than voiceless stops in all positions ($p < b$, $t < d$, $k < g$).

One manifestation of these tendencies is the well-known fact that *p* and *g* are typical gap positions in phonemic systems (see e.g. Sherman 1975: 7f.); Arabic is a famous example of a language which has both these gaps.

If a language has ejectives or "emphatics", they are typically rarer than both the voiceless and the voiced stops: Arabic (not for *q*), Aymara, Georgian, Hausa (not for labials), ancient Hebrew (not for *q*), Ingush, Kabyle, Korean, Oromo (not for labials), Ossetic.

Turning now to Egyptian, we observe that the generalization on aspirates holds very much true for Bohairic: The plain stops *p*, *t*, *k* account for 21.9% of all consonants, the corresponding aspirates p^h, t^h, k^h only for 5.4%.

Earlier Egyptian ‹*p*›, ‹*t*› and ‹*k*› are basically the etymological predecessors of the Bohairic aspirates, which lead me to argue (Peust 1999: 84) that they were distinctly aspirate stops already in the pre-Coptic period as against ‹*d*› and ‹*g*› as plain voiceless stops. However, finding that ‹*p*›, ‹*t*› and ‹*k*› clearly exceed ‹*b*›, ‹*d*›, ‹*g*› and ‹*q*› in fre-

15 German ‹*d*› and (more frequent) ‹*t*› would be counterexamples if they were analysed as /t/ vs. $/t^h/$ (see the comment in the language entry below). While German may be a borderline case, the aspiration of ‹*t*› is certainly less prominent than in such languages where the existence of aspirate stops is undisputed.

quency, I now consider this view improbable. The frequency data rather suggest that ‹*p*›, ‹*t*› and ‹*k*› originally were plain voiceless, not distinctively aspirate stops whereas ‹*b*›, ‹*d*›, ‹*g*› and ‹*q*› may have been characterized by some kind of glottal activity, such as voice (but see section 10 below) or ejective articulation.

6 Voiced and voiceless stops

The frequencies of voiced and voiceless stops can also be compared in a slightly different manner than applied in the preceding section. As was stated there, voiced stops are generally easier to pronounce, and tend to be more frequent, at the fronted places of articulation. The quotient of the frequency of *p* and the frequency of *b*, for which I will henceforward use the shorthand notation $^p/_b$, therefore tends to be a relatively low number. It is, for example, 0.4 in Bengali, or even 0 (since *p* is absent) in Arabic. The quotient $^t/_d$ is typically higher, namely 1.4 in Bengali and 2.2 in Arabic. The quotient $^k/_g$ tends to be still higher, namely 3.6 in Bengali and even ∞ (since *g* is absent) in Arabic.

The inequality $^p/_b < {}^t/_d < {}^k/_g$ is valid for about half of the languages in my sample: Albanian, Arabic, Armenian, Bahasa Indonesia, Bambara, Bengali, Breton, Burushaski, Czech, English, Guaraní, Hausa, ancient Hebrew, Japanese, Kabyle, Latin, Maltese, Persian, Sunwar, Swahili, Turkish, Vietnamese, Wolof, and Yoruba. Most of the remaining languages can be rescued when we only require the weaker generalization $^p/_b < {}^k/_g$, namely Basque, Dholuo, French, Georgian, German, modern Greek, Gujarati, modern Hebrew, Hungarian, Ingush, Oromo, Ossetic, Portuguese, Sami, Tagalog, Thai, and Tok Pisin. The inverse inequality $^p/_b > {}^t/_d > {}^k/_g$ is attested only in a single language, Fagauvea.

The Coptic evidence is difficult to judge here, since it is not known to which extent the signs for voiced plosives, ⲇ and ⲅ, whose use is almost exclusively restricted to Greek loan words, might have had a spirantic pronunciation. For ⲃ, a spirantic pronunciation appears to have been predominant.

Assuming the traditionalist sound values for Egyptian, namely ‹*b*› as /*b*/, ‹*d*› as /*d*/, and ‹*g*› as /*g*/, we would reach the inequality 1.5 < 6.3 < 6.9, which conforms to the expected pattern. The same is true for the Rösslerian reconstruction ‹*b*› = /*b*/, ‹ʿ› = /*d*/, ‹*g*› = /*g*/ (Rössler 1971: 277; Schneider 1997; Kammerzell 2005: 198f.), which would result in the figures 1.5 < 3.1 < 6.9. This piece of quantitative evidence therefore provides no clue, unfortunately, to decide between both reconstructions. I will, however, argue in section 10 that both reconstructions must be considered as unlikely.

7 Stops compared by place of articulation

The voiceless stops at the three principal places of articulation (*p*, *t*, *k*) are employed in languages with noticeable differences in frequency. It turns out that a distribution where *t* is the most and *p* the least frequent of these stops, i.e. *t* > *k* > *p*, clearly predominates within my sample,[16] occurring in almost all European, Semitic and Caucasian languages as well as in Aymara, Bahasa Indonesia, Chinese, Fagauvea, Kabyle, Maori, Oromo, Persian, Tagalog, and Yoruba. Human languages seem to have a "dental bias", a tendency for the dental place of articulation to be the most frequently used and also to be the most varied in manners of articulation.[17]

The next most common type is $k > t > p$, found in Bambara, Bengali, Dholuo, Hausa, Japanese, Korean, Swahili, Telugu, Thai, Turkish, Vietnamese, and Wolof. Other types are rare: $k > p > t$ in Gujarati and Sunwar; $p > k > t$ in Guaraní and Manchu, $t > p > k$ in ancient Greek.

It is interesting to discover that Coptic is a strong representative of the rare type $t > p > k$, otherwise attested only in ancient Greek. The same distribution is already found in Egyptian: ‹*t*› > ‹*p*› > ‹*k*›. One might argue that this disfavours the interpretation of Egyptian ‹*p*›, ‹*t*›, ‹*k*› as plain voiceless stops. If we were, however, rather to assume ‹*b*›, ‹*d*›, ‹*g*› as the plain voiceless stops of earlier Egyptian, we would get the even worse distribution ‹*b*› > ‹*d*› > ‹*g*› (i.e. $p > t > k$) which has no parallel at all within the languages of my corpus (but cf. Tok Pisin which has $p = t > k$).

Only rather inconclusive results can be gained from a similar examination of the voiced stops. The most common distribution is here $d > b > g$[18] (Albanian, Bahasa Indonesia, Basque, Czech, English, French, Hausa, modern Hebrew, Kabyle, Latin, Maltese, Persian, Portuguese, Thai, Turkish, Vietnamese, Wolof). Compared with the dominating type $t > k > p$ in the voiceless stops, the labial and the velar stops change their places, doubtlessly because of the easier articulation of the voiced labials which was discussed above in section 5. But a distribution $d > g > b$ is also well-attested (Armenian, Breton, Dholuo, Georgian, German, ancient Greek, Gujarati, Ingush, Japanese, Ossetic, Sami, Telugu). A third still relatively common distribution is $b > d > g$ (Arabic, Bambara, Bengali, Burushaski, Guaraní, ancient Hebrew, Sunwar, Yoruba).

If we interpreted Egyptian ‹*b*›, ‹*d*›, ‹*g*› as voiced stops, we would get this latter inequality: ‹*b*› > ‹*d*› > ‹*g*›. Under Rössler's hypothesis, we would get ‹ʿ› (=/*d*/) > ‹*b*› > ‹*g*›, which is better but not enough so to allow for any firm conclusion.

8 Affricates

I did not deal with affricates such as *ts*, *dz*, *č*, *ʤ* in the preceding sections although they may be liable for classification as stops in many languages for reasons of symmetry of the phoneme system. In fact, their frequency behaviour is not that of stops, at least not in all languages. We find languages in which, although voiced stops are rarer than the corresponding voiceless stops in all places of articulation, *ʤ* is nevertheless more frequent than *č* (Armenian, Bahasa Indonesia, Gujarati, Tok Pisin), languages

16 Berger (1987: 13), who performed counts on a larger sample of 323 languages, but on smaller text corpora than used here and with a different mode of counting, found *k* to be in the average slightly more frequent than *t*, followed by *p* at some distance.

17 Cf. Melikischwili (1970: 72): "(...) wir kennen keine Sprache, in welcher in der labialen oder der gutturalen Reihe mehr Phoneme vorhanden wären als in der dentalen Reihe."

18 Confirmed by Berger (1987: 13).

which have *ʤ* but no *č* at all (Arabic, Guaraní, Yoruba), or which have *dz* but no *ts* (ancient Greek), languages in which an ejective *tṣ* is more frequent than *dz* or even *ts* (Georgian, ancient Hebrew, Ingush), languages in which an aspirate ts^h or $\check{c}^h$ is more frequent than plain *ts* or *č* (Armenian, Gujarati).

We observe that in the Coffin Texts as well, ‹*ḏ*› occurs more frequently than ‹*ṯ*›, which runs counter to the behaviour of the "real" stops of the corresponding manners of articulation. The frequency of ‹*ḏ*› does not therefore prevent us from assuming that it was the articulatorily "marked" member of the opposition ‹*ḏ*› vs. ‹*ṯ*›, possibly a voiced or (better) ejective sound.

9 Voiceless stops in initial versus non-initial position

In accordance with current conceptions about optimal syllable structure (syllable onsets are occupied preferably by phonemes with a great consonantic strength, see Restle & Vennemann 2001: 1314-1316), it might be expected that plain voiceless stops (*p*, *t*, *k*) are more frequent word-initially than non-initially. We can symbolize this expectation as $p_i > p_{ni}$, $t_i > t_{ni}$, $k_i > k_{ni}$ (*p* initially more frequent than *p* non-initially, etc.). This is indeed found to be true in some languages (Chinese, Dholuo, ancient and modern Greek, Gujarati, Korean, Manchu, Telugu). The reverse system $p_i < p_{ni}$, $t_i < t_{ni}$, $k_i < k_{ni}$ is also possible, but encountered less commonly (Georgian, Sami, Vietnamese).

However, the by far best attested system within my sample is one in which *p* and *k* are indeed frequent in word-initial position, whereas there is a conspicious exception for the dentals, with *t* being particularly frequent in non-initial position: $p_i > p_{ni}$, $t_i < t_{ni}$, $k_i > k_{ni}$. This distribution is attested in the following 18 languages: Albanian, Armenian, Bahasa Indonesia, Breton, English, French, German, Guaraní, modern Hebrew, Hungarian, Ingush, Latin, Maltese, Oromo, Ossetic, Portuguese, Tagalog, and Turkish. In Portuguese, for example, *p* and *k* are approximately four times as frequent in initial as in non-initial position, whereas initial *t* is only about half as frequent as non-initial *t*. I do not know of an explanation for this striking assymetry. A diachronic explanation could be envisaged for modern Hebrew, where earlier *p* and *k* often underwent a spirantization (to *f*, *x*) when not at the beginning of a word, so that their frequency was lowered in this position, whereas the parallel spirantization of *t* to *þ*/*s* did actually occur in (Ashkenazi) varieties of Hebrew but failed to get its way into the dominating dialect of the language. I cannot assess whether the remaining 17 languages experienced similar problems with the sound *þ* at some point in their prehistory, or whether other reasons may have been at work here.[19]

The only other relatively current systems are variations on that one just discussed, namely $p_i < p_{ni}$, $t_i < t_{ni}$, $k_i > k_{ni}$ (ancient Hebrew, Swahili, Tok Pisin, Wolof) and $p_i > p_{ni}$, $t_i < t_{ni}$, $k_i < k_{ni}$ (Aymara, Basque, Czech, Fagauvea, Sunwar, Thai).

---

19 It is assumed that the shift *t* > *þ* once affected Hebrew in general but the instable sound *þ* was subsequently restituted as *t* in some varieties. Further examples of languages which experienced a sound shift *þ* > *t* while retaining voiceless fricatives in the other positions are cited by Kümmel (2007: 147-149).

This system $p_i > p_{ni}$, $t_i < t_{ni}$, $k_i < k_{ni}$ is also that of Egyptian (both Coffin Texts and Bohairic). We may conclude that the interpretation of the Egyptian transcription symbols ‹*p*›, ‹*t*›, ‹*k*› as plain voiceless stops, as argued for in section 5, may not be strongly confirmed, but appears well possible.

10 Voiced stops in initial versus non-initial position

Let us now consider the voiced stops in the same manner as we did for the voiceless stops in the preceding section. Contrary to what at least I expected, voiced stops, despite their resulting from intervocalic lenition in several languages, tend to be particularly frequent in word-initial position just like the voiceless stops. The distribution $b_i > b_{ni}$, $d_i > d_{ni}$, $g_i > g_{ni}$ is even more prevalent than the corresponding distribution for the voiceless stops: It is found in Armenian, Basque, Breton, Fagauvea, German, ancient Greek, Hausa, Maltese, Oromo, Sami, Sunwar, Tok Pisin, and Turkish. The typical exception position is encountered here not in the dentals, but in the velars: A system $b_i > b_{ni}$, $d_i > d_{ni}$, $g_i < g_{ni}$, with *g* being conspiciously rare initially, is found in Albanian, Bahasa Indonesia, Bambara, French, Guaraní, Gujarati, Ingush, Portuguese, and Wolof. Again, I cannot offer any explanation. The next most common system is one in which *d* joins *g* in being exceptional ($b_i > b_{ni}$, $d_i < d_{ni}$, $g_i < g_{ni}$), found in modern Greek, Hungarian, Ossetic, Telugu, and Yoruba.

The Egyptian consonants ‹*b*›, ‹*d*›, ‹*g*›, which seem to represent some kind of stops other than plain voiceless, have been described as voiced stops in traditional philology, though this is based on little evidence (Peust 1999: 80f.). We can observe that ‹*g*› is the only one of them to be particularly frequent initially: What we find is the distribution $‹b›_i < ‹b›_{ni}$, $‹d›_i < ‹d›_{ni}$, $‹g›_i > ‹g›_{ni}$. This is precisely the opposite of what I have just described as the typical behaviour of voiced stops, and it is in fact not attested for voiced stops in any other language of my sample.

A slightly different formulation of the same facts, along the lines of section 6 above, is the quotient notation ${}^{b_i}/_{b_{ni}} > {}^{d_i}/_{d_{ni}} > {}^{g_i}/_{g_{ni}}$, which states that voiced stops tend to prefer the word-initial position as their place of articulation is more fronted. This inequality is fulfilled in Albanian, Bahasa Indonesia, Basque, Guaraní, Gujarati, Hungarian, Ingush, Ossetic, Swahili, Tok Pisin, Wolof, and Yoruba. The majority of the remaining languages agree at least with the weaker formulation ${}^{b_i}/_{b_{ni}} > {}^{g_i}/_{g_{ni}}$: Bambara, Breton, English, French, German, ancient and modern Greek, ancient and modern Hebrew, Kabyle, Maltese, Portuguese, Sami, Tagalog, Telugu, and Turkish. The strictly inverted formula ${}^{b_i}/_{b_{ni}} < {}^{d_i}/_{d_{ni}} < {}^{g_i}/_{g_{ni}}$ is only met in one single language, Sunwar (the increase is not extreme, though: 2.0 < 2.7 < 3.5, and it should be kept in mind that my text corpus of this language is rather small).

When we apply the quotient representation to the Egyptian phonemes $‹b›_i$ / $‹b›_{ni}$, $‹d›_i$ / $‹d›_{ni}$, and $‹g›_i$ / $‹g›_{ni}$, the resulting figures are 0.5 < 0.7 < 3.0, of the Sunwar type which I have described as highly uncommon. Even if we replace ‹*d*› by ‹ʿ› along Rösslerian lines, the numbers do not become any better: 0.5 < 1.3 < 3.0. All this leads to the conclusion that neither Egyptian ‹*b*›, ‹*d*›, ‹*g*› nor ‹*b*›, ‹ʿ›, ‹*g*› can plausibly have been series of voiced stops.

My data are too meagre to make a reasonably well-founded generalization concerning "emphatic" plosives, but it may be noted that the Egyptian distribution is at least similar to that of the "emphatic" (ejective, in Hausa also implosive) stops of Hausa and Oromo, which tend to be rare initially for the labials and frequent initially for the velars. A reconstruction in which Egyptian ‹*b*›, ‹*d*›, ‹*g*› would have formed a series of (in the widest sense) "emphatic" stops therefore appears to be a good possibility. This is not the current reconstruction even by Rösslerians, who assume an emphatic character only for ‹*d*› but maintain the traditional interpretation of ‹*b*› and ‹*g*› as voiced stops.[20]

Since, on the other hand, a voiced interpretation of ‹*b*› remains attractive because of its outcome /*β*/ in Coptic, one could alternatively envisage a system of only two voiced stops for earlier Egyptian, such as ‹*b*› = /*b*/, ‹*d*› = /*d*/, ‹*g*› = /*k*'/[21] or ‹*b*› = /*b*/, ‹ʿ› = /*d*/, ‹*d*› = /*ṭ*/, ‹*g*› = /*k*/.[22]

11 *s* and *z*

/*s*/ is more frequent than /*z*/. This is true for all languages in my sample which at all possess these sounds with the exception only of Hungarian, Persian and Vietnamese. In quite a number of languages, there is only /*s*/ but no /*z*/ (Aymara, Basque, Bengali, Chinese, Danish, Dholuo, Fagauvea, ancient Greek, Guaraní, Japanese, Korean, Latin, Manchu, Sami, Sunwar, Tagalog, Telugu, Thai, Wolof, Yoruba). A reverse example of a language with /*z*/ but no /*s*/ is hardly to be found.[23]

We can further observe that several languages exist in which *s* is particularly frequent in initial position, whereas *z* is particularly frequent in non-initial position (Bambara, Breton, English, French, Hungarian, Ingush, Kabyle, Portuguese, Turkish).

---

20 Schenkel (2002: 31f.) reconstructs ‹*g*›  not as a voiced stop but as a voiceless stop at another place of articulation: [*kʲ*].

21 It is unlikely but not impossible for a language to have only one ejective stop, namely a velar, although a system with ejectives at both the velar and the dental places of articulation is more normal: "A tendency to prefer velar place for ejective stops can be seen in the fact that presence of /*t'/ implies the presence of /*k*'/; it is significant that the 5 languages which have only one place of articulation for their ejective stops all have velars. What is far more salient is that both velar and dental/alveolar places are preferred to bilabial. There are significantly fewer occurrences of /*p*'/ than of either /*k*'/ or /*t'/." (Maddieson 1984: 102f.).

22 Such a system, which implies three manners of stop articulation for Egyptian, conforms to what should be expected from the Afroasiatic perspective, and also to the generalization suggested by Maddieson (1984: 39): "If a language has three stop series it is most likely to have two series with contrasting voice onset time and one 'glottalic' series".
A reconstruction such as ‹*b*› = /*b*/, ‹*d*› = /*ṭ*/, ‹*g*› = /*k*/, ‹ʿ› ≠ /*d*/, in which /*b*/ would be the only voiced stop of the language, would be very exotic, although a few languages with that feature have been reported (Sherman [1975: 4] cites Siriono [South America] and Tzeltal [North America] as examples).

23 Nartey (1979: 8) says "The presence of a voiced primary fricative in a language is highly likely to imply the presence of its voiceless equivalent". His large data sample contains a single language, Chukchi, which appears to have /*z*/ but no /*s*/.

A plausible explanation would be that *-z-* here came about by an intervocalic or post-vocalic lenition of *-s-*, which we actually know was the case in at least some of the languages cited (Breton, English, French, Hungarian, Portuguese).

There are other languages in which this trend is not present or even slightly reversed. In Modern Greek, for example, *s* is noticeably disfavoured in initial position, but *z* is only very slightly so. This makes it improbable here that *-z-* could have arisen from intervocalic *-s-*. In this case we actually know that *z* does not derive from *-s-* but rather from an affricate source (Indo-European **di̯* and sim., probably via *dz* which I assume as the standard sound value of ancient Greek ζ). Similar cases are Czech and Ossetic, whose *z* typically goes back to an Indo-European palatalized **ǵ* or **ǵʰ*, and modern Hebrew, where as well the origin of *z* is assumed in an affricate rather than in *-s-*. Other languages whose *z* is not particularly rare in initial position, and for which we might therefore hypothesize an affricate origin as well, include Georgian and Hausa. A strange and exceptional case is German, where, at least in the standard language, any inherited initial **s-* was shifted either to /z/ or to /š/, which makes /s/ practically absent in this position.

These deliberations could be of interest with respect to the Egyptian phoneme —∞—, which some have interpreted as /z/ and which is commonly transliterated as ‹*z*›. First, ‹*z*› is rarer than ‹*s*›, which would be compatible with a reading /z/. However, ‹*z*› is clearly more frequent initially than elsewhere. The same effect is found, but to a weaker degree, with ‹*s*›. This may suggest that even if ‹*z*› once really was a /z/, this /z/ would only have represented a transitory state in the development from an earlier stronger, affricated or palatalized sound towards *s* as which it finally ends up in Coptic. Since there is no hard evidence for a pronunciation /z/ anyway (Peust 1999: 125f.), I would prefer other reconstructions such as that of Kammerzell (2005: 194) who assumes ‹*z*› as /ts/ or /θ/.

12 Voiced fricatives

Let us consider the voiced fricatives at the labial (*v*), dental (*z*/*ð*), and velar (*γ*) places of articulation. It can generally be said that among these sounds, the labial and the dental are the most common ones and the velar the rarest one. This is certainly comparable to a similar asymmetry in the voiced stops as discussed in section 5.

It is true without exception within my corpus that a language that possesses any voiced fricatives will have at least one front voiced fricative (*v* or *z*) among them. There are a few languages which possess only *v* (Fagauvea, Telugu; marginally, with *v* only used in loans, Japanese and Sunwar; here further belong those Southern German dialects where no /z/ is spoken and, last but not least, Coptic if we consider its ⲃ to be more or less equivalent to a *v*). Many languages possess *v* and *z*, but no *γ* (the English type). A few languages possess *z* and *γ*, lacking *v* (Arabic, Burushaski, Kabyle). There is no language in my sample which would possess only *γ*, or only *z*, or only *v* and *γ*.

It is a related statement to say that /v/ is more frequent than /γ/. This is true for Albanian, Armenian, Breton, Czech, Danish, English, Fagauvea, French, Georgian, German, modern Greek, Guaraní, Gujarati, Ingush, modern Hebrew, Hungarian, Maltese, Persian, Portuguese, Sami, Swahili, Telugu, Tok Pisin, Turkish, and Vietnamese. This list is joined by several languages into which */v/* has crept in as a marginal phoneme in loan words whereas */γ/* has not: Bahasa Indonesia, Bambara, Japanese, Oromo, Sunwar.

Languages in which */γ/* is more common than */v/*, or which have */γ/* but no */v/* at all, are much rarer: Arabic, Burushaski, Kabyle, Ossetic. All these languages possess a voiced dental fricative */z/* in addition.

These generalizations can be pertinent with respect to Egyptian. As far as I am aware, no one has ever reconstructed a phoneme */v/* for Egyptian. We would therefore not expect the existence of any voiced velar fricative. This is even more true as we saw in the preceding section that Egyptian probably also lacked a voiced dental fricative */z/*. The Rösslerian reconstruction, which interprets ‹*ḫ*› as */γ/* (Rössler 1971: 296f.) and posits this as the sole Egyptian voiced fricative, is therefore likely to be flawed.[24]

In order to avoid this typological anomaly, it should be recommended to reconstruct none of the Egyptian velar fricatives, or what is believed to have been velar fricatives, as a voiced sound, but rather to look for some other kind of difference between ‹*ḫ*›, ‹*ẖ*› and ‹*š*›. Another way out could be to keep the reconstruction of ‹*ḫ*› as */γ/*, and to posit one of the remaining Egyptian consonants (perhaps ‹ʿ›?) as */z/*, which would lead to a slightly better voiced fricative system of the Arabic type.

13 *š* and *s*

Let us have a look at the numerical relations of */š/* and */s/*. Most languages appear to employ *s* more frequently than *š*: 28 languages in my sample, joined by 13 more languages which have *s* but no *š* at all, as against a handful of counterexamples (ancient Hebrew [but with uncertainty in the phonetic reconstruction], modern Hebrew, Hungarian, Ingush, Persian, Portuguese, and also Chinese and Vietnamese if their retroflex sibilant *ṣ* were taken as a manifestation of *š*). The predominance of *s* over *š* has already been remarked by Berger (1987: 15), who, however, says that the distribution is reversed in Northern America, a statement which I cannot confirm because I have no data for such languages.

Second, there is a slight tendency for *s* to be relatively preferred in word-initial position and for *š* to be preferred in non-initial position. A perfect illustration of this is Portuguese where an older (Latin) *s* was preserved at the beginning of a word but changed to *š* at the end of a word. In this formulation, the tendency has too many counterexamples to allow for any reasonable prediction, but it can at least be said that languages in which initial *š* is much more (more than double as) frequent than non-initial *š*, i.e. in which $^{\check{s}i}/_{\check{s}ni} > 2.0$, are rare (only Arabic, Georgian, Hausa, Manchu).

Coptic fits neatly into the picture since *š* is clearly rarer than *s* and the quotient $^{\check{s}i}/_{\check{s}ni}$ is as low as 0.5. As for Egyptian, ‹*š*› is rarer than ‹*s*›, too, but ‹*š*› predominates in word-

24 Languages with */γ/* as their only voiced fricative are very rare but not impossible. Four such languages can be culled from the lists in Nartey (1979: 49-54): Irish, Maung, Tiwi, Tolowa.

initial position, the quotient $^{\langle \check{s} \rangle i}/_{\langle \check{s} \rangle ni}$ being 3.0. To be sure, this evidence is far from being decisive, but I take it as an indication that the traditional interpretation of Egyptian ‹š› as /š/ is incorrect, and that ‹š› is rather to be reconstructed as some kind of originally velar fricative along Rösslerian lines (Rössler 1971: 303; Kammerzell 2005: 182-187).

14 *č* and *š*

A similar reasoning can finally be made by considering those languages which possess both the phonemes /č/ and /š/. It is relatively common here for *č* to be more typical in initial position and for *š* to be more typical in non-initial position, which in our formulaic notation can be encoded as $\check{c}_i > \check{c}_{ni}$, $\check{s}_i < \check{s}_{ni}$. This is true for Gujarati, Hungarian, Ossetic, Sami, Swahili, Telugu, Tok Pisin, and Turkish. The reverse situation $\check{c}_i < \check{c}_{ni}$, $\check{s}_i > \check{s}_{ni}$ is only attested from two languages, namely English and Oromo. Furthermore, there are several languages which are neutral with respect to our generalization, showing either $\check{c}_i > \check{c}_{ni}$, $\check{s}_i > \check{s}_{ni}$ (Albanian, Armenian, Bambara, Georgian, Hausa, Manchu) or $\check{c}_i < \check{c}_{ni}$, $\check{s}_i < \check{s}_{ni}$ (Czech, Ingush, Kabyle, Maltese).

The same facts can again be casted into a quotient notation. The formula $^{\check{c}i}/_{\check{c}ni} > {}^{\check{s}i}/_{\check{s}ni}$ holds true for Albanian, Armenian, Bahasa Indonesia, Bambara, Basque, Czech, Georgian, Gujarati, Hungarian, Manchu, Ossetic, Sami, Swahili, Telugu, Tok Pisin, and Turkish (for example, the values are 4.3 > 0.3 for Hungarian, or 3.1 > 0.1 for Telugu); counterexamples are English, Hausa, Ingush, Kabyle, Maltese, and Oromo.

Coptic agrees well with this generalization since *č* is more frequent initially than non-initially, whereas *š* is clearly disfavoured in initial position. If we interpret Earlier Egyptian ‹ṯ› as /č/, which seems fairly certain, and ‹š› as /š/, we get the reverse of our generalization, though: $\langle \underline{t} \rangle_i < \langle \underline{t} \rangle_{ni}$, $\langle \check{s} \rangle_i > \langle \check{s} \rangle_{ni}$, or $0.7 < 3.0$ in the quotient notation. It should be noted that this is quite a gross violation of the expected inequality $^{\check{c}i}/_{\check{c}ni} > {}^{\check{s}i}/_{\check{s}ni}$, stronger than in all the "exceptional" languages mentioned above with the single exception of Oromo where it is even $0.05 < 1.75$ (initial *č* is practically absent in Oromo).

Although the evidence is again not absolutely conclusive due to the existence of counterexamples, it again casts some doubt on the traditional interpretation of Egyptian ‹š› as /š/, at least if this should be taken as its original or inherited sound value.

I believe that certain sound changes of Egyptian – and of other languages – were caused or encouraged by the need to keep the language in agreement with typological universals. We may state, for example, that not long after the originally velar fricative ‹š› was palatalized so that it indeed became /š/, a process dated to ca. 2600 BC by Kammerzell (2005: 195), which may have lead to a temporary clash with the generalizations suggested in the two preceding sections, the well-known late Old Egyptian shift of numerous, particularly of non-initial instances of ‹ṯ› /č/ to ‹t› /t/ (Peust 1999: 123-125) took place, which was one of the factors that helped restitute the typologically normal state of affairs again by the time of Coptic. To view laws of quantitative typology as motivators of language change would be another fascinating line of research, which I do not want to take up here.

I intend to continue this article with a study of Meroitic phoneme frequencies in volume 18 of this journal.

Language details and data sources

Albanian: Articles from the newspaper Ballkan, http://www.ballkan.com/ (ca. 50000 text words).

Arabic: Qur'an Sura's 1–20 from http://www.quranexplorer.com/quran/ (ca. 40000 text words). The assimilation of the article *al-* before sun letters is considered (i.e. *l* is not counted), but not so further assimilations across word boundaries, which are possible in Qur'an recitation. Nunation is, of course, counted. The velarized *ɫ* only appears in certain forms of the word *'allah* but is nevertheless not very rare, which is also due to the genre of this specific text.

Armenian: Four Gospels in Eastern Armenian: http://www.armenianchurchlibrary.com/files/easternarmenianbible/ (ca. 60000 text words; *yisous* and case forms omitted).

Aymara (Bolivia): Various literary texts from http://www.aymara.ucb.edu.bo/ (ca. 70000 text words). I represent orthographical ‹j› as *x* and ‹x› as *χ*.

Bahasa Indonesia: Various texts from the Wikipedia (http://id.wikipedia.org/wiki/) (ca. 30000 text words).[25]

Bambara (Mali): Various literary texts from http://www.bamanan.org/ (ca. 40000 text words).

Basque: Narrations by Iñigo Aranbarri and Eider Rodriguez from http://www.susa-literatura.com/ (ca. 50000 text words). I represent written *j*, whose pronunciation may vary considerably (/i̯/ ~ /ʒ/ ~ /x/), as *ʒ*.

Bengali: Phoneme counts from Ferguson & Chowdhury (1960: 50).

Breton: Articles from the journal Bremañ, http://www.breman.org/ (ca. 50000 text words). For *sh* and *zh*, whose realization differs according to the dialect, I assume a pronunciation as *s* and *z*.

Burushaski: Phoneme counts from Berger (1998: 28). Since the data base is small (8855 consonants), Berger only gives figures rounded to 0.5%. I adopt his figures but ignore *w* and *y*.

Chinese: Phoneme counts of "modern colloquial Chinese of Peking" from Zipf (1932: 6f.), based on a text corpus of 20000 syllables. The dialect described by Zipf is not fully identical with modern Standard Mandarin since, for example, it uses an

25 Frequency data on Indonesian phonemes based on dictionary (type) counts can be found in Altmann *et al.* (2002: 29).

initial ŋ- (e.g. in *ŋo* "I" = Mandarin *wo*). I transliterate Pinyin *x* as *ç*, *j* as *tç*, *sh* as *ʂ*, etc. The criterium of "initiality" is based here on syllables rather than strictly on words (the distinction between words and word compounds is rather difficult to make for Chinese).

Czech: Phoneme counts from Těšitelová *et al.* (1985: plates 1 and 5).

Danish: Phoneme counts generated from a word frequency list of the DanPASS corpus (Danish Phonetically Annotated Spontaneous Speech, http://www.cphling.dk/ng/danpass_webpage/udtaleordbog_med_frekvens.pdf). These materials represent realistic, sometimes blurred or inexact pronunciation. I have reduced the numerous allophones given to the phonemes normally acknowledged for Danish.

Dholuo (Kenya, Nilotic): Texts from the New Testament from http://www.ibs.org/bibles/luo/index.php. For technical reasons, no whole books but numerous disconnected text snippets had to be analyzed (ca. 50000 text words). For reasons of space, I represent written ‹*th*› and ‹*dh*› as *þ* and *ð* respectively although they are predominantly pronounced as fronted stops. Prenasalized stops are considered as biphonematic.

English: Phoneme counts of spoken American English from Roberts (1965: appendices II, XVII).

Fagauvea (New Caledonia, Polynesian): Narrations from the LACITO corpora (Laboratoire de langues et civilisations à tradition orale), http://lacito.vjf.cnrs.fr/archivage/languages/Fagauvea_fr.htm (ca. 7500 text words). My *tʂ* and *dʐ* represent what is written *tr* and *dr* respectively.

French: Phoneme counts from Malécot (1974: plates 1 and 3).

Georgian (ancient): Four Gospel translations from http://armazi.uni-frankfurt.de/framed.htm (ca. 60000 text words; occurrences of *iesu* + case forms omitted). I am grateful to Jost Gippert (Frankfurt) for his helpful comments.

German: Phoneme counts from Meier (1964: 251 and 267) (stops as in *un<u>d</u>*, *sel<u>b</u>st*, *sagte* being considered as voiceless; syllable final *r* still taken as consonantic; written ‹*z*› is treated as biphonematic /*t*/ + /*s*/). Note that the interpretation of the stop system is somewhat controversial: Jessen (2004) suggests that the distinctive feature of *b*, *d*, *g* in German is not [+voiced], as is normally assumed, but rather [–aspirated].

Greek (ancient): Homer, Ilias from Wikisource (http://el.wikisource.org/wiki/) (ca. 70000 text words). I interpret ῥ as /*hr*/. /*h*/ is only written word initially; I assume that there was no -*h*- in internal position although this may be a wrong assumption for Homer's time.

Greek (modern, Dhēmotikē): Articles from the newspaper Ελευθεροτυπία (ca. 50000 text words), http://www.enet.gr/. I interpret written nasal + tenuis (e.g. ντ) within a word as a single voiced stop (e.g. *d*); assimilations over word boundaries are not accounted for.

Guaraní (Paraguay): Various texts from the Wikipedia (http://gn.wikipedia.org/wiki/) (ca. 20000 text words). Guaraní possesses prenasalized voiced stops (written *mb*, *nd*, *ng*) which I simply interpret as *b*, *d*, *g*. Written *b*/*v*, *d*, *g* (*b*, *d* used only in words of Spanish origin) most typically represent fricatives and are represented here as *v*, *ð*, *γ*. The amount of Spanish elements in the texts is quite high; I have eliminated only a part of them, in particular proper names.

Gujarati (India): Phoneme counts from Pandit (1965: appendices 3i, 5 and 7). Among both alternative analyses given by Pandit (b^h etc. taken as mono- or biphonematic), I prefer the first option, from which the second one can easily be derived by merging the counts.

Hausa: Four Gospel translations from http://visionneuse.free.fr/download.htm (ca. 80000 text words; occurrences of *yesu* omitted). I note ‹*ts*› as *ṣ*; the two *r*-phonemes, not distinguished in the orthography, have been disentangled with the help of dictionaries.

Hebrew (ancient): Torah (ca. 80000 text words) from http://bhcv.hebrewtanakh.com/. In accordance with recent approaches in Semitistics, I transliterate ס as *ts*, ז as *dz*, צ as *tṣ*, שׂ as *ś*, שׁ as *š*. I assume that word-final *-h* was not pronounced (*-h* with mappiq excepted). If it were taken as pronounced, *h* would become the most frequent consonant of the language.

Hebrew (modern): Various literary texts in punctuated script from http://www.benyehuda.org/, converted by me semi-automatically into phonemes in agreement with modern Israeli pronunciation (ca. 40000 text words).

Hungarian: Phoneme counts generated from a word frequency list of the Hungarian Webcorpus of the Média Oktatási és Kutató Központ, http://mokk.bme.hu/resources/webcorpus/index.html.

Ingush (NE-Caucasian): Literary texts from http://www.ingushetia.org/culture/bagahbuvcam/ (ca. 180000 text words).

Japanese: Phoneme counts from Tamaoka & Makioka (2004). I consider /*ts*/ and /*f*/ as independent phonemes, a status which they have only aquired by recent European loans. Palatal consonants are taken as combinations C+*y* rather than as separate phonemes. The syllable final nasal is subsumed under *n*.

Kabyle (Berber): Four Gospel translations from http://visionneuse.free.fr/download.htm (ca. 80000 text words; occurrences of *ʕisa* "Jesus" and *sidna* "our Lord" omitted).

Korean: North Korean political texts from http://ndfsk.dyndns.org/ (ca. 70000 text words). In accordance with North Korean pronunciation, written *r* in word initial position and after nasals is assumed to be spoken as such (whereas it merges with *n* in South Korea).

Latin: Virgil, Aenaeis from Project Gutenberg, http://www.gutenberg.org/files/227/227.txt (ca. 60000 text words); ‹c› and ‹qu› subsumed under *k*.

Maltese: Articles from the newspaper it-Torċa, http://www.it-torca.com/ (ca. 30000 text words).

Manchu: Portions of "Secret Chronicle of the Manchu Dynasty" from http://www.anaku.cn/mbrt/Vol1/mbrt(1).htm ff. (ca. 100000 text words). I consider every *n* preceding (written) *i* as a palatal *ɲ*.

Maori: Articles from the newspaper Te Toa Takitini from the 1920ies, from the New Zealand Digital Library, http://nzdl.sadl.uleth.ca/cgi-bin/library/ (ca. 20000 text words).

Oromo (Cushitic): News bulletins from http://www.oromoliberationfront.org/ (ca. 30000 text words).

Ossetic (Iranian): Fairy tales from http://allingvo.ru/FAIRY%20TALES/ (ca. 30000 text words). I do not distinguish between velars (*k*) and labiovelars (k^w).

Persian: Phoneme counts from Moïnfar (1973: 90). His textual base is Shāhnāme ("Book of Kings") by Firdawsī (ca. 1000 AD).

Portuguese: Phoneme counts of European Portuguese from a statistics by the project Frota, Sónia & Vigário, Marina & Martins, Fernando da Assunção: *The FrePOP (Frequency of Phonological Objects in Portuguese) Database, Laboratório de Fonética da FLUL*, which is to be published on the internet in 2009 and for whose communication I am very grateful to Sónia Frota (for the time being, see http://www.fl.ul.pt/LaboratorioFonetica/FreP/FrePProjectFCT06.pdf).

Sami (Northern Sami, Norway): Four Gospel translations from http://www.bibelen.no/ (ca. 60000 text words; occurrences of *jesus* + case forms omitted).

Sunwar (Nepal, Kiranti-language): Narrative texts which Dörte Borchers (Berlin), to whom I am very grateful, provided to me from her field notes (ca. 2500 text words). See her monograph (Borchers 2008) on the language.

Swahili: Four Gospel translations from http://visionneuse.free.fr/download.htm (ca. 60000 text words; occurrences of *yesu* omitted). *mb* etc. considered as phoneme sequences.

Tagalog (Philippines): Four Gospel translations from http://www.biblegateway.com/versions/ (ca. 90000 text words; occurrences of *jesus* omitted).

Telugu (Dravidian): Novel "13–14–15" by Yandamuri Veerendranath from http://www.bharatadesam.com/literature/telugu_novels/telugu_novels.php (ca. 40000 text words). This text reflects modern language with numerous borrowings from English, which may elevate in particular the ratio of retroflex stops (*t* / *d* of English are borrowed as *ṭ* / *ḍ*). In Telugu, a greater or lesser number of phonemes from words of foreign origin may be retained according to the speech style. I assume a speech style which merges *s* and *ś* and, among the aspirates, retains only the two most frequent ones (b^h and d^h). I have attempted to distinguish *ts* and *dz* from *č* and *dž*, as well as *f* from $p^{(h)}$, which are not differentiated in writing. Anusvara preceding palatals and velars (phonetically *ɲ*, *ŋ*) is considered as *n*. The counts for *v* and *m* include the allophones *w* and *w̃* respectively.

Thai: Various literary texts from Wikisource (http://th.wikisource.org/wiki/) (ca. 150000 text words). The transcription of Thai, which is not entirely straightforward, including word separation was performed with the software package "Thai Romanization" available on a web page of Chulalongkorn University (http://www.arts.chula.ac.th/~ling/tts/).

Tok Pisin (English-based creole): Interviews from the project "Remembering the war in New Guinea" of the "Australia-Japan Research Project" (AJRP), http://ajrp.awm.gov.au/ajrp/remember.nsf/Web-Frames/InterviewFrame/ (ca. 30000 text words). In Tok Pisin, English words may retain their original pronunciation to various degrees, which creates several marginal phonemes not used in core Tok Pisin. I assume that terms represented in the texts in English orthography were spoken in the English way, which may be an idealization.

Turkish: "Küçük Prens (The Little Prince)" from http://www.kucukprens.org/kitap/ (occurrences of *küçük prens* + case forms omitted; ca. 10000 text words). I assume *ğ* as still being spoken as /*γ*/ in non-palatal environments.

Vietnamese: Phoneme counts generated from a word frequency list on the "Corpora of Vietnamese Texts (CVT)" page funded by the University of Minnesota, http://vnspeechtherapy.com/vi/CVT/. I note written ‹c› and ‹q› as *k*, ‹ch› as *tɕ*, ‹d› as *dẓ*, ‹đ› as *d*, ‹g› as *γ*, ‹gh› as *g*, ‹gi› as *z*, ‹kh› as *x*, ‹ng(h)› as *ŋ*, ‹nh› as *ɲ*, ‹ph› as *f*, ‹s› as *ṣ*, ‹th› as t^h, ‹tr› as *tṣ*, ‹x› as *s*. In syllable auslaut, the contrast between velar and palatal (*k* – *tɕ* , *ŋ* – *ɲ*) is neutralized. I transcribe /*k*/ and /*ŋ*/ in this case.

Wolof (Senegal): Four Gospel translations from http://www.jesus.org.uk/bible/Wolof+NT/ (ca. 80000 text words; occurrences of *yeesu* omitted).

Yoruba (Nigeria): Various texts from the internet, preferably ones which make a distinction between the phonemes *s* and *š* in writing, most of them from the Wikipedia (http://yo.wikipedia.org/wiki/) (ca. 30000 text words). I opt for a phonemic analysis close to the surface realization and take *l* and *n* as two phonemes, with *n* also including the syllabic nasal. An alternative analysis would consider *n* as a combinatory variant of *l* before nasal vowels and would have to posit the syllabic nasal as a distinct phoneme. Under that analysis, not *n* but *l* would be the most frequent phoneme of Yoruba.

Bibliography

Altmann, Gabriel, Dariusch Bagheri, Hans Goebl, Reinhard Köhler & Claudia Prün. 2002. *Einführung in die quantitative Lexikologie*, Göttinger Linguistische Abhandlungen 5, Göttingen.

Altmann, Gabriel & Werner Lehfeldt. 1980. *Einführung in die Quantitative Phonologie*, Bochum.

Anderson, Gregory D.S. 2005. The Velar Nasal (*ŋ*), in: Haspelmath *et al.* (2005: chapter 9).

Berger, Hermann. 1987. *Einige Ergebnisse der Phonemstatistik*, Abhandlungen der Heidelberger Akademie der Wissenschaften Phil.-Hist. Kl. 1986/1, Heidelberg.

— 1998. *Die Burushaski-Sprache von Hunza und Nager. Teil I: Grammatik*, Neuindische Studien 13, Wiesbaden.

Borchers, Dörte. 2008. *A Grammar of Sunwar: Descriptive Grammar, Paradigms, Texts, and Glossary*, Languages of the Greater Himalayan Region 5/7, Leiden.

Colless, Brian E. 1992. The Byblos Syllabary and the Proto-Alphabet, in: *Abr-Nahrain* 30, 55-102 (non vidi).

Ferguson, Charles A. & Munier Chowdhury. 1960. The Phonemes of Bengali, in: *Language* 36, 22-59.

Gamkrelidze, Thomas V. 1978. On the Correlation of Stops and Fricatives in a Phonological System, in: Joseph H. Greenberg (ed.), *Universals of Human Language*, Stanford, vol. II, 9-46.

Haspelmath, Martin, Matthew S. Dryer, David Gil & Bernard Comrie (eds.). 2005. *The World Atlas of Language Structures*, Oxford (online at http://wals.info/).

Hodge, Carleton T. 1991. Review of K. Petráček, Altägyptisch, Hamitosemitisch und ihre Beziehungen zu einigen Sprachfamilien in Afrika und Asien, in: *Journal of the Americal Oriental Society* 111, 382-384.

— 1997: Review of A. Loprieno, Ancient Egyptian. A Linguistic Introduction, in: *Diachronica* 14, 373-378.

Jessen, Michael. 2004. Instability in the Production and Perception of Intervocalic Closure Voicing as a Cue to *bdg* vs. *ptk* in German, in: *Folia Linguistica* 38, 27-41.

Kammerzell, Frank. 2005. Old Egyptian and Pre-Old Egyptian. Tracing Linguistic Diversity in Archaic Egypt and the Creation of the Egyptian Language, in: Stephan J. Seidlmayer (ed.), *Texte und Denkmäler des ägyptischen Alten Reiches*, Berlin, 165-247.

Krámský, Jiří. 1959. A Quantitative Typology of Languages, in: *Language and Speech* 2, 72-85.

Kümmel, Martin Joachim. 2007. *Konsonantenwandel*, Wiesbaden.

Lepper, Verena. 2006. Language and Text, in: *Lingua Aegyptia* 14, 375-388.

Loprieno, Antonio. 1995. *Ancient Egyptian. A Linguistic Introduction*, Cambridge.

Maddieson, Ian. 1984. *Patterns of Sounds*, Cambridge.

— 2005. Voicing and Gaps in Plosive Systems, in: Haspelmath *et al.* (2005: chapter 5).

Malécot, André. 1974. Frequency of Occurrence of French Phonemes and Consonant Clusters, in: *Phonetica* 29, 158-170.

Meier, Helmut. 1964. *Deutsche Sprachstatistik*, 2nd ed., Hildesheim.

Melikischwili, I.G. 1970. Einige universale Gesetzmäßigkeiten in dem System der Affrikaten, in: László Dezső & Péter Hajdú (eds.), *Theoretical Problems of Typology and the Northern Eurasian Languages*, Amsterdam, 65-73.

Moïnfar, Djafar M. 1973. *Phonologie quantitative du persan*, Paris.

Nartey, Jonas N.A. 1979. *A Study in Phonemic Universals – especially concerning Fricatives and Stops*, UCLA Working Papers in Linguistics 46, Los Angeles.

Osing, Jürgen. 1997. Zum Lautwert von 𓄿 und 𓂝, in: *Studien zur Altägyptischen Kultur* 24, 223-229.

Pandit, Prabodh B. 1965. *Phonemic and Morphemic Frequencies of the Gujarati Language*, Poona.

Peust, Carsten. 1999. *Egyptian Phonology. An Introduction to the Phonology of a Dead Language*, Göttingen (online at http://www.ub.uni-heidelberg.de/helios/fachinfo/www/aegypt/digilit/inschriften_gramm.html).

Restle, David & Theo Vennemann. 2001. Silbenstruktur, in: Haspelmath, Martin *et al.* (eds.), *Language Typology and Language Universals*, Handbücher zur Sprach- und Kommunikationswissenschaft 20, vol. II, 1310-1336.

Roberts, Aaron Hood. 1965. *A Statistical Linguistic Analysis of American English*, The Hague.

Rössler, Otto. 1971. Das Ägyptische als semitische Sprache, in Franz Altheim & Ruth Stiehl (eds.), *Christentum am Roten Meer*, Berlin, vol. I, 263-326.

Schenkel, Wolfgang. 1990. *Einführung in die altägyptische Sprachwissenschaft*, Darmstadt.
—— 2002. Glottalisierte Verschlußlaute, glottaler Verschlußlaut und ein pharyngaler Reibelaut im Koptischen. Rückschlüsse aus den ägyptisch-koptischen Lehnwörtern und Ortsnamen im Ägyptisch-Arabischen, in: *Lingua Aegyptia* 10, 1-57.

Schneider, Thomas. 1997. Beiträge zur sogenannten „Neueren Komparatistik", in: *Lingua Aegyptia* 5: 189-209.

Sherman, Donald. 1975. Stop and Fricative Systems: A Discussion of Paradigmatic Gaps and the Question of Language Sampling, in: *Working Papers on Language Universals* 17, 1-31.

Sokarno Abdel-Hafiz, Ahmed. 2002. Consonant Frequency in Fadicca Nubian, in: *Abḥāṯ al-Yarmūk* 20, 15-36.

Takács, Gábor. 1999ff. *Etymological Dictionary of Egyptian*, Handbuch der Orientalistik I/48, so far 3 vols., Leiden.

Tamaoka, Katsuo & Shogo Makioka. 2004. Frequency of occurrence for units of phonemes, morae, and syllables appearing in a lexical corpus of a Japanese newspaper, in: *Behaviour Research Methods, Instruments, & Computers* 36, 531-547.

Těšitelová, Marie *et al*. 1985. *Kvantitativní charakteristiky současné češtiny*, Praha.

Vycichl, Werner. 1953. Die ägyptischen Pronominalendungen, in: *Le Muséon* 66, 381-389.

Zipf, George Kingsley. 1932. *Selected Studies of the Principle of Relative Frequency in Language*, Cambridge MA.
—— 1935. *The Psycho-Biology of Language*, Cambridge MA.

Zvelebil, Kamil V. 1972. Initial Plosives in Dravidian, in: *Lingua* 30, 216-226.