

# FREQUENCY COUNTS IN COURSE DEVELOPMENT

R. David Zorc

MRM Language Research Center

African Linguistics IV, GURT

March 12, 1994

There are several types of frequency counts, two of which we will be dealing with today:

1. **Lexical counts** represent the standard form one usually thinks of. This is the product of selecting a given amount of textual material and running it through a process of counting all the words that occurred and how many times each word (form) was used.

2. **Grammatical counts** require breaking a word into its constituent morphemes so that every element of word formation can be counted, e.g., friend/li/ness, kind/er, walk/ing, etc. This is by far the most complex procedure, but is useful in determining the importance of grammatical forms in whatever types of text have been chosen.

The value of a frequency count depends on the material upon which it was based. If, for example, children's stories were used exclusively, it would reflect the lexicon and grammar aimed at or useful to such an audience. If, on the other hand, a broad base of literature and/or transcribed text were included, one would begin to get a reasonable characterization of the language in question. One might safely assume that the most (or at least more) important grammatical and lexical items would be represented.

My enthusiasm for this procedure (despite the great amount of tedium involved in developing it and following it though), should not be construed to imply that there are no problems with it.

For one thing, there will always be **uneven representation**. No matter how the results are ordered (alphabetically or by frequency), items that "belong together" (naturally, semantically or grammatically) will not so occur. Pedagogically students need to learn certain vocabulary together, e.g., days of the week, months of the year, antonyms, etc., but are to be found (if at all) here and there on the list.

Secondly, as a direct result of the materials chosen, there is likely to be a **skewed representation** (i.e., topic- or time-related)

dugsiga

the school

[topic related]

jaalle

comrade

[no longer in use]

Thirdly, there invariably will be **no representation** of certain items (a particular grammatical structure, e.g., negative imperative or optative, or even basic vocabulary, e.g., heart, tooth, fingernail, brother, etc.).

Bearing these caveats in mind, one cannot underestimate the pedagogical importance of a frequency count done right. There was a magnificent presentation at GURT (89?) on the results of a survey of a massive amount of French spoken and written material. When the results were compared with the extant textbooks, there was a serious discrepancy between what authors presented and how the French actually talked and wrote. Some constructions of high text frequency (and therefore importance) were not treated until later chapters (or not at all) in some books, while little used or rare constructions were introduced (sometimes very early). I was at that time in the midst of Somali course development, and had used frequency counts. So I related quite enthusiastically throughout to the authoress's statements, while much of the audience appeared bewildered, defensive or even argumentative.

In the preface to the Somali Textbook (Zorc & Issa, 1990), I wrote:

This textbook was designed on the basis of FREQUENCY COUNTS done on both the lexical and grammatical material contained in the folktales and pedagogical works like the *Somali Newspaper Reader* and *Somali Common Expressions*, supplemented by other articles in or on the language. The ordering of the fifty chapters was generally dictated by the highest frequency forms. For example, classifiers (like *baa* [FOCUS] and *waa* [DECLARATIVE]), the definite articles (-*ka* and -*ta*), or the use of the PAST TENSE all occurred hundreds of times and are crucial to the understanding of Somali texts. These are put very early (in chapters 2 through 8). However, the principle of PROGRAMMED INSTRUCTION also played a decisive role in course design, i.e., building upon easier or earlier material. Thus, IMPERATIVE VERBS are not very common in Somali texts, but they are the basis of finding or learning a VERB ROOT, upon which all the tense are built. Hence, these are treated earlier (in Chapter 7), prior to the introduction of the PAST TENSE (Chapter 8). Conversely, NEGATIVE IMPERATIVES may be easy to learn, but they are taken relatively late (in Chapter 28) because they are not so common. Similarly, verbs of Group One (Chapters 8-9) are far more frequent than those of Group Two (Chapter 11), Group Three (Chapter 15), or Group Four (Chapters 35-38). Although some forms of the verb *to be* are very frequent and important (Chapter 14), other are not so (and are delayed until Chapter 42).

When I started out in 1987 on the development of a Somali curriculum, I felt that we needed a decent sample of text and spoken frequency. Hence The *Somali Newspaper Reader* and *Thirty Folktales* represented text samples; *Somali Common Expressions* and *Somali Dialogs* represented spoken language.

FREQUENCY OF WORDS - SOMALI MATERIALS (D. Zorc) - Page 3

FREQUENCY FROM SOMALI DIALOGS			SOMALI COMMON EXPRESSIONS		
WAA	D=183	01	WAA	E=111	01
OO	D=105	02	MA	E=092	02
MA	D=101	03	WALAAL	E=084	03
KA	D=085	04	U	E=077	04
KU	D=079	05	IYO	E=067	05
IYO	D=078	06	KA	E=066	06
SOO	D=070	07	KU	E=064	07
U	D=069	08	OO	E=060	08
LA	D=063	09	BAAN	E=057	09
HAA	D=056	10	WAXAAN	E=053	10
WAX	D=053	11	INAAN	E=043	11
BAA	D=048	12	LA	E=042	12
EE	D=047	13	AH	E=039	13
AH	D=042	14	AAD	E=038	14
BAAD	D=035	15	SOO	E=038	15
BAANNU	D=034	16	INAAD	E=034	16
INAAD	D=030	17	KARTAA	E=034	17
IMMISA	D=027	18	RABAA	E=031	18
MISE	D=027	19	II	E=030	19
SHILIN	D=026	20	BAAD	E=029	20
WAXAANNU	D=026	21	I	E=028	21
BEY	D=024	22	BEY	E=027	22
WANAAGSAN	D=024	23	MIYAAD	E=027	23
KALE	D=024	24	BAA	E=027	24
INAANNU	D=023	25	WAAN	E=024	25
AAD	D=022	26	MAAN	E=022	26
WEY	D=022	27	SOOMAALI	E=020	27
AF	D=021	28	AF	E=019	28
WAANNU	D=020	29	WAX	E=019	29
LABA	D=020	30	IN	E=019	30
YAHAY	D=020	31	KARAA	E=017	31
BADAN	D=019	32	IGA	E=017	32
SHAN	D=018	33	WANAAGSAN	E=017	33
HAGAAG	D=018	34	MAXAY	E=016	34
TAN	D=016	35	IMMISA	E=015	35
BAAN	D=016	36	BUU	E=015	36
LAGU	D=015	37	TAHAY	E=015	37
AYAA	D=015	38	SHAN	E=015	38
DUGSIGA	D=014	39	NAMBAR	E=015	39
I	D=014	40	MUQDISHO	E=015	40
HADDA	D=014	41	KALE	E=015	41
IDIIN	D=014	42	HALKA	E=014	42
TAHAY	D=014	43	MAR	E=014	43
WAXAAN	D=014	44	HAA	E=014	44
AYEY	D=013	45	LABA	E=014	45
IDIN	D=013	46	RAALLI	E=014	46
NA	D=013	47	SOOMAALIYA	E=014	47
KAN	D=013	48	KAN	E=013	48
LAAKIIN	D=013	49	SII	E=013	49
MAREYKAN	D=012	50	AHOW	E=013	50
WAAN	D=011		UGU	E=013	
TAGNO	D=011		MAANTA	E=013	
MAXAAD	D=011		RABTAA	E=012	
MARKA	D=011		HAL	E=012	
TOBAN	D=011		INTEE	E=011	
AAN	D=011		SHILIN	E=011	
LAGA	D=011		MID	E=011	
MAYA	D=011		MAYA	E=011	

<u>SOMALI NEWSPAPER READER</u>			<u>FOLKTALE BY STORY FREQ</u>		
KA	52	R-01	BAA	30	F-01
OO	51	R-02	KU	29	F-02
KU	49	R-03	KA	29	F-03
EE	47	R-04	IYO	28	F-04
IYO	45	R-05	YIRI	28	F-05
AH	44	R-06	U	28	F-06
AY	44	R-07	OO	28	F-07
U	44	R-08	BUU	26	F-08
IN	43	R-09	LA	25	F-09
UU	40	R-10	WAA	24	F-10
LA	37	R-11	SOO	24	F-11
SOO	36	R-12	BEY	19	F-12
AYAA	34	R-13	ISAGOO	17	F-13
WUXUU	26	R-14	NIN	17	F-14
WAXA	25	R-15	UGU	17	F-15
WAXAY	25	R-16	MA	16	F-16
KALE	24	R-17	INUU	16	F-17
WAXAA	24	R-18	EE	15	F-18
DALKA	22	R-19	WUU	15	F-19
LAGU	19	R-20	YAHAY	14	F-20
LOO	19	R-21	WAX	14	F-21
UGU	18	R-22	DABADEEDNA	14	F-22
MID	17	R-23	AYAA	13	F-23
MAGAALADA	16	R-24	WAXAA	13	F-24
HORE	16	R-25	AY	13	F-25
LAGA	16	R-26	AH	12	F-26
AAD	16	R-27	DIB	12	F-27
INUU	16	R-28	IS	12	F-28
UGA	15	R-29	LAHAA	12	F-29
WAX	15	R-30	MARKUU	12	F-30
SI	15	R-31	MID	12	F-31
KALA	15	R-32	WADA	12	F-32
AFRIKA	14	R-33	BAAN	11	F-33
SIDII	14	R-34	BERI	11	F-34
SIDA	14	R-35	BUUNA	11	F-35
DAWLADDA	13	R-36	IN	11	F-36
INEY	13	R-37	ISKU	11	F-37
MADAXWEYNAHA	13	R-38	LIBAAX	11	F-38
DADKA	12	R-39	NINKII	11	F-39
SHEEGAY	12	R-40	UGA	11	F-40
YIRI	12	R-41	WUXUU	11	F-41
URURKA	12	R-42	AAN	10	F-42
MARKII	12	R-43	BAANA	10	F-43
MAREYKANKA	11	R-44	BADAN	10	F-44
NIN	11	R-45	DABADEED	10	F-45
WADA	10	R-46	INTA	10	F-46
MEEL	10	R-47	SI	10	F-47
LAHAA	10	R-48	YIMID	10	F-48
JAALLE	10	R-49			
SHEEGEY	10	R-50			
HASE	10	R-51			
WEYN	10	R-52			

FREQUENCY OF WORDS - SOMALI MATERIALS (D. Zorc)

5

WORD		READER		FOLKTALES		COM. EXPR		DIALOGS	
AAD	very	R- 16	27	F- 05		E=038	14	D=022	26
AAN	not	R- 04		F- 10	42			D=011	
AF	language	R- 01		F- 03		E=019	28	D=021	28
AFRIKA	Africa	R- 14	33	F-NIL					
AH	being [rel]	R- 44	06	F- 12	26	E=039	13	D=042	14
AHOW	be!	R-NIL		F-NIL		E=013	50		
AY	she / they	R- 44	07	F- 13	25			D=006	
AYAA	[FOCUS]	R- 34	13	F- 13	23			D=015	38
AYEY	they [FOC]	R- 03		F-NIL				D=013	45
BAA	[FOCUS]	R- 06		F- 30	01	E=027	24	D=048	12
BAAD	you [FOC]	R-NIL		F- 06		E=029	20	D=035	15
BAAN	I [FOC]	R- 01		F- 11	33	E=057	09	D=016	36
BAANA	[FOC] and	R-NIL		F- 10	43				
BAANNU	we [FOC]	R- 01		F- 01				D=034	16
BADAN	much; very	R- 07		F- 10	44			D=019	32
BERI	day	R- 02		F- 11	34				
BEY	they [FOC]	R- 09		F- 19	12	E=027	22	D=024	22
BUU	he [FOC]	R- 07		F- 26	08	E=015	36		
BUUNA	and he...	R-NIL		F- 11	35				
DABADEED	after	R-NIL		F- 10	45				
DABADEEDNA	and then	R-NIL		F- 14	22				
DAD (KA)	people	R- 12	39						
DAL (KA)	country	R- 22	19						
DAWLAD	government	R- 13	36						
DIB	back	R- 07		F- 12	27				
DOON-	want	R- 20		F- 14		E=010		D=010	
DUGSI	school	R- 01		F-NIL				D=014	39
EE	and [rel]	R- 47	04	F- 15	18			D=047	13
HAA	yes	R-NIL		F-NIL		E=014	44	D=056	10
HADDA	just now	R- 04		F-NIL				D=014	41
HAGAAG	ok!	R-NIL		F-NIL				D=018	34
HALKA	here	R- 06		F- 06		E=014	42	D=010	
HASE	nevertheless	R- 10	51	F- 01					
HORE	before	R- 16	25	F- 03				D=006	
I	me	R- 01		F- 07		E=028	21	D=014	40
IDIIN	for you	R-NIL		F- 02				D=014	42
IDIN	you [obj]	R-NIL		F- 01				D=013	46
IGA	from me	R- 03		F- 05		E=017	32		
II	for me	R- 02		F- 05		E=030	19		
IMMISA	how many?	R-NIL		F-NIL		E=015	35	D=027	18
IN	that	R- 43	09	F- 11	36	E=019	30		
INAAD	that you	R-NIL		F- 01		E=034	16	D=030	17
INAAN	that I	R- 04		F- 08		E=043	11		
INAANNU	that we	R- 02		F- 02				D=023	25
INEY	that they	R- 13	37	F- 08					
INTA	until	R- 05		F- 10	46				
INUU	that he	R- 16	28	F- 16	17				
IS	self	R- 01		F- 12	28				
ISAGOO	while he	R- 06		F- 17	13				
ISKA	oneself	R- 02		F- 09				D=010	
ISKU	together	R- 04		F- 11	37	E=010			
IYO	and	R- 45	05	F- 28	04	E=067	05	D=078	06
JAALLE	comrade	R- 10	49	F-NIL					
KA	from, out	R- 52	01	F- 29	03	E=066	06	D=085	04
KALA	apart	R- 15	32	F- 09				D=006	
KALE	other	R- 24	17	F- 06		E=015	41	D=024	24
KAN	this [m]	R- 02++		F- 04++		E=013	48	D=013	48
KAR---	can, able	R- 10		F- 12		E=034	17	D=006	
KU	on; by	R- 49	03	F- 29	02	E=064	07	D=079	05
LA (1)	with; (2) PRO	R- 37	11	F- 25	09	E=042	12	D=063	09
LAAKIIN	but	R- 02		F-NIL		E=010		D=013	49
LABA	two	R- 06		F- 09		E=014	45	D=020	30
LAGA	= LA + KA	R- 16	26	F- 08				D=011	
LAGU	= LA + KU	R- 19	20	F- 09				D=015	37
LAHAA	he had	R- 10	48	F- 12	29				
LIBAAX	lion	R-NIL		F- 11	38				
LOO	= LA + U	R- 19	21	F- 04					
MA (1)	NEG; (2) QP	R- 05		F- 16	16	E=092	02	D=101	03
MAALIN	day	R- 08		F- 13					
MAAN	I...not	R- 01		F-NIL		E=022	26		

FREQUENCY OF WORDS - SOMALI MATERIALS (D. Zorc)

WORD		READER		FOLKTALES	COM. EXPR		DIALOGS	
MADAXWEYNE	president	R- 13	38	F-NIL				
MAGAALO	city	R- 16	24	F-NIL			D=006	
MAHADSANID	thanks	R-NIL		F-NIL			D=008	
MAR	time	R- 09		F- 01	E=014	43		
MAREYKAN	American	R- 11	44	F-NIL			D=012	50
MARKII	when	R- 12	43	F- 08				
MARKUU	when he...	R- 03		F- 12	30			
MAXAAD	what...you?	R-NIL		F- 06			D=011	
MAXAY	what?	R-NIL		F-NIL		E=016	34	D=007
MAYA	no	R-NIL		F- 01		E=011		D=011
MEEL	place	R- 10	47	F- 06			D=010	
MID	one	R- 17	23	F- 12	31	E=011		D=006
MISE	or...?	R-NIL		F- 01			D=027	19
MIYAAD	do you?	R-NIL		F-NIL		E=027	23	D=005
MUQDISHO	Mogadisho	R- 09		F-NIL		E=015	40	
NA	us	R-NIL		F- 04			D=013	47
NABAD	peace	R- 02		F- 07			D=010	
NAMBAR	number	R-NIL		F-NIL		E=015	39	
NIN	man	R- 11	45	F- 17	14			
NINKII	the man	R- 01		F- 11	39			
OO	that [rel]	R- 51	02	F- 28	07	E=060	08	D=105
RAALLI	willing	R- 03		F- 01		E=014	46	
RAB-	want	R-NIL		F- 02		E=031	18	
SHAN	five	R- 05		F-NIL		E=015	38	D=018
SHEEG-	say; said	R- 29	**	F- 05				33
SHEEGEY	he said	R- 12	40	F- 02				
SHILIN	shilling	R- 01		F-NIL		E=011		D=026
SI	way, manner	R- 15	31	F- 10	47			20
SIDA	like, thus	R- 14	35					D=005
SIDII	the way...	R- 14	34					
SII	thither	R- 09		F- 08		E=013	49	
SOO	hither	R- 36	12	F- 24	11	E=038	15	D=070
SOOMAALI	Somali	R- 09		F-NIL		E=020	27	D=010
SOOMAALIYA	Somalia	R- 05		F-NIL		E=014	47	D=009
TAHAY	you are	R- 09		F- 05		E=015	37	D=014
TAN	this [fem]	R- 06++		F- 04++				43
TOBAN	ten	R- 04						D=016
U	for; to	R- 44	08	F- 28	06	E=077	04	D=011
UGA	= U + KA	R- 15	29	F- 11	40			D=069
UGU	= U + KU	R- 18	22	F- 17	15	E=013		08
URUR	assembly	R- 12	42	F-NIL				
UU	he	R- 40	10					
WAA [DECLARATIVE]		R- 07		F- 24	10	E=111	01	D=183
WAAN	I [am]	R- 01		F- 05		E=024	25	D=011
WAANNU	we [are]	R-NIL		F-NIL				D=020
WADA	together	R- 10	46	F- 12	32			29
WALAAL	"please"	R-NIL		F-NIL		E=084	03	
WANAAGSAN	good	R- 08		F- 01		E=017	33	D=024
WAX	thing	R- 15	30	F- 14	21	E=019	29	D=053
WAXA [CLEFT FOC]		R- 25	15	F- 08				D=007
WAXAA	the one that	R- 24	18	F- 13	24			
WAXAAN	what...I/we	R- 04		F- 03		E=053	10	D=014
WAXAANNU	what...we	R- 03						44
WAXAY	what...she	R- 25	16	F- 07				D=026
WEY	she [is]	R- 01		F- 05		E=011		D=007
WEYN	large	R- 10	52					D=022
WUU	he [is]	R- 02		F- 15	19			27
WUXUU	what...he	R- 26	14	F- 11	41			D=008
XAGGEE	where?	R-NIL		F- 01				D=006
XUN	bad	R- 03		F- 02		E=010		
YAHAY	he is	R- 09		F- 14	20			D=020
YAR	small	R- 03		F- 07				31
YIHIIN	they are	R- 07		F- 02				D=008
YIMID	he came	R- 04		F- 10	48			D=005
YIRI	he said	R- 12	41	F- 28	05			

R- Occurrence in N out of 51 selections (i.e., not total number)  
 F- Occurrence in N out of 30 folktales (i.e., not a total number)  
 E= Total number of occurrences in the Somali Common Expressions  
 D= Total number of occurrences in the Somali Dialogs

SAMPLE GRAMMAR PARSING (D. Zorc)

SOMALI NEWSPAPER READER - SELECTION 02

= = = = =

MUDANE ARAP MOI OO DHAMBAAL SALAAN AH U SOO  
DIR/AY JAALLE SIYAAD

Muqdisho, Abriil 4 (SONNA) -- Madaxweyna/ha Keenya, Mudane Daniel arap Moi, ah/na Guddoomiya/ha UMA (Urur/ka Midow/ga Afrika) ee sannad/kan, isag/oo mar/aya hawa/da dal/ka JDS (Jamhuuriyad/da Dimoqraadi/ga Soomaali/yeed), ku/na sii jeed/a Indiya, ayaa shalay dhambaal salaam ah u soo dir/ay Madaxweyna/ha JDS, Jaalle Maxamed Siyaad Barre.

= = = = =

A	L23*	[FOCUS VERB]	KA	L04	the [masc]
ABRIIL	L33	April	KAN	L05	this
AFRIKA		Africa	KEENYA	L02	Kenya
AH	L19	being [rel]	KU	L11	to [PREP]
AY	L08	-ed [PAST]	MADAXWEYNA	L49	president
AYA	L23*	-ing [PROG]	MAR		pass (by)
AYAA	L23	[FOCUS]	MAXAMED	L03	Mohammed
DA	L06	the [fem]	MIDOW		unity
DAL		country	MUDANE		Honorable
DHAMBAAL		message	MUQDISHO		Mogadisho
DIMOQRAADI		Democratic	NA	L21	and
DIR		send	OO	L23	while [conj]
EE	L24	[relative]	SALAAN		greetings
GA	L05	the [masc]	SANNAD		year
GUDDOOMIYA		chairman	SHALAY	L33	yesterday
HA	L05	the [masc]	SII	L11	that way
HAWA		air(space)	SONNA [abr]		Somali News
INDIYA		India	SOO	L11	this way
ISAGOO	L23	while he	SOOMAALI		Somali
JAALLE		comrade	U	L11	to [PREP]
JAMHUURIYAD		republic	UMA [abr]		Org.Afr.Unity
JDS [abr]		Som.Dem.Rep.	URUR		organization
JEED		head towards	YEED	L31	of [N-GEN]

READING	<u>COURSE</u>	ENTRIES	WORDS	LINES	HIGHEST GRM.LEV
R01	38*	27	38	04	47
R02	23*	42	64	07	33
R03	24*	42	76	07	46
R04	47*	75	114	11	48
R05	36*	68	98	09	41
R06	40*	70	100	10	46
R07	46*	94	151	15	48
R08	44*	75	107	10	51
R09	30*	59	98	09	50
R10	32*	61	94	09	50
R11	37*	77	111	11	37
R12	43*	89	153	13	48
R13	31*	64	116	12	48
R14	49*	81	124	12	49
R15	29*	64	112	12	48
R16	34*	74	112	09	41
R17	36*	55	83	09	47
R18	38*	93	183	19	49
R19	37*	64	99	10	50
R20	29*	79	133	12	33
R21	30*	70	127	12	33
R22	31*	65	109	13	48
R23	35*	82	139	13	48
R24	40*	83	128	13	48
R25	35*	65	100	11	48
R26	42*	77	108	13	50
R27	27*	73	124	12	49
R28	39*	92	150	16	48
R29	41*	93	150	15	47
R30	50*	104	166	15	50
R31	50*	74	119	13	50
R32	25*	49	70	07	47
R33	33*	61	107	10	48
R34	34*	91	142	13	42
R35	47*	68	130	14	48
R36	42*	81	127	14	42
R37	48*	68	107	11	48
R38	26*	41	66	06	46
R39	47*	183	329	32	47
R40	32*	99	159	15	48
R41	48*	115	183	21	48
R42	44*	164	292	30	51
R43	39*	98	147	16	48
R44	33*	263	433	44	
R45a		184	276	24	
R45b		227	347	31	
R46		132	208	21	
R47		86	136	12	
R48	47*	80	131	14	48
R49		89	135	13	
R50	25	81	122	12	32
R51a		294	534	55	
R51b		303	554	55	
--	28*				
--	45*				

NO NEGATIVE IMPERATIVE IN READER

NO POTENTIAL VERBS or SOW/SHOW IN READER