Appendix A

Table A. Verbs with highest total lemma frequency: Top60s, listed in decreasing order of total verb frequency per corpus. (The list only includes verbs with at least one finite occurrence.) Auxiliary and modal verbs printed in bold.

English	Dutch	German
be	**zijn**	**sein**
have	**hebben**	**haben**
do	gaan	**koennen**
know	**kunnen**	**werden**
get	**moeten**	gehen
think	zeggen	**müssen**
go	doen	machen
shall/should/will/would	**zullen**	sagen
can/could	**worden**	fahren
mean	weten	**sollen**
see	denken	**wollen**
say	komen	passen
guess	vinden	denken
like	zitten	nehmen
take	**willen**	geben
want	zien	wissen
make	staan	sehen
work	kijken	treffen
come	maken	aussehen
use	krijgen	fliegen
try	**mogen**	vorschlagen
talk	geven	**mögen**
look	laten	buchen
start	horen	brauchen
live	liggen	finden
need	werken	kommen
seem	blijven	glauben
put	beginnen	lassen
feel	bedoelen	schauen
pay	vragen	ausmachen
hear	houden	liegen
may/might	lezen	grüßen
find	lopen	kosten
read	kennen	meinen

buy	gebeuren	tun
let	spelen	heißen
tell	zetten	kümmern
keep	nemen	dauern
watch	lijken	bleiben
call	proberen	reservieren
give	vertellen	halten
enjoy	praten	festhalten
happen	halen	losfahren
play	kopen	freuen
spend	eten	vereinbaren
sound	geloven	anhören
remember	spreken	freihaben
stay	bellen	ankommen
believe	heten	ausschauen
love	schrijven	reichen
change	noemen	kennen
move	hoeven	anbieten
run	vallen	erkundigen
agree	gebruiken	klingen
suppose	brengen	planen
help	voelen	unternehmen
understand	rijden	besprechen
sit	wonen	sprechen
leave	leren	anrufen
drive	betalen	überlegen

Appendix B

Table B. Finite verbs with highest weighted Main-Fnt and weighted Sub-Fnt scores, listed in order of decreasing scores. For instance, the verb *guess* has the highest proportion in the set of all weighted Main-Fnt proportions in the English corpus; and the verb *ought* has the highest proportion in the set of all weighted Sub-Fnt English verbforms. Auxiliary and modal verbs printed in bold.

Main-Fnt			Sub-Fnt		
English	**Dutch**	**German**	**English**	**Dutch**	**German**
guess	menen	grüßen	ought	betreffen	laufen
mean	snappen	bedanken	**can**	thuiskomen	erinnern
know	geloven	danken	need	danken	losfliegen
hope	lijken	anhören	**shall/will**	terugkomen	**mögen**
sound	klinken	glauben	**may**	binnenkomen	**wollen**
bet	hoeven	heißen	dance	behoren	hinkommen
must	schelen	stimmen	deserve	overblijven	kriegen
wish	uitmaken	klingen	exist	aankunnen	losgehen
suspect	gelden	annehmen	want	vergissen	**dürfen**
love	afhangen	denken	mention	dreigen	liegen
seem	denken	freuen	affect	weggaan	zurückkommen
think	betekenen	halten	happen	meehebben	bestehen
hate	uitzien	aussehen	**be**	aanhebben	kennen
agree	vinden	meinen	end	bezighouden	kommen
wonder	vrezen	vorhaben	win	lesgeven	losfahren
appreciate	hopen	betragen	act	toekomen	abfahren
figure	meevallen	verbleiben	go out	beseffen	**können**
forget	aankijken	dabeihaben	start out	voldoen	wohnen
scare	afvragen	**müssen**	commit	bestaan	bekommen
tend	doorhebben	freihaben	graduate	overhebben	interessieren
do	weten	ausschauen	require	aangaan	auskennen
be	schijnen	**werden**	**have**	afkomen	ankommen
may	heten	hoffen	come	tegenkomen	brauchen
shall/will	ophebben	wünschen	tend	dienen	**sollen**
end up	kruipen	**sollen**	own	**zullen**	**haben**
prefer	**mogen**	schätzen	retire	teruggaan	geben
quit	**zijn**	**sein**	end up	uitkomen	freihaben
believe	opschieten	**haben**	produce	kloppen	dabeihaben
laugh	vermoeden	wissen	purchase	aantrekken	abholen
depend	bedoelen	auskennen	die	terechtkomen	zahlen
assume	**zullen**	**können**	fall	afgaan	arbeiten

can	**hebben**	kosten	move	opvallen	**müssen**
startout	**moeten**	brauchen	open	**willen**	wünschen
enjoy	kloppen	bestehen	belong	aanspreken	treffen
have	blijken	kennen	come out	regenen	zurückfahren
manage	toegeven	geben	turn	omgaan	beginnen
like	schatten	**dürfen**	cost	optreden	wissen
turn out	voelen	stehen	bother	overkomen	dauern
feel	kennen	gehen	assume	bedoelen	hingehen
wind up	**kunnen**	lassen	**do**	herkennen	**werden**
mix	opgaan	warten	put out	ophebben	stehen
decide	**willen**	festhalten	must	plaatsvinden	finden
get up	kosten	aufschreiben	start	wonen	stattfinden
pour	duren	festmachen	earn	passeren	**sein**
jump	inhouden	eintragen	live	kijken	klappen
miss	afgaan	**mögen**	say	voorkomen	bringen
grow up	zitten	melden	charge	rondlopen	vorhaben
belong	pleiten	tun	take off	aansluiten	anschauen
keep	aanhebben	**wollen**	place	geraken	anrufen
use	wegen	sehen	contribute	ervaren	sparen
put in	beweren	nehmen	report	**worden**	bevorzugen
rent	schrikken	dauern	choose	opgaan	wegfahren
help out	ruiken	schauen	ask	**kunnen**	probieren
live	bestaan	gönnen	figure	eindigen	sehen
say	staan	vorbeikommen	turnout	**hebben**	zurückfliegen
come on	gaan	liegen	waste	leiden	ausschauen
find	tegenkomen	finden	wind up	liggen	hinfahren
subscribe	begrijpen	bleiben	feel	uitgaan	fahren
take up	herhalen	anfangen	come up	heten	übernachten
understand	overhebben	übernehmen	dump	**moeten**	ansehen

Appendix C: Summaries of the Betareg models

This Appendix contains essential properties of the output produced by Betareg. (For information on Betareg within R, see https://CRAN.R-project.org/package=betareg). Explanations of terms not reserved by Betareg:

- *Language* codes (for treatment coding): en (English), de (German), nl (Dutch). English was the reference language.
- *NormTotFreq*: the "raw" frequency count of a lemma divided by the total number of verbforms per corpus (see rightmost column in Table 4).
- *WMain, WSub*: weighted main proportions, weighted sub proportions (dependent variables in first and second models below).
- *DiffWMainWSub*: WMain proportion minus WSub proportion (dependent variable in the third model).

For all three dependent variables (DVs: WMain, WSub, and DiffWMainWSub) we tested the following models:

```
mDV         ← betareg(WMain ~ 0 + NormTotFreq * Language, data = dat)
mDV.hetero ← betareg(WMain ~ 0 + NormTotFreq * Language | Language, data=dat)
mDV.loglog ← betareg(WMain ~ 0 + NormTotFreq * Language, data = dat, link = "loglog")
```

The input formulae include a zero term, which forces every level of the categorical independent variable (here: every language) to receive its own intercept. Therefore, the main effects are the per-language intercepts, and the slope for the frequency effect of each target language is not included in the interactions. The interaction terms reflect the differences in the slope for each language, i.e. the offset from the reference language (English).

The results of the Betareg analyses are presented on the following three pages.

R reports for the three Betareg analyses

First analysis

```
> mWMain <- betareg(WMain  ~ 0 + NormTotFreq * Language, data = dat)
> mWMain

Call:
betareg(formula = WMain ~ 0 + NormTotFreq * Language, data = dat)

Phi coefficients (precision model with identity link):
(phi)
213.9

> summary(mWMain)

Call:
betareg(formula = WMain ~ 0 + NormTotFreq * Language, data = dat)

Standardized weighted residuals 2:
    Min      1Q     Median     3Q      Max
-10.2344  -0.1074   0.1421   0.4016   3.2397

Coefficients (mean model with logit link):
                          Estimate Std. Error z value Pr(>|z|)
NormTotFreq               20.37866    0.78062  26.106  < 2e-16 ***
Languageen                -6.86133    0.07942 -86.395  < 2e-16 ***
Languagenl                -6.90968    0.07486 -92.306  < 2e-16 ***
Languagede                -6.31671    0.08631 -73.187  < 2e-16 ***
NormTotFreq:Language[T.nl]  4.67678    1.09409   4.275 1.92e-05 ***
NormTotFreq:Language[T.de]  3.42252    1.12684   3.037  0.00239 **

Phi coefficients (precision model with identity link):
      Estimate Std. Error z value Pr(>|z|)
(phi)   213.88      15.71   13.62   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Type of estimator: ML (maximum likelihood)
Log-likelihood:  6681 on 7 Df
Pseudo R-squared: 0.06807
Number of iterations: 82 (BFGS) + 6 (Fisher scoring)
```

Second analysis

```
> mWSub <- betareg(WSub  ~ 0 + NormTotFreq * Language, data = dat)
> mWSub

Call:
betareg(formula = WSub ~ 0 + NormTotFreq * Language, data = dat)

Phi coefficients (precision model with identity link):
(phi)
792.5

> summary(mWSub)

Call:
betareg(formula = WSub ~ 0 + NormTotFreq * Language, data = dat)

Standardized weighted residuals 2:
  Min      1Q    Median    3Q     Max
-7.4068 -0.0223  0.2256  0.4871  3.4204

Coefficients (mean model with logit link):
                           Estimate Std. Error  z value Pr(>|z|)
NormTotFreq                21.82391    0.56390   38.702   <2e-16 ***
Languageen                 -7.88847    0.07520 -104.896   <2e-16 ***
Languagenl                 -8.11729    0.07344 -110.528   <2e-16 ***
Languagede                 -8.29445    0.09644  -86.003   <2e-16 ***
NormTotFreq:Language[T.nl]  0.88753    0.97254    0.913    0.361
NormTotFreq:Language[T.de] -0.68686    1.28687   -0.534    0.594

Phi coefficients (precision model with identity link):
     Estimate  Std. Error z value  Pr(>|z|)
(phi)   792.53      58.01   13.66   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Type of estimator: ML (maximum likelihood)
Log-likelihood:  7761 on 7 Df
Pseudo R-squared: 0.03978
Number of iterations: 86 (BFGS) + 10 (Fisher scoring)
```

Third analysis

```
> mDiffWMainWSub <- betareg(DiffWMainWSub  ~ 0 + NormTotFreq * Language, data = dat)
> mDiffWMainWSub

Call:
betareg(formula = DiffWMainWSub ~ 0 + NormTotFreq * Language, data = dat)

Phi coefficients (precision model with identity link):
(phi)
28431

> summary(mDiffWMainWSub)

Call:
betareg(formula = DiffWMainWSub ~ 0 + NormTotFreq * Language, data = dat)

Standardized weighted residuals 2:
   Min      1Q     Median      3Q      Max
-11.1289  -0.0787  -0.0245   0.0338  20.6661

Coefficients (mean model with logit link):
                             Estimate Std. Error  z value Pr(>|z|)
NormTotFreq                 2.3014736  0.0545653    42.18   <2e-16 ***
Languageen                 -2.1966648  0.0011182 -1964.48   <2e-16 ***
Languagenl                 -2.1983881  0.0009566 -2298.01   <2e-16 ***
Languagede                 -2.1979136  0.0015543 -1414.05   <2e-16 ***
NormTotFreq:Language[T.nl]  2.3755478  0.0784999    30.26   <2e-16 ***
NormTotFreq:Language[T.de]  3.6286611  0.0780530    46.49   <2e-16 ***

Phi coefficients (precision model with identity link):
      Estimate  Std. Error z value Pr(>|z|)
(phi)    28431       1319   21.55   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Type of estimator: ML (maximum likelihood)
Log-likelihood:  4559 on 7 Df
Pseudo R-squared: 0.9384
Number of iterations: 609 (BFGS) + 5 (Fisher scoring)
```