Table 3. Quantitative techniques and their usage in corpus-based Cognitive Linguistics

Technique	**Type**	**Object** collocation	**Example of application**	**Explanation**
T-score, Z-score, MI score	measure	strength	identifying multi-word patterns – Biber (2009) identifying constructional variants – Wong (2009)	Evert (2009), Biber & Jones (2010)
Chi-squared test, Fisher's exact Test	univariate	probability / independence	synonymy, constructional – Wulff (2006) polysemy, lexical – Robinson (2010)	Dalgaard (2008), Everitt & Hothorn (2009), Gries (this volume)
Collostructional analysis	univariate	collocation strength	synonymy, constructional – Hilpert (2008) synonymy, constructional – Gilquin (2010)	Stefanowitsch & Gries (2003), Gries & Stefanowitsch (2004a), Hilpert (this volume)
Hierarchical cluster analysis	multivariate	associations btw. objects of single variable	polysemy, lexical – Gries (2006) synonymy, lexical – Divjak (2010a)	Baayen (2008), Everitt *et al.* (2011), Divjak & Fieller (this volume)
Multidimensional scaling	multivariate	associations btw. objects of multiple variables	synonymy, morphological – Croft & Poole (2008) relations between variants Berthele (2010)	Baayen (2008), Izenman (2008), Everitt & Hothorn (2010)
Correspondence analysis	multivariate	associations btw. objects of multiple variables	synonymy, concepts – Szelid & Geeraerts (2008) polysemy, constructional – Glynn (2009)	Le Roux & Rouanet (2010), Husson *et al.* (2011), Glynn (this volume)
Configural frequency analysis	multivariate	associations btw. objects of multiple variables	polysemy, constructional – Hilpert (2009) synonymy, constructional – Hoffmann (2011)	von Eye (2002), Tabachnick & Fidell (2007), Gries (2009b)
Discriminant analysis	multivariate	identify factors that lead to an outcome / prediction	synonymy, constructional – Gries (2003) synonymy, lexical – Divjak (2010a)	Tabachnick & Fidell (2007), Baayen (2008), Maindonald & Braun (2010)
Classification tree analysis	multivariate	identify factors that lead to an outcome / prediction	polysemy, lexical – Robinson (2012a) synonymy, lexical – Levshina *et al.* (this vol.)	Venables & Ripley (2002), Everitt & Hothorn (2010), Maindonald & Braun (2010)
Loglinear analysis	multivariate	predict correlation multiple response variables	synonymy, constructional – Krawczak & Glynn (in press) polysemy, lexical – Glynn (forthc.)	Kroonenberg (2008), Hilbe (2011), Smith (2011)
Binary logistic regression	multivariate	predict outcome binary response variable	synonymy, constructional – Szmrecsanyi (2003) synonymy, lexical – Speelman & Geeraerts (2010)	Orme & Combs-Orme (2009), Everitt & Hothorn (2010), Speelman (this volume)
Ordinal logistic regression	multivariate	predict outcome ranked response variable	synonymy. lexico–constructional – Klavan (2012) synonymy, constructional – Glynn & Krawczak (forthc.)	Baayen (2008), Tarling (2009), Orme & Combs-Orme (2009)
Multinomial logistic regression	multivariate	predict outcome multiple response variables	synonymy, lexical Arppe (2008) synonymy, lexical – Krawczak (2014a)	Long & Freese (2006), Tarling (2009), Orme & Combs-Orme (2009)
Mixed-effects logistic regression	multivariate	predict outcome include random variables	synonymy, constructional – Divjak (2010b) synonymy, constructional – Levshina *et al.* (this vol.)	Baayen (2008), Maindonald & Braun (2010), Smith (2011)

Table 4. Functions and packages for categorical multivariate statistics in R

Technique	Function	Package	R code tutorial
Hierarchical cluster analysis	`hclust`	stats*	Crawley (2007: 738ff.); Zhao (2013)
	`agnes`	cluster	Kaufman & Rousseeuw (2005); Maechler (2013)
	`pvclust`	pvclust	Suzuki & Hidetoshi (2006); Suzuki (2013)
K-means cluster analysis	`kmeans`	stats*	Crawley (2007: 742ff.); Zhao (2013)
	`clara`	cluster	Kaufman & Rousseeuw (2005); Maechler (2013)
	`pamk`	fpc	Hennig (2013)
Binary correspondence analysis	`corresp`	MASS*	Venables & Ripley (2002: 326ff.); Ripley (2013)
	`ca`	ca	Greenacre (2007); Neandić & Greenacre (2007)
	`anacor`	anacor	de Leeuw & Mair (2009a, 2013a)
Multiple correspondence analysis	`mca`	MASS*	Venables & Ripley (2002: 329f.); Ripley (2013)
	`mjca`	ca	Greenacre (2007); Neandić & Greenacre (2007)
	`MCA`	FactoMineR	Lê *et al.* (2008); Husson *et al.* (2013)
Multidimensional scaling	`cmdscale`	stats*	Baayen (2008: 136ff.); Johnson (2008: 208ff.)
	`sammon`	MASS*	Maindonald & Baun (2010: 284f.)
	`smacofSym`	smacof	de Leeuw & Mair (2009b, 2013b)
Configural frequency analysis	`cfa`	cfa	Funke *et al.* (2007); von Eye & Mair (2008)
	`hcfa`	cfa	Gries (2010: 248ff.); von Eye *et al.* (2010: 265ff.)
	`cfa2`	cfa2[4]	No tutorials available, cf. Schönbrodt (2013)
Linear discriminant Analysis	`lda`	MASS*	Baayen (2008: 167ff.); Maindonald & Braun (2010: 385ff.)
	`discrim`	ade4	Chessel *et al.* (2004); Chessel & Dufour (2013)
	`rda`	klaR	Roever *et al.* (2013)
Classification tree analysis / Random forest classification	`rpart`	rpart	Zhao (2012: 32ff.); Therneau *et al.* (2013)
	`tree`	tree	Venables & Ripley (2002: 266)
	`ctree`	party	Zhao (2013: 29ff.)
	`cforest`	party	Strobl *et al.* (2009a, 2009b)
	`randomForest`	randomForest	Maindonald & Braun (2010: 351ff.); Liaw & Wiener (2002)
Loglinear analysis	`glm`	MASS*	Maindonald & Braun (2010: 258ff.); Baguley (2012)
	`loglm`	MASS*	Thompson (2009: 142ff.); Baguley (2012)
	`quasipois`	aod	Lesnoff & Lancelot (2013)

4. The package `cfa2` is not currently in the CRAN repository for R but can be found in the RForge repository. This repository is typically used for packages still under development. A simple command listed on the Rforge site for the package will install a package as effortlessly as installation using the 'normal' method in R.

Table 4. (*continued*)

Technique	Function	Package	R code tutorial
Binary logistic regression	`glm`	MASS*	Baayen (2008: 195ff.); Everitt & Hothorn (2010: 122ff.)
	`lrm`	rms†	Harrell (2001: 257ff.; 2012: 221ff.); Baayen (2008: 195ff.)
	`MCMClogit`	MCMCpack	Martin *et al.* (2010)
Ordinal logistic regression	`polr`	MASS*	Faraway (2006: 117ff.); Maindonald & Braun (2010: 270ff.)
	`lrm`	rms†	Baayen (2008); Johnson (2008)
	`clm`	ordinal	Christensen (2012)
Multinomial logistic regression	`multinom`	nnet	Faraway (2006: 106); Thompson (2009: 118)
	`polytomous`	polytomous	Arppe (2014)
	`mlogit`	mlogit	Field *et al.* (2012: 325); Croissant (2013)
Multilevel logistic regression (mixed effects)	`lmer`	lme4	Baayen (2008: 278ff.); Bates (forthc.: 1ff.)
	`glmmPQL`[5]	MASS*	Johnson (2008: 255ff.); Thompson (2009: 179ff.)
	`MCMCglmm`	MCMCglmm	Hadfield (2010)

* Recommended package for the base installation. This means it comes 'pre-installed'.

† In older textbooks and tutorials, this package is called `Design`. The package `rms` is simply a new version. The command line to use the package is unchanged and so older descriptions remain helpful.

from what is covered in this volume on all three fronts – developing knowledge of R, the basic statistical principles and tests, as well as advanced statistical analysis.

Gries' (2009a) *Quantitative Corpus Linguistics with R* is another book to consider. Although an excellent book, it is designed more for corpus linguistics *per se* than multivariate analysis. More in line with the focus of this volume is Gries' (2009b) *Statistics for Linguistics Using R*. It covers the basics thoroughly and introduces some multivariate statistical techniques. A new edition, Gries (2013), expands the chapter on precisely the techniques covered in this volume

Johnson's (2008) *Quantitative Methods in Linguistics* is good for a debutant level statistics textbook using R – it explains both the command line and statistics lucidly and concisely. However, it 'orders' the different statistical techniques relative to different subfields of linguistics. This could be misleading for the novice and not particularly logical for the reader with some knowledge in the field, since most of the techniques are not at all restricted to the subfield Johnson ascribes to them. However, the expla-

5. The function `glmmPQL` uses a so-called penalised quasi-likelihood, which has lost favour in the research community (Crawley 2007: 655). Although the functions `lmer` and `MCMCglmm` are more up to date in this regard, `glmmPQL` in the MASS package still works perfectly well, especially when learning since some of the command line is closer to other regression functions a learner may have already mastered.