Processing Lexical Bundles

Lensink, S.E.

Appendix A

trigram	translation	frequency ABC
aan de man	to the man	15373
aan de universiteit	at (the) university	4073
aan de vooravond	on the eve	7798
aan het begin	at the beginning	89579
aan het eind	at the end	150873
aan het werk	working	1285050
aan te pakken	to deal with	39314
aan te passen	to adapt	41295
achter het raam	behind the window	9642
begin dit jaar	at the beginning of the year	13635
bij grote bedrijven	at big companies	969
bij hem thuis	at his house	10643
bij hun moeder	at their mother's	1754
daar gaan we	there we go	241060
dat blijkt uit	that appears from	1461
dat is altijd	that is always	67085
dat is jammer	that is a pity	42773
de aanpak van	the approach of	10757
de aanslagen van	the (terrorist) attacks of	1152
de actie van	the action of	19982
de afgelopen jaren	the past years	41346
de afgelopen maanden	the past months	17046
de Amerikaanse president	the American president	2542
de Amerikaanse regering	the American government	875
de andere kant	the other side	426100
de bacterie is	the bacteria is	45
de besten van	the best of	989
de bouw van	the construction of	37885
de buurt van	the neighborhood of	217614
de dag dat	the day that	194752
de discussie over	the discussion on	13560
de dood van	the death of	42027
de economie van	the economy of	3566
de eerste plaats	the first place	34783
de finale van	the finals of	103195
de foto van	the picture of	68385

de gevolgen van	the consequences of	30608
de halve finale	the semi-finals of	81520
de handen vol	hands full	1988
de hele dag	the whole day	1946081
de hele wereld	the whole world	249719
de helft van	half of	417353
de inval in	the invasion of	666
de jaren negentig	the nineties	2893
de jaren twintig	the twenties	387
de kans dat	the chance that	41581
de keuze van	the choice of	12330
de komende jaren	the coming years	48187
de komende vier	the next four	6470
de komst van	the arrival of	48370
de kwaliteit van	the quality of	45240
de laatste twintig	the last twenty	507
de markt is	the market is	6906
de mensen hier	the people here	11779
de mogelijkheid om	the possibility to	31704
de moord op	the murder on	24169
de nabijheid van	the proximity of	2100
de nationale ploeg	the national team	3493
de ontvangst van	the reception of	2864
de oorlog in	into the war	5740
de oorlog tegen	the war against	1573
de organisatie van	the organization of	24050
de ploeg van	the team of	6725
de politie had	the police had	957
de positie van	the position of	11485
de presentatie van	the presentation of	37769
de prijs van	the price of	59208
de rand van	the edge of	66531
de rechtbank in	the court in	17346
de rest van	the rest of	782197
de resultaten van	the results of	16923
de rol van	the role of	55053
de Russische president	the Russian president	1116
de sociale zekerheid	the social security	2577
de strijd tegen	the battle against	34540
de trainer van	the coach of	10399
de tweede helft	the second half	59770
de tweede plaats	the second place	11487
de tweede ronde	the second round	22796
de universiteit van	the university of	5810
de vader van	the father of	68116
de vleugels van	the wings of	2756
de vraag is	the question is	62321
de website van	the website of	107326
de woorden van	the words of	24221
de zoon van	the son of	28840
door het ministerie	by the ministry	704
door onze correspondent	by our correspondent	105
drie jaar geleden	three years ago	9994
een aantal weken	a couple of weeks	16249
een actie van	an action by	11974
een belangrijke rol	an important role	10667

een bezoek aan	a visit to	47188
een brief aan	a letter to	9675
een deel van	a part of	130958
een film van	a movie of	20778
een gesprek met	a conversation with	97516
een groot aantal	a large number	16102
een groot deel	a big part	49737
een half jaar	half a year	220540
een half uur	half an hour	1031387
een hoger niveau	a higher level	11297
een idee van	an idea of	9561
een jaar eerder	a year earlier	4725
een jaar geleden	a year ago	123936
een kans om	a chance to	17502
een kwart van	a quarter of	17557
een kwestie van	a question of	80116
een moment dat	a moment that	17900
een onderzoek naar	an investigation into	16947
een opkomst van	a rise of	1276
een paar dagen	a couple of days	399097
een paar jaar	a couple of years	155619
een paar weken	a couple of weeks	177710
een tentoonstelling van	an exhibition of	1062
een vorm van	a form of	43031
een winst van	a profit of	3386
een woordvoerder van	a spokesman for	2266
eerder deze week	previously this week	13746
einde van het	end of the	77611
elke keer weer	every time again	92297
en de manier	and the way	6222
en te weinig	and too little	18505
euro per maand	euro per month	47736
genieten van een	enjoying a	132200
ging het mis	it went wrong	64252
het afgelopen jaar	the past year	35986
het begin van	the beginning of	212462
het belang van	the importance of	51523
het bestuur van	the board of	17895
het centrum van	the center of	128503
het centrum voor	the center for	2134
het eerst sinds	for the first time since	91715
het eerste kwartaal	the first quarter	19414
het functioneren van	the functioning of	2732
het gebied van	the area of	128659
het gebruik van	the use of	63198
het gevoel van	the feeling of	29980
het ging om	it was about	15042
het herstel van	the recovery of	5251
het is geen	it is no	155183
het is niet	it is not	512730
het kader van	the framework of	128254
het laatste kwartier	the last quarter	4456
het ministerie van	the ministry of	19317
het moment dat	the moment that	158686
het najaar van	the fall of	2818
het plan van	the plan of	10331

het spel van	the game of	13382
het tweede kwartaal	the second quarter	14558
het vertrek van	the departure of	19881
het werk van	the work of	45904
het zoeken naar	the search for	27018
ik denk dat	I think that	1770156
in de aanloop	in the run-up	9802
in de auto	in the car	1520967
in de buurt	in the neighborhood/close	944910
in de eredivisie	in the premier division	39637
in de hoek	at the corner	53419
in de krant	in the newspaper	224065
in de lucht	in the air	320247
in de ogen	in the eyes	62735
in de omgeving	in the neighborhood	73136
in de partij	in the party	3097
in de politiek	in politics	47360
in de praktijk	in practice	75131
in de regio	in the region	175792
in de rij	in line	202447
in de stad	in the city	1106466
in de strijd	in the battle	52453
in de wereld	in the world	207594
in de zomer	during summer	335094
in de zorg	in health care	125713
in dit gebouw	in this building	2176
in een open	in an open	5010
in eigen land	in your own country	32385
in Europa zijn	being in Europe	2226
in handen van	in the hands of	22157
in het begin	at the beginning	133282
in het boek	in the book	51266
in het centrum	in the center	143835
in het kader	in the framework	78640
in het land	in the country	93987
in het noorden	in the north	58157
in het openbaar	in public	67556
in het verhaal	in the story	12894
in het ziekenhuis	in the hospital	350649
in ons land	in our country	41727
is een beetje	is a little	337220
is er ook	is there too	347370
is niet alleen	is not alone	75895
is niet nodig	is not necessary	68369
is nog steeds	still is	448018
is zo goed	is so good	77105
kans op een	chance of a	358386
kijken naar de	to look at the	105645
maakt niet uit	does not matter	1306258
maar het was	but it was	260391
met de mededeling	with the announcement	9055
met de speler	with the player	466
met een optie	with an option	2562
met een winst	with a profit	1453
moet ook nog	also has to	333650
na de oorlog	after the war	5851

na de pauze	after the break	28459
naar het buitenland	abroad	46869
net als hij	just like him	1766
niet al te	not too	195025
niet de enige	not the only one	432110
niet te veel	not too much	209056
niet te vergeten	not to forgot	116559
niets weten van	knowing nothing of	2986
nog een seizoen	yet another season	7423
nog niet bekend	not yet known	62111
nog ver weg	still far away	14766
nu is dat	now is that	19928
oktober vorig jaar	October last year	1739
om dit jaar	to ... this year	8176
om het leven	to the life	52336
om te buigen	in order to bend	2712
om te overleven	in order to survive	15291
ook wel eens	every now and then	332971
op dat moment	at that moment	103274
op de bank	on the couch	3286884
op de beurs	at the stock market	46130
op de dag	on the day	159050
op de eerste	at the first	126026
op de markt	on the market	176993
op de schouder	on the shoulder	5291
op de website	on the website	229875
op die manier	in this way	92344
op dit moment	at this moment	1037494
op een aanslag	at an attack	503
op het moment	at the moment	222221
op het nieuwe	at/on the new	34288
op het werk	at work	386510
op langere termijn	in the long run	3658
op te lossen	to solve	104502
op zijn bureau	at his desk	1171
op zijn minst	at least	27093
over de geschiedenis	about the history	8633
over de manier	about the way	4767
over de toekomst	about the future	43455
over het land	about the country	9189
paar jaar geleden	couple of years ago	57132
pas sinds kort	only recently	3790
raad van bestuur	board of directors	7008
sinds lange tijd	since a long time	26597
te beseffen dat	to realize that	16136
te staan in	to stand in	7740
te weten dat	to know that	106478
te zien zijn	to be visible	32458
terug te komen	to return	46567
tijdens de dictatuur	during the dictatorship	30
tot nu toe	until now	632637
twee weken op	two weeks on	6750
uit de grond	from the ground	31069
uit de ploeg	from the team	874
uit een boek	from a book	5952
uit eigen ervaring	from my own experience	8863

uit het Engels	from English	1175
uit te leggen	to explain	172789
uit te nodigen	to invite	16723
van de aandelen	of the shares	2190
van de aarde	from the earth	29344
van de beurs	from the stock market	8974
van de bevolking	from the population	18109
van de drie	out of three	33303
van de economie	of the economy	8738
van de families	of the families	263
van de gemeente	of the municipality	102650
van de jaren	of the years	15061
van de mogelijkheden	of the possibilities	4651
van de overheid	from the government	31026
van de tien	from the ten	11905
van de trainer	from the trainer	6949
van de twee	from the two	50064
van de volgende	from the next	17977
van de vorige	from the previous	43525
van de wereld	from/of the world	447158
van dit jaar	of this year	181143
van het aanbod	of the offer	2090
van het Britse	of the British	3003
van het kabinet	from the cabinet	18244
van het land	from/of the country	116573
van het museum	from/of the museum	4352
van het onderzoek	from/of the research	8406
van het publiek	from the public	10216
van het seizoen	from/of the season	202065
van onze redactie	from our editorial staff	848
van zijn proces	of his process	208
verdacht van fraude	suspected of fraud	1469
volgens het boekje	according to the rules	3106
voor de bescherming	for the protection	1022
voor eigen publiek	for your own audience	7444
voor het eerst	for the first time	1025405
voor het leven	for life	106138
voor iemand die	for someone who	89362
wel of niet	yes or no	286718
wordt beschuldigd van	is being accused of	1593
zo lang geleden	so long ago	40677

Table 1: Stimuli used in the eye-tracking experiment of Chapter 2. Translations in English are provided in the second column. Phrasal frequencies (freqABC) are listed in the righter-most column.

Appendix B

trigram	translation	condition	frequency ABC	frequency C
aan de beurt	turn	MWU	1743	7458
aan de prins	to the prince	Control	80	12030
aan het eind	at the end	MWU	6395	33356
aan het bed	at the bed	Control	140	24021
aan te passen	to adapt	MWU	2307	10335
aan te steken	to strike	Control	103	7385
de andere kant	the other side	MWU	8113	30948
de andere groep	the other group	Control	108	38917
de eerste plaats	the first place	MWU	5853	80379
de eerste vraag	the first question	Control	237	63831
de hele dag	the whole day	MWU	5243	112638
de hele weg	the whole way	Control	142	101363
een belangrijke rol	an important role	MWU	2875	33764
een belangrijke vorm	an important shape	Control	21	33661
een groot deel	a big part	MWU	5427	69946
een groot kind	a big child	Control	26	46095
een paar dagen	a couple of days	MWU	4809	56414
een paar miljoen	a couple of million	Control	159	58205
er is geen	there is no	MWU	5264	384494
er is wat	there is something	Control	101	525923
in de praktijk	in practice	MWU	5764	11017
in de uitspraak	in the statement	Control	79	9997
in dit geval	in this case	MWU	4333	55463
in dit verhaal	in this story	Control	240	34648
in het centrum	in the center	MWU	3284	15338
in het vertrek	in the room	Control	73	8615
is de kans	is the chance	MWU	1224	30148
is de druk	is the pressure	Control	78	33051
mee te maken	to experience	MWU	2548	143877
mee te komen	to join	Control	90	130534
met de auto	by car	MWU	1302	28882
met de partij	with the party	Control	117	35702
na te denken	to think about	MWU	2349	42304
na te vragen	to inquire	Control	28	47636
nog een keer	again	MWU	5217	91196
nog een week	another week	Control	324	76650

om te kijken	to watch	MWU	1987	50552
om te nemen	to take	Control	18	58350
op dat moment	at that moment	MWU	6441	48502
op dat idee	on that idea	Control	58	37396
op de markt	on the market	MWU	5900	24936
op de brief	on the letter	Control	159	17057
op korte termijn	at short-notice	MWU	2060	14004
op korte afstand	at short distance	Control	104	13658
op te lossen	to solve	MWU	2974	4409
op te drukken	to push up	Control	23	4308
van de bevolking	from the population	MWU	6022	20512
van de discussie	from the discussion	Control	319	17251
van het jaar	of the year	MWU	6150	294718
van het nu	from nowadays	Control	75	358059
voor de toekomst	for the future	MWU	1926	23744
voor de liefde	for the love	Control	127	19180

Table 1: Stimuli used in the experiment. On the left all frequent multi-word units are listed, directly followed by their matched control. Translations in English are provided in the second column. Phrasal frequencies (freqABC) and frequencies of the third word (freqC) are listed in the two righter-most columns next to the stimuli.

Bibliography

Arnold, D., Tomaschek, F., Sering, K., Lopez, F., and Baayen, R. H. (2017). Words from spontaneous conversational speech can be recognized with human-like accuracy by an error-driven learning algorithm that discriminates between meanings straight from smart acoustic features, bypassing the phoneme as recognition unit. *PloS one*, 12(4):e0174623.

Arnon, I. and Cohen Priva, U. (2013). More than words: The effect of multi-word frequency and constituency on phonetic duration. *Language and Speech*, 56(3):349–371.

Arnon, I. and Priva, U. C. (2014). Time and again: The changing effect of word and multi-word frequency on phonetic duration for highly frequent sequences. *The Mental Lexicon*, 9(3):377–400.

Arnon, I. and Ramscar, M. (2012). Granularity and the acquisition of grammatical gender: How order-of-acquisition affects what gets learned. *Cognition*, 122(3):292–305.

Arnon, I. and Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, 62(1):67–82.

Baayen, H., Vasishth, S., Kliegl, R., and Bates, D. (2017a). The cave of shadows: Addressing the human factor with generalized additive mixed models. *Journal of Memory and Language*, 94:206–234.

Baayen, R. and Ramscar, M. (2015). Abstraction, storage and naive discriminative learning. *Handbook of Cognitive Linguistics*, 39:100–120.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press.

Baayen, R. H. (2010). Demythologizing the word frequency effect: A discriminative learning perspective. *The Mental Lexicon*, 5(3):436–461.

Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4):390–412.

Baayen, R. H., Hendrix, P., and Ramscar, M. (2013). Sidestepping the combinatorial explosion: An explanation of n-gram frequency effects based on naive discriminative learning. *Language and Speech*, 56(3):329–347.

Baayen, R. H., Milin, P., Đurđević, D. F., Hendrix, P., and Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review*, 118(3):438.

Baayen, R. H., Sering, T., Shaoul, C., and Milin, P. (2017b). Language comprehension as a multiple label classification problem.

Baayen, R. H., Shaoul, C., Willits, J., and Ramscar, M. (2016a). Comprehension without segmentation: A proof of concept with naive discriminative learning. *Language, Cognition and Neuroscience*, 31(1):106–128.

Baayen, R. H., van Rij, J., de Cat, C., and Wood, S. N. (2016b). Autocorrelated errors in experimental data in the language sciences: Some solutions offered by generalized additive mixed models. *arXiv preprint arXiv:1601.02043*.

Baisa, V., Michelfeit, J., Medved, M., and Jakubícek, M. (2016). European union language resources in sketch engine. In *LREC*.

Bannard, C. and Matthews, D. (2008). Stored word sequences in language learning: The effect of familiarity on children's repetition of four-word combinations. *Psychological Science*, 19(3):241–248.

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, 68(3):255–278.

Bates, D., Kliegl, R., Vasishth, S., and Baayen, H. (2015). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*.

Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., Holland, J., Ke, J., Larsen-Freeman, D., and Schoenemann, T. (2009). Language is a complex adaptive system: Position paper. *Language Learning*, 59(s1):1–26.

Biber, D., Johansson, S., Leech, G., Conrad, S., and Finegan, E. (1999). *Longman grammar of spoken and written English*. Pearson Education Ltd.,-1999.–1204 p.

Boersma, P. and Weenink, D. (2016). Praat: Doing phonetics by computer.[computer program]. version 6.0. 19.

Box, G. E. and Draper, N. R. (1987). *Empirical model-building and response surfaces*. John Wiley & Sons.

Brady, T. F., Konkle, T., Alvarez, G. A., and Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences*, 105(38):14325–14329.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

Breiman, L. et al. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical science*, 16(3):199–231.

Brink, D. v. d. and Hagoort, P. (2004). The influence of semantic and syntactic context constraints on lexical selection and integration in spoken-word comprehension as revealed by erps. *Journal of Cognitive Neuroscience*, 16(6):1068–1084.

Bybee, J. (2010). *Language, usage and cognition*. Cambridge University Press.

Bybee, J. L. (2006). From usage to grammar: The mind's response to repetition. *Language*, 82(4):711–733.

Carrol, G. and Conklin, K. (2015). Eye-tracking multi-word units: Some methodological questions. *Journal of Eye Movement Research*, 7(5).

Conklin, K. and Schmitt, N. (2012). The processing of formulaic language. *Annual Review of Applied Linguistics*, 32:45–61.

Connolly, J. F. and Phillips, N. A. (1994). Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *Journal of Cognitive Neuroscience*, 6(3):256–266.

Coulson, S., King, J. W., and Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and cognitive processes*, 13(1):21–58.

Croft, W. (2001). *Radical construction grammar: Syntactic theory in typological perspective*. Oxford University Press on Demand.

Dąbrowska, E. (2014). Recycling utterances: A speaker's guide to sentence processing.

De Cat, C., Klepousniotou, E., and Baayen, R. H. (2015). Representational deficit or processing effect? An electrophysiological study of noun-noun compound processing by very advanced L2 speakers of English. *Frontiers in psychology*, 6.

Dehaene, S. (2009). *Reading in the brain: The new science of how we read*. Penguin.

Dikker, S. and Pylkkänen, L. (2013). Predicting language: Meg evidence for lexical preactivation. *Brain and language*, 127(1):55–64.

Durrant, P. and Doherty, A. (2010). Are high-frequency collocations psychologically real? investigating the thesis of collocational priming. *Corpus Linguistics and Linguistic Theory*, 6(2):125–155.

Erman, B. and Warren, B. (2000). The idiom principle and the open choice principle. *Text-Interdisciplinary Journal for the Study of Discourse*, 20(1):29–62.

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, 44(4):491–505.

Fisher, K., Bassok, M., and Osterhout, L. (2010). When two plus two does not equal four: Event-related potential responses to semantically incongrous arithmetic word problems. In *Proceedings of the Cognitive Science Society*, volume 32.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in cognitive sciences*, 6(2):78–84.

Friederici, A. D. (2012). The cortical language circuit: from auditory perception to sentence comprehension. *Trends in cognitive sciences*, 16(5):262–268.

Geeraert, K., Newman, J., and Baayen, R. H. (2017). Idiom variation: Experimental data and a blueprint of a computational model. *Topics in Cognitive Science*.

Goldberg, A. E. (2003). Constructions: A new theoretical approach to language. *Trends in Cognitive Sciences*, 7(5):219–224.

Green, C. (2017). Usage-based linguistics and the magic number four. *Cognitive Linguistics*, 28(2):209–237.

Hagoort, P. (2003). How the brain solves the binding problem for language: a neurocomputational model of syntactic processing. *Neuroimage*, 20:S18–S29.

Hagoort, P. and Brown, C. M. (2000). ERP effects of listening to speech: semantic ERP effects. *Neuropsychologia*, 38(11):1518–1530.

Han, S. (2015). *Processing formulaic sequences by native and nonnative speakers of English: Evidence from reading aloud*. PhD thesis, Northern Arizona University.

Harley, T. A. (2013). *The psychology of language: From data to theory*. Psychology press.

Hastie, T. and Tibshirani, R. (1990). *Generalized additive models*. Wiley Online Library.

Hauk, O., Davis, M. H., Ford, M., Pulvermüller, F., and Marslen-Wilson, W. D. (2006). The time course of visual word recognition as revealed by linear regression analysis of ERP data. *Neuroimage*, 30(4):1383–1400.

Hauk, O. and Pulvermüller, F. (2004). Effects of word length and frequency on the human event-related potential. *Clinical Neurophysiology*, 115(5):1090–1103.

Hendrix, P., Bolger, P., and Baayen, H. (2017). Distinct ERP signatures of word frequency, phrase frequency, and prototypicality in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(1):128.

Holcomb, P. J. and Neville, H. J. (1991). Natural speech processing: An analysis using event-related brain potentials. *Psychobiology*, 19(4):286–300.

Hothorn, T., Hornik, K., and Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics*, 15(3):651–674.

Jakubíček, M., Kilgarriff, A., Kovář, V., Rychlỳ, P., and Suchomel, V. (2013). The tenten corpus family. In *7th International Corpus Linguistics Conference CL*, pages 125–127.

Janssen, N. and Barber, H. A. (2012). Phrase frequency effects in language production. *PloS one*, 7(3):e33202.

Jiang, N. A. and Nekrasova, T. M. (2007). The processing of formulaic sequences by second language speakers. *The Modern Language Journal*, 91(3):433–445.

Just, M. A. and Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87(4):329.

Kaan, E. and Swaab, T. Y. (2003). Repair, revision, and complexity in syntactic analysis: An electrophysiological differentiation. *Journal of cognitive neuroscience*, 15(1):98–110.

Keuleers, E., Stevens, M., Mandera, P., and Brysbaert, M. (2015). Word knowledge in the crowd: Measuring vocabulary size and word prevalence in a massive online experiment. *The Quarterly Journal of Experimental Psychology*, 68(8):1665–1692.

King, J. W. and Kutas, M. (1995). Who did what and when? using word-and clause-level erps to monitor working memory usage in reading. *Journal of cognitive neuroscience*, 7(3):376–395.

Kircher, T. T., Brammer, M., Andreu, N. T., Williams, S. C., and McGuire, P. K. (2001). Engagement of right temporal cortex during processing of linguistic context. *Neuropsychologia*, 39(8):798–809.

Kluender, R. and Kutas, M. (1993). Bridging the gap: Evidence from ERPs on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience*, 5(2):196–214.

Koehn, P. (2005). Europarl: A parallel corpus for statistical machine translation. In *MT summit*, volume 5, pages 79–86.

Kryuchkova, T., Tucker, B. V., Wurm, L. H., and Baayen, R. H. (2012). Danger and usefulness are detected early in auditory lexical processing: Evidence from electroencephalography. *Brain and Language*, 122(2):81–91.

Kuiper, K. (1996). Smooth talkers. *The linguistic performance of auctioneers and sportscasters. Mawah, NJ: Lawrence Erlabum Associates.*

Kuiper, K., Van Egmond, M.-E., Kempen, G., and Sprenger, S. (2007). Slipping on superlemmas: Multi-word lexical items in speech production. *The Mental Lexicon*, 2(3):313–357.

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain research*, 1146:23–49.

Kuperman, V., Schreuder, R., Bertram, R., and Baayen, R. H. (2009). Reading polymorphemic dutch compounds: Toward a multiple route model of lexical processing. *Journal of Experimental Psychology: Human Perception and Performance*, 35(3):876.

Kutas, M. and Van Petten, C. (1994). Psycholinguistics electrified. *Handbook of psycholinguistics*, pages 83–143.

Landauer, T. K. and Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104(2):211.

Lau, E. F., Phillips, C., and Poeppel, D. (2008). A cortical network for semantics:(de) constructing the n400. *Nature Reviews Neuroscience*, 9(12):920.

Laubrock, J., Kliegl, R., and Engbert, R. (2006). SWIFT explorations of age differences in eye movements during reading. *Neuroscience & Biobehavioral Reviews*, 30(6):872–884.

Lazaridou, A., Marelli, M., Zamparelli, R., and Baroni, M. (2013). Compositional-ly derived representations of morphologically complex words in distributional semantics. In *ACL (1)*, pages 1517–1526.

Lensink, S. E., Schiller, N. O., and Verhagen, A. (submitted). Old and young: How language experiences (do not) shape the reading of lexical bundles.

Lensink, S. E., Verdonschot, R., Schiller, N. O., and Tamaoka, K. (in preparation). Reading japanese lexical bundles.

Lensink, S. E., Verhagen, A., Schiller, N. O., and Baayen, R. H. (submitted). Keeping it apart: On using a discriminative approach to study the nature and processing of multi-word units.

Lin, X. and Zhang, D. (1999). Inference in generalized additive mixed modelsby using smoothing splines. *Journal of the royal statistical society: Series b (statistical methodology)*, 61(2):381–400.

Linke, M., Bröker, F., Ramscar, M., and Baayen, H. (2017). Are baboons learning "orthographic" representations? Probably not. *PloS one*, 12(8):e0183876.

Lõo, K., Järvikivi, J., and Baayen, R. H. (2017). Whole-word frequency and inflectional paradigm size facilitate estonian case-inflected noun processing. *Cognition*.

Lõo, K., Järvikivi, J., Tomaschek, F., Tucker, B. V., and Baayen, R. H. (2018). Production of estonian case-inflected nouns shows whole-word frequency and paradigmatic effects. *Morphology*, pages 1–27.

Luce, P. A. and Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, 19(1):1.

Lund, K. and Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers*, 28(2):203–208.

Martín-Loeches, M., Muñoz, F., Casado, P., Melcon, A., and Fernández-Frías, C. (2005). Are the anterior negativities to grammatical violations indexing working memory? *Psychophysiology*, 42(5):508–519.

McClelland, J. L. and Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive psychology*, 18(1):1–86.

McGowan, V. A. and Reichle, E. D. (2018). The "risky" reading strategy revisited: New simulations using ez reader. *Quarterly Journal of Experimental Psychology*, 71(1):179–189.

McWhinney, S. R., Tremblay, A., Chevalier, T. M., Lim, V. K., and Newman, A. J. (2016). Using cforest to analyze diffusion tensor imaging data: A study of white matter integrity in healthy aging. *Brain connectivity*, 6(10):747–758.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119.

Milin, P., Feldman, L. B., Ramscar, M., Hendrix, P., and Baayen, R. H. (2017). Discrimination in lexical decision. *PLoS One*, 12(2):e0171935.

Milin, P., Kuperman, V., Kostic, A., and Baayen, R. H. (2009). Paradigms bit by bit: An information theoretic approach to the processing of paradigmatic structure in inflection and derivation. *Analogy in grammar: Form and acquisition*, pages 214–252.

Mitchell, J. and Lapata, M. (2008). Vector-based models of semantic composition. In *ACL*, pages 236–244.

Miwa, K., Libben, G., and Ikemoto, Y. (2017). Visual trimorphemic compound recognition in a morphographic script. *Language, Cognition and Neuroscience*, 32(1):1–20.

Mueller, J. L., Hahne, A., Fujii, Y., and Friederici, A. D. (2005). Native and nonnative speakers' processing of a miniature version of Japanese as revealed by ERPs. *Journal of Cognitive Neuroscience*, 17(8):1229–1244.

Müller, H. M., King, J. W., and Kutas, M. (1997). Event-related potentials elicited by spoken relative clauses. *Cognitive Brain Research*, 5(3):193–203.

Norris, D. and McQueen, J. M. (2008). Shortlist B: a Bayesian model of continuous speech recognition. *Psychological Review*, 115(2):357.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, 9(1):97–113.

Oostdijk, N., Reynaert, M., Hoste, V., and Schuurman, I. (2013). The construction of a 500-million-word reference corpus of contemporary written Dutch. In *Essential speech and language technology for Dutch*, pages 219–247. Springer.

Pawley, A. and Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. *Language and Communication*, 191:225.

Penolazzi, B., Hauk, O., and Pulvermüller, F. (2007). Early semantic context integration and lexical access as revealed by event-related brain potentials. *Biological Psychology*, 74(3):374–388.

Pinker, S. and Ullman, M. T. (2002). The past and future of the past tense. *Trends in cognitive sciences*, 6(11):456–463.

Pollatsek, A., Hyönä, J., and Bertram, R. (2000). The role of morphological constituents in reading finnish compound words. *Journal of Experimental Psychology: Human Perception and Performance*, 26(2):820.

Pylkkänen, L., Feintuch, S., Hopkins, E., and Marantz, A. (2004). Neural correlates of the effects of morphological family frequency and family size: an meg study. *Cognition*, 91(3):B35–B45.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ramscar, M. (2013). Suffixing, prefixing, and the functional order of regularities in meaningful strings. *Psihologija*, 46(4):377–396.

Ramscar, M., Dye, M., and McCauley, S. M. (2013). Error and expectation in language learning: The curious absence of mouses in adult speech. *Language*, 89(4):760–793.

Ramscar, M., Hendrix, P., Shaoul, C., Milin, P., and Baayen, H. (2014). The myth of cognitive decline: Non-linear dynamics of lifelong learning. *Topics in cognitive science*, 6(1):5–42.

Ramscar, M. and Port, R. (2015). Categorization (without categories). In Dąbrowska, E. and Divjak, D., editors, *Handbook of Cognitive Linguistics*, pages 75–99. De Gruyter, Berlin.

Ramscar, M. and Yarlett, D. (2007). Linguistic self-correction in the absence of feedback: A new approach to the logical problem of language acquisition. *Cognitive Science*, 31(6):927–960.

Ramscar, M., Yarlett, D., Dye, M., Denny, K., and Thorpe, K. (2010). The effects of feature-label-order and their implications for symbolic learning. *Cognitive Science*, 34(6):909–957.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, 124(3):372.

Rayner, K., Castelhano, M. S., and Yang, J. (2009). Eye movements and the perceptual span in older and younger readers. *Psychology and aging*, 24(3):755.

Rayner, K., Reichle, E. D., Stroud, M. J., Williams, C. C., and Pollatsek, A. (2006). The effect of word frequency, word predictability, and font difficulty on the eye movements of young and older readers. *Psychology and aging*, 21(3):448.

Rayner, K., Yang, J., Schuett, S., and Slattery, T. J. (2014). The effect of foveal and parafoveal masks on the eye movements of older and younger readers. *Psychology and aging*, 29(2):205.

Reichle, E. D., Pollatsek, A., Fisher, D. L., and Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological review*, 105(1):125.

Reichle, E. D., Rayner, K., and Pollatsek, A. (1999). Eye movement control in reading: Accounting for initial fixation locations and refixations within the ez reader model. *Vision research*, 39(26):4403–4411.

Reichle, E. D., Rayner, K., and Pollatsek, A. (2012). Eye movements in reading versus nonreading tasks: Using ez reader to understand the role of word/stimulus familiarity. *Visual cognition*, 20(4-5):360–390.

Reifegerste, J., Meyer, A. S., and Zwitserlood, P. (2017). Inflectional complexity and experience affect plural processing in younger and older readers of dutch and german. *Language, Cognition and Neuroscience*, 32(4):471–487.

Rescorla, R. A., Wagner, A. R., et al. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical conditioning II: Current research and theory*, 2:64–99.

Roehm, D., Bornkessel-Schlesewsky, I., Rösler, F., and Schlesewsky, M. (2007). To predict or not to predict: Influences of task and strategy on the processing of semantic relations. *Journal of Cognitive Neuroscience*, 19(8):1259–1274.

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386.

Samek, W., Wiegand, T., and Müller, K.-R. (2017). Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. *arXiv preprint arXiv:1708.08296*.

Sereno, S. C., Brewer, C. C., and O'Donnell, P. J. (2003). Context effects in word recognition: Evidence for early interactive processing. *Psychological Science*, 14(4):328–333.

Sereno, S. C., Rayner, K., and Posner, M. I. (1998). Establishing a time-line of word recognition: evidence from eye movements and event-related potentials. *Neuroreport*, 9(10):2195–2200.

Shafto, M. A. and Tyler, L. K. (2014). Language in the aging brain: the network dynamics of cognitive decline and preservation. *Science*, 346(6209):583–587.

Shaoul, C., Baayen, R. H., and Westbury, C. F. (2014a). N-gram probability effects in a cloze task. *The Mental Lexicon*, 9(3):437–472.

Shaoul, C., Schilling, N., Bitschnau, S., Arppe, A., Hendrix, P., and Baayen, R. (2014b). NDL2: Naive discriminative learning. R package version 1.901, development version available upon request.

Shaoul, C. and Westbury, C. (2010). Exploring lexical co-occurrence space using hidex. *Behavior Research Methods*, 42(2):393–413.

Shaoul, C. and Westbury, C. (2011). Formulaic sequences: Do they exist and do they matter? *The Mental Lexicon*, 6(1):171–196.

Shaoul, C., Westbury, C. F., and Baayen, H. R. (2013). The subjective frequency of word n-grams. *Psihologija*, 46(4):497–537.

Siyanova-Chanturia, A. (2013). Eye-tracking and ERPs in multi-word expression research: A state-of-the-art review of the method and findings. *The Mental Lexicon*, 8(2):245–268.

Siyanova-Chanturia, A. (2015). On the "holistic" nature of formulaic language. *Corpus Linguistics and Linguistic Theory*, 11(2):285–301.

Siyanova-Chanturia, A., Conklin, K., and Schmitt, N. (2011a). Adding more fuel to the fire: An eye-tracking study of idiom processing by native and non-native speakers. *Second Language Research*, 27(2):251–272.

Siyanova-Chanturia, A., Conklin, K., and Van Heuven, W. J. (2011b). Seeing a phrase "time and again" matters: The role of phrasal frequency in the processing of multi-word sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(3):776.

Snider, N. and Arnon, I. (2012). A unified lexicon and grammar? Compositional and non-compositional phrases in the lexicon. *Frequency Effects in Language Representation*, pages 127–163.

Sosa, A. V. and MacFarlane, J. (2002). Evidence for frequency-based constituents in the mental lexicon: Collocations involving the word of. *Brain and Language*, 83(2):227–236.

Sprenger, S. A., Levelt, W. J., and Kempen, G. (2006). Lexical access during the production of idiomatic phrases. *Journal of Memory and Language*, 54(2):161–184.

Steinhauer, K., Connolly, J. F., Stemmer, B., and Whitaker, H. (2008). Event-related potentials in the study of language. *Concise Encyclopedia of Brain and Language*, pages 91–104.

Steinhauer, K. and Drury, J. E. (2012). On the early left-anterior negativity (ELAN) in syntax studies. *Brain and Language*, 120(2):135–162.

Steinhauer, K., Drury, J. E., Portner, P., Walenski, M., and Ullman, M. T. (2010). Syntax, concepts, and logic in the temporal dynamics of language comprehension: Evidence from event-related potentials. *Neuropsychologia*, 48(6):1525–1542.

Strobl, C., Malley, J., and Tutz, G. (2009). An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological methods*, 14(4):323.

Tagliamonte, S. A. and Baayen, R. H. (2012). Models, forests, and trees of york english: Was/were variation as a case study for statistical practice. *Language variation and change*, 24(2):135–178.

Tomasello, M. (2009). *Constructing a language*. Harvard university press.

Tremblay, A., Asp, E., Johnson, A., Migdal, M. Z., Bardouille, T., and Newman, A. J. (2016). What the networks tell us about serial and parallel processing. *The Mental Lexicon*, 11(1):115–160.

Tremblay, A., Baayen, H., Derwing, B., Libben, G., Tucker, B. V., and Westbury, C. (2012). Empirical evidence for an inflationist lexicon. *Yearbook Phraseoly*, 3:109–126.

Tremblay, A. and Baayen, R. H. (2010). Holistic processing of regular four-word sequences: A behavioral and ERP study of the effects of structure, frequency, and probability on immediate free recall. *Perspectives on formulaic language: Acquisition and communication*, pages 151–173.

Tremblay, A., Derwing, B., and Libben, G. (2009). Are lexical bundles stored and processed as single units? *Working Papers of the Linguistics Circle*, 19(1):258.

Tremblay, A., Derwing, B., Libben, G., and Westbury, C. (2011). Processing advantages of lexical bundles: evidence from self-paced reading and sentence recall tasks. *Language Learning*, 61(2):569–613.

Tremblay, A. and Tucker, B. V. (2011). The effects of n-gram probabilistic measures on the recognition and production of four-word sequences. *The Mental Lexicon*, 6(2):302–324.

Underwood, G., Schmitt, N., and Galpin, A. (2004). The eyes have it. *Formulaic sequences: Acquisition, processing, and use*, 9:153.

Van Den Brink, D., Brown, C. M., and Hagoort, P. (2001). Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. *Journal of Cognitive Neuroscience*, 13(7):967–985.

Van Den Brink, D. and Hagoort, P. (2004). The influence of semantic and syntactic context constraints on lexical selection and integration in spoken-word comprehension as revealed by ERPs. *Journal of Cognitive Neuroscience*, 16(6):1068–1084.

Van Rij, J., Hollebrandse, B., and Hendriks, P. (2016). Children's eye gaze reveals their use of discourse context in object pronoun resolution. *Experimental Perspectives on Anaphora Resolution. Information Structural Evidence in the Race for Salience. Boston: De Gruyter*, pages 267–293.

Vespignani, F., Canal, P., Molinaro, N., Fonda, S., and Cacciari, C. (2010). Predictive mechanisms in idiom comprehension. *Journal of Cognitive Neuroscience*, 22(8):1682–1700.

Vigneau, M., Beaucousin, V., Herve, P.-Y., Duffau, H., Crivello, F., Houde, O., Mazoyer, B., and Tzourio-Mazoyer, N. (2006). Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing. *Neuroimage*, 30(4):1414–1432.

Vigneau, M., Beaucousin, V., Hervé, P.-Y., Jobard, G., Petit, L., Crivello, F., Mellet, E., Zago, L., Mazoyer, B., and Tzourio-Mazoyer, N. (2011). What is right-hemisphere contribution to phonological, lexico-semantic, and sentence processing?: Insights from a meta-analysis. *Neuroimage*, 54(1):577–593.

Vitu, F., McConkie, G. W., Kerr, P., and O'Regan, J. K. (2001). Fixation location effects on fixation durations during reading: An inverted optimal viewing position effect. *Vision Research*, 41(25-26):3513–3533.

Widrow, B. and Hoff, M. E. (1960). Adaptive switching circuits. Technical report, STANFORD UNIV CA STANFORD ELECTRONICS LABS.

Wieling, M., Montemagni, S., Nerbonne, J., and Baayen, R. H. (2014). Lexical differences between Tuscan dialects and standard Italian: Accounting for geographic and sociodemographic variation using generalized additive mixed modeling. *Language*, 90(3):669–692.

Winter, B. and Wieling, M. (2016). How to analyze linguistic change using mixed models, Growth Curve Analysis and Generalized Additive Modeling. *Journal of Language Evolution*, 1(1):7–18.

Wood, S. (2006). *Generalized additive models: an introduction with R*. CRC press.

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(1):3–36.

Wood, S. N., Goude, Y., and Shaw, S. (2015). Generalized additive models for large data sets. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 64(1):139–155.

Wray, A. (2012). What do we (think we) know about formulaic language? An evaluation of the current state of play. *Annual Review of Applied Linguistics*, 32:231–254.

Wurm, L. H. and Fisicaro, S. A. (2014). What residualizing predictors in regression analyses does (and what it does not do). *Journal of Memory and Language*, 72:37–48.

Yates, M., Friend, J., and Ploetz, D. M. (2008). The effect of phonological neighborhood density on eye movements during reading. *Cognition*, 107(2):685–692.