



Universiteit  
Leiden  
The Netherlands

## Processing Lexical Bundles

Lensink, S.E.

### Citation

Lensink, S. E. (2020, June 4). *Processing Lexical Bundles*. LOT dissertation series. LOT, Amsterdam. Retrieved from <https://hdl.handle.net/1887/92886>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/92886>

**Note:** To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/92886> holds various files of this Leiden University dissertation.

**Author:** Lensink, S.E.

**Title:** Processing Lexical Bundles

**Issue Date:** 2020-06-04

# Processing Lexical Bundles

Published by

LOT  
Kloveniersburgwal 48  
1012 CX Amsterdam  
The Netherlands

phone: +31 20 525 2461  
e-mail: [lot@uva.nl](mailto:lot@uva.nl)  
<http://www.lotschool.nl>

The illustration on the cover shows a fragment of an alluvial diagram of all trigrams from the text of this dissertation beginning with 'hearing', 'reading', and 'speaking', the three big themes of this dissertation. The full illustration is printed on the back side of the propositions. To create the image, the author used the *alluvial* package in R, and the code is available upon request.

ISBN: 978-94-6093-350-9  
NUR: 616

Copyright © 2020 Saskia E. Lensink. All rights reserved.

# Processing Lexical Bundles

## Proefschrift

ter verkrijging van  
de graad van Doctor aan de Universiteit Leiden,  
op gezag van Rector Magnificus prof. mr. C.J.J.M. Stolker,  
volgens besluit van het College voor Promoties  
te verdedigen op donderdag 4 juni 2020  
klokke 15:00 uur

door

Saskia E. Lensink

geboren 7 mei 1989  
te Woudenberg, Nederland



To all the parts that make me whole.





---

## Contents

---

Acknowledgements . . . . .	xi
<b>1 Introduction</b>	<b>1</b>
1.1 The What . . . . .	2
1.2 The Why . . . . .	2
1.3 The How . . . . .	4
1.3.1 Reading . . . . .	4
1.3.2 Listening . . . . .	7
1.3.3 Speaking . . . . .	7
1.4 Quantifying processing . . . . .	9
1.4.1 Statistical modeling . . . . .	9
1.4.2 Computational modeling . . . . .	13
1.5 This dissertation . . . . .	15
<b>2 Reading</b>	<b>17</b>
2.1 Introduction . . . . .	18
2.1.1 Language processing in younger and older adults . . . . .	19
2.1.2 Reading lexical bundles in younger and older adults . . . . .	20
2.1.3 The present study . . . . .	20
2.2 Materials and Methods . . . . .	21
2.2.1 Participants . . . . .	21
2.2.2 Stimuli . . . . .	21
2.2.3 Procedure . . . . .	22
2.2.4 Generalized Additive Mixed-Effects models . . . . .	22
2.3 Results . . . . .	23
2.3.1 First fixation durations . . . . .	23
2.3.2 Second fixation durations . . . . .	26
2.3.3 Number of fixations . . . . .	29
2.4 General discussion . . . . .	31
2.4.1 The Inverted Frequency Effect . . . . .	32

2.4.2	Limitations and future directions . . . . .	33
2.5	Conclusion . . . . .	35
<b>3</b>	<b>Listening</b>	<b>37</b>
3.1	Introduction . . . . .	38
3.1.1	Previous work on the time course of multi-word unit processing . . . . .	39
3.1.2	Current study . . . . .	41
3.2	Materials and Methods . . . . .	41
3.2.1	Participants . . . . .	41
3.2.2	Stimuli . . . . .	42
3.2.3	Procedure . . . . .	43
3.2.4	EEG recordings . . . . .	43
3.2.5	Conditional inference random forest analysis . . . . .	44
3.3	ERP results . . . . .	45
3.4	CForest modeling results . . . . .	47
3.4.1	Results Conditional Inference Tree . . . . .	48
3.4.2	Stage 1: Prediction and bottom-up information, 0-270 ms . . . . .	49
3.4.3	Stage 2: Inhibition and competition, 270-402 ms . . . . .	56
3.4.4	Stage 3: Lexical integration, 402-800 ms . . . . .	58
3.5	General discussion . . . . .	63
<b>4</b>	<b>Reading and speaking</b>	<b>67</b>
4.1	Introduction . . . . .	68
4.1.1	Including multi-word units in models of lexical access . . . . .	69
4.1.2	Computational modeling of multi-word units . . . . .	70
4.2	NDL model . . . . .	71
4.2.1	How the model works . . . . .	71
4.2.2	NDL measures of lexical access . . . . .	73
4.3	Generalized additive mixed models . . . . .	76
4.4	Eye-tracking experiment . . . . .	77
4.4.1	Materials . . . . .	78
4.4.2	Design . . . . .	78
4.4.3	Participants . . . . .	79
4.4.4	Procedure . . . . .	79
4.4.5	Analyses . . . . .	79
4.4.6	First Fixation Durations . . . . .	80
4.4.7	First Pass Reading Times . . . . .	81
4.4.8	Number of fixations . . . . .	84
4.4.9	Discussion eye-tracking data . . . . .	86
4.5	Production experiment . . . . .	88
4.5.1	Materials . . . . .	88
4.5.2	Design . . . . .	88
4.5.3	Participants . . . . .	89
4.5.4	Procedure . . . . .	89

4.5.5	Analyses . . . . .	89
4.5.6	Production onset latencies . . . . .	89
4.5.7	Production durations . . . . .	90
4.5.8	Discussion production data . . . . .	92
4.6	General discussion . . . . .	95
4.6.1	Lexical access of multi-word units . . . . .	96
<b>5</b>	<b>Conclusion</b>	<b>99</b>
5.1	Reading . . . . .	100
5.2	Listening . . . . .	101
5.3	Reading and speaking . . . . .	102
5.4	Overall conclusions . . . . .	104
5.5	Useful models - an outlook . . . . .	105
	Appendix A . . . . .	107
	Appendix B . . . . .	113
	Bibliography . . . . .	115
	Nederlandse samenvatting . . . . .	129
	About the author . . . . .	137



---

## Acknowledgements

---

Completing a PhD thesis is often considered a solitary and lonely task. However, looking back at the past years, I realize that I was by no means lonely, and I am grateful for all the support and companionship I received from so many people.

First and foremost, I would like to thank my supervisors Arie Verhagen and Niels Schiller, without whose supervision and support this dissertation would not exist. Thank you for your creative insights, interesting discussions, and giving me the freedom to chose my own directions.

I am grateful to the Netherlands National Graduate School of Linguistics (LOT) for awarding me a scholarship to fund my PhD. Together with 'LOT-genoten' Jeroen Breteler, Sophie Villerius, Brechje Van Osch, LOT directors Frank Wijnen and Henriette de Swart, I had regular meetings where we discussed our progress, academia, ambitions, and life.

Martijn Wieling and Jacolien van Rij introduced me to statistical modeling. I am greatly indebted to them for this as they convinced me that I would be able to learn how to apply these models myself, even though I understood almost nothing at first. This has set me on a path of further developing my machine learning skills. I furthermore owe special thanks to Antoine Tremblay, who patiently introduced me to ensemble machine learning models, and whose enthusiasm and creative insights have been of invaluable importance in the research reported in Chapter 3.

My dissertation would have never been complete without the inspiring collaboration with the quantitative linguistics group at the university of Tübingen. I am especially grateful to Harald Baayen and Tineke Baayen Oudshoorn, who have invited me over to Tübingen a couple of times and made me feel right at home. It was an honor to have had the opportunity to work with Harald and applying some of my ideas within the NDL framework. Besides Harald and Tineke, I had a great time with Petar Milin, Tino Sering, Fabian Tomaschek, June Hendrix-Sun, Peter Hendrix, Karlina Denistia, and Denis Arnold.

In 2015, I spent the summer at the university of Chicago, where I attended

the LSA Summer Institute — one of the highlights of my time as a PhD candidate. I learned so much from the courses and teachers, and made great friends. Thank you Amos Teo, Grant Berry, Darcy Rose, Kailen Shantz, Bodo Winter, and Andy Wedel for the memories, inspiring talks, and many beers we shared. My time at Leiden was filled with laughs, ideas, and support of my colleagues from LUCL, in particular my roommates Maaïke van Naerssen and Evelyn Bosma. Thank you for the great times, Jessy Nixon, Alex Reuneker, Véronique Verhagen, Elly Dutton, Bobby Ruijgrok, Olga Kepinska, Cesko Voeten, Aliza Glasbergen-Plas, Leticia Pablos Robles, Xander Vertegaal, Stefan Norbruis, Bob Schoemaker, Andreas Krogull, Martine Bruil, Kate Bellamy, María Carme Parafita Couto, Katja Lubina, and Marijn van 't Veer. Besides my colleagues, I am indebted to all people who participated in my experiments, all the students I had the pleasure to meet teaching different courses, Maxime Tulling, Lilian Ye, for their endless efforts in helping me collecting data, their critical questions, and creative energy, and especially Jos Pacilly without whom none of my experiments would have been a success.

Special thanks go to Rinus Verdonschot, with whom I had the pleasure to collaborate on several projects, even before I started my PhD. He introduced me to Katsuo Tamaoka, to whom I am grateful for the interesting conversations we had and the collaboration we set up. Thanks to Rinus I got the chance to share my knowledge on statistical modeling at Waseda University, which I consider another highlight of my PhD.

Finishing a PhD without the help and support of friends, and the liters of coffee, tea, wine, and beer we shared, is impossible. I am thankful to Marte Boonen, Pauline van Deursen, Majelle Verbraak, Anoushka Kloosterman, Laura Berveling, Mariska Veenendaal, Gert Poot, Daša Abtová, Stefan Penders, Joey Weidema, Fokelien Kootstra, Carlos Merchán Pulgarín, Pip Schuijt, Nina Ouddeken, Jurgen van den Heuvel, Benjamin Suchard, Hilde Gunnink, Bas Clercx, and Redmer Kronemeijer. My buddies both above and below water, especially Bart Braun, Nathalie Raats, Mariska Beijer, and Marit van Elsas, have helped me to relax and enjoy (underwater) life.

Wilma en Joop, dank jullie wel dat jullie altijd voor me klaarstonden, vooral op momenten dat het echt nodig was. Van jullie heb ik geleerd om hard te werken en nooit op te geven. Ik dacht dat een PhD afronden zwaar was, tot het noodlot toesloeg en we leerden wat écht moeilijk is. Papa, ik mis je elke dag.

Mi querido chanchi wawi, 'amortje' Diego Vásquez Villaseca, mil gracias por siempre estar conmigo, creyendo en mi, apoyandome en las buenas y en las malas. Tu amor y apoyo incondicional fueron un gran soporte para llegar a este logro. Te amo.