

Fluency in Native and Nonnative English speech:

Theory, description, implications

by

Sandra Götz

B.A. (University of Bayreuth, Germany)

M.A. (Justus Liebig University Giessen, Germany)

This thesis is presented for the degree of Doctor of Philosophy

Department of International Studies

Faculty of Arts

Macquarie University, Sydney

October 2011

To my mother

Table of Contents

Abstract	v
Declaration	vii
Acknowledgements	ix
List of Abbreviations and Acronyms	xi
List of Tables	xiii
List of Figures	xv
 CHAPTER 1: INTRODUCTION AND PREVIEW.....	1
1.1 INTRODUCTION.....	1
1.2 FLUENCY IN NATIVE AND NON-NATIVE ENGLISH SPEECH: PREVIEW	2
 CHAPTER 2: FLUENCY IN NATIVE ENGLISH SEECH	5
2.1 INVESTIGATED VARIABLES	6
2.2 FLUENCEMES OF PRODUCTION.....	8
2.2.1 <i>Temporal variables</i>	9
2.2.1.1 <i>Speech rate</i>	10
2.2.1.2 <i>The mean length of runs</i>	11
2.2.1.3 <i>Unfilled pauses</i>	13
2.2.1.4 <i>The phonation/time ratio</i>	14
2.2.1.5 <i>Summary</i>	15
2.2.2 <i>Formulaic sequences</i>	15
2.2.2.1 <i>Characteristics of formulaic sequences</i>	16
2.2.2.2 <i>Formulaic sequences and fluency</i>	20
2.2.2.3 <i>Analyzing formulaic sequences</i>	21
2.2.2.4 <i>Summary</i>	22
2.2.3 <i>Performance phenomena</i>	23
2.2.3.1 <i>Speech management strategies and fluency enhancement strategies</i>	24
2.2.3.2 <i>Discourse markers and smallwords</i>	27
2.2.3.3 <i>Summary</i>	28
2.3 FLUENCEMES OF PERCEPTION.....	28
2.3.1 <i>Accuracy</i>	29
2.3.2 <i>Idiomacity</i>	30
2.3.3 <i>Intonation</i>	32
2.3.4 <i>Accent</i>	34
2.3.5 <i>Pragmatic features</i>	35
2.3.6 <i>Lexical diversity/register</i>	36
2.3.7 <i>Sentence structure</i>	38
2.4 NON-VERBAL VARIABLES OF FLUENCY.....	38
2.5 A THEORETICAL APPROACH TO FLUENCY IN SPEECH.....	39
 CHAPTER 3: FLUENCY IN NON-NATIVE ENGLISH SPEECH.....	43
3.1 LEARNING ENGLISH IN GERMANY	44
3.2 FLUENCEMES OF PRODUCTION.....	46
3.2.1 <i>Temporal variables</i>	47
3.2.1.1 <i>Speech rate</i>	49
3.2.1.2 <i>The mean length of runs</i>	50
3.2.1.3 <i>Unfilled pauses</i>	50

3.2.1.4 The phonation/time ratio.....	51
3.2.1.5 Summary	52
3.2.2 Formulaic sequences.....	53
3.2.3 Performance phenomena.....	55
3.2.3.1 Speech management strategies and fluency enhancement strategies.....	57
3.2.3.2 Discourse Markers and smallwords.....	59
3.2.3.3 Summary	60
3.2.4 Summary of fluencemes of productive fluency	60
3.3 FLUENCEMES OF PERCEPTION	61
3.3.1 Accuracy.....	62
3.3.2 Idiomaticity	64
3.3.3 Intonation	65
3.3.4 Accent.....	67
3.3.5 Pragmatic features	71
3.3.6 Lexical diversity.....	74
3.3.7 Sentence structure/register.....	75
3.4 NON-VERBAL VARIABLES OF FLUENCY	77
3.5 A THEORETICAL APPROACH TO FLUENCY IN NON-NATIVE SPEECH	78
CHAPTER 4: DATABASE AND METHODOLOGY	81
4.1 LEARNER CORPUS RESEARCH.....	81
4.2 CORPORA AND TOOLS	83
4.2.1 LINDSEI-GE & LOCNEC.....	83
4.2.2 Corpus analysis Software: WordSmith Tools V. 4.0	84
4.2.3 Statistical methods	85
4.3 METHODOLOGY	86
4.3.1 Pilot Study.....	86
4.3.1.1 General information	87
4.3.1.2 Analyzed variables.....	87
4.3.1.3 Findings.....	88
4.3.1.4 Summary of findings	94
4.3.2 Consequences for the PhD project: A refined approach to fluency	95
4.3.3 Methodology of the main study: Corpus analysis plus NS ratings.....	96
4.3.3.1 Productive fluency: Quantitative and statistical data analysis	96
4.3.3.2 Perceptive fluency: Native-speaker perception of five prototypical learners.....	98
CHAPTER 5: DATA ANALYSIS OF PRODUCTIVE FLUENCY IN LINDSEI-GE VS. LOCNEC	103
5.1 PRODUCTIVE FLUENCY: LEARNER DATA VS. NATIVE-SPEAKER DATA	103
5.1.1 Speech rate	103
5.1.2 Mean length of runs	104
5.1.3 Unfilled pauses.....	107
5.1.3.1 Total number of unfilled pauses.....	107
5.1.3.2 Unfilled pauses within clauses (UPWCL)	108
5.1.3.3 Unfilled pauses within constituents (UPWCON)	109
5.1.4 Formulaic sequences.....	111
5.1.4.1 Overview of findings	111
5.1.4.2 3-grams in LINDSEI-GE vs. LOCNEC.....	112
5.1.4.3 4-grams in LINDSEI-GE vs. LOCNEC.....	114
5.1.5 Repeats	116
5.1.6 Filled pauses	120
5.1.6.1 Total number of filled pauses (FPs).....	120
5.1.6.2 Filled pauses within clauses (FPWCL).....	121
5.1.6.2 Filled pauses within constituents (FPWONL).....	122
5.1.7 Discourse markers	124
5.1.7.1 well	126
5.1.7.2 you know	127
5.1.7.3 like	128
5.1.7.4 Variation in discourse marker use	129

5.1.8 Smallwords	130
5.1.8.1 Overall frequencies of smallwords.....	130
5.1.8.2 sort of/sorta	131
5.1.8.3 kind of/kinda	132
5.1.8.4 quite	133
5.1.8.5 Variation in smallword use	134
5.1.9 Summary, caveats and implications.....	135
5.2 CORRELATIONS OF FLUENCES AND DIFFERENT FLUENCY TYPES IN LINDSEI-GE AND LOCNEC:	
BRINGING TOGETHER THE FLUENCES	138
5.2.1 Correlations of individual fluences in LOCNEC	138
5.2.2 Correlations of individual fluences in LINDSEI-GE.....	140
5.2.3 Fluency groups in LOCNEC.....	142
5.2.4 Fluency groups in LINDSEI-GE.....	146
5.2.5 Summary.....	149
5.3 IMPLICATIONS	150
5.3.1 Implications for classroom-methodology	151
5.3.2 Implications for teacher education	154
5.3.3 Implications for teaching and learning materials.....	155
5.3.4 Implications for assessment of spoken output.....	157
5.4 Summary.....	158
CHAPTER 6: PERCEPTIVE FLUENCY OF PROTOTYPICAL LEARNER TYPES IN LINDSEI-GE.....	160
6.1 SELECTION OF THE FIVE LEARNERS	160
6.2 THE QUESTIONNAIRE	161
6.3 THE RATERS	166
6.4 THE RATINGS.....	168
6.4.1 Learner 1, ger024	168
6.4.2 Learner 2, ger027	170
6.4.3 Learner 3, ger028	171
6.4.4 Learner 4, ger041	173
6.4.5 Learner 5, ger001	174
6.4.7 Combinations and correlations of ratings with individual variables	175
6.4.6 Feedback on the survey	179
6.5 THE FOCUS GROUPS	180
6.6 SUMMARY, CAVEATS AND IMPLICATIONS.....	184
6.6.1 Caveats	184
6.6.2 Implications for ELT.....	185
6.6.3 Summary.....	187
CHAPTER 7: SUMMARY AND PROSPECTS FOR FUTURE RESEARCH.....	188
7.1 SUMMARY OF MAJOR FINDINGS	188
7.2 PROSPECTS FOR FUTURE RESEARCH.....	191
REFERENCES	194
APPENDIX	216
APPENDIX A1: QUESTIONNAIRE OF PILOT STUDY	216
APPENDIX A2: LISTS OF 3-GRAMS AND 4-GRAMS IN LOCNEC AND LINDSEI-GE	216
APPENDIX A3: FINAL ETHICS APPROVAL.....	216

Abstract

Most of the fluency-related research so far has focused on one of the following aspects: (1) temporal variables in speech production, (2) the use of formulaic language or (3) certain performance phenomena which contribute to a perception of naturalness in speech, such as discourse markers. The mainstream linguistic approach to fluency focuses on temporal variables of speech production and is generally accepted as being the best indicator of a learner's fluency. Recent studies have shown that temporal variables as well as the number of errors correlate highly with native speaker assessments of the learners' overall oral proficiency. However, they have focussed on a small amount of raters and neglected other variables that are equally responsible for a perception of oral proficiency on the sides of the listeners, such as accent, idiomaticity, lexical diversity, register, sentence structure, intonation, or pragmatic features.

Thus, the present study will take into consideration a combination of these approaches: Firstly, a quantitative analysis of the error-tagged version of the 90,000-word German component of the Louvain International Database of Spoken English Interlanguage (LINDSEI-Ger) as compared to the Louvain Corpus of Native English Conversation (LOCNEC) will be presented. Thus, areas will be revealed, in which, on the one hand, advanced German learners of English still deviate strongly from the native target norm and in which they have already approximated to the target norm on the other. Secondly, based on the quantitative findings of the corpus analysis, five learners, that represent certain prototypical accuracy and fluency learner types, are subject to 50 native speakers' ratings of (1) their overall oral proficiency, and (2) seven other perceptive fluency variables. The overall ratings will be correlated with the ratings for the individual variables in order to reveal which of the variables has the strongest impact on an overall perception of fluency.

Finally, some language-pedagogical implications for the improvement of the oral proficiency in learner language derived from these findings will be presented.

Declaration

I certify that the research described in this dissertation has not already been submitted for any other degree. I certify that to the best of my knowledge all sources used and any help received in the preparation of this dissertation have been acknowledged. Ethics committee approval has been obtained for this research project.¹

Sandra Götz

14 October 2011

Publications:

The following publications have emanated from this study:

Brand, Christiane & Sandra Götz (2011): "Fluency versus accuracy in advanced spoken learner language: A multi-method approach", *International Journal of Corpus Linguistics* 16 (2), 255-275.

Götz, Sandra (forthcoming): "Temporal fluency variables in native and non-native English speech: Corpus findings and language-pedagogical implications". In: James Thomas and Alex Boulton (eds.), *Proceedings from the 8th Teaching and Language Corpora Conference 2010 in Brno, Czech Republic*, July 2010. Brno: Masarky University press.

Götz, Sandra (forthcoming): "How fluent are advanced German learners of English (perceived to be)? Corpus findings vs. native-speaker perception". In: Magnus Huber & Joybrato Mukherjee (Hrsg.), *Corpus Linguistics and Variation in English: Focus on non-native Englishes*. University of Helsinki: Varieng Electronic Series.

¹ Macuarie University, Ethics Review Committee (Human research): HE03MAY2008-D05765

Acknowledgements

I owe a huge debt of gratitude to a number of people who helped or encouraged me throughout my doctoral research and without who this project would never have been completed. It is my greatest pleasure to take the opportunity to thank some of them.

First and foremost, I would like to thank my supervisors Martina Möllering and Joybrato Mukherjee: I am very grateful to Martina Möllering for being so enthusiastic about my project, which highly motivated me throughout the whole process. She warmly welcomed me at her department for two research stays in Sydney and gave generously of her time to discuss the major steps of the experimental analysis of the present study in detail. She always gave the most helpful feedback and extremely nice, supportive and motivating comments. I am also very grateful to Joybrato Mukherjee, who has been nothing but supportive of me ever since I started this project. He always made time to discuss and shape the major ideas and concepts and gave me the most helpful advice. My work has greatly benefited from his thorough editing, constructive feedback, comments and suggestions. Both of my supervisors have always believed in me and my work and always found the time and a way to make things possible. I am highly appreciative of this. Without their enduring support, this project would not have been completed.

I am very grateful to the CECL team in Louvain-la-Neuve for providing me with the transcripts and recordings of LOCNEC on which the data analysis of this thesis is based. I also owe a great debt to the raters who took part in the survey and without who this project would not have been possible. I am very grateful to Johannes Herrmann for his expert consultations with the statistical calculations performed in this thesis. He patiently supervised all the statistical analyses that were performed with SPSS. I am also very grateful to Rosemary Bock's native-speaker expertise. She read this manuscript in its entirety and provided me with fantastic feedback and suggestions. Thank you very much, too, to Christiane Brand for giving me invaluable stylistic advice on large parts of this manuscript.

My research was supported by an international Macquarie Research Excellence Scholarship (iMQRES) and a grant from the *German Academic Exchange Service* (DAAD). This financial support enabled me to conduct two research stays in Sydney in order to perform the experimental analysis of the present project. I am very grateful for that. Also, I received various grants from the *International Graduate Centre for the Study of Culture* (GCSC) in Giessen, and from the *German Research Foundation* (DFG) which gave me the possibility to attend several conferences abroad where I presented the key components of the present research project and received valuable feedback. I am very grateful for this support.

I would also like to thank Magnus Huber for his kindness and supportive words over the years. I owe a very special thank you to Dietmar Rösler, whose kind and warm-hearted personality and supportive, motivational comments have accompanied my studies and research in the most supportive way from day one. I am very appreciative of his always offering an open door and ear to give me advice or support. I am also grateful to Michael Legutke for sharing his enthusiasm and making research so enjoyable. I would also like to thank my colleagues at the linguistics department at the University of Giessen: Christiane Brand for always having an open door for me to share and discuss all kinds of things and welcoming me with open arms, Marco Schilk for being a very

knowledgeable office mate and for critically discussing various concepts and ideas, Stefanie Dose for many insightful discussions and for her overall support and enthusiasm, Rosemary Bock for her kindness and support on all possible levels, Christopher Koch and Dieter Laufer for technical support and Thorsten Brato, Svetla Rogatcheva, Annette Obert-Sochor, Sven Saage, Kathryn Khairi-Taraki, Viktoria Künstler, Janina Werner, Patrick Maiwald, Susanne Kämmerer, Tobias Bernaisch and Simone Müller for providing a very friendly and pleasant work atmosphere in the department. I am also very grateful to my colleagues at Macquarie University, Sydney, Michelle Roberts, Ursula Ibaraki and Jennifer Heward, who were always very friendly, helpful and supportive. I would also like to thank the members of the TechAG for providing the most motivating and comfortable research and discussion forum and for the most pleasant research and dinner meetings. Thank you Nora Benitt, Katrin Biebighäuser, Silke Jahns, Olga Kamarouskaya, Michael Legutke, Gabriela Marques-Schäfer, Eva Platten, Csilla Puskás, Dietmar Rösler, Torben Schmidt, Anne Schröter and Marja Zibelius.

I am particularly grateful for everybody who made living in Giessen and Sydney a very nice and pleasant experience. Thank you for sharing so many memorable moments to Christi, David, Iva, Julia, Kerstin, Marie, Meike, Nina, Pom, Steffi, Tobi & Jazz, Luke, Louise, Marja, Steve, Tawfik, Wafaa & William.

I am particularly grateful to my closest friends and sisters who have been keeping my life in a warm, loving and caring context. Thank you ever so much for being in my life Alina, Carmen, Christiane, Doro, Eva, July, Kaija, Mary, Steffi, Viveka and Wafaa.

Finally, I owe the greatest debt to my mother. Without a lifetime of her love, support, care, understanding and encouragement, this thesis would never have seen the daylight.

List of Abbreviations and Acronyms

ACCESS	<i>Automatization in Communicative Contexts of Essential Speech Segments</i>
3G	3-gram
4G	4-gram
AR	articulation rate
CA	contrastive analysis
cf.	<i>confer</i>
CIA	contrastive interlanguage analysis
CL	corpus linguistics
CLC	computer learner corpora
CLT	communicative language teaching
CUP	Common Underlying Proficiency
DCT	discourse completion task
DDL	data-driven learning
DPU	corpora for delayed pedagogical use
e.g.	<i>exempli gratia</i>
ed.	editor
eds.	editors
EFL	English as a foreign language
ELF	English as a <i>lingua franca</i>
ELT	English language teaching
ENL	English as a native language
ERR	error
ESL	English as a second language
et al.	<i>et alii</i>
etc.	<i>et cetera</i>
FLT	foreign language teaching
FP	filled pause
FPWCL	filled pause within clause
FPWCON	filled pause within constituent
FRR	formula/run ratio
HRWC	highly recurring word combination
i.e.	<i>id est</i>
ibid.	<i>ibidem</i>
IELTS	International English Language Testing System
IPU	corpora for immediate pedagogical use
KWIC	keyword in context
L1	first language
L2	second language
LCR	learner corpus research
LINDSEI	<i>Louvain International Database of Spoken English Interlanguage</i>
LINDSEI-GE	<i>Louvain International Database of Spoken English Interlanguage – The German component</i>
LL	log likelihood
LOCNEC	<i>The Louvain Corpus of Native English Conversations</i>

MLR	mean length of runs
MON	months spent abroad in an English-speaking country
NNS	non-native speaker
nph	number of phonemes per time unit
NS	native speaker
Perl	<i>Practical Extraction and Report Language</i>
phw	per hundred words
pmw	per million words
PTR	phonation/time ratio
ptw	per thousand words
SC	self-correction
SLA	Second Language Acquisition
spm	syllables per minute
sps	syllables per second
SPSS	<i>Statistical Package for the Social Sciences</i>
spw	syllables per word
SR	speech rate
TEFL	teaching English as a foreign language
TF	temporal fluency
TTR	type token ratio
UIEX	unidiomatic expression
UP	unfilled pause
UPWCL	unfilled pause within clause
UPWCON	unfilled pause within constituent
wpc	words per clause
wpm	words per minute
wps	words per second

List of Tables

Table 2-1: Range of speech rates by words per minute for interviews in British English speech (cf. Tauroza & Allison 1990: 102).....	11
Table 4-1: Overview of fluencemes analyzed in pilot study.....	87
Table 4-2: Overall findings of the corpus-based data-analysis of productive fluenceme performance in the picture story retelling task by three selected learners of LINDSEI-GE.....	88
Table 4-3: Speech rate: Corpus analysis of speech rate in wpm vs. NS-ratings.....	90
Table 4-4: Proportion of unfilled pauses of three selected learners of LINDSEI-GE: Corpus analysis vs. NS-ratings.....	90
Table 4-5: Proportion of filled pauses of three selected learners of LINDSEI-GE: Corpus analysis vs. NS-ratings.....	90
Table 4-6: Use of discourse markers of three selected learners of LINDSEI-GE: Corpus analysis vs. NS-ratings.....	91
Table 4-7: Use of repetitions of three selected learners of LINDSEI-GE: Corpus analysis vs. NS-ratings.....	92
Table 4-8: Self-corrections of three selected learners of LINDSEI-GE: Corpus analysis vs. NS-ratings.....	92
Table 4-9: Accuracy and accent of three selected learners of LINDSEI-GE: NS-ratings.....	93
Table 4-10: Idiomaticity of three selected learners of LINDSEI-GE: NS-ratings.....	93
Table 5-1: Total number of unfilled pauses in LINDSEI-GE in relation to the LOCNEC-mean.....	108
Table 5-2: UPWCL in LINDSEI-GE in relation to the LOCNEC-mean.....	109
Table 5-3: UPWCON in LINDSEI-GE in relation to the LOCNEC mean.....	110
Table 5-4: Overview of n-gram use in LINDSEI-GE and LOCNEC.....	112
Table 5-5: Total number of 3-grams used by individual speakers in LINDSEI-GE vs. LOCNEC-mean.....	113
Table 5-6: Frequencies of 4-grams used by individual speakers in LINDSEI-GE vs. LOCNEC mean.....	114
Table 5-7: Overview of investigated groups and types of repeats (cf. Biber et al. 1999).....	116
Table 5-8: Overview of repeats phw in LINDSEI-GE and LOCNEC.....	117
Table 5-9: Frequencies of repeats of individual speakers of LINDSEI-GE vs. LOCNEC mean.....	119
Table 5-10: Frequencies of FPs used by individual speakers of LINDSEI-GE vs. LOCNEC mean.....	120
Table 5-11: Frequencies of FPWCL of individual speakers of LINDSEI-GE vs. LOCNEC mean.....	122
Table 5-12: Frequencies of FPWCON of individual speakers of LINDSEI-GE vs. LOCNEC mean.....	123
Table 5-13: Overview of overall frequencies of discourse markers <i>well</i> , <i>you know</i> and <i>like</i> in LINDSEI-GE and LOCNEC.....	124
Table 5-14: Use of the discourse marker <i>well</i> by individual speakers of LINDSEI-GE vs. LOCNEC mean.....	126
Table 5-15: Use of the discourse marker <i>you know</i> by individual speakers of LINDSEI-GE vs. LOCNEC mean.....	127
Table 5-16: Use of the discourse marker <i>like</i> by individual speakers in LINDSEI-GE vs. LOCNEC mean.....	128
Table 5-17: Overview of overall frequencies of the smallwords <i>sort of/sorta</i> , <i>kind of/kinda</i> and <i>quite</i> in LINDSEI-GE and LOCNEC.....	130
Table 5-18: Frequencies of <i>sort of/sorta</i> by individual speakers in LINDSEI-GE vs. LOCNEC mean.....	132
Table 5-19: Frequencies of <i>kind of/kinda</i> used by individual speakers of LINDSEI-GE vs. LOCNEC.....	133
Table 5-20: Frequencies of use of <i>quite</i> by individual speakers of LINDSEI-GE vs. LOCNEC.....	134

Table 5-21: Overview of findings for temporal fluencemes in LINDSEI-GE vs. LOCNEC	135
Table 5- 22: Overview of findings for fluency enhancing fluencemes in LINDSEI-GE.....	137
Table 5-23: Correlations of TF-score and fluency enhancement strategies.....	139
Table 5-24: Correlations of TF-Score and fluency enhancement strategies in LINDSEI-GE	141
Table 5-25: Means of the fluenceme performances in the three fluency groups in LOCNEC	144
Table 5-26: Summary of the major characteristics of the fluency groups of LOCNEC	145
Table 5-27: Means of the fluenceme performances in three fluenceme groups in LINDSEI-GE	148
Table 5- 28: Summary of major characteristics of three fluency groups of LINDSEI-GE.....	148
Table 6-1: Fluency and accuracy performance of five selected learners of LINDSEI-GE (cf. Brand & Götz 2011).....	161
Table 6-2: Means of analyzed variables and native speaker ratings of all variables (N=50, icc=0.89)....	175
Table 6-3: Correlations of overall perception of oral proficiency with investigated fluencemes	177

List of Figures

Figure 2-1: Formulaic Sequences in native speech	19
Figure 2-2: Fluency in native speech	40
Figure 4-1: Means of rating of the overall degree of fluency of the learners	89
Figure 5-1: Speech rate in words per minute in LINDSEI-GE (cf. Brand & Götz 2011)	104
Figure 5- 2: Mean length of runs in LINDSEI-GE vs. LOCNEC	106
Figure 5- 3: Number and frequency of 3-grams phw in LINDSEI-GE vs. LOCNEC mean	113
Figure 5- 4: Number and frequency of 4-grams phw in LINDSEI-GE vs. LOCNEC	115
Figure 5-5: Distribution of the discourse markers well, you know and like across LOCNEC and LINDSEI-GE	125
Figure 5-6: Performance profiles across the discourse markers well, you know and like of individual learners in LINDSEI-GE phw vs. the LOCNEC mean	129
Figure 5-7: Distribution of the smallwords sort of/sorta, kind of/kinda and quite in LOCNEC and LINDSEI-GE	131
Figure 5-8: Smallword profiles of individual speakers in LINDSEI-GE	135
Figure 5-9: Dendrogram obtained from hierarchical cluster analysis: Similarity of fluenceme performance of speakers of LOCNEC (cluster algorithm: Ward)	144
Figure 5-10: Dendrogram obtained from hierarchical cluster analysis: Comparison of fluenceme performance of speakers of LINDSEI-GE (cluster algorithm: Ward)	147
Figure 6-1: Raters' self-evaluations of their native language	167
Figure 6-2: Means of NS-ratings of the overall oral proficiency and seven perceptive fluencemes of learner 1 (ger024)	169
Figure 6-3: Means of NS-ratings of the overall oral proficiency and seven perceptive fluencemes of learner 2 (ger027)	170
Figure 6-4: Means of NS-ratings of the overall oral proficiency and seven perceptive fluencemes of learner 3 (ger028)	172
Figure 6- 5: Means of NS-ratings of the overall oral proficiency and seven perceptive fluencemes of learner 4 (ger041)	173
Figure 6-6: Means of NS-ratings of the overall oral proficiency and seven perceptive fluencemes of learner 5 (ger001)	174
Figure 6-7: Means of ratings for feedback on the survey	179

Chapter 1: Introduction and preview

1.1 Introduction

Every native speaker of a language, it seems, can fill time with talk easily, speak smoothly, appropriately, correctly, with ease and effortlessness. But it appears extremely difficult to pinpoint where this perception comes from. Language researchers have offered countless different aspects that contribute to defining fluency as an overall oral proficiency in speech. However, previous research on fluency has also shown that describing the linguistic substance of fluency "with a degree of consensus is notoriously difficult" (Hasselgren 2002: 147). This is because firstly, various definitions of the concept of fluency co-exist and, secondly, fluency is an epiphenomenon to which many individual (and interrelated) factors contribute. Chambers (1997: 535) states that "[s]moothness, ease and effortlessness are qualitative judgements, often based on a **global impression**" (my emphasis). Maybe this difficulty of categorization is the reason why for many teachers, learners and researchers, their overall goal in English language teaching is to prepare language learners to be able to engage in "successful communication" (Kennedy & Trofimovich 2008: 460) in the target language. In this context, therefore, "[f]luency is mainly of interest because it is related to communicative effectiveness" (Bygate 2009: 409). In this approach, the clear focus is on the intelligibility and comprehensibility of the learners' utterances, instead of speakers striving for nativelike communicational behavior (e.g. Tench 1981; Kenworthy 1987; Jenkins 2000). While this approach is the most applicable one for the great majority of English language learners around the world and an enormous body of research has been devoted to the topic, the present study will take a different stance and take into account the comparatively very small proportion of highly advanced English language learners who strive for nativelike fluency in their speech. Consequently, for investigations of highly advanced learners, intelligibility and comprehensibility can be taken for granted and the major focus of research is thus to investigate these areas in which even advanced learners deviate from a native speaker's performance.

Accordingly, in the present study, I will give an overview of previous approaches to fluency and compare these different understandings of fluency and then join them. I will include the major different aspects of features of spoken language that are relevant for a native speaker's fluency.

1.2 Fluency in native and non-native English speech: preview

The present study includes seven chapters. The following chapter will provide the framework of the concept of fluency and provides a general description of how the subject has been treated by researchers so far. Here, I will give an overview of previous research of fluency in speech applied to native speech. A major conceptual distinction will be made between productive and perceptive fluency, the former including all those features that are caused by characteristics of speech production processes on the part of the speakers, whereas the latter refers to those features that contribute to a perception of fluency on the part of the listeners. Finally, a summary of the most influential factors of productive and perceptive fluency will form the framework for the present study.

Chapter 3 will apply this framework to non-native speech and the peculiarities of highly advanced German learners of English.

Chapter 4 introduces the descriptive apparatus, the corpus data and the experimental set-up of the present study. Since the study is based on two separate methodological procedures to investigate productive vs. perceptive fluency, namely a corpus-based quantitative analysis and a native-speaker rating, the different methodological considerations will be described in detail. The second part of Chapter 4 introduces the findings of a pilot study that was carried out before the main study and the findings of which led to a refined approach to the investigation and the measurement of fluency.

In Chapter 5 the findings of the contrastive analysis of productive fluency in LINDSEI-GE vs. LOCNEC will be presented. In doing so, the individual variables of fluency of production will be investigated and deviations from as well as approximations to the native target norm will be discussed. In the second part of Chapter 5, the fluencemes will be looked at in combination in order to investigate the extent to which (1) there are correlations between the investigated fluencemes and (2) certain speakers show similar fluency performances and can be summarized in distinctive native and non-native fluency groups. Some language-pedagogical implications derived from these findings will be presented in the last part of Chapter 5.

Chapter 6 will present the findings of the analysis of perceptive fluency for five prototypical fluency learner types gathered from the corpus analysis. Firstly, the findings of the overall oral proficiency and seven other perceptive variables will be presented for each learner individually. Findings from focus group interviews will complement these findings. Secondly, the ratings for the five learners and the individual perceptive fluencemes will be combined and the overall impressions of the native

speakers are correlated with their perception of the individual fluencemes. The third part of Chapter 6 will summarize some language-pedagogical implications derived from these findings.

A brief conclusion of the present analysis and an outlook for future research into fluency in native and non-native English speech will be given in Chapter 7.

Pages 5-186 of this thesis have been removed due to copyright reason.

Chapter 7: Summary and prospects for future research

7.1 Summary of major findings

The present study has opened up new perspectives on the concept of fluency as a holistic phenomenon in native and non-native speech. Specifically, I combined different previous approaches researchers had taken on fluency. The mainstream approach of temporal correlates of speech production (e.g. Lennon 1990, Riggenbach 1991) was complemented by taking into consideration other features which influence cognitive processes of speech production, i.e. a speaker's use of formulaic language or their use of fluency-enhancement strategies, such as performance phenomena. These speaker-dependent cognitive correlates of fluency were categorized as variables of productive fluency. It was then shown how certain features of speech production are responsible for the perception of fluency on the part of listeners. These hearer-dependent variables were categorized as variables of perceptive fluency. Non-verbal fluency features were added as the third area of fluency in speech. Each of these three areas was set to include a variety of fluencemes, i.e. abstracted, distinctive, functional and identifiable categories of speech production that contribute to the production or perception of fluency in speech. This led to a refined and tripartite model of productive, perceptive and non-verbal fluency (see chapter 2).

These theoretical considerations of the description of fluency in native speech were then applied to non-native English speech by taking into account the peculiarities of the speech of highly advanced German learners of English who aim at a native target norm and who have learned English as a foreign language in institutionalized settings such as schools and universities and the majority of whom have been exposed to natural and authentic input by native speakers during a stay abroad. Here, previous research on individual productive and perceptive fluencemes and on previous studies that include native-speaker ratings were presented and applied to the context of the present study (see chapter 3).

The fluenceme-based model of productive and perceptive fluency was then tested on corpus data. The findings of a pilot study led to refined methodological considerations which included taking two opposing steps in the analysis. Firstly, the productive fluency model was tested by a contrastive analysis of 86,000 word German learner corpus LINDSEI-GE and the 118,000 word native English control corpus LOCNEC. Each fluenceme of production was analyzed quantitatively for each speaker in both corpora in

isolation and in relation to each other. In a second step, the perceptive fluency model was tested by way of an experimental set-up. Five prototypical learner types selected from the productive fluency analysis were subjected to native speaker ratings. 50 raters evaluated their overall perception of fluency as well as their perception of the learners' performance according to the perceptive fluencemes identified in chapter 2 (see chapter 4).

The quantitative productive fluenceme analysis revealed that none of the highly advanced learners' performance approximated to the native target norm. The learners had significantly poorer performances than the native speakers across all the temporal fluencemes and in their use of 3-grams and 4-grams. With respect to their use of fluency enhancing strategies, they were shown to deviate from the native target norm by overusing hesitation phenomena and underusing a broad variety of discourse markers and smallwords as alternative fluency enhancing strategies that the native speakers have at their disposal (see section 5.1). When correlations of native and non-native speakers' temporal fluencemes with their use of 3-grams, 4-grams and their fluency enhancing strategies were tested, no significant correlations could be found in either of the speaker groups. These findings led to the assumption of the existence of certain allo-fluencemes, i.e. concrete realizations of fluencemes, which showed, for example, in speakers' preferences for certain fluency-enhancing strategies over others. This hypothesis was strengthened by the findings of cluster analyses, which were performed for both speaker groups. The cluster analyses revealed preferences of certain speaker groups to use certain fluencemes over others in order to establish spoken fluency. However, the learner group displayed different patterns of fluency groups than the native speakers did (see section 5.2). In the light of these findings, some suggestions for theoretical and language-pedagogical implications with respect to classroom-methodology, materials design and teacher education were suggested (see section 5.3).

The experimental perceptive fluenceme analysis showed that none of the prototypical learner types approximated to the perception of a nativelike fluency. However, despite the great importance that has been attributed to productive fluency by previous studies, this was not confirmed by the ratings in the present study. In fact, there were even negative correlations of the learners' productive fluency performance and the native speakers' perception of the learners' overall level of proficiency. Instead, the native speakers' ratings of the individual perceptive fluencemes revealed positive correlations of the native speakers' perceived accuracy of the learners (i.e. not the actual number of

errors they committed, but the overall impression of correctness), the ratings for the learners' accent and for their performance with respect to pragmatic features (see chapter 6).

In the light of these findings, the present study has raised some interesting points that provide valuable insights:

- for the study of fluency in linguistics;
- for methodological issues in combining corpus linguistic approaches with experimental set-ups;
- for second language acquisition research;
- for English language teaching and learning.

For decades, fluency research has focussed on temporal variables of fluency alone and has neglected the variety of other features of speech that are equally important to a holistic analysis in fluency research. The present study revealed that there are quite a number of other fluencemes which are more important for native-speaker perception of what constitutes fluency in learner language. The present study has proposed one possible model of how to tackle the concept holistically and yet ensure its analyzability by way of defining its individual components as fluencemes. The notion of fluencemes serves as a relevant concept for future studies, as well.

The methodological framework of the present project, based on the integration of corpus-based methods into experimental set-ups has proven to provide sound and novel results in fluency research. It uncovered discrepancies of an objectively measured and analyzable productive fluency of a speaker which deviates considerably from the overall perception of native-speaking listeners. It also revealed how other areas of speech production become much more important in the perception of fluency.

The present study contributed further knowledge to the area of SLA research, where it has been stated that "[i]t is clear that the speech of second-language learners is deviant with respect to native speaker models; it is less clear what the nature and systematicity of that deviance is" (Bialystok & Smith 1985: 116). With the findings of the present study, I hope to have contributed to finding an answer to that question. It has been stated that contrastive corpus analyses are "most useful as a step between intuition and hypothesis" (Cobb 2003: 419). An initial step has been taken in this direction by the present study in the area of fluency research.

The findings of the present study gave rise to various language-pedagogical implications. The quantitative findings of the productive fluenceme analysis proved quantitatively and statistically that even highly advanced learners show significant deviances from the native target norm. A paradigm shift in syllabus design is needed to make teaching and learning processes lean more towards processes of first language acquisition, namely to teach language in chunks and formulaic sequences rather than stick to the strict divide of grammar and vocabulary exercises. Also, some ideas have been suggested as to how the teaching of fluency enhancement strategies may be included explicitly in daily teaching routines to increase learners' automaticity while they speak in a foreign language. New materials and methods were introduced that could be made use of in the classroom.

The findings of the native speaker ratings revealed the importance of a variety of other features for the perception of fluency on the part of native-speaking listeners. Here, clear correlations could be shown between the speakers' ratings of the overall degree of oral proficiency of speakers and their performance with respect to their perceived level of accuracy, their use of pragmatic features and their accent. Some suggestions were made as to how teaching these features may be integrated into the classroom on all levels.

7.2 Prospects for future research

However, while the results of the present study are promising, the field is both complex and vast and I can only claim to have provided an initial attempt at describing the process of a comprehensive and holistic fluency analysis in speech production and perception. Various strands of research need to be pursued in the future, and the following suggestions are by no means to be seen as an exhaustive list:

- The present study exclusively focussed on the frequencies of occurrence of the investigated fluencemes of production. Although this quantitative perspective uncovered important facts about the communicative characteristics of native speakers and deviances in the learners' fluency performance, future studies may benefit greatly from also including the variety of formal, functional and contextual differences of the fluencemes in their analyses.
- Owing to the raters' time constraints and the complexity and variety of features to be rated, the present study could only include five prototypical fluency types in the

survey. However, ratings of more learners with different profiles may lead to more diverse findings as to what becomes perceptively most important for a native-speaking listener. A greater variability of speakers would ensure that ratings are made objectively with respect to the learners' actual performances in the investigated variables. To this end, it would also be most valuable to have the fluencemes rated in isolation and ensure an equal proportion of native speakers with a greater balance of different English L1s.

- In order to ensure a truly valuable contribution of learner corpus research to a better understanding of late SLA processes, a variety of L1s need to be covered by further contrastive interlanguage analyses. The result may be that there are indeed features that are common to learner groups of different backgrounds. This would shed light on which of the deviances account for genuine SLA processes and which are variant-specific with regard to the learners' L1. In doing so, late interlanguage could be filled with empirical, quantifiable data of what has been described as "the outcome of mental functioning which attributes to the learner specific limitations in the aspects of mental processing" (Bialystok & Smith 1985: 116).
- Due to a lack of data the present study did not include the analysis of non-verbal fluencemes. Since the great importance and clear communicative functions of non-verbal features in speech have been shown on a variety of levels, future research on fluency should include video materials in ratings in order to arrive at more fine-grained results.
- The present study has focussed exclusively on spoken language. How speaking and writing differ from each other regarding how fluency is produced and perceived is yet another question worth being addressed by future studies. The spoken learner corpus LINDSEI and its written counterpart, LOCNESS, along with their respective native-speaker counterparts, provide a gold mine for contrastive (interlanguage) analyses of fluency in the spoken and written medium.

In the foreword to a volume by Riggensbach (ed., 2000) that takes different perspectives on fluency, Schmidt (2000: v) has claimed that "[a]sking what fluency is may not result in a definite answer". While the present study has not found a definite answer to this question, either, it has offered a new perspective on fluency by suggesting a fluenceme-based model.

REFERENCES

- Abrahamsson, N. & K. Hyltenstam (2009): "Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny", *Language Learning* 59 (2), 249-306.
- Adolphs, S. & N. Schmitt (2003): "Lexical coverage of spoken discourse", *Applied Linguistics* 24 (4), 425-438.
- Aijmer, K. (2002): *English discourse particles – evidence from a corpus*. Amsterdam: John Benjamins.
- Altenberg, B. & M. Eeg-Olofsson (1990): "Phraseology in spoken English: Presentation of a project", *Theory and practice in corpus linguistics*, eds. B. Aarts and W. Meijs. Amsterdam: Rodopi. 1-26.
- Arevart, S. & P. Nation (1991): "Fluency improvement in a second language", *RELC Journal* 22 (1), 84-94.
- Barron, A. (2002): *Acquisition in interlanguage pragmatics: Learning how to do things with words*. Amsterdam: John Benjamins.
- Bavelas, J.B. (2000): "Nonverbal aspects of fluency", *Perspectives on fluency*, ed. H. Riggensbach. Ann Arbor: The University of Michigan Press. 91-127.
- Bavelas, J.B. (1994): "Gestures as part of speech: Methodological implications", *Research on Language and Social Interaction* 27 (3), 201-221.
- Bavelas, J.B. & N. Chovil (1997): "Faces in dialogue", *The Psychology of Facial Expression*, eds. J. Russell & J.-M. Fernandez-Dols. Cambridge: Cambridge University Press. 334-346.
- Beardsmore, H.B. (1974): "Testing oral fluency", *International Review of Applied Linguistics* 12 (4), 317-326.
- Bialystok, E., & M.S. Smith (1985): "Interlanguage is not a state of mind: An evaluation of the construct for second-language acquisition", *Applied Linguistics* 6 (2), 101-117.
- Biber, D., S. Johansson, G. Leech, S. Conrad & E. Finegan (1999): *Longman grammar of spoken and written English*. Harlow: Pearson Education.
- Biber, D., S. Conrad & V. Cortes (2004): "If you look at...: Lexical bundles in university teaching and textbooks", *Applied Linguistics* 25 (3), 371-405.
- Biber, D. (2009): "A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing", *International Journal of Corpus Linguistics* 14 (3), 275-311.
- Bley-Vroman, R. (1983): "The comparative fallacy in interlanguage studies: The case of systematicity", *Language Learning* 33, 1-17.

- Blum-Kulka, S. & E. Olshtain (1986): "Too many words: Length of utterance and pragmatic failure", *Studies in Second Language Acquisition* 8 (2), 165-180.
- Blum-Kulka, S., J. House & G. Kasper (eds.) (1989): *Cross-cultural pragmatics: Requests and apologies*. Norwood, NJ: Ablex Publishing Corporation.
- Bongaerts T., C. van Summeren, B. Planken & E. Schils (1997): "Age and ultimate attainment in the pronunciation of a foreign language", *Studies in Second Language Acquisition* 19 (4), 447-465.
- Boomer, D.S. (1965): "Hesitation and grammatical encoding", *Language and Speech* 8 (3), 148-158.
- Borg, S. (1999): "The use of grammatical strategy in the second language classroom: A qualitative study of teachers' practice and cognition", *Applied Linguistics* 20 (1), 95-126.
- Borin, L. & K. Prütz (1996): "New wine in old skins? A corpus investigation of L1 syntactic transfer in learner language", *Corpora and language learners*, eds. G. Aston, S. Bernadini & D. Stewart. Amsterdam: John Benjamins. 67-88.
- Brainerd, B. (1982): "The type-token relation in the works of S. Kierkegaard", *Computing in the humanities*, ed. R.W. Bailey. Amsterdam: North Holland. 97-109.
- Brand, C. & S. Götz (2011): "Fluency versus accuracy in advanced spoken learner language: A multi-method approach", *International Journal of Corpus Linguistics* 16 (2), in press.
- Brand, C. & S. Kämmerer (2006): "The Louvain International Database of Spoken English Interlanguage (LINDSEI): Compiling the German component", *Corpus technology and language pedagogy: New resources, new tools, new methods*, eds. S. Braun, K. Kohn & J. Mukherjee. Frankfurt am Main: Peter Lang. 127-140.
- Braun, S. (2005): "From pedagogically relevant corpora to authentic language learning purposes", *ReCALL* 17 (1), 47-64.
- Breuer, A. & R. Geluykens (2007): "Variation in British and American English requests: A contrastive analysis", *Cross-cultural pragmatics and interlanguage English*, eds. B. Kraft & R. Geluykens. München: Lincom. 107-126.
- Brinton, L. (1990): "The development of discourse markers in English", *Historical linguistics and philology*, ed. J. Fisiak. Berlin: Mouton de Gruyter. 45-71.
- Brown, P. & S.C. Levinson (1987): *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press.
- Burgschmidt, E. (2005): "Was ist eigentlich fluency?", *Linguistische Dimensionen des Fremdsprachenunterrichts*, ed. T. Herbst. Würzburg: Königshausen & Neumann. 135-140.
- Burgschmidt, E. & D. Götz (1972): "Kontrastive Phonologie Deutsch-Englisch und Mundartinterferenz", *Linguistik und Didaktik* 11, 209-225.

- Burt, M. & C. Kiparsky (1974): "Global and local mistakes", *New frontiers in second language learning*, eds. J. Schumann & N. Stenson. Rowley, MA: Newbury House. 71-80.
- Butcher, A. (1980): "Pause and syntactic structure", *Temporal variables in speech*, eds. H.W. Dechert & M. Raupach. The Hague: Mouton de Gruyter. 85-90.
- Butterworth, B. (1975): "Hesitation and semantic planning in speech", *Journal of Psycholinguistic Research* 4 (1), 75-87.
- Butterworth, B. (1980): "Evidence from pauses in speech", *Language production: Speech and talk*, ed. B. Butterworth. London: Academic Press. 155-177.
- Bygate, M. (1987): *Speaking*. Oxford: Oxford University Press.
- Bygate, M. (2009): "Teaching the spoken foreign language", *Handbook of foreign language communication and learning*, eds. K. Knapp & B. Seidelhofer. Berlin, New York: Mouton de Gruyter. 401-438.
- Cameron, L. (2001): *Teaching Languages to young Learners*. Cambridge: Cambridge University Press.
- Canale, M. & M. Swain (1980): "Theoretical basis of communicative approaches to second language teaching and testing", *Applied Linguistics* 1 (1), 1-47.
- Chafe, W. (1980): *The pear stories: Cognitive, cultural and linguistic aspects of narrative production*. Norwood: Ablex.
- Chafe, W. (1985): "Some reasons for hesitating", *Perspectives on silence*, eds. D. Tannen & M. Saville-Troike. Norwood: Ablex. 21-30.
- Chambers, F. (1997): "What do we mean by fluency?", *System* 25 (4), 535-544.
- Chambers, F. & B.J. Richards (1995): "The free conversation and the assessment of oral proficiency", *The Language Learning Journal* 11 (1), 6-10.
- Cobb, T. (2003): "Analyzing late interlanguage with learner corpora: Quebec replications of three European studies", *The Canadian Modern Language Review/La Revue canadienne des langues vivantes* 59 (3), 393-423.
- Cohen, A.D. (1996): "Developing the ability to perform speech acts", *Studies in Second Language Acquisition* 18 (2), 253-267.
- Cohen, A.D. (2007): "Teaching and assessing L2 pragmatics: What can we expect from learners?", *Language Teaching* 41 (2), 213-235.
- Conklin, K. & N. Schmitt (2008): "Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers?", *Applied Linguistics* 29 (1), 72-89.
- Corder, S.P. (1971): "Idiosyncratic dialects and error analysis", *International Review of Applied Linguistics* 9 (2), 147-159.
- Corder, S.P. (1974): "Error Analysis", *Techniques in Applied Linguistics*, eds. J.P.B. Allen & S.P. Corder. London: Oxford University Press. 122-154.

- Corder, S.P. (1978): "Strategies of communication", *Publications de l'Association Finlandaise de Linguistique Appliquée* 23, 7-12.
- Corder, S.P. (1984): "The significance of learners' errors", *Error analysis: Perspectives on second language acquisition*, ed. J.C. Richards. Essex: Longman. 19-27.
- Council of Europe (2001): *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Crystal, D. (1987): *The Cambridge Encyclopedia of Language*. Cambridge: Cambridge University Press.
- Crystal, D. (1988): "Another look at, well, you know...", *English Today* 4 (1), 47-49.
- Cucchiaroni, C., H. Strik & L. Boves (2000): "Quantitative assessment of second language learners' fluency by means of automatic speech recognition technology", *Journal of the Acoustical Society of America* 107 (2), 989-999.
- Cucchiaroni, C., H. Strik & L. Boves (2002): "Quantitative assessment of second language learners' fluency: Comparisons between read and spontaneous speech", *Journal of the Acoustical Society of America* 111 (6), 2862-2873.
- Cummins, J. & M. Swain (1986): *Bilingualism in education*. Harlow: Longman.
- Cutting, J. (2006): "Spoken grammar: Vague language and EAP", *Spoken English, TESOL and applied linguistics: Challenges for theory and praxis*, ed. R. Hughes. Houndmills: Palgrave Macmillan. 159-181.
- Cutting, J. (2008) [2002]: *Pragmatics and discourse: A resource book for students*. 2nd edition. London: Routledge.
- Czwenar, I. (2004): "Oral proficiency of Polish EFL students: Corpus-based analysis", *Practical applications in language and computers, PALC 2003*, ed. Barbara Lewandowska-Tomaszyk. Frankfurt am Main: Peter Lang. 391-399.
- Dagneaux, E., S. Denness, S. Granger, F. Meunier, J. Neff & J. Thewissen (2005): *Error Tagging Manual Version 1.2*. Université catholique de Louvain.
- Dagut, M. & B. Laufer (1985): "Avoidance of phrasal verbs: A case for contrastive analysis", *Studies in Second Language Acquisition* 7 (1), 73-79.
- Davies, A. (2003): *The native speaker: Myth and reality*. Clevedon: Multilingual Matters.
- Dechert, H.-W. (1983): "How a story is done in a second language", *Strategies in interlanguage communication*, eds. C. F{rch & G. Kasper. London: Longman. 175-195.
- De Cock S. (1998): "A recurrent word combination approach to the study of formulae in the speech of native and non-native speakers of English", *International Journal of Corpus Linguistics*, 3 (1), 59-80
- De Cock, S. (2000): "Repetitive phrasal chunkiness and advanced EFL speech and writing", *Corpus linguistics and linguistic theory. Papers from the twentieth International Conference on English Language Research on Computerized*

- Corpora (ICAME 20)*, eds. C. Mair & M. Hundt. Amsterdam, Atlanta: Rodopi. 51-68.
- De Cock, S., S. Granger, G. Leech & T. McEnery (1998): "An automated approach to the phrasicon of EFL learners", *Learner English on computer*, ed. S. Granger. Harlow: Longman. 67-79.
- Deese, J. (1980): "Pauses, prosody, and the demands of production in language", *Temporal variables in speech: Studies in honour of Frieda Goldman-Eisler*, eds. H.-W. Dechert & M. Raupach. The Hague: Mouton de Gruyter. 69-84.
- Derwing, T.M., M.J. Munro, R.I. Thomson (2007): "A longitudinal study of ESL learners' fluency and comprehensibility development", *Applied Linguistics* 29 (3), 359-380.
- Derwing, T. M., M.J. Munro, R.I. Thomson & M.J. Rossiter (2009): "The relationship between L1 fluency and L2 fluency development", *Studies in Second Language Acquisition* 31 (4), 533-557.
- Deschamps, A. (1980): "The syntactical distribution of pauses in English spoken as a second language by French students", *Temporal variables in speech: Studies in Honour of Frieda Goldman-Eisler*, eds. H.-W. Dechert & M. Raupach. The Hague: Mouton de Gruyter. 255-262.
- Días-Campos, M. (2006): "The effect of style in second language phonology: An analysis of segmental acquisition in study abroad and regular-classroom students", *Selected proceedings of the 7th conference on the acquisition of Spanish and Portuguese as first and second languages*, eds. C. A. Klee & T. L. Face. Somerville, MA: Cascadilla Proceedings Project. 26-39.
- Dose, S. (forthcoming): "Flipping the script: A Corpus of American Television Series (CATS) for corpus-based language learning and teaching", *Corpus linguistics and variation in English: Focus on non-native Englishes*, eds. M. Huber & J. Mukherjee. University of Helsinki: Varieng Electronic Series.
- Dretzke, B. (1985): *Fehlerbewertung im Aussprachebereich*. Hamburg: Buske.
- Dretzke, B. (1998): *Modern British and American pronunciation*. Paderborn: Schöningh.
- Drommel, R.H. (1980): "Towards a subcategorization of speech pauses", *Temporal variables in Speech: Studies in honour of Frieda Goldman-Eisler*, eds. H.-W. Dechert & M. Raupach. The Hague: Mouton de Gruyter. 227-238.
- Dulay, H. & M. Burt (1974): "Natural sequences in child second language acquisition", *Language Learning* 24 (1), 37-53.
- Dunning, T. (2003): "Accurate Methods for the Statistics of Surprise and Coincidence", *Computational Linguistics* 19 (1), 61-74.
- Ejzenberg, R. (1997): "The role of task structure in oral fluency assessment", Paper presented at the *Annual Meeting of the Teachers of English to Speakers of other Languages* (28th Baltimore, MD, March 8-12 1994) and at the *Annual Meeting of the American Association for Applied Linguistics* (17th, Long Beach, CA, March 25-28,

- 1995). Available on <http://www.eric.ed.gov/PDFS/ED406827.pdf>. (Accessed on 9 March 2011).
- Elder, C. (2001): "Assessing the language proficiency of teachers: Are there any border controls?", *Language Testing* 18 (2), 149-170.
- Ellis, R. (1992): "Learning to communicate in the classroom", *Studies in Second Language Acquisition* 14 (1), 1-23.
- Erman, B. (1986): "Some pragmatic expressions in English conversation", *English speech and writing: A symposium*, eds. G. Tottie & I. Bäcklund. Stockholm: Almqvist & Wiksell. 131-147.
- Erman, B. (1987): *Pragmatic expressions in English*. Stockholm: Almqvist & Wiksell.
- Erman, B. & B. Warren (2000): "The idiom principle and the open-choice principle", *Text* 20, 29-62.
- Erman, B. (2007): "Cognitive processes as evidence of the idiom principle", *International Journal of Corpus Linguistics* 12 (1), 25-53.
- Esser, J. (1992)[1979]: *Englische Prosodie. Eine Einführung*. 3rd edition. Tübingen: Narr.
- Esser, J. (1993): *English linguistic stylistics*. Tübingen: Niemeyer.
- Firth, C. & G. Kasper (eds.) (1983): *Strategies in interlanguage communication*. London: Longman.
- Fant, G., A. Kruckenberg & L. Nord (1991): "Durational correlates of stress in Swedish, French and English", *Journal of Phonetics* 19, 351-365.
- Fathman, A. (1980): "Repetition and correction as an indication of speech planning and execution processes among second language learners", *Towards a cross-linguistic assessment of speech production*, eds. H.-W. Dechert & M. Raupach. Frankfurt am Main: Peter Lang. 77-86.
- Fernando, C. (1996): *Idioms and idiomaticity*. Oxford: Oxford University Press.
- Field, A. (2009)[2000]: *Discovering statistics using SPSS – Introducing statistic methods*. 3rd edition. London: Sage Publications Ltd.
- Fillmore, C. (1979): "On fluency", *Individual differences in language ability and language behavior*, eds. C. Fillmore, D. Kempler & W. Wang. New York: New York Academic Press. 85-102.
- Firth, A. (1988): *Modes of interaction: A contrastive analysis of advanced Danish learners' modes of interaction in English*. Unpublished M.A. Thesis, University of Birmingham.
- Folse, K.S. (2006): *The art of teaching speaking: Research and pedagogy for the ESL/EFL classroom*. Ann Arbor: University of Michigan Press.
- Foster, P. (2001): "Rules and routines: A consideration of their role in the task-based language production of native and non-native speakers", *Researching pedagogic*

- tasks: Second language learning, teaching and testing*, eds. M. Bygate, P. Skehan & M. Swain. Oxford: Oxford University Press. 145-160.
- Foster, P. & P. Tavakoli (2009): "Native speakers and task performance: Comparing effects on complexity, fluency, and lexical diversity", *Language Learning* 59 (4), 866-896.
- Francis, G., S. Hunston & E. Manning (1996): *Collins cobuild grammar patterns I: Verbs*. London: Collins CoBUILD.
- Francis, G., S. Hunston & E. Manning (1998): *Collins cobuild grammar patterns II: Nouns and adjectives*. London: Collins CoBUILD.
- Freed, B.F., N. Segalowitz, D.P. Dewey (2004): "Context of learning and second language fluency in French: Comparing regular classroom, study abroad, and intensive domestic immersion programs", *Studies in Second Language Acquisition* 26 (2), 275-301.
- Fulcher, G. (1996): "Does thick description lead to smart tests? A data-based approach to rating-scale development", *Language Testing* 13 (2), 208-238.
- Gass, S. & L. Selinker (1994): *Second language acquisition: An introductory course*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Gass, S. & L. Selinker (2001) [1994]: *Second language acquisition: An introductory course*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Gass, S. & E. Varonis (1984): "The effect of familiarity on the comprehensibility of nonnative speech", *Language Learning* 34 (1), 65-89.
- García, P. (2004): "Pragmatic comprehension of high and low level language learners", *Teaching English as a Second or Foreign Language* 8 (2). Available on <http://www-writing.berkeley.edu/TESL-EJ/ej30/a1.html> (accessed on 7 January 2010).
- Gatbonton, E. & N. Segalowitz (2005): "Rethinking communicative language teaching: A focus on access to fluency", *The Canadian Modern Language Review/La Revue canadienne des langues vivantes* 61 (3), 325-353.
- Gilquin, G. (2002): "Automatic retrieval of syntactic structures: The quest for the Holy Grail", *International Journal of Corpus Linguistics* 7 (2), 183-214.
- Gilquin, G. (2008): "Hesitation markers among EFL learners: Pragmatic deficiency or difference?", *Pragmatics and corpus linguistics: A mutualistic entente*, ed. J. Romero-Trillo. Berlin, Heidelberg, New York: Mouton de Gruyter. 119-149.
- Gilquin, G. & M. Paquot (2008): "Too chatty: Learner academic writing and register variation", *English Text Construction* 1 (1), 41-61.
- Gilquin, G., S. De Cock & S. Granger (2010): *The Louvain International Database of Spoken English Interlanguage. Handbook and CD-ROM*. Louvain-la-Neuve: Presses Universitaires de Louvain.
- Goldman-Eisler, F. (1968): *Psycholinguistics: Experiments in spontaneous speech*. London: Academic Press.

- Götz, D. (1986): "Intonation an der Hochschule", *Die Neuen Sprachen* 85 (5/6), 465-477.
- Götz, S. (2007): "Performanzphänomene in gesprochenem Lernerenglisch: Eine korpusbasierte Pilotstudie", *Zeitschrift für Fremdsprachenforschung* 18 (1), 67-84.
- Götz, S. & M. Schilk (2011): "Formulaic sequences in spoken ENL, ESL and EFL: Focus on British English, Indian English and learner English of advanced German learners", *Exploring second-language varieties of English and learner Englishes: Bridging a paradigm gap*, eds. J. Mukherjee & M. Hundt. Amsterdam: John Benjamins. 79-100.
- Götz-Votteler, K. & T. Herbst (2009): "Innovation in advanced learners' dictionaries of English", *Lexicographica* 25, 47-66.
- Granger, S. (1996): "From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora", *Languages in contrast: Text-based cross-linguistic studies*, eds. K. Aijmer, B. Altenberg & M. Johansson. Lund: Lund University Press. 37-51.
- Granger, S. (2004a): "Computer learner corpus research: Current state and future prospects", *Applied corpus linguistics: A multidimensional perspective*, eds. U. Connor & T.A. Upton. Amsterdam, New York: Rodopi. 123-145.
- Granger, S. (2004b): "Practical applications of learner corpora", *Practical applications in language and computers*, PALC 2003, ed. B. Levandowska-Tomaszyk. Frankfurt am Main: Peter Lang. 291-301.
- Granger, S. (2009): "The contribution of learner corpora to second language acquisition and foreign language teaching", *Corpora and language teaching*, ed. K. Aijmer. Amsterdam: John Benjamins. 13-32.
- Granger, S. & S. Tyson (eds.) (1996): *Extending the scope of corpus-based research: New applications, new challenges*. Amsterdam: Rodopi.
- Greenbaum, Sidney (1974): "Verb-intensifier collocations in American and British English", *American Speech* 49 (1/2), 79-89.
- Grice, H.P. (1975): "Logic and conversation", *Syntax and semantics*, eds. P. Cole & J. Morgan. New York: Academic Press. 41-58.
- Gries, S.Th. (2008): "Phraseology and linguistic theory", *Phraseology: An interdisciplinary perspective*, eds. S. Granger & F. Meunier. Amsterdam, Philadelphia: John Benjamins. 3-25.
- Gries, S. Th. & J. Mukherjee (2010): "Lexical gravity across varieties of English: an ICE-based study of n-grams in Asian Englishes", *International Journal of Corpus Linguistics* 15 (4), 520-548.
- Griffiths, R. (1991): "Pausological research in an L2 context: A rationale and review of selected studies", *Applied Linguistics* 12 (4), 345-364.
- Grosjean, F. & A. Deschamps (1973): "Analyse des variables temporelles du français spontané II. Comparaison du français oral dans la description avec l'anglais (description) et avec le français (interview radio-phonique)", *Phonetica* 28, 191-226.

- Grosjean, F. & A. Deschamps (1975): "Analyse des variables temporelles de l'anglais et du français: Vitesse de parole et variables composantes, phénomènes d'hésitation", *Phonetica* 31, 144-184.
- Grosjean, F. (1980): "Linguistic structures and performance structures: Studies in pause distribution", *Temporal variables in speech: Studies in honour of Frieda Goldman-Eisler*, eds. H.-W. Dechert & M. Raupach. The Hague: Mouton de Gruyter. 91-106.
- Guillot, M.-N. (1999): *Fluency and its teaching*. Clevedon: Multilingual Matters.
- Gut, U. (2009): *Non-native speech: A corpus-based analysis of phonological and phonetic properties of L2 English and German*. Frankfurt am Main: Peter Lang.
- Halliday, M.A.K. (1967): *Intonation and grammar in British English*. The Hague: Mouton.
- Halliday, M.A.K. (1970): *A course in spoken English. Intonation*. London: Oxford University Press.
- Hasselgren, A. (1994): "Lexical teddy bears and advanced learners: A study into the ways Norwegian students cope with English vocabulary", *International Journal of Applied Linguistics* 4 (2), 237-260.
- Hasselgren, A. (2002): "Learner corpora and language testing: smallwords as markers of learner fluency", *Computer learner corpora, second language acquisition and foreign language teaching*, eds. S. Granger, J. Hung & S. Petch-Tyson. Amsterdam: John Benjamins. 143-173.
- Herbst, T. (1992): "Pro-Nunciation: Zur Bedeutung einer guten Aussprache in der Fremdsprache", *Die Neueren Sprachen* 91 (1): 2-18.
- Herbst, T. & M. Klotz (2008): "Chunks statt Chomsky", *Fortschritte im frühen Fremdsprachenlernen: Ausgewählte Tagungsbeiträge Nürnberg 2007*, ed. H. Böttger. München: Domino-Verlag. 336-347.
- Herbst, T., D. Heat, I.F. Roe & D. Götz (2004): *A valency dictionary of English. A corpus-based analysis of the complementation patterns of English verbs, nouns and adjectives*. Berlin: Mouton de Gruyter.
- Hewings, M. (2004): *Pronunciation practice activities: A resource book for teaching English pronunciation*. Cambridge: Cambridge University Press.
- Hincks, R. (2008): "Presenting in English or Swedish: Differences in speaking rate", *Proceedings Fonetik 2008: The XXIst Swedish phonetics conference June 11-13 2008*, eds. A. Eriksson & J. Lindh. Gothenburg: Reprocentralen, Humanisten, University of Gothenburg. 21-24. Available on http://www.ling.gu.se/konferenser/fonetik2008/papers/Proc_fonetik_2008.pdf. (last accessed on 5 May 2011).
- Hójen, A. (2000): "Detection of Danish accent in English by native speakers of Danish", *Proceedings of New Sounds 2000*, ed. A. & J. Leather. Klagenfurt: University of Klagenfurt. 80-86. Available on <http://www.andershojen.dk/papers/NewSounds2000.pdf>. (last accessed on 5 May 2011).

- Holliday, A. (2010): "Analyzing qualitative data", *Continuum companion to research methods in applied linguistics*, ed. B. Paltridge & A. Phakiti. London: Continuum. 98-110.
- Holme, R. (2010): "A construction grammar for the classroom", *International Review of Applied Linguistics in Language Teaching* 48 (4), 355-377.
- House, J. (1984): "Some methodological problems and perspectives in contrastive discourse analysis", *Applied Linguistics* 5 (3), 255-254.
- House, J. (1996): "Developing pragmatic fluency in English as a foreign language: Routines and metapragmatic awareness", *Studies in Second Language Acquisition* 18 (2), 225-252.
- Hughes, R. (1996): *English in speech and writing: Investigating language and literature*. London: Routledge.
- Hughes, R. (2010): "Researching speaking", *Continuum companion to research methods in applied linguistics*, eds. B. Paltridge & A. Phakiti. London: Continuum. 145-159.
- Jackendoff, R. (1995): "The boundaries of the lexicon", *Idioms: Structural and psychological perspectives*, eds. M. Evaraert & E-J. van der Linden. Hillsdale: Erlbaum. 133-166.
- Jahns, S. & A. Schröter (forthcoming): "Förderung der englischen Sprach- und Aussprachekompetenz durch authentische Webaufgaben", *Aufgaben 2.0: Konzepte, Materialien und Methoden für das Fremdsprachenlehren und -lernen mit digitalen Medien*, eds. K. Biebighäuser, M. Zibelius & T. Schmidt. Tübingen: Narr.
- Jenkins, J. (2000): *The phonology of English as an international language*. Oxford: Oxford University Press.
- Jilka, M. (2007): "Different manifestations and perceptions of foreign accent in intonation", *Non-native prosody: Phonetic description and teaching practice*, eds. J. Trouvain & U. Gut. 77-96.
- Kachru, B.B. (1996): "World Englishes: Agony and ecstasy", *Journal of Aesthetic Education* 30 (2), 135-155.
- Kämmerer, S. (2009): "Error-tagging spoken features of (learner) language: The UCL Error Editor 'revised'", Paper presented at the 30th annual conference of the *International Computer Archive of Modern and Medieval English (ICAME30)*. Lancaster.
- Kang, O., D. Robin & L. Pickering (2010): "Suprasegmental measures of accentedness and judgements of language learner proficiency in oral English", *The Modern Language Journal* 94 (4), 554-566.
- Kasper, G. (1982): "Teaching-induced aspects of interlanguage discourse", *Studies in Second Language Acquisition* 4 (2), 99-113.
- Kasper, G. & K.R. Rose (2002): *Pragmatic development in a second language*. Oxford: Blackwell.

- Kennedy, S. & P. Trofimovich (2008): "Intelligibility, comprehensibility, and accentedness of L2 speech: The role of listener experience and semantic context", *The Canadian Modern Language Review* 64 (3), 459-489.
- Kenworthy, J. (1987): *Teaching English pronunciation*. New York: Longman.
- Kieweg, M. & W. Kieweg (2000): "Praxiserprobte Dialogtechniken", *Der fremdsprachliche Unterricht Englisch* 34 (5), 17-24.
- Kim, M. (2001): "Detecting DIF across the different language groups in a speaking test", *Language Testing* 18 (1), 89-114.
- Kjellmer, G. (1991): "A mint of phrases", *English corpus linguistics*, eds. K. Aijmer & B. Altenberg. London: Longman. 111-127.
- Kjellmer, G. (2003): "Hesitation. In defence of ER and ERM", *English Studies* 84 (2), 170-198.
- Klee, T. (1992): "Developmental and diagnostic characteristics of quantitative measures of children's language production", *Topics in Language Disorders* 12, 28-41.
- Klippel, F. (1994) [1984]: *Keep talking: Communicative fluency activities for language teaching*. Cambridge: Cambridge University Press.
- Koponen, M. & H. Riggensbach (2000): "Overview: Varying perspectives on fluency", *Perspectives on Fluency*, ed. H. Riggensbach. Ann Arbor: The University of Michigan Press. 5-25.
- Kormos, J. & M. Dénes (2004): "Exploring measures and perceptions of fluency in the speech of second language learners", *System* 32 (2), 145-164.
- Kraft, B. & R. Geluykens (eds.) (2007a): *Cross-cultural pragmatics and interlanguage English*. München: Lincom.
- Kraft, B. & R. Geluykens (2007b): "Cross-cultural and interlanguage pragmatics", *Cross-cultural pragmatics and interlanguage English*, eds. B. Kraft & R. Geluykens. München: Lincom. 3-20.
- Kramsch, C. (1998): "The privilege of the intercultural speaker", *Language learning in intercultural perspective: Approaches through drama and ethnography*, eds. M. Byram & M. Fleming. Cambridge: Cambridge University Press. 16-31.
- Krashen, S. (1982): *Principles and Practice in Second Language Acquisition*. New York: Pergamon.
- Kucharek, R. (1988): *Untersuchungen zur Englischen Schüleraussprache an Gymnasien in Oberbayern*. University of Augsburg: Augsburger I- & -I-Schriften.
- Lamendella, J. (1977): "General principles of neurofunctional organization and their manifestation in primary and nonprimary language acquisition", *Language Learning* 27 (1), 155-196.
- Larsen-Freeman, D. (2006): "The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English", *Applied Linguistics* 27 (4), 590-619.

- Leech, G. (2000): "Grammars of spoken English: New outcomes of corpus-oriented research", *Language Learning* 50 (4), 675-724.
- Leeson, R. (1975): *Fluency and language teaching*. London: Longman.
- Lenk, U. (1998): *Marking discourse coherence: Functions of discourse markers in spoken English*. Tübingen: Narr.
- Lennon, P. (1987): *Second language acquisition of advanced German learners*. Unpublished doctoral dissertation. University of Reading: Berkshire.
- Lennon, P. (1990a): "Investigating fluency in EFL: A quantitative approach", *Language Learning* 40 (3), 387-417.
- Lennon, P. (1990b): "The advanced learner at large in the L2 community", *International Review of Applied Linguistics in Language Teaching* 28 (4), 309-324.
- Lennon, P. (1998): "Approaches to the teaching of idiomatic language", *International Review of Applied Linguistics in Language Teaching* 36 (1), 11-30.
- Lennon, P. (2000): "The lexical element in spoken second language fluency", *Perspectives on Fluency*, ed. H. Riggenbach. Ann Arbor, MI: The University of Michigan Press. 25-42.
- Levelt, W.J.M. (1989): *Speaking: From intention to articulation*. Cambridge: MIT Press.
- Levis, J.M. (2006): "Pronunciation and the Assessment of Spoken Language", *Spoken English, TESOL and applied linguistics*, ed. R. Hughes. Houndmills: Palgrave Macmillan. 245-270.
- Liao, Y. & Y.J. Fukuya (2004): "Avoidance of phrasal verbs: The case of Chinese learners of English", *Language Learning* 54 (2), 193-226.
- Lightbown, P. & N. Spada (2006): *How languages are learned*. Oxford: Oxford University Press.
- Long, M. (1990): "Maturation constraints on language development", *Studies in Second Language Acquisition* 12 (3), 251-285.
- Lorenz, G. R. (1999): *Adjective Intensification - Learners versus native speakers: A corpus study of argumentative writing*. Amsterdam: Rodopi.
- Lounsbury, F. (1954): "Transitional probability, linguistic structure and systems of habit-family hierarchies", *Psycholinguistics: A survey of theory and research problems*, eds. C. Osgood & T. Sebek. Bloomington: Indiana University Press. 93-101.
- Luoma, S. (2004): *Assessing speaking*. Cambridge: Cambridge University Press.
- Maclay, H. & C.E. Osgood (1959): "Hesitation phenomena in spontaneous English speech", *Word* 1, 19-44.
- Malcolm, I.G. & Rochecouste, J. (2000): "Event and story schemas in Australian Aboriginal English discourse", *English World-Wide* 21 (2), 95-123.

- Mair, C. (1997): "Parallel corpora: A real-time approach to the study of language change in progress", *Corpus-based studies in English. Papers from the seventeenth International Conference on English Language Research on Computerized Corpora (ICAME 17)*, ed. M. Ljung. Amsterdam: Rodopi. 195-209.
- Major, R. (1987): "English voiceless stop production by speakers of Brazilian Portuguese", *Journal of Phonetics* 15 (2), 197-202.
- Major, R. (2007): "Identifying a foreign accent in an unfamiliar language", *Studies in Second Language Acquisition* 29 (4), 539-556.
- Malvern, D.D., & B.J. Richards (1997): "A new measure of lexical diversity", *Evolving models of language. Papers from the annual meeting of the British Association of Applied Linguists held at the University of Wales, Swansea, September 1996*, ed. A. Ryan & A. Wray. Clevedon: Multilingual Matters. 58-71.
- Malvern, D.D., B.J. Richards, N. Chipere & P. Durán (2008): *Lexical diversity and language development: Quantification and assessment*. Houndmills: Palgrave Macmillan.
- Marshall, R.C. (2000): "Speech fluency and aphasia", *Perspectives on fluency*, ed. H. Riggensbach. Ann Arbor: The University of Michigan Press. 74-88.
- Marslen-Wilson, W. (1973): "Linguistic structure and speech shadowing at very short latencies", *Nature* 244, 522-523.
- McCafferty, S.G. (2006): "Gesture and the materialization of second language prosody", *International Review of Applied Linguistics in Language Teaching* 44 (2), 197-209.
- McCarthy, M. & R. Carter (1995): "Spoken grammar: What is it and how can we teach it?", *ELT Journal* 49 (3), 207-218.
- McCarthy, M.J. (1999): "What is a basic spoken vocabulary?", *FELT Newsletter* 1(4), 7-9.
- McCarthy, M.J. (2006) [2002]: "Fluency and confluence: What fluent speakers do", *Explorations in Corpus Linguistics*, ed. M.J. McCarthy. Cambridge: Cambridge University Press. 1-6.
- McCarthy, M.J. & R. Carter (2006) [2002]: "This that and the other: Multi-word clusters in spoken English as visible patterns of interaction", *Explorations in Corpus Linguistics*, ed. M. McCarthy. Cambridge: Cambridge University Press. 7-26.
- McCarthy, M.J.. & J. Scott (2007): "'vocd': A theoretical and empirical evaluation", *Language Testing* 24 (4), 459-488.
- McGarr, N. (1981): "The effect of context on the intelligibility of hearing and deaf children's speech", *Language and Speech* 24 (3), 255-264.
- McKee, G., D. Malvern & B. Richards (2000): "Measuring vocabulary diversity using dedicated software", *Literary and Linguistic Computing* 15 (3), 323-337.

- McKelvie, D. (1998): "The syntax of disfluency in spontaneous spoken language", *Corpus Linguistics: Reading in a widening discipline*, eds. G. Sampson & D. McCarthy. London, New York: Continuum. 404-420.
- Michel, M., F. Kuiken & I. Vedder (2007): "The influence of complexity in monologic versus dialogic tasks in Dutch L2", *International Review of Applied Linguistics in Language Teaching* 45 (3), 241-259.
- Miller, J.F. (1991): "Quantifying productive language disorders", *Research on child language disorders: A decade of process*, eds. J.F. Miller. Austin, TX: Pro-Ed. 211-220.
- Mittmann, B. (2004): *Mehrwort-Cluster in der Englischen Alltagskonversation*. Tübingen: Gunter Narr.
- Möhle, D. (1984): "Analysis and evaluation of the second language speech production of different native speakers", *Second language productions*, eds. H.-W. Dechert, D. Möhle & M. Raupach. Tübingen: Narr. 199-209.
- Möhle, D. & M. Raupach (1993): "Ausdrucksschwierigkeiten als Merkmal von Lernersprache: Sprachproduktion fortgeschrittener Lerner des Französischen", *Fremdsprachen Lehren und Lernen* 22, 109-128.
- Möllering, M. (2001): "Teaching German modal particles: A corpus-based approach", *Language Learning & Technology* 5 (3), 130-151.
- Möllering, M. (2004): *The acquisition of German modal particles: A corpus-based approach*. Bern: Peter Lang.
- Monroe, J. (1975): "Measuring and enhancing syntactic fluency in French", *French Review* 48, 1023-1031.
- Monsela, E. (2010): "Pragmatische Äquivalenzen des Präteritums und des Perfekts für das Sprachpaar Deutsch - Neugriechisch", *Norms in educational linguistics – Normen in Educational Linguistics: Linguistic, didactic and cultural perspectives – Linguistische, didaktische und kulturwissenschaftliche Perspektiven*, eds. S. Dose, S. Götz, T. Brato & C. Brand. Frankfurt am Main: Peter Lang. 159-166.
- Moon, R. (1998): *Fixed Expressions and Idioms in English*, Oxford: Clarendon Press.
- Müller, S. (2005): *Discourse markers in native and non-native English discourse*. Amsterdam: John Benjamins.
- Müller-Hartmann, A. & M. Schocker-von Ditfurth (2004): *Introduction to English language teaching*. Berlin: Klett.
- Mukherjee, J. (2000): "Speech is silver, but silence is golden: Some remarks on the function(s) of pauses", *Anglia* 118 (4), 571-584.
- Mukherjee, J. (2002): *Korpuslinguistik und Englischunterricht: Eine Einführung*. Frankfurt am Main: Peter Lang.
- Mukherjee, J. (2005): "The native speaker is alive and kicking: Linguistic and language-pedagogical perspectives", *Anglistik* 16 (2), 7-23.

- Mukherjee, J. (2009): "The grammar of conversation in advanced spoken learner English: learner corpus data and language-pedagogical implications", *Corpora and language teaching*, ed. K. Aijmer. Amsterdam: John Benjamins. 203-230.
- Mukherjee, J. & J.-M. Rohrbach (2006): "Rethinking applied corpus linguistics from a language-pedagogical perspective: New departures in learner corpus research", *Planing, gluing and painting corpora: Inside the applied corpus linguist's workshop*, eds. B. Kettemann & G. Marko. Frankfurt am Main: Peter Lang. 205-232.
- Mumford, S. (2009): "An analysis of spoken grammar: The case of free production", *ELT Journal* 63 (2), 137-144.
- Munro M., T. Derwing & S. Morton (2006): "The mutual intelligibility of L2 speech", *Studies in Second Language Acquisition* 28 (1), 111-131.
- Nakatani, Y. (2010): "Identifying strategies that facilitate EFL learners' oral communication: A classroom study using multiple data collection procedures", *The Modern Language Journal* 94 (1), 116-136.
- Nation, P. (1990): *Teaching and learning vocabulary*. New York: Newbury House.
- Nation, P. (1993): "Measuring readiness for simplified material: A test of the first 1,000 words of English", *Simplification: Theory and application*, ed. M.L. Tickoo. Singapore: SEAMEO Regional Language Center. 193-203.
- Nattinger, J.R. & J.S. DeCarrico (1992): *Lexical phrases and language teaching*. Oxford: Oxford University Press.
- Nolasco, R. & L. Arthur (1987): *Conversation*. Oxford: Oxford University Press.
- Norrish, J. (1983): *Language learners and their errors*. London: Macmillan.
- Oakes, M.P. (1998): *Statistics for corpus linguistics*. Edinburgh: Edinburgh University Press.
- O'Connel, D.C. & S. Kowal (2004): "The history of research on the filled pause as evidence of The Written Language Bias in Linguistics", *Journal of Psycholinguistic Research* 33 (4), 459-474.
- O'Connor di Vito, N. (1991): "Incorporating native speaker norms in second language materials", *Applied Linguistics* 12 (4), 383-396.
- Osborne, J. (2011a): "Oral learner corpora and assessment of fluency in the Common European Framework", *New Trends in corpora and language learning*, eds. A. Frankenberg-Garcia, L. Flowerdew, G. Aston. London: Continuum. 181-197.
- Osborne, J. (2011b): "Fluency, complexity and informativeness in native and non-native speech ", *International Journal of Corpus Linguistics* 16 (2), in press.
- Paikeday, T.M. (1985): *The native speaker is dead!*, Toronto: Paikeday.
- Palmer, G.B. (1996): *Toward a theory of cultural linguistics*. Austin, TX: University of Texas Press.
- Patkowski, M. (1990): "Age and accent in a foreign language: A reply to James Emil Flege", *Applied Linguistics* 11 (1), 73-89.

- Pawley, A. (1991): "Saying things in Kalam", *Man and a half: Essays in pacific anthropology and ethnography in honour of Ralph Bulmer*, ed. A. Pawley. Auckland: The Polynesian Society, 432-44.
- Pawley, A. & F.H. Syder (1983): "Two puzzles for linguistic theory: Native-like selection and native-like fluency", *Language and communication*, eds. J. Richards & R. Schmidt. London: Longman. 191-226.
- Pawley, A. & F.H. Syder (2000): "The one-clause-at-a-time hypothesis", *Perspectives on fluency*, ed. H. Riggensbach. Ann Arbor: The University of Michigan Press. 163-199.
- Pennington, M.C. (1998): "The teachability of phonology in adulthood: A re-examination", *International Review of Applied Linguistics in Language Teaching* 36 (4), 323-341.
- Piske, T. (2001): *Artikulatorische Muster im frühen Laut- und Lexikonerwerb*. Tübingen: Gunter Narr.
- Piller, I. (2001): "Who, if anyone, is a native speaker?", *Anglistik* 12 (2), 109-121.
- Pimsleur, P., C. Hancock & P. Furey (1977): "Speech rate and listening comprehension", *Viewpoints on English as a second language*, eds. M. Burt, H. Dulay & M. Finocchiaro. New York: Regents. 27-34.
- Porter, P. (1986): "How learners talk to each other: input and interaction in task-centred discussions", *Talking to learn: Conversation in second language acquisition*, ed. R. Day. Rowley, MA: Newbury House. 200-222.
- Poulish, N. (1999): *Slips of the tongue: Speech errors in first and second language production*. Amsterdam, Philadelphia: John Benjamins.
- Prodromou, L. (2007): "Is ELF a variety of English?", *English Today* 23 (2), 47-53.
- Puccinelli, N. (2010): "Nonverbal communicative competence", *Handbook of communication competence*, eds. G. Rickheit & H. Strohner. Berlin: Walter de Gruyter. 257-275.
- Raupach, M. (1980): "Temporal variables in first and second language speech production", *Temporal variables in speech: Studies in honour of Frieda Goldman-Eisler*, eds. H. W. Dechert & M. Raupach. The Hague: Mouton de Gruyter. 263-270.
- Raupach, M. (1983): "Analysis and evaluation of communication strategies", *Strategies in interlanguage communication*, eds. C. F{rch & G. Kasper. London: Longman. 199-209.
- Raupach, M. (1987): "Procedural learning in advanced learners of a foreign language", *The advanced language learner*, eds. J.A. Coleman & R. Towell. London: Centre for Information on Language Teaching and Research. 123-155.
- Rayson, P. & R. Garside (2000): "Comparing corpora using frequency profiling", *Proceedings of the workshop on comparing corpora, held in conjunction with the 38th annual meeting of the Association for Computational Linguistics (ACL 2000)*, Hong Kong, eds. A. Kilgariff & T. Berber Sardina. Hong Kong: The Association for Computational Linguistics. 1-6.

- Read, J. & P. Nation (2004): "Measurement of formulaic sequences", *Formulaic Sequences*, ed. N. Schmitt. Amsterdam, Philadelphia: John Benjamins. 23-35.
- Richards, B. (1983): "Type/token ratios: What do they really tell us?", *Journal of Child Language* 14 (2), 201-209.
- Ridha Ben Maad, M. (2008): "Time pressure and within-task variation in EFL oral performance", *Electronic Journal of Foreign Language Teaching* 5 (1), 5-12.
- Riding, R.J. & D.J.T. Vincent (1980): "Listening Comprehension: The effects of sex, age, passage structure and speech rate", *Educational Review* 32 (3), 259-266.
- Riggenbach, H. (1991): "Toward an understanding of fluency: A microanalysis of nonnative speaker conversations", *Discourse Processes* 14 (4), 423-441.
- Riggenbach, H. (ed.) (2000): *Perspectives on fluency*. Ann Arbor: The University of Michigan Press.
- Riley, P. (1989): "'Well don't blame me!' On the interpretation of pragmatic errors", *Contrastive pragmatics*, ed. W. Olesky. Amsterdam: John Benjamins. 231-249.
- Riney T. & N. Tagaki (1999): "Global foreign accent and voice onset time among Japanese EFL speakers", *Language Learning* 49 (1), 275-302.
- Ringbom, H. (1999): "High frequency verbs in the ICLE corpus", *Explorations in corpus linguistics*, ed. A. Renouf. Amsterdam: Rodopi. 191-200.
- Rivers, W.M. (1981): *Teaching foreign language skills*, 2nd Edition, Chicago: University of Chicago Press.
- Römer, U. (2005): *Progressives, patterns, pedagogy: A corpus-driven approach to English progressive forms, functions, contexts and didactics*. Amsterdam: John Benjamins.
- Rossiter, M.J. (2009): "Perceptions of L2 fluency by native and non-native speakers of English", *The Canadian Modern Language Review/La Revue canadienne des langues vivantes* 65 (3), 395-412.
- Rossiter, M.J., T.M. Derwing, L.G. Manimtim, R.I. Thomson (2010): "Oral Fluency: The neglected component in the communicative language classroom", *The Canadian Modern Language Review/La Revue canadienne des langues vivantes* 66 (4), 583-606.
- Rühlemann, C. (2006): "Coming to terms with conversational grammar: 'Dislocation' and 'dysfluency'", *International Journal of Corpus Linguistics* 11(4), 385-409.
- Rühlemann, C. (2007): *Conversation in context: A corpus-driven approach*. London, New York: Continuum.
- Sato, K. & R.C. Kleinsasser (1999): "Communicative language teaching (CLT): Practical understandings", *Modern Language Journal* 83 (4), 494-517.
- Saz Rubio, M. M. D. (2007): *English discourse markers of reformulation*. Bern: Peter Lang.

- Schauer, G.A. & S. Adolphs (2006): "Expressions of gratitude in corpus and DCT data: Vocabulary, formulaic sequences, and pedagogy", *System* 34 (1), 119-134.
- Schauer, G.A. (2007): "Finding the right words in the study abroad context: The development of German learners' use of external modifiers in English", *Intercultural Pragmatics* 4 (2), 193-220.
- Scheuer, S. (2002): "What makes foreign accent sound foreign?", *Proceedings of New Sounds 2000*, eds. A. James & J. Leather. Klagenfurt: University of Klagenfurt. 306-314.
- Schilperoord, J. (1996): *It's about time: Temporal aspects of cognitive processes in text production*. Amsterdam: Rodopi.
- Schmidt, R. (1992): "Psychological mechanisms underlying second language fluency", *Studies in Second Language Acquisition* 14 (4), 357-385.
- Schmidt, R. (2000): "Foreword", *Perspectives on fluency*, ed. H. Riggensbach. Ann Arbor: The University of Michigan Press. v-viii.
- Schmitt, N. (2000): *Vocabulary and language teaching*. Cambridge: Cambridge University Press.
- Schmitt, N. & R. Carter (2004): "Formulaic sequences in action: An introduction", *Formulaic sequences*, ed. N. Schmitt. Amsterdam: John Benjamins. 1-22.
- Schmitt, N., S. Grandage & S. Adolphs (2004): "Are corpus-derived recurrent clusters psycholinguistically valid?", *Formulaic sequences*, ed. N. Schmitt. Amsterdam: John Benjamins. 127-151.
- Scott, M. (1998): *WordSmith Tools*. Version 4.00. Oxford: Oxford University Press.
- Scovel, T. (1988): *A time to speak. A psycholinguistic inquiry to the critical period for human speech*. Rowley, MA: Newbury House.
- Segalowitz, N. (2010): *Cognitive bases of second language fluency*. New York: Routledge.
- Segalowitz, N. & B. Freed (2004): "Context, contact and cognition in oral fluency acquisition", *Studies in Second Language Acquisition* 26 (2), 173-199.
- Selinker, L. (1972): "Interlanguage", *International Review of Applied Linguistics* 10 (1-4), 209-231.
- Shiffrin, D. (1987): *Discourse Markers*. Cambridge: Cambridge University Press.
- Shin, D.H. (1989): *Effect of formal vs. informal environments and Krashen's Monitor Model*, Unpublished M.A. Thesis, University of Birmingham.
- Seliger, H. (1980): "Utterance planning and correction behaviour: Its function in the grammar construction process for second language learners", *Towards a cross-linguistic assessment of speech production*, ed. H.-W. Dechert & M. Raupach. Frankfurt am Main: Peter Lang. 87-100.
- Sichel, H.S. (1986): "Word frequency distributions and type-token characteristics", *Mathematical Scientist* 11, 45-72.

- Sicola, L. (2009): *No, they won't 'just sound like each other'. NNS-NNS negotiated interaction and attention to phonological form on targeted L2 pronunciation tasks*. Frankfurt am Main: Peter Lang.
- Siepmann, D. (2007): "Wortschatz und Grammatik: zusammenbringen, was zusammengehört", *Beiträge zur Fremdsprachenvermittlung* 46, 59-80.
- Sinclair, J. (1991): *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Skehan, P. (1998): *A cognitive approach to language learning*. Oxford: Oxford University Press.
- Soler, E.A. (2008): "Investigating pragmatic language learning in foreign language classrooms", *International Review of Applied Linguistics in Language Teaching* 46 (2), 173-195.
- Sperber, D. & D. Wilson (1986): *Relevance*. Oxford: Blackwell.
- Stefanowitsch, A. & S.Th. Gries (2003): "Collostructions: Investigating the interaction of words and constructions", *International Journal of Corpus Linguistics* 8 (2), 209-243.
- Stenström, A.-B. (1994): *An introduction to spoken interaction*. London: Longman.
- Stirk, H., D. Binnenpoorte & C. Cucchiaroni (2005): "Multiword expressions in spontaneous speech: Do we really speak like that?", *Proceedings of Interspeech 2005*, 1161-1164. Available on <http://lands.let.kun.nl/literature/stirk.2005.2.pdf> (last accessed on 9 May 2011).
- Suenobu, M., K. Kanzaki, S. Yamane & R. Young (1986): "Listening comprehension and the process of information acquisition by non-native speakers of English", *International Review of Applied Linguistics in Language Teaching* 24 (3), 239-248.
- Svartvik, J. (1990): "The TESS project", *The London Lund Corpus of Spoken English: Description and research*, ed. J. Svartvik. Amsterdam: Rodopi. 203-218.
- Taguchi, N. (2007): "Task difficulty in oral speech act production", *Applied Linguistics* 28 (1), 113-135.
- Tajima, K., R. Port & J. Dalby (1997): "Effects of temporal correction on intelligibility of foreign-accented English", *Journal of Phonetics* 25 (1), 1-24.
- Tarone, E. (1977): "Conscious communication strategies in interlanguage", *On TESOL 77: Teaching and learning English as a second language: Trends in research and practice*, eds. H.D. Brown, C. Yorio & R. Crymes. Washington: TESOL. 194-203.
- Tarone, E. (1981): "Some thoughts on the notion of communication strategy", *TESOL Quarterly* 15 (3), 285-295.
- Tarone, E. (1983): "On the variability of interlanguage systems", *Applied Linguistics* 4 (2), 142-164.
- Tauroza, S. & D. Allison (1990): "Speech Rates in British English", *Applied Linguistics*, 11 (1), 90-105.

- Tavakoli, P. & P. Foster (2008): "Task design and second language performance: The effect of narrative type on learner output", *Language Learning* 58 (2), 439-473.
- Tench, P. (1981): *Pronunciation skills*. London: Macmillan.
- Trenchs-Parera, M. (2009): "Effects of formal instruction and a stay abroad on the acquisition of native-like oral fluency", *The Canadian Modern Language Review/La Revue canadienne des langues vivantes* 65 (3), 365-393.
- Thomas, J. (1983): "Cross-cultural pragmatic failure", *Applied Linguistics* 4 (2), 91-112.
- Thornbury, S. & D. Slade (2006): *Conversation: From description to pedagogy*. Cambridge: Cambridge University Press.
- Toivanen, J. (2007): "Tone choice in the English intonation of Finns", *Cross-cultural Pragmatics and Interlanguage English*, eds. B. Kraft & R. Geluykens. München: Lincom. 243-260.
- Tottie, G. & S. Svaldus (2009): "Er/uh, erm/um: planners in varieties and registers of English", Paper presented at the 30th ICAME conference, Lancaster.
- Towell, R. (1987): "Approaches to the analysis of the oral language development of the advanced learner", *The advanced language learner*, ed. J. Coleman & R. Towell. London: AFLS/CILT. 157-183.
- Towell, R. (2002): "Relative degrees of fluency. A comparative case study of advanced learners of French", *International Review of Applied Linguistics in Language* 40 (2), 117-150.
- Towell, R., R. Hawkins & N. Bazergui (1996): "The development of fluency in advanced learners of French", *Applied Linguistics* 17 (1), 84-119.
- Trofimovich, P. & W. Baker (2006): "Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech", *Studies in Second Language Acquisition* 28 (1), 1-30.
- Vellenga, H. (2004). "Learning pragmatics from ESL & EFL textbooks: How likely?" *Teaching English as a Second or foreign Language – Electronic Journal*, 8 (2). Available on <http://tesl-ej.org/ej30/a3.html> (last accessed on 9 May 2011).
- Van Els, T. & K. de Bot (1987): "The role of intonation in foreign accent", *The Modern Language Journal* 71 (2), 147-155.
- Van Gelderen, A. (1994): "Prediction of global ratings of fluency and delivery in narrative discourse by linguistic and phonetic measures – oral performances of students aged 11-12 years", *Language Testing* 11 (3), 291-319.
- Van Moere, A. (2006): "Validity evidence in a university group oral test", *Language Testing* 23 (4), 411-440.
- Vygotsky, L. (1978): *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Wagner, E. (2010): "Survey research", *Continuum companion to research methods in applied linguistics*, eds. B. Paltridge & A. Phakiti. London: Continuum. 22-37.

- Wells, J.C. (1982): *Accents of English : An introduction*. Cambridge: Cambridge University Press.
- Wells, J.C. (2006): *English intonation: An introduction*. Cambridge: Cambridge University Press.
- Wennestrom, A. (2001): *The music of everyday speech: Prosody and discourse analysis*. Oxford: Oxford University Press.
- Wichmann, A. (2000): *Intonation in text and discourse: Beginnings, middles and ends*. London: Pearson.
- Widdowson, H.G. (1990): *Aspects of language teaching*. Oxford: Oxford University Press.
- Wierzbicka, A. (1991): *Cross-cultural pragmatics: The semantics of human interaction*. Berlin: Mouton de Gruyter.
- Wiktorsson, M. (2001): "Register differences between prefabs in native and EFL English", *The Department of English in Lund: Working Papers in Linguistics (Vol. 1)*, eds. S. Manninen & C. Paradis. Lund: Lund University. 85-94.
- Willems, N. (1982): *English Intonation from a Dutch point of view*. Ambach: Intercontinental Graphics.
- Wood, D. (2006): "Uses and functions of formulaic sequences in second language speech: An exploration of the foundations of fluency", *The Canadian Modern Language Review/La Revue canadienne des langues vivantes*, 63 (1), 13-33.
- Wood, D. (2010): *Formulaic language and second language speech fluency: Background, evidence and classroom applications*. London: Continuum.
- Wray, A. & M.R. Perkins (2000): "The functions of formulaic language: An integrated model", *Language and Communication* 20 (1), 1-28.
- Wray, A. (2002), *Formulaic language and the lexicon*, Cambridge: Cambridge University Press.
- Wray, A. (2008), *Formulaic language: Pushing the boundaries*, Oxford: Oxford University Press.
- Wulff, S. (2008): *Rethinking idiomaticity: A usage-based approach*. London, New York: Continuum.
- Yavaş, Mehmet (2009): *Applied English phonology*, Malden, MA: Blackwell.

Appendix

Appendix A1: Questionnaire of pilot study

(3) Ratings:

Please have a look at the questions and read the instructions below. Then, listen to the sound files and rate the overall degree of fluency for each learner (please check (x) from very low (1) to very high (10) based on your personal perception). Then, specify if you based your judgement on any of the given features (please check (x) yes or no). After your rating, please give feedback on this questionnaire and/or the study.

Learner 1:

- A. What is your overall perception of degree of fluency of this learner (from 1 (lowest) to 10 (highest))?

1	2	3	4	5	6	7	8	9	10

- B. Did you base your judgement on any of the following features?
Please check (x) yes or no

1. Temporal Variables

- a. **Speech Rate ("speed")**
O No

O Yes

If **yes**, please specify how "native-like"/natural you perceive the learner's speech rate from 1 (far from native-like) to 10 (native-like)

1	2	3	4	5	6	7	8	9	10

- b. **Overall number of unfilled pauses ("silences")** O Yes

O No

If **yes**, please specify how "native-like"/natural you perceive the overall number of silences in the learner's performance from 1 (far more silences than in native speech) to 10 (equal number of silences as in native speech)

1	2	3	4	5	6	7	8	9	10

- c. Overall number of filled pauses (e.g. *erm, uh, er*) ☐ Yes ☐ No

If **yes**, please specify how "native-like"/natural you perceive the learner's number of filled pauses from 1 (far more filled pauses than in native speech) to 10 (equal number of filled pauses as in native speech)

1	2	3	4	5	6	7	8	9	10

2. Performance Phenomena/Naturalness of speech

- d. The use of discourse markers (e.g. *you know, like*) ☐ Yes ☐ No

If **yes**, please specify how "native-like" you perceive the learner's range and distribution of discourse markers from 1 (far from native-like) to 10 (native-like)

1	2	3	4	5	6	7	8	9	10

- e. The number of repetitions (e.g. *The the painter...*) ☐ Yes ☐ No

If **yes**, please specify how "native-like" you perceive the learner's number and position of repetitions from 1 (far from native-like) to 10 (native-like)

1	2	3	4	5	6	7	8	9	10

- f. The number self-corrections or completions by the listener (e.g. *She sit . er she sits*) ☐ Yes ☐ No

If **yes**, please specify how "native-like" you perceive the learner's performance from 1 (far from native-like) to 10 (native-like)

1	2	3	4	5	6	7	8	9	10

3. Accuracy

- g. Number of errors ☐ Yes ☐ No

If **yes**, please specify how "native-like"/natural you perceive the learner's number of errors from 1 (far from native-like) to 10 (native-like)

1	2	3	4	5	6	7	8	9	10

h. German Accent**O Yes O No**

If yes, please specify how "native-like"/natural you perceive the learner's accent from 1 (far from native-like) to 10 (native-like)

1	2	3	4	5	6	7	8	9	10

3. Idiomaticity**i. The adequate vocabulary for spoken language (register) O Yes O No**

If yes, please specify how "native-like" you perceive the learner's performance from 1 (far from native-like) to 10 (native-like)

1	2	3	4	5	6	7	8	9	10

j. The number of unidiomatic expressions (i.e. things that are not wrong, but sound odd to you) O Yes O No

If yes, please specify how "native-like"/natural you perceive the learner's performance from 1 (far from native-like) to 10 (native-like)

1	2	3	4	5	6	7	8	9	10

k. The number of prefabricated units the learner uses (e.g. *the thing is that, to put it in a nutshell*, etc.) O Yes O No

If yes, please specify how "native-like"/natural you perceive the learner's performance from 1 (far from native-like) to 10 (native-like)

1	2	3	4	5	6	7	8	9	10

4. Other Features?**l. Other (please specify) _____**

Please specify how "native-like" you perceive the learner's performance from 1 (far from native-like) to 10 (native-like)

1	2	3	4	5	6	7	8	9	10

m. Other (please specify) _____

Please specify how "native-like" you perceive the learner's performance from 1 (far from native-like) to 10 (native-like)

1	2	3	4	5	6	7	8	9	10

n. Other (please specify) _____

Please specify how "native-like" you perceive the learner's performance from 1 (far from native-like) to 10 (native-like)

1	2	3	4	5	6	7	8	9	10

Learner 2:

A. What is your overall perception of degree of fluency of this learner (from 1 (lowest) to 10 (highest))?

1	2	3	4	5	6	7	8	9	10

B. Did you base your judgement on any of the following features?
Please check (x) yes or no

1. Temporal Variables

a. **Speech Rate ("speed")**

O Yes

O No

If yes, please specify how "native-like"/natural you perceive the learner's speech rate from 1 (far from native-like) to 10 (native-like)

1	2	3	4	5	6	7	8	9	10

b. **Overall number of unfilled pauses ("silences")**

O Yes

O No

If yes, please specify how "native-like"/natural you perceive the overall number of silences in the learner's performance from 1 (far more silences than in native speech) to 10 (equal number of silences as in native speech)

1	2	3	4	5	6	7	8	9	10

- c. Overall number of filled pauses (e.g. *erm, uh, er*) ☐ Yes ☐ No

If **yes**, please specify how "native-like"/natural you perceive the learner's number of filled pauses from 1 (far more filled pauses than in native speech) to 10 (equal number of filled pauses as in native speech)

1	2	3	4	5	6	7	8	9	10

2. Performance Phenomena/Naturalness of speech

- d. The use of discourse markers (e.g. *you know, like*) ☐ Yes ☐ No

If **yes**, please specify how "native-like" you perceive the learner's range and distribution of discourse markers from 1 (far from native-like) to 10 (native-like)

1	2	3	4	5	6	7	8	9	10

- e. The number of repetitions (e.g. *The the painter...*) ☐ Yes ☐ No

If **yes**, please specify how "native-like" you perceive the learner's number and position of repetitions from 1 (far from native-like) to 10 (native-like)

1	2	3	4	5	6	7	8	9	10

- f. The number self-corrections or completions by the listener (e.g. *She sit . er she sits*) ☐ Yes ☐ No

If **yes**, please specify how "native-like" you perceive the learner's performance from 1 (far from native-like) to 10 (native-like)

1	2	3	4	5	6	7	8	9	10

3. Accuracy

- g. Number of errors ☐ Yes ☐ No

If **yes**, please specify how "native-like"/natural you perceive the learner's number of errors from 1 (far from native-like) to 10 (native-like)

1	2	3	4	5	6	7	8	9	10

h. German Accent**O Yes O No**

If **yes**, please specify how "native-like"/natural you perceive the learner's accent from 1 (far from native-like) to 10 (native-like)

1	2	3	4	5	6	7	8	9	10

3. Idiomaticity**i. The adequate vocabulary for spoken language (register) O Yes O No**

If **yes**, please specify how "native-like" you perceive the learner's performance from 1 (far from native-like) to 10 (native-like)

1	2	3	4	5	6	7	8	9	10

j. The number of unidiomatic expressions (i.e. things that are not wrong, but sound odd to you) O Yes O No

If **yes**, please specify how "native-like"/natural you perceive the learner's performance from 1 (far from native-like) to 10 (native-like)

1	2	3	4	5	6	7	8	9	10

k. The number of prefabricated units the learner uses (e.g. *the thing is that, to put it in a nutshell*, etc.) O Yes O No

If **yes**, please specify how "native-like"/natural you perceive the learner's performance from 1 (far from native-like) to 10 (native-like)

1	2	3	4	5	6	7	8	9	10

4. Other Features?**l. Other (please specify) _____**

Please specify how "native-like" you perceive the learner's performance from 1 (far from native-like) to 10 (native-like)

1	2	3	4	5	6	7	8	9	10

m. Other (please specify) _____

Please specify how "native-like" you perceive the learner's performance from 1 (far from native-like) to 10 (native-like)

1	2	3	4	5	6	7	8	9	10

n. Other (please specify) _____

Please specify how "native-like" you perceive the learner's performance from 1 (far from native-like) to 10 (native-like)

1	2	3	4	5	6	7	8	9	10

Learner 3:

A. What is your overall perception of degree of fluency of this learner (from 1 (lowest) to 10 (highest))?

1	2	3	4	5	6	7	8	9	10

B. Did you base your judgement on any of the following features?
Please check (x) yes or no

1. Temporal Variables**a. Speech Rate ("speed")****O Yes****O No**

If **yes**, please specify how "native-like"/natural you perceive the learner's speech rate from 1 (far from native-like) to 10 (native-like)

1	2	3	4	5	6	7	8	9	10

b. Overall number of unfilled pauses ("silences")**O Yes****O No**

If **yes**, please specify how "native-like"/natural you perceive the overall number of silences in the learner's performance from 1 (far more silences than in native speech) to 10 (equal number of silences as in native speech)

1	2	3	4	5	6	7	8	9	10

- c. Overall number of filled pauses (e.g. *erm, uh, er*) ☐ Yes ☐ No

If **yes**, please specify how "native-like"/natural you perceive the learner's number of filled pauses from 1 (far more filled pauses than in native speech) to 10 (equal number of filled pauses as in native speech)

1	2	3	4	5	6	7	8	9	10

2. Performance Phenomena/Naturalness of speech

- d. The use of discourse markers (e.g. *you know, like*) ☐ Yes ☐ No

If **yes**, please specify how "native-like" you perceive the learner's range and distribution of discourse markers from 1 (far from native-like) to 10 (native-like)

1	2	3	4	5	6	7	8	9	10

- e. The number of repetitions (e.g. *The the painter...*) ☐ Yes ☐ No

If **yes**, please specify how "native-like" you perceive the learner's number and position of repetitions from 1 (far from native-like) to 10 (native-like)

1	2	3	4	5	6	7	8	9	10

- f. The number self-corrections or completions by the listener (e.g. *She sit . er she sits*) ☐ Yes ☐ No

If **yes**, please specify how "native-like" you perceive the learner's performance from 1 (far from native-like) to 10 (native-like)

1	2	3	4	5	6	7	8	9	10

3. Accuracy

- g. Number of errors ☐ Yes ☐ No

If **yes**, please specify how "native-like"/natural you perceive the learner's number of errors from 1 (far from native-like) to 10 (native-like)

1	2	3	4	5	6	7	8	9	10

h. German Accent**O Yes O No**

If **yes**, please specify how "native-like"/natural you perceive the learner's accent from 1 (far from native-like) to 10 (native-like)

1	2	3	4	5	6	7	8	9	10

3. Idiomaticity**i. The adequate vocabulary for spoken language (register) O Yes O No**

If **yes**, please specify how "native-like" you perceive the learner's performance from 1 (far from native-like) to 10 (native-like)

1	2	3	4	5	6	7	8	9	10

j. The number of unidiomatic expressions (i.e. things that are not wrong, but sound odd to you) O Yes O No

If **yes**, please specify how "native-like"/natural you perceive the learner's performance from 1 (far from native-like) to 10 (native-like)

1	2	3	4	5	6	7	8	9	10

k. The number of prefabricated units the learner uses (e.g. *the thing is that, to put it in a nutshell*, etc.) O Yes O No

If **yes**, please specify how "native-like"/natural you perceive the learner's performance from 1 (far from native-like) to 10 (native-like)

1	2	3	4	5	6	7	8	9	10

4. Other Features?**l. Other (please specify) _____**

Please specify how "native-like" you perceive the learner's performance from 1 (far from native-like) to 10 (native-like)

1	2	3	4	5	6	7	8	9	10

m. Other (please specify) _____

Please specify how "native-like" you perceive the learner's performance from 1 (far from native-like) to 10 (native-like)

1	2	3	4	5	6	7	8	9	10

n. Other (please specify) _____

Please specify how "native-like" you perceive the learner's performance from 1 (far from native-like) to 10 (native-like)

1	2	3	4	5	6	7	8	9	10

(4) Feedback:

- (a) How easy was it for you to rate the overall degree of fluency? Please check (x) from 1 (very easy) to 10 (very hard)?

1	2	3	4	5	6	7	8	9	10

Comment:

- (b) How easy was it for you to rate the given features? Please check (x) from 1 (very easy) to 10 (very hard)?**

1	2	3	4	5	6	7	8	9	10

Please name the features which were unclear/difficult/inappropriate for you to rate and why you had problems with them.

- (c) Do you think anything is missing in this questionnaire? Is there anything you would add to it?**

(d) Do you have any other comments on the questionnaire/project?

THANK YOU VERY MUCH FOR PARTICIPATING IN THIS STUDY !!!



Appendix A2
List of 3-grams and 4-grams in
LOCNEC and LINDSEI-GE

Table A1: List of all 3-grams in LOCNEC with a minimum frequency of 5

N	Word	Freq.	ptw	N	Word	Freq	ptw
1	I DON'T KNOW	209	1.76	871	ME AND I	7	0.06
2	A LOT OF	129	1.09	872	MEAN I DON'T	7	0.06
3	YEAH YEAH YEAH	111	0.94	873	MEAN I WAS	7	0.06
4	AND IT WAS	94	0.79	874	MM YEAH I	7	0.06
5	I MEAN I	86	0.73	875	MORE OF A	7	0.06
6	IT WAS A	80	0.67	876	NICE TO HAVE	7	0.06
7	IT WAS JUST	66	0.56	877	NO I DIDN'T	7	0.06
8	I'D LIKE TO	62	0.52	878	NOT SO MUCH	7	0.06
9	DON'T KNOW I	61	0.51	879	NOW AND THEN	7	0.06
10	YEAH YEAH I	59	0.50	880	OF HER AND	7	0.06
11	I WENT TO	57	0.48	881	OF IT I	7	0.06
12	THINGS LIKE THAT	57	0.48	882	OF THEM ARE	7	0.06
13	AND I WAS	56	0.47	883	OH RIGHT YEAH	7	0.06
14	YEAH IT WAS	56	0.47	884	ONE OF THEM	7	0.06
15	YOU HAVE TO	56	0.47	885	PART OF THE	7	0.06
16	I THINK I	53	0.45	886	PEOPLE YOU KNOW	7	0.06
17	IT WAS REALLY	53	0.45	887	PLEASED WITH IT	7	0.06
18	I THINK IT'S	52	0.44	888	REALLY GOOD FILM	7	0.06
19	A BIT OF	50	0.42	889	REALLY GOOD I	7	0.06
20	AT THE MOMENT	50	0.42	890	RIGHT AT THE	7	0.06
21	ONE OF THE	50	0.42	891	SAYS NO NO	7	0.06
22	SORT OF LIKE	50	0.42	892	SEE IT AND	7	0.06
23	AND THEN I	49	0.41	893	SHE DOESN'T LOOK	7	0.06
24	IT WAS IT	49	0.41	894	SHE'S NOT VERY	7	0.06
25	WAS IT WAS	49	0.41	895	SHOWS IT TO	7	0.06
26	YOU KNOW YOU	49	0.41	896	SO WE WENT	7	0.06
27	YOU KNOW I	48	0.40	897	SORT OF YEAH	7	0.06
28	WHEN I WAS	47	0.40	898	TEACHING ENGLISH AS	7	0.06
29	AND YOU KNOW	46	0.39	899	THAT AS WELL	7	0.06
30	I DON'T THINK	46	0.39	900	THAT ERM I	7	0.06
31	I HAD TO	45	0.38	901	THAT WAS LIKE	7	0.06
32	SORT OF THING	44	0.37	902	THAT YOU CAN	7	0.06
33	SO IT WAS	42	0.35	903	THE CENTRE OF	7	0.06
34	TO GO TO	42	0.35	904	THE PEOPLE WHO	7	0.06
35	THAT'S RIGHT YEAH	41	0.35	905	THE SECOND PICTURE	7	0.06
36	IT WAS LIKE	40	0.34	906	THEN AND THEN	7	0.06
37	YEAH I MEAN	40	0.34	907	THEN YOU GET	7	0.06
38	YEAH THAT'S RIGHT	40	0.34	908	THERE FOR A	7	0.06
39	IT WAS QUITE	39	0.33	909	THERE WAS ONE	7	0.06
40	I THINK IT	38	0.32	910	THERE WAS THERE	7	0.06
41	I WANTED TO	38	0.32	911	THEY DID IT	7	0.06
42	YOU KNOW IT'S	37	0.31	912	THEY HAVE A	7	0.06
43	ALL THE TIME	36	0.30	913	THINK SO YEAH	7	0.06
44	BUT I MEAN	36	0.30	914	TIME I WAS	7	0.06
45	ERM I DON'T	36	0.30	915	TO LIVE IN	7	0.06
46	AND THINGS LIKE	35	0.30	916	TO WORK WITH	7	0.06
47	I END OF	35	0.30	917	UP TO THE	7	0.06
48	IT WAS VERY	34	0.29	918	USED TO GO	7	0.06
49	YEAH I THINK	34	0.29	919	USED TO IT	7	0.06

50	A COUPLE OF	33 0.28	920	WANTED TO SEE	7 0.06
51	BUT I THINK	33 0.28	921	WAS A REALLY	7 0.06
52	I WANT TO	31 0.26	922	WAS THE FIRST	7 0.06
53	SO THAT WAS	31 0.26	923	WE DIDN'T HAVE	7 0.06
54	WE WENT TO	30 0.25	924	WE GO TO	7 0.06
55	AND I THINK	29 0.24	925	WE HAVE A	7 0.06
56	BUT I DON'T	29 0.24	926	WE SORT OF	7 0.06
57	I WAS THERE	29 0.24	927	WE WERE IN	7 0.06
58	IT IT WAS	29 0.24	928	WELL I DID	7 0.06
59	A SORT OF	28 0.24	929	WELL IN THE	7 0.06
60	BUT IT WAS	28 0.24	930	WHAT I WANTED	7 0.06
61	SOMETHING LIKE THAT	28 0.24	931	WHAT IT'S LIKE	7 0.06
62	YEAH YEAH IT'S	28 0.24	932	WHERE I WAS	7 0.06
63	AND THAT WAS	26 0.22	933	WHICH IS GOOD	7 0.06
64	END OF THE	26 0.22	934	WHICH IS THE	7 0.06
65	LIKE THAT AND	26 0.22	935	WOULD LIKE TO	7 0.06
66	THERE WAS A	26 0.22	936	YEAH A LOT	7 0.06
67	TO SORT OF	26 0.22	937	YEAH ERM AND	7 0.06
68	WAS REALLY GOOD	26 0.22	938	YEAH I'D LIKE	7 0.06
69	YEAH ERM I	26 0.22	939	YEAH IT'S JUST	7 0.06
70	YEAH I DON'T	26 0.22	940	YEAH IT'S VERY	7 0.06
71	YEAH YEAH IT	26 0.22	941	YEAH SO ERM	7 0.06
72	ALL HER FRIENDS	25 0.21	942	YEAH SO YEAH	7 0.06
73	I MEAN IT'S	25 0.21	943	YEAR AND I	7 0.06
74	I USED TO	25 0.21	944	YES IT IS	7 0.06
75	MM YEAH YEAH	25 0.21	945	YOU CAN'T REALLY	7 0.06
76	THE FACT THAT	25 0.21	946	YOU KNOW A	7 0.06
77	THOUGHT IT WAS	25 0.21	947	YOU KNOW ALL	7 0.06
78	YOU KNOW THE	25 0.21	948	YOU KNOW AS	7 0.06
79	A LOT MORE	24 0.20	949	YOU KNOW ERM	7 0.06
80	I SORT OF	24 0.20	950	YOU KNOW WHERE	7 0.06
81	IN THE FIRST	24 0.20	951	A A A	6 0.05
82	IN THE MORNING	24 0.20	952	A BIT TOO	6 0.05
83	IT'S A BIT	24 0.20	953	A BIT WORRIED	6 0.05
84	LIKE TO GO	24 0.20	954	A FEW MONTHS	6 0.05
85	SOME OF THE	24 0.20	955	A FEW TIMES	6 0.05
86	THAT SORT OF	24 0.20	956	A FEW YEARS	6 0.05
87	YEAH YEAH THAT'S	24 0.20	957	A HALF HOURS	6 0.05
88	YOU KNOW IT	24 0.20	958	A LOT YEAH	6 0.05
89	A LITTLE BIT	23 0.19	959	A YEAR AND	6 0.05
90	ERM I THINK	23 0.19	960	ABOUT IT SO	6 0.05
91	ERM WELL I	23 0.19	961	ABOUT THE SAME	6 0.05
92	I MEAN IT	23 0.19	962	AGAIN AND HE	6 0.05
93	SORT OF THE	23 0.19	963	ALL YOU KNOW	6 0.05
94	THAT WAS A	23 0.19	964	AND ALSO THE	6 0.05
95	WANTED TO DO	23 0.19	965	AND AND THEN	6 0.05
96	YEAH YEAH AND	23 0.19	966	AND ER IT	6 0.05
97	AND I I	22 0.19	967	AND ER JUST	6 0.05
98	ERM IT WAS	22 0.19	968	AND ERM THE	6 0.05
99	I CAN'T REMEMBER	22 0.19	969	AND ERM WE	6 0.05
100	I I DON'T	22 0.19	970	AND HAD A	6 0.05
101	I WAS IN	22 0.19	971	AND HE HE	6 0.05

102	I'M NOT SURE	22 0.19	972	AND HE WAS	6 0.05
103	IT WAS THE	22 0.19	973	AND I LIKE	6 0.05
104	NO I DON'T	22 0.19	974	AND I REALLY	6 0.05
105	NO NO NO	22 0.19	975	AND IT'S VERY	6 0.05
106	STUFF LIKE THAT	22 0.19	976	AND SEE IT	6 0.05
107	WANT TO DO	22 0.19	977	AND SO IT	6 0.05
108	WAS A BIT	22 0.19	978	AND STAY WITH	6 0.05
109	AND I THOUGHT	21 0.18	979	AND THE FIRST	6 0.05
110	AS WELL SO	21 0.18	980	AND THE THE	6 0.05
111	BE ABLE TO	21 0.18	981	AND THEN A	6 0.05
112	BIT OF A	21 0.18	982	AND THEN WHEN	6 0.05
113	GO TO THE	21 0.18	983	AND THEY ALL	6 0.05
114	I DON'T REALLY	21 0.18	984	AND THEY DON'T	6 0.05
115	I HAD A	21 0.18	985	AND THEY'RE NOT	6 0.05
116	MOST OF THE	21 0.18	986	AND YEAH I	6 0.05
117	QUITE A LOT	21 0.18	987	AS A SORT	6 0.05
118	SHE DOESN'T LIKE	21 0.18	988	AS FAR AS	6 0.05
119	SO I WAS	21 0.18	989	AS GOOD AS	6 0.05
120	THINK IT WAS	21 0.18	990	AS PART OF	6 0.05
121	TO DO IT	21 0.18	991	AS SOON AS	6 0.05
122	A BIT MORE	20 0.17	992	AS WELL BECAUSE	6 0.05
123	AND STUFF LIKE	20 0.17	993	AS WELL COS	6 0.05
124	AND THEY WERE	20 0.17	994	AS WELL THAT	6 0.05
125	I DON'T LIKE	20 0.17	995	AT THE BOTTOM	6 0.05
126	IN A WAY	20 0.17	996	AT THE PICTURE	6 0.05
127	IT'S SORT OF	20 0.17	997	AT THE TOP	6 0.05
128	KNOW WHAT I	20 0.17	998	BACK TO IT	6 0.05
129	NO IT WAS	20 0.17	999	BECAUSE YOU HAVE	6 0.05
130	WENT TO SEE	20 0.17	1000	BECAUSE YOU YOU	6 0.05
131	YEAH I KNOW	20 0.17	1001	BUT ER AND	6 0.05
132	YOU KNOW AND	20 0.17	1002	BUT ER YEAH	6 0.05
133	AND THEN SHE	19 0.16	1003	BUT I HAVEN'T	6 0.05
134	ER IT WAS	19 0.16	1004	BUT I JUST	6 0.05
135	I DON'T I	19 0.16	1005	BUT I REALLY	6 0.05
136	I HAVE TO	19 0.16	1006	BUT I'D LIKE	6 0.05
137	THAT KIND OF	19 0.16	1007	BUT WHEN I	6 0.05
138	THE SORT OF	19 0.16	1008	BUT YOU CAN	6 0.05
139	TO GO AND	19 0.16	1009	CAME TO UNIVERSITY	6 0.05
140	TO HAVE A	19 0.16	1010	COULD DO IT	6 0.05
141	WAS A LOT	19 0.16	1011	COULD HAVE BEEN	6 0.05
142	YEAH YEAH ERM	19 0.16	1012	DO IT AND	6 0.05
143	YOU KNOW YEAH	19 0.16	1013	DO YOU WANT	6 0.05
144	A PICTURE OF	18 0.15	1014	DOESN'T LIKE THE	6 0.05
145	AND AND THE	18 0.15	1015	DON'T KNOW HOW	6 0.05
146	AND THERE WAS	18 0.15	1016	DON'T KNOW I'M	6 0.05
147	AS WELL I	18 0.15	1017	DON'T KNOW WHY	6 0.05
148	DON'T KNOW IF	18 0.15	1018	DON'T LIKE IT	6 0.05
149	HAVE TO GO	18 0.15	1019	ER AND I	6 0.05
150	I MEAN YOU	18 0.15	1020	ER I MEAN	6 0.05
151	I THINK THAT	18 0.15	1021	ER I WENT	6 0.05
152	I THINK THE	18 0.15	1022	ER IT'S NOT	6 0.05
153	I THOUGHT IT	18 0.15	1023	ERM AND IT	6 0.05

154	LIKE THAT YEAH	18 0.15	1024 ERM BUT THE	6	0.05
155	ON MY OWN	18 0.15	1025 ERM I DID	6	0.05
156	ONE OF MY	18 0.15	1026 ERM I MEAN	6	0.05
157	OR SOMETHING LIKE	18 0.15	1027 ERM IN THE	6	0.05
158	RIGHT YEAH YEAH	18 0.15	1028 ERM I'VE BEEN	6	0.05
159	SO I THINK	18 0.15	1029 ERM THAT WAS	6	0.05
160	TO BE A	18 0.15	1030 ERM YEAH YEAH	6	0.05
161	TO DO A	18 0.15	1031 GET ON WITH	6	0.05
162	TO SEE IT	18 0.15	1032 GO BACK TO	6	0.05
163	AND I HAD	17 0.14	1033 GOING ON IN	6	0.05
164	AND I JUST	17 0.14	1034 GOING TO DO	6	0.05
165	AND THEN YOU	17 0.14	1035 GOING TO GO	6	0.05
166	AT THE TIME	17 0.14	1036 GOING TO THE	6	0.05
167	COS I WAS	17 0.14	1037 GOOD ERM I	6	0.05
168	FOR A YEAR	17 0.14	1038 HAD A LOT	6	0.05
169	GO AND SEE	17 0.14	1039 HAVE A LOT	6	0.05
170	HAVE TO DO	17 0.14	1040 HE WAS A	6	0.05
171	I LIKE THE	17 0.14	1041 HERE YOU KNOW	6	0.05
172	I WAS JUST	17 0.14	1042 I CAN DO	6	0.05
173	IT WOULD BE	17 0.14	1043 I CAN'T REALLY	6	0.05
174	IT'S IT'S A	17 0.14	1044 I COULDN'T BELIEVE	6	0.05
175	I'VE BEEN TO	17 0.14	1045 I DECIDED TO	6	0.05
176	LOT OF PEOPLE	17 0.14	1046 I DID ENGLISH	6	0.05
177	NO NO I	17 0.14	1047 I DIDN'T HAVE	6	0.05
178	OF IT AND	17 0.14	1048 I DON'T MIND	6	0.05
179	SO I MEAN	17 0.14	1049 I END BUT	6	0.05
180	SUPPOSED TO BE	17 0.14	1050 I GOT A	6	0.05
181	THE PICTURE AND	17 0.14	1051 I HAD I	6	0.05
182	WAS SORT OF	17 0.14	1052 I HAVE A	6	0.05
183	YOU KNOW THEY	17 0.14	1053 I HAVEN'T BEEN	6	0.05
184	AND ERM I	16 0.13	1054 I HAVEN'T REALLY	6	0.05
185	AND IT'S JUST	16 0.13	1055 I I LIKE	6	0.05
186	AND SO I	16 0.13	1056 I I MEAN	6	0.05
187	BUT ER I	16 0.13	1057 I I REALLY	6	0.05
188	ERM BUT I	16 0.13	1058 I JUST DIDN'T	6	0.05
189	ERM I WAS	16 0.13	1059 I LIKE IT	6	0.05
190	I ER THE	16 0.13	1060 I LIKE TO	6	0.05
191	I WAS GOING	16 0.13	1061 I LIVE WITH	6	0.05
192	I WAS TEACHING	16 0.13	1062 I MEAN AND	6	0.05
193	I WOULD HAVE	16 0.13	1063 I MEAN ERM	6	0.05
194	IN THE MIDDLE	16 0.13	1064 I MEAN SORT	6	0.05
195	IN THE SUMMER	16 0.13	1065 I REALLY LIKE	6	0.05
196	IT WAS GOOD	16 0.13	1066 I REALLY LIKED	6	0.05
197	I'VE GOT A	16 0.13	1067 I SUPPOSE IT	6	0.05
198	KNOW I THINK	16 0.13	1068 I THINK ER	6	0.05
199	KNOW IT WAS	16 0.13	1069 I THINK I'LL	6	0.05
200	SO I DON'T	16 0.13	1070 I THINK SHE	6	0.05
201	SO YEAH I	16 0.13	1071 I THOUGHT WAS	6	0.05
202	THE FIRST TIME	16 0.13	1072 I WAS INTERESTED	6	0.05
203	THINK I THINK	16 0.13	1073 I WAS SORT	6	0.05
204	TO GO BACK	16 0.13	1074 I WENT I	6	0.05
205	TO SEE THE	16 0.13	1075 I WENT WITH	6	0.05

206	WHICH IS A	16 0.13	1076 I WHEN I	6 0.05
207	WOULD HAVE BEEN	16 0.13	1077 I WORKED IN	6 0.05
208	YEAH BUT I	16 0.13	1078 I WOULDN'T HAVE	6 0.05
209	YEAH WELL I	16 0.13	1079 IF I DON'T	6 0.05
210	YEAH YEAH BUT	16 0.13	1080 I'M NOT GONNA	6 0.05
211	YOU CAN GO	16 0.13	1081 I'M NOT REALLY	6 0.05
212	A LONG TIME	15 0.13	1082 IN THE CLASSROOM	6 0.05
213	AND SORT OF	15 0.13	1083 IN THE PICTURE	6 0.05
214	AND THEN HE	15 0.13	1084 IN THE WORLD	6 0.05
215	AND YOU HAVE	15 0.13	1085 IN YOU KNOW	6 0.05
216	BUT I I	15 0.13	1086 IS WHAT I	6 0.05
217	BUT YOU KNOW	15 0.13	1087 IS YOU KNOW	6 0.05
218	DO YOU KNOW	15 0.13	1088 IT ALL AND	6 0.05
219	DON'T I DON'T	15 0.13	1089 IT AND SHE	6 0.05
220	HAVE TO BE	15 0.13	1090 IT AND SHE'S	6 0.05
221	I I WAS	15 0.13	1091 IT AS WELL	6 0.05
222	I THINK YEAH	15 0.13	1092 IT DOESN'T REALLY	6 0.05
223	IT SORT OF	15 0.13	1093 IT ERM I	6 0.05
224	KIND OF THING	15 0.13	1094 IT IT IT	6 0.05
225	OF THE TIME	15 0.13	1095 IT REALLY IS	6 0.05
226	OUT OF THE	15 0.13	1096 IT SO I	6 0.05
227	SO I I	15 0.13	1097 IT TO ALL	6 0.05
228	SO I THOUGHT	15 0.13	1098 IT TO HER	6 0.05
229	SO YOU KNOW	15 0.13	1099 IT WAS MY	6 0.05
230	THAT IT WAS	15 0.13	1100 IT WELL I	6 0.05
231	THAT WAS THE	15 0.13	1101 IT'S IN THE	6 0.05
232	THAT YOU KNOW	15 0.13	1102 IT'S IT'S ALL	6 0.05
233	THERE'S A LOT	15 0.13	1103 IT'S JUST THE	6 0.05
234	THEY SORT OF	15 0.13	1104 IT'S LIKE A	6 0.05
235	WAS JUST LIKE	15 0.13	1105 IT'S LIKE YOU	6 0.05
236	WE HAD TO	15 0.13	1106 I'VE ALWAYS WANTED	6 0.05
237	WELL I MEAN	15 0.13	1107 I'VE THOUGHT ABOUT	6 0.05
238	YEAH I WAS	15 0.13	1108 JUST FOR THE	6 0.05
239	YEAH IT IS	15 0.13	1109 JUST KIND OF	6 0.05
240	YEAH YEAH WELL	15 0.13	1110 JUST LIKE OH	6 0.05
241	YOU CAN GET	15 0.13	1111 JUST TO GET	6 0.05
242	A REALLY GOOD	14 0.12	1112 JUST WANTED TO	6 0.05
243	AND A HALF	14 0.12	1113 KNOW IT'S IT'S	6 0.05
244	BEFORE I CAME	14 0.12	1114 KNOW IT'S JUST	6 0.05
245	ER I THINK	14 0.12	1115 KNOW SORT OF	6 0.05
246	ERM AND THEN	14 0.12	1116 LIKE OH NO	6 0.05
247	I CAME HERE	14 0.12	1117 LIKE THAT BUT	6 0.05
248	I DIDN'T REALLY	14 0.12	1118 LIVE ON CAMPUS	6 0.05
249	I MEAN THE	14 0.12	1119 LIVED IN A	6 0.05
250	I WAS I	14 0.12	1120 LOOK LIKE HER	6 0.05
251	IN THE SECOND	14 0.12	1121 LOOKING FORWARD TO	6 0.05
252	IT AND I	14 0.12	1122 MEAN I I	6 0.05
253	IT I THINK	14 0.12	1123 MEAN IT'S A	6 0.05
254	IT WAS SO	14 0.12	1124 MEAN SORT OF	6 0.05
255	IT YEAH YEAH	14 0.12	1125 MM MM YEAH	6 0.05
256	JUST SORT OF	14 0.12	1126 NO ERM I	6 0.05
257	LIKE THAT SO	14 0.12	1127 NO I I	6 0.05

258	NO NO NOT	14 0.12	1128 NO I THINK	6 0.05
259	NOT TOO BAD	14 0.12	1129 NO NO I'M	6 0.05
260	OH YEAH YEAH	14 0.12	1130 NOT A LOT	6 0.05
261	SO YEAH YEAH	14 0.12	1131 NOT VERY HAPPY	6 0.05
262	THE FIRST YEAR	14 0.12	1132 OF A A	6 0.05
263	WAS THAT WAS	14 0.12	1133 OF ALL THE	6 0.05
264	WELL I DON'T	14 0.12	1134 OF IT SO	6 0.05
265	WELL I THINK	14 0.12	1135 OF MY LIFE	6 0.05
266	YEAH AND I	14 0.12	1136 OF ONE OF	6 0.05
267	YEAH I I	14 0.12	1137 OF THE PLAY	6 0.05
268	YEAH OH YEAH	14 0.12	1138 OF THE THINGS	6 0.05
269	YOU KNOW WHAT	14 0.12	1139 OF THING BUT	6 0.05
270	A LONG WAY	13 0.11	1140 OH I DON'T	6 0.05
271	A YEAR OUT	13 0.11	1141 OH IT WAS	6 0.05
272	ALL OVER THE	13 0.11	1142 ON A BUS	6 0.05
273	AND I MEAN	13 0.11	1143 ON T V	6 0.05
274	AND THEN WE	13 0.11	1144 ONE OF THOSE	6 0.05
275	BUT ERM YEAH	13 0.11	1145 ONE OR TWO	6 0.05
276	BUT IT IS	13 0.11	1146 OUT OF IT	6 0.05
277	DON'T KNOW WHAT	13 0.11	1147 PEOPLE WHO CAN	6 0.05
278	DON'T WANT TO	13 0.11	1148 PICTURE OF HER	6 0.05
279	ERM AND I	13 0.11	1149 PLACES LIKE THAT	6 0.05
280	ERM YEAH I	13 0.11	1150 PROUD OF IT	6 0.05
281	ERM YOU KNOW	13 0.11	1151 QUITE A BIT	6 0.05
282	FOR A WHILE	13 0.11	1152 QUITE A GOOD	6 0.05
283	I I THINK	13 0.11	1153 QUITE A LONG	6 0.05
284	I MEAN I'M	13 0.11	1154 REALLY GOOD FUN	6 0.05
285	I MEAN I'VE	13 0.11	1155 REALLY IT WAS	6 0.05
286	I MEAN THEY	13 0.11	1156 REALLY LIKED IT	6 0.05
287	I THOUGHT THAT	13 0.11	1157 REALLY WANTED TO	6 0.05
288	I WAS DOING	13 0.11	1158 REALLY YEAH YEAH	6 0.05
289	I WAS LIVING	13 0.11	1159 SHE LOOKS AT	6 0.05
290	IT YOU KNOW	13 0.11	1160 SHE WAS VERY	6 0.05
291	IT'S NICE TO	13 0.11	1161 SHE'S SHOWING IT	6 0.05
292	IT'S NOT TOO	13 0.11	1162 SHOWING IT TO	6 0.05
293	KNOW I DON'T	13 0.11	1163 SINCE I WAS	6 0.05
294	LIKE THAT I	13 0.11	1164 SO AND I	6 0.05
295	LOOK AT IT	13 0.11	1165 SO BUT I	6 0.05
296	OF THE FILM	13 0.11	1166 SO HE DOES	6 0.05
297	THAT I WAS	13 0.11	1167 SO I DIDN'T	6 0.05
298	THEN I WENT	13 0.11	1168 SO IT'S JUST	6 0.05
299	TO GO INTO	13 0.11	1169 SO IT'S QUITE	6 0.05
300	WANT TO GO	13 0.11	1170 SO WE HAD	6 0.05
301	WAS VERY GOOD	13 0.11	1171 SO YOU CAN	6 0.05
302	WHAT I MEAN	13 0.11	1172 SO YOU DON'T	6 0.05
303	YEAH NO I	13 0.11	1173 SOME KIND OF	6 0.05
304	YEAH NO NO	13 0.11	1174 SOMETHING TO DO	6 0.05
305	YEAH THAT WAS	13 0.11	1175 SORT OF IT	6 0.05
306	YOU WANT TO	13 0.11	1176 SORT OF IT'S	6 0.05
307	ABOUT IT AND	12 0.10	1177 TALK ABOUT THE	6 0.05
308	AND AND I	12 0.10	1178 THAT AND THEN	6 0.05
309	AND EVERYTHING AND	12 0.10	1179 THAT I COULD	6 0.05

310	AND I DON'T	12 0.10	1180 THAT I THINK	6 0.05
311	AND THEN THE	12 0.10	1181 THAT I THOUGHT	6 0.05
312	AND THINGS AND	12 0.10	1182 THAT WAS GOOD	6 0.05
313	AND YOU CAN	12 0.10	1183 THAT WAS IT	6 0.05
314	AS WELL AND	12 0.10	1184 THAT WE WERE	6 0.05
315	BECAUSE I WAS	12 0.10	1185 THAT'S THE THING	6 0.05
316	BUT I WAS	12 0.10	1186 THE FILM IS	6 0.05
317	BUT THAT WAS	12 0.10	1187 THE MAIN THING	6 0.05
318	DIDN'T WANT TO	12 0.10	1188 THE PEOPLE I	6 0.05
319	DOESN'T LIKE IT	12 0.10	1189 THE TIME WHEN	6 0.05
320	DON'T REALLY KNOW	12 0.10	1190 THEN SHE SHOWS	6 0.05
321	HAD TO DO	12 0.10	1191 THERE FOR ABOUT	6 0.05
322	HER FRIENDS AND	12 0.10	1192 THERE IS A	6 0.05
323	I DIDN'T KNOW	12 0.10	1193 THERE YOU KNOW	6 0.05
324	I MEAN THERE'S	12 0.10	1194 THEY SAID OH	6 0.05
325	I THINK SO	12 0.10	1195 THEY WERE ALL	6 0.05
326	I THINK THEY	12 0.10	1196 THEY'VE GOT THE	6 0.05
327	I WENT BACK	12 0.10	1197 THOUGHT ABOUT IT	6 0.05
328	IS A BIT	12 0.10	1198 TO DO IN	6 0.05
329	IT AND YOU	12 0.10	1199 TO SEE AND	6 0.05
330	IT IS QUITE	12 0.10	1200 TO TALK TO	6 0.05
331	IT LOOKS LIKE	12 0.10	1201 TO THE CINEMA	6 0.05
332	IT'S JUST A	12 0.10	1202 TWO YEARS AGO	6 0.05
333	IT'S JUST LIKE	12 0.10	1203 USED TO HAVE	6 0.05
334	IT'S NOT A	12 0.10	1204 VERY HAPPY WITH	6 0.05
335	IT'S ONE OF	12 0.10	1205 WANT TO BE	6 0.05
336	I'VE GOT TO	12 0.10	1206 WANTED TO BE	6 0.05
337	KNOW I MEAN	12 0.10	1207 WAS A NIGHTMARE	6 0.05
338	LIVING IN TOWN	12 0.10	1208 WAS I MEAN	6 0.05
339	NO I MEAN	12 0.10	1209 WAS JUST AMAZING	6 0.05
340	NO I'M NOT	12 0.10	1210 WAS KIND OF	6 0.05
341	NO NOT REALLY	12 0.10	1211 WAS THERE FOR	6 0.05
342	SO ERM I	12 0.10	1212 WAS VERY INTERESTING	6 0.05
343	SO I HAD	12 0.10	1213 WAS WHEN I	6 0.05
344	SO I WENT	12 0.10	1214 WAS YOU KNOW	6 0.05
345	SORT OF A	12 0.10	1215 WE DON'T HAVE	6 0.05
346	THAT WAS THAT	12 0.10	1216 WE GOT A	6 0.05
347	THE MIDDLE OF	12 0.10	1217 WE HAVE TO	6 0.05
348	THE REST OF	12 0.10	1218 WELL I HAD	6 0.05
349	TO GET A	12 0.10	1219 WELL IT'S NOT	6 0.05
350	WAS LIKE A	12 0.10	1220 WELL THAT WAS	6 0.05
351	WHEN I WENT	12 0.10	1221 WELL THERE'S A	6 0.05
352	WHERE I LIVE	12 0.10	1222 WELL YEAH YEAH	6 0.05
353	WHICH IS QUITE	12 0.10	1223 WENT DOWN TO	6 0.05
354	YEAH IT'S IT'S	12 0.10	1224 WENT TO A	6 0.05
355	YEAH MM YEAH	12 0.10	1225 WE'RE GOING TO	6 0.05
356	YOU CAN DO	12 0.10	1226 WHAT IT WAS	6 0.05
357	YOU KNOW BUT	12 0.10	1227 WHEN I WHEN	6 0.05
358	YOU KNOW LIKE	12 0.10	1228 WHICH IS VERY	6 0.05
359	YOU YOU KNOW	12 0.10	1229 WITH IT AND	6 0.05
360	YOU'VE GOT TO	12 0.10	1230 YEAH AND AND	6 0.05
361	A VERY GOOD	11 0.09	1231 YEAH AND ER	6 0.05

362	AN AWFUL LOT	11	0.09	1232	YEAH ERM IT'S	6	0.05
363	AND ER I	11	0.09	1233	YEAH I SUPPOSE	6	0.05
364	AND HE PAINTS	11	0.09	1234	YEAH I WOULD	6	0.05
365	AND I WENT	11	0.09	1235	YEAH I'VE GOT	6	0.05
366	AND SHE WAS	11	0.09	1236	YEAH NO IT	6	0.05
367	AND THEN ER	11	0.09	1237	YEAH NO IT'S	6	0.05
368	AND WE WENT	11	0.09	1238	YEAH THERE'S A	6	0.05
369	AND WE WERE	11	0.09	1239	YEAH WE WENT	6	0.05
370	AND YOU JUST	11	0.09	1240	YEAH YEAH A	6	0.05
371	AS WELL YEAH	11	0.09	1241	YEAH YEAH DEFINITELY	6	0.05
372	AT A LEVEL	11	0.09	1242	YEAH YEAH I'M	6	0.05
373	AT ALL YEAH	11	0.09	1243	YEAH YEAH THAT	6	0.05
374	BECAUSE OF THE	11	0.09	1244	YES YEAH I	6	0.05
375	BUT ERM I	11	0.09	1245	YES YEAH YEAH	6	0.05
376	BUT IT'S IT'S	11	0.09	1246	YES YES I	6	0.05
377	GO TO A	11	0.09	1247	YOU CAN YOU	6	0.05
378	I DO LIKE	11	0.09	1248	YOU CAN'T DO	6	0.05
379	I DON'T WANT	11	0.09	1249	YOU DON'T GET	6	0.05
380	I I I	11	0.09	1250	YOU GET THE	6	0.05
381	I KNOW I	11	0.09	1251	YOU KNOW SORT	6	0.05
382	I MEAN WE	11	0.09	1252	YOU KNOW THERE'S	6	0.05
383	I REALLY ENJOYED	11	0.09	1253	YOU KNOW WHICH	6	0.05
384	I THINK I'D	11	0.09	1254	YOU'VE GOT THE	6	0.05
385	I THINK THAT'S	11	0.09	1255	A BIT AND	5	0.04
386	I THINK WE	11	0.09	1256	A BIT DIFFICULT	5	0.04
387	I WAS ABOUT	11	0.09	1257	A BIT SORT	5	0.04
388	I WAS LIKE	11	0.09	1258	A FILM I	5	0.04
389	I WAS QUITE	11	0.09	1259	A GO AT	5	0.04
390	I WAS VERY	11	0.09	1260	A GOOD IDEA	5	0.04
391	IF I CAN	11	0.09	1261	A GOOD TIME	5	0.04
392	IF YOU KNOW	11	0.09	1262	A GREAT DEAL	5	0.04
393	IN THE SAME	11	0.09	1263	A HUNDRED PERCENT	5	0.04
394	IT BUT I	11	0.09	1264	A LEVEL I	5	0.04
395	IT I MEAN	11	0.09	1265	A LEVELS AND	5	0.04
396	IT IS YEAH	11	0.09	1266	A LOT A	5	0.04
397	IT WAS I	11	0.09	1267	A LOT BETTER	5	0.04
398	IT WAS IN	11	0.09	1268	A NURSING HOME	5	0.04
399	IT YEAH I	11	0.09	1269	A WAY BUT	5	0.04
400	IT'S NOT REALLY	11	0.09	1270	A WHILE AND	5	0.04
401	KNOW I KNOW	11	0.09	1271	ALL MY FRIENDS	5	0.04
402	O'CLOCK IN THE	11	0.09	1272	ALL RIGHT BUT	5	0.04
403	OF MY FRIENDS	11	0.09	1273	ALL THE DIFFERENT	5	0.04
404	OF SORT OF	11	0.09	1274	ALL THE PEOPLE	5	0.04
405	OF THE DAY	11	0.09	1275	AND A FEW	5	0.04
406	PICTURE OF A	11	0.09	1276	AND ALL THIS	5	0.04
407	SORT OF ERM	11	0.09	1277	AND AND IT	5	0.04
408	THAT WAS QUITE	11	0.09	1278	AND AND THAT	5	0.04
409	THAT WAS REALLY	11	0.09	1279	AND ER BUT	5	0.04
410	THAT WAS VERY	11	0.09	1280	AND ER SO	5	0.04
411	THE FILM SOCIETY	11	0.09	1281	AND ERM THEN	5	0.04
412	THE FIRST OFFICER	11	0.09	1282	AND EVERYTHING BUT	5	0.04
413	TO ALL HER	11	0.09	1283	AND HE DIDN'T	5	0.04

414	WAS A VERY	11	0.09	1284	AND I COULDN'T	5	0.04
415	WAS GOING TO	11	0.09	1285	AND I SAID	5	0.04
416	WAS QUITE GOOD	11	0.09	1286	AND IT IS	5	0.04
417	WE HAD A	11	0.09	1287	AND LIKE I	5	0.04
418	WENT TO THE	11	0.09	1288	AND ONE OF	5	0.04
419	WHAT'S GOING ON	11	0.09	1289	AND SEE A	5	0.04
420	WHEN I CAME	11	0.09	1290	AND SHE GOES	5	0.04
421	YEAH THAT'S IT	11	0.09	1291	AND SHE SAID	5	0.04
422	YEAH YEAH MM	11	0.09	1292	AND SHE'S SHOWING	5	0.04
423	YOU DON'T KNOW	11	0.09	1293	AND THAT I	5	0.04
424	A GOOD FILM	10	0.08	1294	AND THAT KIND	5	0.04
425	A PORTRAIT OF	10	0.08	1295	AND THAT SORT	5	0.04
426	AND ALL THE	10	0.08	1296	AND THEN ERM	5	0.04
427	AND ER AND	10	0.08	1297	AND THEN GET	5	0.04
428	AND I DIDN'T	10	0.08	1298	AND THEN IN	5	0.04
429	AND IT'S A	10	0.08	1299	AND THEN THIS	5	0.04
430	AND IT'S IT'S	10	0.08	1300	AND THERE WERE	5	0.04
431	AND SHE DOESN'T	10	0.08	1301	AND THINGS SO	5	0.04
432	AND YOU CAN'T	10	0.08	1302	AND THINGS YEAH	5	0.04
433	AND YOU GET	10	0.08	1303	AND WE GOT	5	0.04
434	AS WELL IT'S	10	0.08	1304	AND WHAT THEY	5	0.04
435	AT ALL AND	10	0.08	1305	AND YOU HAD	5	0.04
436	AT IT AND	10	0.08	1306	AS AN ASSISTANT	5	0.04
437	BUT I'M NOT	10	0.08	1307	AS I SAY	5	0.04
438	BUT IT'S NOT	10	0.08	1308	AS YOU CAN	5	0.04
439	ER YOU KNOW	10	0.08	1309	AT ALL BUT	5	0.04
440	ERM I I	10	0.08	1310	AT NIGHT AND	5	0.04
441	FOR A WEEK	10	0.08	1311	AT ONE POINT	5	0.04
442	HAD TO GO	10	0.08	1312	AT THE BACK	5	0.04
443	HAPPY WITH IT	10	0.08	1313	BACK TO WORK	5	0.04
444	I HAVEN'T GOT	10	0.08	1314	BE A BIT	5	0.04
445	I MEAN THAT	10	0.08	1315	BE A LOT	5	0.04
446	I THINK IF	10	0.08	1316	BE YOU KNOW	5	0.04
447	I THOUGHT I	10	0.08	1317	BECAUSE ERM I	5	0.04
448	I THOUGHT WELL	10	0.08	1318	BEFORE SO I	5	0.04
449	I WAS A	10	0.08	1319	BEING ON CAMPUS	5	0.04
450	I WAS GONNA	10	0.08	1320	BIT SORT OF	5	0.04
451	I WAS REALLY	10	0.08	1321	BUT ER BUT	5	0.04
452	IF I WAS	10	0.08	1322	BUT ER IT'S	5	0.04
453	IT DOESN'T LOOK	10	0.08	1323	BUT I DO	5	0.04
454	IT WAS ERM	10	0.08	1324	BUT I LIKE	5	0.04
455	IT WOULD HAVE	10	0.08	1325	BUT IF YOU	5	0.04
456	IT'S A GOOD	10	0.08	1326	BUT IT DOESN'T	5	0.04
457	IT'S IT'S NOT	10	0.08	1327	BUT IT IT	5	0.04
458	KIND OF A	10	0.08	1328	BUT THEN YOU	5	0.04
459	LIKE TO DO	10	0.08	1329	CAMPUS LAST YEAR	5	0.04
460	LIKE TO SEE	10	0.08	1330	CAN DO IT	5	0.04
461	LIKE TO TEACH	10	0.08	1331	CAN'T REMEMBER WHAT	5	0.04
462	LIKE YOU KNOW	10	0.08	1332	CAN'T REMEMBER WHERE	5	0.04
463	LOT OF THE	10	0.08	1333	COME BACK AND	5	0.04
464	MY FIRST YEAR	10	0.08	1334	COS I THINK	5	0.04
465	OF THING AND	10	0.08	1335	DIDN'T HAVE TO	5	0.04

466	OF YOU KNOW	10 0.08	1336 DIDN'T LIKE THE	5 0.04
467	OR SOMETHING AND	10 0.08	1337 DIDN'T REALLY KNOW	5 0.04
468	REALLY ENJOYED IT	10 0.08	1338 DIFFICULT TO GET	5 0.04
469	SO IT'S IT'S	10 0.08	1339 DO A LOT	5 0.04
470	SOME OF THEM	10 0.08	1340 DO IT FOR	5 0.04
471	THAT THAT WAS	10 0.08	1341 DOESN'T LOOK LIKE	5 0.04
472	THE SECOND YEAR	10 0.08	1342 DON'T HAVE ANY	5 0.04
473	THE TIME AND	10 0.08	1343 DON'T KNOW NO	5 0.04
474	THE TIME YEAH	10 0.08	1344 DON'T LIKE THAT	5 0.04
475	THE WAY THAT	10 0.08	1345 DOWN THE ROAD	5 0.04
476	THERE AND I	10 0.08	1346 ENJOYED IT I	5 0.04
477	THOUGHT THAT WAS	10 0.08	1347 ER I I	5 0.04
478	TO COME BACK	10 0.08	1348 ER I WAS	5 0.04
479	TO DO SOMETHING	10 0.08	1349 ER YEAH I	5 0.04
480	TO DO THAT	10 0.08	1350 ERM AND ALSO	5 0.04
481	TO DO WITH	10 0.08	1351 ERM AND WE	5 0.04
482	TO GO OUT	10 0.08	1352 ERM BUT IT	5 0.04
483	TO HER FRIENDS	10 0.08	1353 ERM I DIDN'T	5 0.04
484	USED TO BE	10 0.08	1354 ERM I HAD	5 0.04
485	WAS I WAS	10 0.08	1355 ERM I HAVE	5 0.04
486	WAS IN THE	10 0.08	1356 ERM I'D LIKE	5 0.04
487	WELL IT WAS	10 0.08	1357 ERM IT IT	5 0.04
488	WENT BACK TO	10 0.08	1358 ERM IT'S A	5 0.04
489	WE'VE GOT A	10 0.08	1359 ERM NO NO	5 0.04
490	WHAT I WANT	10 0.08	1360 ERM WE WENT	5 0.04
491	WHEN YOU GO	10 0.08	1361 ERM WE WERE	5 0.04
492	YEAH IT IT	10 0.08	1362 ERM WELL THE	5 0.04
493	YEAH IT'S A	10 0.08	1363 ERM WHAT ELSE	5 0.04
494	YEAH YEAH YOU	10 0.08	1364 ERM YEAH IT'S	5 0.04
495	YOU CAN JUST	10 0.08	1365 ESPECIALLY IN THE	5 0.04
496	YOU KNOW SO	10 0.08	1366 EXACTLY THE SAME	5 0.04
497	A FOREIGN LANGUAGE	9 0.08	1367 FIRST TIME I	5 0.04
498	A LOOK AT	9 0.08	1368 FOR A A	5 0.04
499	AND ALL THAT	9 0.08	1369 FOR ABOUT A	5 0.04
500	AND ER YEAH	9 0.08	1370 FOR THE LAST	5 0.04
501	AND IN THE	9 0.08	1371 FRENCH AND GERMAN	5 0.04
502	AND IT JUST	9 0.08	1372 GET THE BUS	5 0.04
503	AND IT'S NOT	9 0.08	1373 GET TO SEE	5 0.04
504	AND THEN AND	9 0.08	1374 GET TO THE	5 0.04
505	AND THEN IT	9 0.08	1375 GO BACK AND	5 0.04
506	AND WHEN I	9 0.08	1376 GO FOR A	5 0.04
507	AND YEAH YEAH	9 0.08	1377 GOOD THING ABOUT	5 0.04
508	AND YOU DON'T	9 0.08	1378 GOOD YEAH AND	5 0.04
509	AND YOU'VE GOT	9 0.08	1379 GOOD YEAH YEAH	5 0.04
510	AS A FOREIGN	9 0.08	1380 HAD TO GET	5 0.04
511	AS WELL ERM	9 0.08	1381 HAVE SOMETHING TO	5 0.04
512	BECAUSE IT WAS	9 0.08	1382 HAVE TO WORK	5 0.04
513	BUT I CAN'T	9 0.08	1383 HAVE TO WRITE	5 0.04
514	BUT YEAH I	9 0.08	1384 HAVEN'T GOT A	5 0.04
515	COS IT WAS	9 0.08	1385 HE DOES IT	5 0.04
516	DOESN'T LOOK VERY	9 0.08	1386 HE SAYS OH	5 0.04
517	DON'T KNOW ERM	9 0.08	1387 HE SORT OF	5 0.04

518	DON'T KNOW IT	9 0.08	1388 HER LOOK MORE	5 0.04
519	EACH OTHER AND	9 0.08	1389 HER SO HE	5 0.04
520	ENGLISH AS A	9 0.08	1390 HERE AND THEN	5 0.04
521	ER I DON'T	9 0.08	1391 I APPLIED TO	5 0.04
522	ERM AND SO	9 0.08	1392 I CAME BACK	5 0.04
523	ERM AND THEY	9 0.08	1393 I CHOSE TO	5 0.04
524	I COULD DO	9 0.08	1394 I COME FROM	5 0.04
525	I DID A	9 0.08	1395 I DIDN'T GET	5 0.04
526	I DON'T HAVE	9 0.08	1396 I DON'T GET	5 0.04
527	I FIND IT	9 0.08	1397 I DON'T UNDERSTAND	5 0.04
528	I KNOW THAT	9 0.08	1398 I FIND THAT	5 0.04
529	I LIVE IN	9 0.08	1399 I HAD NO	5 0.04
530	I LIVED IN	9 0.08	1400 I HAVEN'T SEEN	5 0.04
531	I MEAN IN	9 0.08	1401 I I THOUGHT	5 0.04
532	I SAW IT	9 0.08	1402 I JUST WANTED	5 0.04
533	I SUPPOSE IT'S	9 0.08	1403 I KIND OF	5 0.04
534	I THINK YOU	9 0.08	1404 I KNOW IT	5 0.04
535	I THOUGHT I'D	9 0.08	1405 I LOVED IT	5 0.04
536	I THOUGHT OH	9 0.08	1406 I MEAN JUST	5 0.04
537	I WAS THINKING	9 0.08	1407 I MEAN THERE	5 0.04
538	I YOU KNOW	9 0.08	1408 I MEAN YEAH	5 0.04
539	IF YOU WANT	9 0.08	1409 I MOVED AWAY	5 0.04
540	I'LL HAVE TO	9 0.08	1410 I NEED TO	5 0.04
541	I'M A BIT	9 0.08	1411 I ONLY ONE	5 0.04
542	I'M GOING TO	9 0.08	1412 I ONLY PLACE	5 0.04
543	IN A HOUSE	9 0.08	1413 I ONLY THING	5 0.04
544	IN IT AND	9 0.08	1414 I REALLY WANTED	5 0.04
545	IN MY FIRST	9 0.08	1415 I SHOULD HAVE	5 0.04
546	IT A LOT	9 0.08	1416 I SUPPOSE YEAH	5 0.04
547	IT AT ALL	9 0.08	1417 I THINK FOR	5 0.04
548	IT DEPENDS ON	9 0.08	1418 I THINK IS	5 0.04
549	IT IS A	9 0.08	1419 I WAS ONLY	5 0.04
550	IT IS IT'S	9 0.08	1420 I WAS SAYING	5 0.04
551	IT IT'S IT'S	9 0.08	1421 I WAS STAYING	5 0.04
552	IT WAS ABOUT	9 0.08	1422 I WENT THERE	5 0.04
553	IT WAS ACTUALLY	9 0.08	1423 I'D LOVE TO	5 0.04
554	IT WAS ALL	9 0.08	1424 IF I COULD	5 0.04
555	IT WAS AND	9 0.08	1425 IF I WANTED	5 0.04
556	I'VE NOT REALLY	9 0.08	1426 IF YOU WERE	5 0.04
557	JUST LIKE TO	9 0.08	1427 I'M GONNA BE	5 0.04
558	KNOW YEAH YEAH	9 0.08	1428 I'M GONNA GO	5 0.04
559	LIKE THAT ERM	9 0.08	1429 I'M INTERESTED IN	5 0.04
560	LOOK A BIT	9 0.08	1430 I'M NOT A	5 0.04
561	MAKES HER LOOK	9 0.08	1431 I'M QUITE INTERESTED	5 0.04
562	MANAGED TO GET	9 0.08	1432 IN A SCHOOL	5 0.04
563	MY A LEVELS	9 0.08	1433 IN IN LANCASTER	5 0.04
564	NO IT'S NOT	9 0.08	1434 IN MY SECOND	5 0.04
565	NO NO IT'S	9 0.08	1435 IN THE BOOK	5 0.04
566	NOT REALLY NO	9 0.08	1436 IN THE CENTRE	5 0.04
567	OF THE PEOPLE	9 0.08	1437 IN THE HOUSE	5 0.04
568	ON THE WAY	9 0.08	1438 IN THE LAST	5 0.04
569	QUITE INTERESTED IN	9 0.08	1439 IN THE SOUTH	5 0.04

570	REALLY NICE AND	9 0.08	1440 IN THE WAY	5 0.04
571	REALLY WANT TO	9 0.08	1441 IN TOWN SO	5 0.04
572	SO I JUST	9 0.08	1442 IS ONE OF	5 0.04
573	SORT OF YOU	9 0.08	1443 IT AND HE	5 0.04
574	THAT YEAH YEAH	9 0.08	1444 IT AT THE	5 0.04
575	THAT'S WHAT I	9 0.08	1445 IT IS REALLY	5 0.04
576	THERE YEAH YEAH	9 0.08	1446 IT IT IT'S	5 0.04
577	THINK IT'S A	9 0.08	1447 IT IT'S A	5 0.04
578	TO GET TO	9 0.08	1448 IT MUST BE	5 0.04
579	TO YOU KNOW	9 0.08	1449 IT NO NO	5 0.04
580	TWO OR THREE	9 0.08	1450 IT OFF TO	5 0.04
581	WANTED TO GO	9 0.08	1451 IT WAS ER	5 0.04
582	WAS A A	9 0.08	1452 IT WAS INTERESTING	5 0.04
583	WAS QUITE A	9 0.08	1453 IT WAS ONLY	5 0.04
584	WAS THERE WAS	9 0.08	1454 IT WAS SORT	5 0.04
585	WE WERE THERE	9 0.08	1455 IT WAS THAT	5 0.04
586	WELL I I	9 0.08	1456 IT WAS WELL	5 0.04
587	WELL I WAS	9 0.08	1457 IT WAS WITH	5 0.04
588	WHILE I WAS	9 0.08	1458 IT WAS YOU	5 0.04
589	WORKED IN A	9 0.08	1459 IT WHEN I	5 0.04
590	YEAH AND THEN	9 0.08	1460 IT'S I DON'T	5 0.04
591	YEAH ER I	9 0.08	1461 IT'S IT'S REALLY	5 0.04
592	YEAH ERM WELL	9 0.08	1462 IT'S JUST TO	5 0.04
593	YEAH ERM YEAH	9 0.08	1463 IT'S LIKE THE	5 0.04
594	YEAH SO I	9 0.08	1464 IT'S NOT IT'S	5 0.04
595	YEAH YEAH NO	9 0.08	1465 IT'S NOT SO	5 0.04
596	YEAH YEAH OH	9 0.08	1466 IT'S ONLY A	5 0.04
597	YEAH YEAH SO	9 0.08	1467 IT'S QUITE NICE	5 0.04
598	YOU DON'T HAVE	9 0.08	1468 IT'S SOMETHING THAT	5 0.04
599	YOU HAD TO	9 0.08	1469 I'VE GOT THE	5 0.04
600	YOU KNOW IF	9 0.08	1470 I'VE GOT TWO	5 0.04
601	YOU KNOW I'M	9 0.08	1471 I'VE I'VE NOT	5 0.04
602	YOU KNOW TO	9 0.08	1472 I'VE ONLY BEEN	5 0.04
603	YOU NEED TO	9 0.08	1473 I'VE SORT OF	5 0.04
604	YOU SORT OF	9 0.08	1474 KNOW I I	5 0.04
605	A KIND OF	8 0.07	1475 KNOW I MIGHT	5 0.04
606	A LOT TO	8 0.07	1476 KNOW IF YOU	5 0.04
607	ALWAYS WANTED TO	8 0.07	1477 KNOW THAT I	5 0.04
608	AND A LOT	8 0.07	1478 KNOW THEY WERE	5 0.04
609	AND ER THE	8 0.07	1479 KNOW WHAT IT'S	5 0.04
610	AND ERM AND	8 0.07	1480 KNOW WHAT TO	5 0.04
611	AND HAVE A	8 0.07	1481 LIKE A A	5 0.04
612	AND HE SAYS	8 0.07	1482 LIKE GOING TO	5 0.04
613	AND IT'S LIKE	8 0.07	1483 LIKE IT ERM	5 0.04
614	AND OF COURSE	8 0.07	1484 LIKE THAT JUST	5 0.04
615	AND SO HE	8 0.07	1485 LIKE THE WAY	5 0.04
616	AND SO YOU	8 0.07	1486 LIVED IN TOWN	5 0.04
617	AND THEN THEY	8 0.07	1487 LIVING AT HOME	5 0.04
618	AND THEY HAD	8 0.07	1488 LIVING IN A	5 0.04
619	AND THEY HAVE	8 0.07	1489 LONG AS YOU	5 0.04
620	AS WELL AS	8 0.07	1490 LOOK AT THE	5 0.04
621	AT HOME AND	8 0.07	1491 LOOKED AT THE	5 0.04

622	AWFUL LOT OF	8 0.07	1492 LOT OF WORK	5 0.04
623	BUT AT THE	8 0.07	1493 MEAN IT'S NOT	5 0.04
624	BUT IN THE	8 0.07	1494 MEAN THAT WAS	5 0.04
625	DO THAT AND	8 0.07	1495 MEAN YOU YOU	5 0.04
626	DON'T KNOW YEAH	8 0.07	1496 MEMBER OF THE	5 0.04
627	DON'T NEED TO	8 0.07	1497 MIDDLE OF THE	5 0.04
628	DURING THE DAY	8 0.07	1498 MUST HAVE BEEN	5 0.04
629	END OF IT	8 0.07	1499 NEVER BEEN TO	5 0.04
630	ENGLISH LANGUAGE AND	8 0.07	1500 NICE YEAH YEAH	5 0.04
631	ER AND THEN	8 0.07	1501 NO I WAS	5 0.04
632	ERM AND THE	8 0.07	1502 NO IT'S IT'S	5 0.04
633	ERM I WENT	8 0.07	1503 NO NO YEAH	5 0.04
634	ERM SO I	8 0.07	1504 NO NOT AT	5 0.04
635	ERM THERE WAS	8 0.07	1505 NO WELL I	5 0.04
636	GO TO UNIVERSITY	8 0.07	1506 NO YEAH YEAH	5 0.04
637	GONNA HAVE TO	8 0.07	1507 NOT I MEAN	5 0.04
638	HAVE A LOOK	8 0.07	1508 NOT REALLY A	5 0.04
639	HAVE TO YOU	8 0.07	1509 NOT YOU KNOW	5 0.04
640	HER AND SHE	8 0.07	1510 OF A BIT	5 0.04
641	I CAME TO	8 0.07	1511 OF A LADY	5 0.04
642	I DID I	8 0.07	1512 OF A WOMAN	5 0.04
643	I DIDN'T WANT	8 0.07	1513 OF HER FRIENDS	5 0.04
644	I DON'T WANNA	8 0.07	1514 OF IT ERM	5 0.04
645	I I DO	8 0.07	1515 OF IT IS	5 0.04
646	I JUST I	8 0.07	1516 OF IT WAS	5 0.04
647	I MEAN IF	8 0.07	1517 OF LIKE I	5 0.04
648	I MEAN LIKE	8 0.07	1518 OF THE BEST	5 0.04
649	I OTHER SIDE	8 0.07	1519 OF THE YEAR	5 0.04
650	I REALLY ENJOY	8 0.07	1520 OF THEM WERE	5 0.04
651	I THINK I'M	8 0.07	1521 OF THING YEAH	5 0.04
652	I WAS ON	8 0.07	1522 OH MY GOD	5 0.04
653	I WAS SO	8 0.07	1523 OH YOU KNOW	5 0.04
654	I WAS THE	8 0.07	1524 ON CAMPUS LAST	5 0.04
655	I WAS WORKING	8 0.07	1525 ON HER FACE	5 0.04
656	IF YOU GO	8 0.07	1526 OR SOMETHING BUT	5 0.04
657	IN FRONT OF	8 0.07	1527 OR WHATEVER AND	5 0.04
658	IN IN THE	8 0.07	1528 OR WHATEVER SO	5 0.04
659	IN TERMS OF	8 0.07	1529 OUT AT NIGHT	5 0.04
660	INTERESTED IN THE	8 0.07	1530 OUT THERE FOR	5 0.04
661	IS IT IS	8 0.07	1531 PEOPLE IN THE	5 0.04
662	IT AND IT	8 0.07	1532 PICTURE AND SHE	5 0.04
663	IT IS IT	8 0.07	1533 PREFER TO GO	5 0.04
664	IT IT IS	8 0.07	1534 REALLY GOOD AND	5 0.04
665	IT IT'S NOT	8 0.07	1535 REALLY GOOD ERM	5 0.04
666	IT SO HE	8 0.07	1536 REALLY GOOD YEAH	5 0.04
667	IT WAS GREAT	8 0.07	1537 RIGHT YEAH I	5 0.04
668	IT'S A LONG	8 0.07	1538 SAID YOU KNOW	5 0.04
669	IT'S A LOT	8 0.07	1539 SAT IN A	5 0.04
670	IT'S A VERY	8 0.07	1540 SAYING YOU KNOW	5 0.04
671	IT'S ALL RIGHT	8 0.07	1541 SECOND YEAR AND	5 0.04
672	IT'S DIFFICULT TO	8 0.07	1542 SEE A LOT	5 0.04
673	IT'S KIND OF	8 0.07	1543 SEE IF I	5 0.04

674	IT'S NOT AS	8 0.07	1544 SHE HAD A	5 0.04
675	IT'S NOT LIKE	8 0.07	1545 SHE SHOWS IT	5 0.04
676	IT'S REALLY GOOD	8 0.07	1546 SHE SITS DOWN	5 0.04
677	I'VE NEVER BEEN	8 0.07	1547 SIDE OF THE	5 0.04
678	KNOW BUT I	8 0.07	1548 SO BUT ERM	5 0.04
679	KNOW YOU KNOW	8 0.07	1549 SO I DECIDED	5 0.04
680	LAST YEAR AND	8 0.07	1550 SO I'D LIKE	5 0.04
681	LIKE IT WAS	8 0.07	1551 SO IT IS	5 0.04
682	MM I DON'T	8 0.07	1552 SO IT'S LIKE	5 0.04
683	MY SECOND YEAR	8 0.07	1553 SO MM YEAH	5 0.04
684	NOT AT ALL	8 0.07	1554 SO SHE SITS	5 0.04
685	OF THE THE	8 0.07	1555 SO SO IT	5 0.04
686	OF THING SO	8 0.07	1556 SO WE WERE	5 0.04
687	OH NO NO	8 0.07	1557 SO YEAH IT	5 0.04
688	OVER THE PLACE	8 0.07	1558 SO YOU HAVE	5 0.04
689	PAINTING A PICTURE	8 0.07	1559 SOME OF MY	5 0.04
690	QUITE A FEW	8 0.07	1560 SORT OF ER	5 0.04
691	REALLY SORT OF	8 0.07	1561 SORT OF I	5 0.04
692	REALLY YOU KNOW	8 0.07	1562 SORT OF MORE	5 0.04
693	SO ER YEAH	8 0.07	1563 T V AND	5 0.04
694	SO IT'S NOT	8 0.07	1564 TALK TO YOU	5 0.04
695	SO NO I	8 0.07	1565 THAT AND I	5 0.04
696	SOME SORT OF	8 0.07	1566 THAT AND YOU	5 0.04
697	SORT OF QUITE	8 0.07	1567 THAT I I	5 0.04
698	THAT I DON'T	8 0.07	1568 THAT WAS JUST	5 0.04
699	THAT I MEAN	8 0.07	1569 THAT WAS NICE	5 0.04
700	THAT WOULD BE	8 0.07	1570 THAT WAS WHEN	5 0.04
701	THAT'S IT YEAH	8 0.07	1571 THAT'S ALL RIGHT	5 0.04
702	THE FIRST PICTURE	8 0.07	1572 THAT'S WHY I	5 0.04
703	THE SAME TIME	8 0.07	1573 THE CAPTAIN SAID	5 0.04
704	THE STORY OF	8 0.07	1574 THE DAY AND	5 0.04
705	THE TIME I	8 0.07	1575 THE GOOD THING	5 0.04
706	THE WAY I	8 0.07	1576 THE LAST ONE	5 0.04
707	THERE AND THEN	8 0.07	1577 THE LIBRARY AND	5 0.04
708	THING ABOUT IT	8 0.07	1578 THE MORNING AND	5 0.04
709	THINGS AS WELL	8 0.07	1579 THE PROBLEM WAS	5 0.04
710	THIS IS A	8 0.07	1580 THE SAME THING	5 0.04
711	TO BE ABLE	8 0.07	1581 THE SUMMER AND	5 0.04
712	TO DO AND	8 0.07	1582 THE THING IS	5 0.04
713	TO DO ERM	8 0.07	1583 THE TIME YOU	5 0.04
714	TO DO THE	8 0.07	1584 THE USE OF	5 0.04
715	TO DO YEAH	8 0.07	1585 THE WAY THEY	5 0.04
716	TO GET OUT	8 0.07	1586 THEN I WAS	5 0.04
717	TO LOOK AT	8 0.07	1587 THEN IT WAS	5 0.04
718	TO TALK ABOUT	8 0.07	1588 THERE AND ERM	5 0.04
719	WAS A GOOD	8 0.07	1589 THERE ARE SOME	5 0.04
720	WAS GOING ON	8 0.07	1590 THERE'S LOADS OF	5 0.04
721	WAS JUST A	8 0.07	1591 THERE'S LOTS OF	5 0.04
722	WAS LIKE OH	8 0.07	1592 THERE'S QUITE A	5 0.04
723	WE STAYED IN	8 0.07	1593 THERE'S SO MUCH	5 0.04
724	WE USED TO	8 0.07	1594 THEY DON'T HAVE	5 0.04
725	WERE SORT OF	8 0.07	1595 THEY SEEM TO	5 0.04

726	WHAT I WAS	8 0.07	1596 THEY USED TO	5 0.04
727	WHEN I GOT	8 0.07	1597 THEY WERE VERY	5 0.04
728	WHICH WAS A	8 0.07	1598 THINGS AND YOU	5 0.04
729	WITH IT SO	8 0.07	1599 THINGS YOU KNOW	5 0.04
730	WITH MY PARENTS	8 0.07	1600 THINK ABOUT IT	5 0.04
731	YEAH AND ERM	8 0.07	1601 THINK IT'S JUST	5 0.04
732	YEAH I DID	8 0.07	1602 THINK THAT WAS	5 0.04
733	YEAH I WENT	8 0.07	1603 THINK YOU KNOW	5 0.04
734	YEAH YOU CAN	8 0.07	1604 THIRD YEAR I	5 0.04
735	YOU DON'T REALLY	8 0.07	1605 THIS IS LIKE	5 0.04
736	YOU GET A	8 0.07	1606 THIS IS ME	5 0.04
737	YOU GET TO	8 0.07	1607 THIS SORT OF	5 0.04
738	YOU GO TO	8 0.07	1608 THIS YEAR I	5 0.04
739	YOU KNOW THAT	8 0.07	1609 TIME I WENT	5 0.04
740	YOU YOU CAN	8 0.07	1610 TO A CERTAIN	5 0.04
741	A LOT AND	7 0.06	1611 TO BE IN	5 0.04
742	A LOT DIFFERENT	7 0.06	1612 TO DO BUT	5 0.04
743	A WEEK AND	7 0.06	1613 TO DO FOR	5 0.04
744	ABOUT A MONTH	7 0.06	1614 TO DO SOME	5 0.04
745	ABOUT A YEAR	7 0.06	1615 TO ENGLAND AND	5 0.04
746	ABOUT IT ERM	7 0.06	1616 TO GET THE	5 0.04
747	ABOUT IT YEAH	7 0.06	1617 TO GO ON	5 0.04
748	ALL THE WAY	7 0.06	1618 TO GO SOMEWHERE	5 0.04
749	AN M A	7 0.06	1619 TO HAVE THE	5 0.04
750	AND AND AND	7 0.06	1620 TO I MEAN	5 0.04
751	AND ER IT'S	7 0.06	1621 TO IT AND	5 0.04
752	AND ER WE	7 0.06	1622 TO LIVE ON	5 0.04
753	AND EVERYTHING SO	7 0.06	1623 TO SEE THAT	5 0.04
754	AND I DID	7 0.06	1624 TO TEACH ENGLISH	5 0.04
755	AND I GOT	7 0.06	1625 TO TRY AND	5 0.04
756	AND IF YOU	7 0.06	1626 TO UNIVERSITY I	5 0.04
757	AND IT IT	7 0.06	1627 TO WORK FOR	5 0.04
758	AND IT WOULD	7 0.06	1628 TOOK A YEAR	5 0.04
759	AND LIKE THE	7 0.06	1629 TRY AND GET	5 0.04
760	AND SOME OF	7 0.06	1630 TWO AND A	5 0.04
761	AND STUFF AND	7 0.06	1631 UP ALL THE	5 0.04
762	AND THE PEOPLE	7 0.06	1632 UP AT THE	5 0.04
763	AND THERE'S A	7 0.06	1633 UP IN THE	5 0.04
764	AND THEY SAID	7 0.06	1634 VERY PLEASED WITH	5 0.04
765	AND THEY THEY	7 0.06	1635 VERY PROUD OF	5 0.04
766	AND YOU COULD	7 0.06	1636 VERY WELL AND	5 0.04
767	AND YOU YOU	7 0.06	1637 WANT TO TEACH	5 0.04
768	ANYTHING LIKE THAT	7 0.06	1638 WAS A BIG	5 0.04
769	AS LONG AS	7 0.06	1639 WAS A REAL	5 0.04
770	AS WELL BUT	7 0.06	1640 WAS DOING A	5 0.04
771	AT ALL I	7 0.06	1641 WAS INTERESTED IN	5 0.04
772	AT ALL NO	7 0.06	1642 WAS JUST REALLY	5 0.04
773	AT FIRST I	7 0.06	1643 WAS LIVING IN	5 0.04
774	AT SCHOOL AND	7 0.06	1644 WAS ON CAMPUS	5 0.04
775	AT THE BEGINNING	7 0.06	1645 WAS TALKING TO	5 0.04
776	AT THE SAME	7 0.06	1646 WE MANAGED TO	5 0.04
777	AWAY FROM HOME	7 0.06	1647 WE WENT OUT	5 0.04

778	BACK AND I	7 0.06	1648 WE WENT UP	5 0.04
779	BACK TO ENGLAND	7 0.06	1649 WE WERE ALL	5 0.04
780	BE NICE TO	7 0.06	1650 WE WERE JUST	5 0.04
781	BECAUSE I MEAN	7 0.06	1651 WELL I THOUGHT	5 0.04
782	BECAUSE I THINK	7 0.06	1652 WELL SO I	5 0.04
783	BUT ER IT	7 0.06	1653 WELL YEAH I	5 0.04
784	BUT I DIDN'T	7 0.06	1654 WELL YOU KNOW	5 0.04
785	BUT IT IT'S	7 0.06	1655 WHAT ELSE DID	5 0.04
786	C S E	7 0.06	1656 WHAT HAPPENED WAS	5 0.04
787	CAN'T I CAN'T	7 0.06	1657 WHAT SORT OF	5 0.04
788	COME BACK TO	7 0.06	1658 WHAT YOU WANT	5 0.04
789	COS THEY WERE	7 0.06	1659 WHAT YOU'RE DOING	5 0.04
790	COUPLE OF WEEKS	7 0.06	1660 WHEN I FIRST	5 0.04
791	COUPLE OF YEARS	7 0.06	1661 WHEN IT'S FINISHED	5 0.04
792	DIDN'T KNOW WHAT	7 0.06	1662 WHERE YOU CAN	5 0.04
793	DON'T HAVE TO	7 0.06	1663 WHICH IS LIKE	5 0.04
794	DON'T KNOW IT'S	7 0.06	1664 WHICH WAS QUITE	5 0.04
795	DON'T THINK I	7 0.06	1665 WITH ALL THE	5 0.04
796	DON'T THINK SO	7 0.06	1666 WITH HIM AND	5 0.04
797	ELSE DID WE	7 0.06	1667 WITH ONE OF	5 0.04
798	ER IN THE	7 0.06	1668 WITH SORT OF	5 0.04
799	ERM I SUPPOSE	7 0.06	1669 WITH WHAT HE'S	5 0.04
800	ERM NO I	7 0.06	1670 WORK IN THE	5 0.04
801	ERM ONE OF	7 0.06	1671 WORKING IN THE	5 0.04
802	ERM SORT OF	7 0.06	1672 WOULD BE A	5 0.04
803	FOR A COUPLE	7 0.06	1673 WOULD HAVE THOUGHT	5 0.04
804	FOR A FEW	7 0.06	1674 YEAH BUT ER	5 0.04
805	FOR TWO YEARS	7 0.06	1675 YEAH BUT IT'S	5 0.04
806	G C S	7 0.06	1676 YEAH BUT YOU	5 0.04
807	GET A BIT	7 0.06	1677 YEAH ERM NO	5 0.04
808	GET AWAY FROM	7 0.06	1678 YEAH I LIKE	5 0.04
809	GOOD FILM AND	7 0.06	1679 YEAH I'M GOING	5 0.04
810	GOT TO GET	7 0.06	1680 YEAH IT'S ER	5 0.04
811	HALF AN HOUR	7 0.06	1681 YEAH IT'S ERM	5 0.04
812	HAVE YOU SEEN	7 0.06	1682 YEAH IT'S NICE	5 0.04
813	I CAN SEE	7 0.06	1683 YEAH IT'S NOT	5 0.04
814	I CAN'T I	7 0.06	1684 YEAH MHM YEAH	5 0.04
815	I COULD I	7 0.06	1685 YEAH NO NOT	5 0.04
816	I DID THAT	7 0.06	1686 YEAH RIGHT YEAH	5 0.04
817	I DIDN'T LIKE	7 0.06	1687 YEAH SORT OF	5 0.04
818	I END I	7 0.06	1688 YEAH THAT'S TRUE	5 0.04
819	I ENJOY IT	7 0.06	1689 YEAH WELL YEAH	5 0.04
820	I ENJOYED IT	7 0.06	1690 YEAH WHICH IS	5 0.04
821	I ERM THE	7 0.06	1691 YEAH YEAH WE	5 0.04
822	I FOUND THAT	7 0.06	1692 YEAR AND THEN	5 0.04
823	I GO TO	7 0.06	1693 YEARS AGO AND	5 0.04
824	I I DIDN'T	7 0.06	1694 YES YEAH IT	5 0.04
825	I I HAD	7 0.06	1695 YOU CAN HAVE	5 0.04
826	I I JUST	7 0.06	1696 YOU CAN SEE	5 0.04
827	I JUST DON'T	7 0.06	1697 YOU CAN TAKE	5 0.04
828	I JUST THOUGHT	7 0.06	1698 YOU CAN'T GET	5 0.04
829	I KNOW BUT	7 0.06	1699 YOU DO IT	5 0.04

830	I KNOW YEAH	7 0.06	1700 YOU DON'T NEED	5 0.04
831	I MEAN MY	7 0.06	1701 YOU GO IN	5 0.04
832	I MEAN OBVIOUSLY	7 0.06	1702 YOU HAVE A	5 0.04
833	I MEAN THAT'S	7 0.06	1703 YOU KNOW BECAUSE	5 0.04
834	I SUPPOSE I	7 0.06	1704 YOU KNOW HE	5 0.04
835	I THINK I'VE	7 0.06	1705 YOU KNOW HOW	5 0.04
836	I WAS AT	7 0.06	1706 YOU KNOW I'VE	5 0.04
837	I WOULD LIKE	7 0.06	1707 YOU KNOW JUST	5 0.04
838	I WOULDN'T MIND	7 0.06	1708 YOU KNOW NO	5 0.04
839	IF IT WAS	7 0.06	1709 YOU KNOW SHE	5 0.04
840	I'M HOPING TO	7 0.06	1710 YOU KNOW THERE	5 0.04
841	IN THE FILM	7 0.06	1711 YOU KNOW THIS	5 0.04
842	IN THE FUTURE	7 0.06	1712 YOU KNOW WHEN	5 0.04
843	IN THIS COUNTRY	7 0.06	1713 YOU WANNA GO	5 0.04
844	IS A LOT	7 0.06	1714 YOU YOU JUST	5 0.04
845	IS A VERY	7 0.06	1715 YOU'RE GOING TO	5 0.04
846	IS SORT OF	7 0.06	1716 YOU'VE GOT A	5 0.04
847	IT AND ER	7 0.06	1717 AT THE I END	44 0.37
848	IT AND THEN	7 0.06	1718 THE I END OF	35 0.30
849	IT AS A	7 0.06	1719 IN THE I END	16 0.13
850	IT COULD BE	7 0.06	1720 THE I ER THE	16 0.13
851	IT COULD HAVE	7 0.06	1721 THE I THE I	11 0.09
852	IT I DON'T	7 0.06	1722 THE I OTHER SIDE	8 0.07
853	IT WAS MORE	7 0.06	1723 I THINK THE I	7 0.06
854	IT WAS NEAR	7 0.06	1724 ONE OF THE I	7 0.06
855	IT WAS NICE	7 0.06	1725 THE I END I	7 0.06
856	IT WAS YEAH	7 0.06	1726 ON THE I OTHER	6 0.05
857	IT'S A REALLY	7 0.06	1727 THAT'S THE I ONLY	6 0.05
858	IT'S GOT A	7 0.06	1728 THE I END BUT	6 0.05
859	IT'S IT'S IT'S	7 0.06	1729 THE I ERM THE	6 0.05
860	IT'S NOT VERY	7 0.06	1730 TOWARDS THE I END	6 0.05
861	IT'S QUITE A	7 0.06	1731 WAS THE I ONLY	6 0.05
862	IT'S YOU KNOW	7 0.06	1732 YOU KNOW THE I	6 0.05
863	JUST LIKE A	7 0.06	1733 BY THE I END	5 0.04
864	KNOW I JUST	7 0.06	1734 IT WAS THE I	5 0.04
865	KNOW IT'S NOT	7 0.06	1735 OF THE I ENGLISH	5 0.04
866	KNOW YOU CAN	7 0.06	1736 SORT OF THE I	5 0.04
867	LAST YEAR I	7 0.06	1737 THE I ONLY ONE	5 0.04
868	LIKE HER BUT	7 0.06	1738 THE I ONLY PLACE	5 0.04
869	LOOKS AT THE	7 0.06	1739 THE I ONLY THING	5 0.04
870	LOT OF TIME	7 0.06	1740 THE I SORT OF	5 0.04

Table A2: List of all 3-grams in LINDSEI-GE with a minimum frequency of 5

No.	Word	Freq.	ptw	No.	Word	Freq.	ptw
1	I DON'T KNOW	284	3.30	559	THE ON THE	7	0.08
2	A LOT OF	71	0.82	560	THE PICTURE SHE	7	0.08
3	AND ERM I	67	0.78	561	THERE WERE SO	7	0.08
4	IN THE I	65	0.75	562	THEY ARE VERY	7	0.08
5	THE I OTHER	59	0.68	563	THEY SEEM TO	7	0.08
6	YEAH YEAH YEAH	58	0.67	564	THEY THEY THEY	7	0.08
7	YOU HAVE TO	55	0.64	565	THEY WERE REALLY	7	0.08
8	A LITTLE BIT	54	0.63	566	THINGS LIKE THAT	7	0.08
9	AND I THINK	52	0.60	567	THINK I WOULD	7	0.08
10	I THINK IT'S	52	0.60	568	THIS IS NOT	7	0.08
11	AND IT WAS	47	0.55	569	TO EACH OTHER	7	0.08
12	AND ERM YEAH	43	0.50	570	TO GET TO	7	0.08
13	I THINK I	43	0.50	571	TWO OR THREE	7	0.08
14	ERM I THINK	41	0.48	572	TWO YEARS AGO	7	0.08
15	TO GO TO	40	0.46	573	VERY MUCH AND	7	0.08
16	# SEC BREAK	39	0.45	574	WAS A GOOD	7	0.08
17	AND THEN I	39	0.45	575	WAS A LITTLE	7	0.08
18	I HAVE TO	39	0.45	576	WAS ERM I	7	0.08
19	ERM I DON'T	38	0.44	577	WENT TO THE	7	0.08
20	THE I END	37	0.43	578	WHAT WAS GOING	7	0.08
21	I WENT TO	35	0.41	579	YEAH AND THE	7	0.08
22	I THINK IT	34	0.39	580	YEAH BUT I	7	0.08
23	WOULD LIKE TO	34	0.39	581	YEAH MHM YEAH	7	0.08
24	I WANT TO	33	0.38	582	YEAH WELL ERM	7	0.08
25	I WOULD SAY	32	0.37	583	YEAH WELL I	7	0.08
26	IT WAS A	32	0.37	584	YES I HAVE	7	0.08
27	IT WAS REALLY	32	0.37	585	YES I THINK	7	0.08
28	BUT I THINK	31	0.36	586	YOU DON'T KNOW	7	0.08
29	DON'T KNOW IF	31	0.36	587	YOU GO TO	7	0.08
30	YEAH I THINK	31	0.36	588	YOU HAVE THE	7	0.08
31	AND I WAS	29	0.34	589	YOU KNOW IT	7	0.08
32	ERM IT WAS	29	0.34	590	YOU YOU HAVE	7	0.08
33	I THINK THE	29	0.34	591	A COUNTRY I'VE	6	0.07
34	I WOULD LIKE	28	0.32	592	A GOOD JOB	6	0.07
35	IT WAS ERM	28	0.32	593	A GOOD THING	6	0.07
36	WHEN I WAS	28	0.32	594	A REALLY NICE	6	0.07
37	ERM YEAH I	27	0.31	595	A WEEK AND	6	0.07
38	I HAD TO	27	0.31	596	A WOMAN AND	6	0.07
39	I THINK THAT	27	0.31	597	A YEAR AND	6	0.07
40	THE I ERM	27	0.31	598	ABOUT IT BUT	6	0.07
41	THE PICTURE AND	27	0.31	599	ALSO WENT TO	6	0.07
42	AND IN THE	26	0.30	600	AND AFTER THAT	6	0.07
43	ERM THE I	26	0.30	601	AND ALL THAT	6	0.07
44	TO HER FRIENDS	26	0.30	602	AND ER WE	6	0.07
45	I THINK SO	25	0.29	603	AND ERM AFTER	6	0.07
46	I WANTED TO	25	0.29	604	AND ERM ER	6	0.07
47	I WAS IN	25	0.29	605	AND ERM TO	6	0.07
48	ON THE I	25	0.29	606	AND I COULD	6	0.07
49	AND ERM WELL	24	0.28	607	AND I JUST	6	0.07
50	BUT ERM I	24	0.28	608	AND I WENT	6	0.07
51	I MEAN I	23	0.27	609	AND IT'S IT'S	6	0.07
52	IN A WAY	23	0.27	610	AND IT'S JUST	6	0.07
53	DON'T KNOW I	22	0.26	611	AND IT'S REALLY	6	0.07
54	I REALLY LIKED	22	0.26	612	AND IT'S YEAH	6	0.07

55	IT WAS QUITE	22	0.26	613	AND SHE LOOKS	6	0.07
56	SO IT WAS	22	0.26	614	AND SO ERM	6	0.07
57	YEAH IT WAS	22	0.26	615	AND STUFF AND	6	0.07
58	AND I DON'T	21	0.24	616	AND THE PEOPLE	6	0.07
59	AND THE I	21	0.24	617	AND THE THE	6	0.07
60	BUT IT WAS	21	0.24	618	AND TO THE	6	0.07
61	I THINK THEY	21	0.24	619	AND YEAH I	6	0.07
62	I'D LIKE TO	21	0.24	620	AS FAR AS	6	0.07
63	IT WAS JUST	21	0.24	621	AS LONG AS	6	0.07
64	IT WAS VERY	21	0.24	622	AT LEAST I	6	0.07
65	SO I THINK	21	0.24	623	BACK TO THE	6	0.07
66	SOMETHING LIKE THAT	21	0.24	624	BECAUSE ER I	6	0.07
67	TO TALK ABOUT	21	0.24	625	BECAUSE I HAD	6	0.07
68	AND ER I	20	0.23	626	BECAUSE YOU KNOW	6	0.07
69	ERM I I	20	0.23	627	BEFORE I WENT	6	0.07
70	I I THINK	20	0.23	628	BIT OF A	6	0.07
71	THE I THE	20	0.23	629	BUT ERM THE	6	0.07
72	THE UNITED STATES	20	0.23	630	BUT I WAS	6	0.07
73	AND THAT WAS	19	0.22	631	CLOSE TO THE	6	0.07
74	ERM IN THE	19	0.22	632	COUNTRY AND I	6	0.07
75	ERM WELL I	19	0.22	633	COUPLE OF YEARS	6	0.07
76	I DON'T THINK	19	0.22	634	DEPENDS ON THE	6	0.07
77	I I I	19	0.22	635	DIDN'T WANT TO	6	0.07
78	IT WAS NOT	19	0.22	636	DO YOU CALL	6	0.07
79	ON THE PICTURE	19	0.22	637	DON'T KNOW BUT	6	0.07
80	A COUPLE OF	18	0.21	638	DON'T KNOW WHETHER	6	0.07
81	AND SO ON	18	0.21	639	DON'T KNOW YET	6	0.07
82	AND THIS IS	18	0.21	640	EACH OTHER AND	6	0.07
83	I MEAN IT'S	18	0.21	641	ER I I	6	0.07
84	I THINK ERM	18	0.21	642	ER THE FIRST	6	0.07
85	LIKE TO GO	18	0.21	643	ERM AFTER A	6	0.07
86	THINK IT WAS	18	0.21	644	ERM AND THEN	6	0.07
87	AND THEN ERM	17	0.20	645	ERM ERM I	6	0.07
88	BECAUSE IT WAS	17	0.20	646	ERM HOW DO	6	0.07
89	FIRST OF ALL	17	0.20	647	ERM LET'S SAY	6	0.07
90	GO TO THE	17	0.20	648	ERM THERE IS	6	0.07
91	HAVE TO BE	17	0.20	649	ERM THEY ARE	6	0.07
92	I THINK THAT'S	17	0.20	650	ERM THEY THEY	6	0.07
93	IN THE BEGINNING	17	0.20	651	ERM WELL SHE	6	0.07
94	MOST OF THE	17	0.20	652	ERM YOU HAVE	6	0.07
95	OR SOMETHING LIKE	17	0.20	653	EVEN IF I	6	0.07
96	SHOWS IT TO	17	0.20	654	FOR ME TO	6	0.07
97	THE WAY I	17	0.20	655	FOR THE FIRST	6	0.07
98	WAS REALLY NICE	17	0.20	656	FROM THE I	6	0.07
99	AND ERM THE	16	0.19	657	GET A JOB	6	0.07
100	AND SO I	16	0.19	658	GOING ON THERE	6	0.07
101	DON'T KNOW THE	16	0.19	659	GOT TO KNOW	6	0.07
102	ERM I WAS	16	0.19	660	HAVE THE I	6	0.07
103	MORE OR LESS	16	0.19	661	HE CHANGED THE	6	0.07
104	SO THAT WAS	16	0.19	662	HE'S A BIT	6	0.07
105	VERY NICE AND	16	0.19	663	I CAME BACK	6	0.07
106	YEAH YEAH I	16	0.19	664	I CAME TO	6	0.07
107	AND ER THE	15	0.17	665	I CAN SAY	6	0.07
108	AND ER YEAH	15	0.17	666	I CAN'T REALLY	6	0.07
109	AND I REALLY	15	0.17	667	I COME FROM	6	0.07
110	AND THEN THE	15	0.17	668	I DID MY	6	0.07

111	GET TO KNOW	15	0.17	669	I GUESS THAT'S	6	0.07
112	I DON'T REALLY	15	0.17	670	I HAVE THE	6	0.07
113	I REALLY LIKE	15	0.17	671	I HOPE I	6	0.07
114	IN IN THE	15	0.17	672	I HOPE THAT	6	0.07
115	IN THE PICTURE	15	0.17	673	I I LIKE	6	0.07
116	IT TO HER	15	0.17	674	I LIKE ERM	6	0.07
117	IT WOULD BE	15	0.17	675	I LIKE THAT	6	0.07
118	NOT SO MUCH	15	0.17	676	I LIVED WITH	6	0.07
119	SHE WANTS TO	15	0.17	677	I MEAN IN	6	0.07
120	SO ERM I	15	0.17	678	I THINK YOU	6	0.07
121	SO I DON'T	15	0.17	679	I THOUGHT I	6	0.07
122	THERE IS A	15	0.17	680	I THOUGHT WELL	6	0.07
123	A BIT OF	14	0.16	681	I TRY TO	6	0.07
124	A PAINTER AND	14	0.16	682	I WAS QUITE	6	0.07
125	AND ERM THEY	14	0.16	683	I WAS REALLY	6	0.07
126	AT THE I	14	0.16	684	I WATCHED IT	6	0.07
127	AT THE MOMENT	14	0.16	685	IF YOU LOOK	6	0.07
128	DON'T KNOW WHAT	14	0.16	686	IF YOU SEE	6	0.07
129	ER I DON'T	14	0.16	687	IN GERMANY I	6	0.07
130	ERM I HAD	14	0.16	688	IN IN A	6	0.07
131	HER FRIENDS AND	14	0.16	689	IN PRIMARY SCHOOL	6	0.07
132	I LIKED THE	14	0.16	690	IN THE LAST	6	0.07
133	IN FRONT OF	14	0.16	691	IN THE MIDDLE	6	0.07
134	ONE OF THE	14	0.16	692	IN THE WHOLE	6	0.07
135	SEEMS TO BE	14	0.16	693	IT DOESN'T LOOK	6	0.07
136	THE PICTURE IS	14	0.16	694	IT IN THE	6	0.07
137	THERE WAS A	14	0.16	695	IT WAS BECAUSE	6	0.07
138	TO BE A	14	0.16	696	IT WAS I	6	0.07
139	TO ERM TO	14	0.16	697	IT WAS KIND	6	0.07
140	TO THE I	14	0.16	698	IT WAS NICE	6	0.07
141	WELL I THINK	14	0.16	699	IT'S A BIT	6	0.07
142	YEAH AND I	14	0.16	700	IT'S EASY TO	6	0.07
143	YEAH I DON'T	14	0.16	701	IT'S IT'S REALLY	6	0.07
144	YOU CAN DO	14	0.16	702	IT'S NOT SO	6	0.07
145	AND I HAD	13	0.15	703	LIKE IN THE	6	0.07
146	AND I I	13	0.15	704	LIKE THAT YEAH	6	0.07
147	AND THEN SHE	13	0.15	705	LIKE THIS AND	6	0.07
148	AND THEN YOU	13	0.15	706	LIKE TO WORK	6	0.07
149	BECAUSE I WAS	13	0.15	707	LOT OF THINGS	6	0.07
150	ER THE I	13	0.15	708	LOT OF TIME	6	0.07
151	I DON'T HAVE	13	0.15	709	ME AND ERM	6	0.07
152	I DON'T LIKE	13	0.15	710	MHM # SEC	6	0.07
153	I OTHER HAND	13	0.15	711	MHM YEAH YEAH	6	0.07
154	I THE I	13	0.15	712	NO NOT REALLY	6	0.07
155	I THINK HE	13	0.15	713	OF COURSE AND	6	0.07
156	I WAS A	13	0.15	714	OF COURSE BUT	6	0.07
157	IT AND I	13	0.15	715	OF MY LIFE	6	0.07
158	IT WAS THE	13	0.15	716	OF THE PEOPLE	6	0.07
159	LIKE THE I	13	0.15	717	OH I DON'T	6	0.07
160	THE SECOND PICTURE	13	0.15	718	OKAY ERM I	6	0.07
161	THE THE THE	13	0.15	719	ON THE FIRST	6	0.07
162	TO GO BACK	13	0.15	720	OR AT LEAST	6	0.07
163	WANTS TO BE	13	0.15	721	OR WHATEVER AND	6	0.07
164	WE HAD A	13	0.15	722	OUT OF IT	6	0.07
165	WE WENT TO	13	0.15	723	OUT OF THE	6	0.07
166	AND ERM AND	12	0.14	724	OVER THE WORLD	6	0.07

167	AND ERM WE	12	0.14	725	PAINTER AND ERM	6	0.07
168	AND I DIDN'T	12	0.14	726	PORTRAIT OF A	6	0.07
169	AND IF YOU	12	0.14	727	REALLY LIKED THE	6	0.07
170	AND THEY ARE	12	0.14	728	SEC BREAK WELL	6	0.07
171	AS WELL AND	12	0.14	729	SHE LOOKS A	6	0.07
172	BUT I DON'T	12	0.14	730	SHE LOOKS LIKE	6	0.07
173	BUT I I	12	0.14	731	SO AND ERM	6	0.07
174	DON'T KNOW IT'S	12	0.14	732	SO ERM IT	6	0.07
175	ERM BUT I	12	0.14	733	SO I MEAN	6	0.07
176	ERM YEAH AND	12	0.14	734	SO I WENT	6	0.07
177	HAVE TO GO	12	0.14	735	SO MANY THINGS	6	0.07
178	I DIDN'T KNOW	12	0.14	736	THAN SHE ACTUALLY	6	0.07
179	I HAD THE	12	0.14	737	THAT ERM I	6	0.07
180	I MEAN YOU	12	0.14	738	THAT I HAVE	6	0.07
181	I STARTED TO	12	0.14	739	THAT THE I	6	0.07
182	I THINK SHE	12	0.14	740	THAT THERE IS	6	0.07
183	IT AND ERM	12	0.14	741	THAT THIS IS	6	0.07
184	IT WAS SO	12	0.14	742	THAT WAS PRETTY	6	0.07
185	LIKE THAT AND	12	0.14	743	THAT WAS THE	6	0.07
186	MM I THINK	12	0.14	744	THE I AFTERNOON	6	0.07
187	REALLY NICE AND	12	0.14	745	THE I AMERICAN	6	0.07
188	SHE DOESN'T LIKE	12	0.14	746	THE I AREA	6	0.07
189	SOME KIND OF	12	0.14	747	THE I EXPERIENCE	6	0.07
190	THAT WAS REALLY	12	0.14	748	THE ONE HAND	6	0.07
191	THE COUNTRY AND	12	0.14	749	THE SAME WITH	6	0.07
192	THE WAY SHE	12	0.14	750	THE TIME AND	6	0.07
193	THINK IT'S A	12	0.14	751	THE U S	6	0.07
194	TO DO IT	12	0.14	752	THEN ERM I	6	0.07
195	TO HAVE A	12	0.14	753	THERE AND I	6	0.07
196	WANT TO GO	12	0.14	754	THERE ARE MANY	6	0.07
197	WENT TO A	12	0.14	755	THERE IS NO	6	0.07
198	YEAH AND ERM	12	0.14	756	THERE IS ONE	6	0.07
199	YEAH OF COURSE	12	0.14	757	THERE WAS ERM	6	0.07
200	YEAH THAT WAS	12	0.14	758	THERE WILL BE	6	0.07
201	ALL THE TIME	11	0.13	759	THEY THEY ARE	6	0.07
202	AND ERM SO	11	0.13	760	THEY THEY WERE	6	0.07
203	AND ERM YES	11	0.13	761	THEY WERE SO	6	0.07
204	AND I HAVE	11	0.13	762	THINK ABOUT IT	6	0.07
205	AND THEY WERE	11	0.13	763	THOUGHT IT WAS	6	0.07
206	DIDN'T LIKE IT	11	0.13	764	THREE OR FOUR	6	0.07
207	ER IN THE	11	0.13	765	TO DO THIS	6	0.07
208	ERM WHEN I	11	0.13	766	TO MAKE A	6	0.07
209	HAVE TO DO	11	0.13	767	TO SAY I	6	0.07
210	I I DON'T	11	0.13	768	TO SEE ALL	6	0.07
211	I IT WAS	11	0.13	769	TO TALK TO	6	0.07
212	I KNOW THAT	11	0.13	770	TO THE STATES	6	0.07
213	I LIKED IT	11	0.13	771	VERY INTERESTING AND	6	0.07
214	I WHEN I	11	0.13	772	WANT TO DO	6	0.07
215	I WOULD HAVE	11	0.13	773	WAS A REALLY	6	0.07
216	I'M GOING TO	11	0.13	774	WAS REALLY REALLY	6	0.07
217	I'M NOT QUITE	11	0.13	775	WAY OF LIFE	6	0.07
218	IN IN IN	11	0.13	776	WE HAD TO	6	0.07
219	IN THE UNITED	11	0.13	777	WERE SO MANY	6	0.07
220	IT WAS ER	11	0.13	778	WHAT IMPRESSED ME	6	0.07
221	IT WAS IN	11	0.13	779	WHEN I STARTED	6	0.07
222	KNOW HOW TO	11	0.13	780	WHERE YOU CAN	6	0.07
223	NOT QUITE SURE	11	0.13	781	WHICH WAS REALLY	6	0.07

224	OF THE I	11	0.13	782	WITH THE RESULT	6	0.07
225	PICTURE AND SHE	11	0.13	783	WOULD BE NICE	6	0.07
226	PICTURE OF HER	11	0.13	784	WOULD HAVE BEEN	6	0.07
227	SHE SHOWS IT	11	0.13	785	YEAH AND ER	6	0.07
228	THAT WAS A	11	0.13	786	YEAH I LIKE	6	0.07
229	THAT WAS VERY	11	0.13	787	YEAH WHAT ELSE	6	0.07
230	THE I OLD	11	0.13	788	YEAH YEAH AND	6	0.07
231	THE IN THE	11	0.13	789	YEAH YEAH IT	6	0.07
232	THE PICTURE TO	11	0.13	790	YEARS AGO AND	6	0.07
233	THE THE I	11	0.13	791	YES ERM I	6	0.07
234	THEY HAVE TO	11	0.13	792	YOU CAN GO	6	0.07
235	THEY WERE ALL	11	0.13	793	YOU GET TO	6	0.07
236	TO DO THAT	11	0.13	794	YOU KNOW YOU	6	0.07
237	TO GET A	11	0.13	795	A BIG CITY	5	0.06
238	WAS IT WAS	11	0.13	796	A BIT MORE	5	0.06
239	YEAH YEAH ERM	11	0.13	797	A DIFFICULT QUESTION	5	0.06
240	A PORTRAIT OF	10	0.12	798	A FOREIGN LANGUAGE	5	0.06
241	AFTER A WHILE	10	0.12	799	A GROUP OF	5	0.06
242	AND A HALF	10	0.12	800	A HALF YEARS	5	0.06
243	AND ERM BUT	10	0.12	801	A LOT ERM	5	0.06
244	AND ERM THERE	10	0.12	802	A LOT TO	5	0.06
245	AND SHE IS	10	0.12	803	A NEW PICTURE	5	0.06
246	AND SHE WAS	10	0.12	804	A PAINTER WHO	5	0.06
247	AND THERE WAS	10	0.12	805	A T V	5	0.06
248	AND YEAH AND	10	0.12	806	A VERY NICE	5	0.06
249	AND YOU CAN	10	0.12	807	A VERY VERY	5	0.06
250	AND YOU HAVE	10	0.12	808	A YEAR OR	5	0.06
251	AT THAT TIME	10	0.12	809	ALL KINDS OF	5	0.06
252	BECAUSE I THINK	10	0.12	810	ALL OF THEM	5	0.06
253	BUT I'M NOT	10	0.12	811	AN AU PAIR	5	0.06
254	DON'T REALLY KNOW	10	0.12	812	AND AT THE	5	0.06
255	DON'T WANT TO	10	0.12	813	AND ER ER	5	0.06
256	FOR ME AND	10	0.12	814	AND ER IT	5	0.06
257	GO TO A	10	0.12	815	AND ER SO	5	0.06
258	HAD TO GO	10	0.12	816	AND ER WELL	5	0.06
259	HERE IN GERMANY	10	0.12	817	AND ERM ERM	5	0.06
260	I DIDN'T HAVE	10	0.12	818	AND ERM MY	5	0.06
261	I DON'T I	10	0.12	819	AND ERM THAT	5	0.06
262	I HAVE A	10	0.12	820	AND ERM THIS	5	0.06
263	I KNOW I	10	0.12	821	AND HE HE	5	0.06
264	I LIKE THE	10	0.12	822	AND I ALSO	5	0.06
265	I OTHER ONE	10	0.12	823	AND IT'S VERY	5	0.06
266	I THINK YEAH	10	0.12	824	AND ON THE	5	0.06
267	I WAS LIKE	10	0.12	825	AND SHE DOESN'T	5	0.06
268	I'M NOT SURE	10	0.12	826	AND SHE SHOWS	5	0.06
269	IN THE IN	10	0.12	827	AND SO HE	5	0.06
270	IT WAS IT	10	0.12	828	AND THERE IS	5	0.06
271	IT'S KIND OF	10	0.12	829	AND THERE WERE	5	0.06
272	LITTLE BIT MORE	10	0.12	830	AND WE HAD	5	0.06
273	ON THE ONE	10	0.12	831	AND WHEN I	5	0.06
274	SATISFIED WITH THE	10	0.12	832	AND YEAH IT	5	0.06
275	SEEM TO BE	10	0.12	833	AND YEAH YEAH	5	0.06
276	SHE DIDN'T LIKE	10	0.12	834	AND YOU YOU	5	0.06
277	SO ERM YEAH	10	0.12	835	ARE A LOT	5	0.06
278	THAT SHE LOOKS	10	0.12	836	AS GOOD AS	5	0.06
279	THAT WAS QUITE	10	0.12	837	AS WELL YEAH	5	0.06

280	THAT'S WHAT I	10	0.12	838	AT IT AND	5	0.06
281	THE I ENGLISH	10	0.12	839	AT SCHOOL ERM	5	0.06
282	THE I ONLY	10	0.12	840	AT THE BEGINNING	5	0.06
283	THE I OTHERS	10	0.12	841	BE A LITTLE	5	0.06
284	THEY ARE NOT	10	0.12	842	BE A TEACHER	5	0.06
285	THIS IS A	10	0.12	843	BECAUSE ER THE	5	0.06
286	TO GO THERE	10	0.12	844	BECAUSE ERM IT	5	0.06
287	WAS VERY NICE	10	0.12	845	BECAUSE HE WAS	5	0.06
288	WHAT I REALLY	10	0.12	846	BECAUSE I DIDN'T	5	0.06
289	WHEN I WHEN	10	0.12	847	BECAUSE I MEAN	5	0.06
290	YEAH ERM YEAH	10	0.12	848	BECAUSE SHE LOOKS	5	0.06
291	YOU DON'T HAVE	10	0.12	849	BECAUSE THERE ARE	5	0.06
292	YOU SEE THE	10	0.12	850	BREAK OKAY ERM	5	0.06
293	A FRIEND OF	9	0.10	851	BUT ERM IF	5	0.06
294	A LOT ABOUT	9	0.10	852	BUT I DIDN'T	5	0.06
295	ALL THE I	9	0.10	853	BUT I GUESS	5	0.06
296	AND ERM WHEN	9	0.10	854	BUT IN THE	5	0.06
297	AND OF COURSE	9	0.10	855	BUT THEY WERE	5	0.06
298	AND THEN HE	9	0.10	856	BUT THIS IS	5	0.06
299	AND THEY HAVE	9	0.10	857	COS IT WAS	5	0.06
300	AND THIS WAS	9	0.10	858	COUNTRY I'VE VISITED	5	0.06
301	BECAUSE ERM I	9	0.10	859	DIDN'T HAVE A	5	0.06
302	BECAUSE I I	9	0.10	860	DIDN'T LIKE THE	5	0.06
303	BECAUSE THEY	9	0.10	861	DO IT AGAIN	5	0.06
	WERE						
304	BECOME A	9	0.10	862	DO THAT AND	5	0.06
	TEACHER						
305	BUT ON THE	9	0.10	863	DOESN'T LIKE IT	5	0.06
306	DON'T KNOW ERM	9	0.10	864	DOESN'T LOOK LIKE	5	0.06
307	DON'T KNOW HOW	9	0.10	865	DOESN'T SEEM TO	5	0.06
308	DON'T KNOW	9	0.10	866	DON'T HAVE A	5	0.06
	MAYBE						
309	DON'T KNOW	9	0.10	867	DON'T I DON'T	5	0.06
	REALLY						
310	ER IT WAS	9	0.10	868	DON'T KNOW THEY	5	0.06
311	ERM I WOULD	9	0.10	869	DON'T KNOW YEAH	5	0.06
312	ERM THAT WAS	9	0.10	870	EASY TO GET	5	0.06
313	ERM YES I	9	0.10	871	EH I DON'T	5	0.06
314	G D R	9	0.10	872	ER THE THE	5	0.06
315	HE PAINTS HER	9	0.10	873	ERM A LITTLE	5	0.06
316	HOW DO YOU	9	0.10	874	ERM AT THE	5	0.06
317	I ERM I	9	0.10	875	ERM ENGLISH AND	5	0.06
318	I HAD A	9	0.10	876	ERM I CAN'T	5	0.06
319	I LIKE TO	9	0.10	877	ERM I REALLY	5	0.06
320	I LIVED IN	9	0.10	878	ERM I STARTED	5	0.06
321	I MEAN ERM	9	0.10	879	ERM I STAYED	5	0.06
322	I MEAN THEY	9	0.10	880	ERM I WENT	5	0.06
323	I THINK IN	9	0.10	881	ERM I'M NOT	5	0.06
324	I THINK WE	9	0.10	882	ERM IT'S A	5	0.06
325	I THOUGHT ABOUT	9	0.10	883	ERM IT'S QUITE	5	0.06
326	I WAS THERE	9	0.10	884	ERM THE THE	5	0.06
327	I WAS VERY	9	0.10	885	ERM TO BE	5	0.06
328	I WENT THERE	9	0.10	886	ERM TO GET	5	0.06
329	IF YOU IF	9	0.10	887	ERM YEAH SHE	5	0.06
330	IN THE THIRD	9	0.10	888	ERM YEAH WHAT	5	0.06
331	IS A PAINTER	9	0.10	889	ERM YOU CAN	5	0.06
332	IT'S HARD TO	9	0.10	890	ESPECIALLY IN THE	5	0.06
333	IT'S NOT THAT	9	0.10	891	EXPRESSION ON HER	5	0.06

334	LIKE I DON'T	9	0.10	892	FOR A COUPLE	5	0.06
335	LIKE TO DO	9	0.10	893	FOR EXAMPLE AND	5	0.06
336	MM I DON'T	9	0.10	894	FOR FOUR WEEKS	5	0.06
337	OF IT AND	9	0.10	895	FOR HIM AND	5	0.06
338	OF THE TIME	9	0.10	896	FOR ME IT	5	0.06
339	QUITE A LOT	9	0.10	897	FRIENDS AND I	5	0.06
340	REALLY LIKE TO	9	0.10	898	GO THERE AND	5	0.06
341	REALLY LIKED IT	9	0.10	899	HAD TO DO	5	0.06
342	SEC BREAK OKAY	9	0.10	900	HAD TO ERM	5	0.06
343	SHE IS NOT	9	0.10	901	HAPPY WITH IT	5	0.06
344	SHE SEEMS TO	9	0.10	902	HAS TO BE	5	0.06
345	SO I HAD	9	0.10	903	HAVE A VERY	5	0.06
346	THAT AND ERM	9	0.10	904	HAVE TO ADMIT	5	0.06
347	THE FIRST PICTURE	9	0.10	905	HAVE TO LEARN	5	0.06
348	THE FIRST TIME	9	0.10	906	HAVE TO SAY	5	0.06
349	THE I ARTIST	9	0.10	907	HER AND SHE	5	0.06
350	THE THIRD PICTURE	9	0.10	908	HER IN A	5	0.06
351	THERE AND ERM	9	0.10	909	HER THE WAY	5	0.06
352	THERE IS THIS	9	0.10	910	HERE AND ERM	5	0.06
353	THEY HAVE A	9	0.10	911	I CAME THERE	5	0.06
354	TO BE ERM	9	0.10	912	I CAN ONLY	5	0.06
355	TO BECOME A	9	0.10	913	I CAN'T SEE	5	0.06
356	WAS ERM THE	9	0.10	914	I CHOSE THE	5	0.06
357	WAS NOT REALLY	9	0.10	915	I DECIDED TO	5	0.06
358	WHAT DO YOU	9	0.10	916	I DID A	5	0.06
359	WHICH IS NOT	9	0.10	917	I DIDN'T THINK	5	0.06
360	WITH THE PICTURE	9	0.10	918	I DO HAVE	5	0.06
361	YEAH I GUESS	9	0.10	919	I DON'T REMEMBER	5	0.06
362	YEAH I WOULD	9	0.10	920	I DON'T WANNA	5	0.06
363	YOU WANT TO	9	0.10	921	I DON'T WANT	5	0.06
364	A IN A	8	0.09	922	I GO SWIMMING	5	0.06
365	A KIND OF	8	0.09	923	I HAD SOME	5	0.06
366	A PICTURE OF	8	0.09	924	I HAVE I	5	0.06
367	ABOUT IT AND	8	0.09	925	I I REALLY	5	0.06
368	ALL HER FRIENDS	8	0.09	926	I JUST I	5	0.06
369	ALL OVER THE	8	0.09	927	I KNOW BUT	5	0.06
370	AND AND AND	8	0.09	928	I MEAN IT	5	0.06
371	AND ERM HE	8	0.09	929	I MEAN THERE	5	0.06
372	AND ERM SHE	8	0.09	930	I SAID OKAY	5	0.06
373	AND EVERYTHING WAS	8	0.09	931	I SAW THAT	5	0.06
374	AND I GUESS	8	0.09	932	I THINK I'M	5	0.06
375	AND I LIKED	8	0.09	933	I THOUGHT THAT	5	0.06
376	AND THE PICTURE	8	0.09	934	I WAS SO	5	0.06
377	AND THERE ARE	8	0.09	935	IF I CAN	5	0.06
378	AND THEY JUST	8	0.09	936	IF IF YOU	5	0.06
379	AS WELL BECAUSE	8	0.09	937	IF YOU ARE	5	0.06
380	AT THE PICTURE	8	0.09	938	IF YOU COMPARE	5	0.06
381	BECAUSE OF THE	8	0.09	939	I'M NOT REALLY	5	0.06
382	BUT IT'S NOT	8	0.09	940	I'M NOT THAT	5	0.06
383	DON'T HAVE TO	8	0.09	941	I'M SUPPOSED TO	5	0.06
384	DON'T KNOW IT	8	0.09	942	IN I DON'T	5	0.06
385	ER I THINK	8	0.09	943	IN NINETEEN NINETY	5	0.06
386	ERM ERM YEAH	8	0.09	944	IN THE BACK	5	0.06
387	ERM I GUESS	8	0.09	945	IN THE MOVIE	5	0.06
388	ERM THERE WERE	8	0.09	946	IN TWO THOUSAND	5	0.06
389	ERM YEAH ERM	8	0.09	947	IS A BIT	5	0.06

390	ERM YEAH IT'S	8	0.09	948	IS DIFFERENT AND	5	0.06
391	ERM YEAH THE	8	0.09	949	IS NOT REALLY	5	0.06
392	FOR EXAMPLE ERM	8	0.09	950	IS NOT SATISFIED	5	0.06
393	FRIEND OF MINE	8	0.09	951	IS QUITE A	5	0.06
394	HALF A YEAR	8	0.09	952	IT A LOT	5	0.06
395	HAVE TO PAY	8	0.09	953	IT AND HE	5	0.06
396	I DIDN'T LIKE	8	0.09	954	IT AND SHE	5	0.06
397	I HAVE NO	8	0.09	955	IT BUT IT	5	0.06
398	I I WOULD	8	0.09	956	IT DEPENDS ON	5	0.06
399	I THINK AND	8	0.09	957	IT IS A	5	0.06
400	I THOUGHT IT	8	0.09	958	IT WAS ABOUT	5	0.06
401	IN A IN	8	0.09	959	IT YEAH I	5	0.06
402	IN ORDER TO	8	0.09	960	IT'S ER IT'S	5	0.06
403	IN THE SECOND	8	0.09	961	IT'S ERM IT'S	5	0.06
404	IT BUT I	8	0.09	962	IT'S IT WAS	5	0.06
405	IT I THINK	8	0.09	963	IT'S IT'S THE	5	0.06
406	IT IT WAS	8	0.09	964	IT'S IT'S VERY	5	0.06
407	IT SO I	8	0.09	965	IT'S JUST A	5	0.06
408	IT WAS LIKE	8	0.09	966	IT'S REALLY NICE	5	0.06
409	IT'S A GOOD	8	0.09	967	IT'S THE SAME	5	0.06
410	IT'S IT'S A	8	0.09	968	I'VE GOT A	5	0.06
411	I'VE BEEN THERE	8	0.09	969	KNOW I DON'T	5	0.06
412	JUST HAVE TO	8	0.09	970	KNOW THE I	5	0.06
413	KIND OF A	8	0.09	971	KNOW WHAT I	5	0.06
414	KNOW IF I	8	0.09	972	LIKE THAT BUT	5	0.06
415	LOOK AT THE	8	0.09	973	LIKE THAT I	5	0.06
416	LOOKS A BIT	8	0.09	974	LIKE THAT SO	5	0.06
417	MOST OF THEM	8	0.09	975	LIKE TO BE	5	0.06
418	NICE AND ERM	8	0.09	976	LOT ABOUT IT	5	0.06
419	NOT AT ALL	8	0.09	977	MAYBE THEY ARE	5	0.06
420	ON T V	8	0.09	978	NICE IT WAS	5	0.06
421	SHOWS THE PICTURE	8	0.09	979	NO I THINK	5	0.06
422	SUPPOSED TO BE	8	0.09	980	NO NO I	5	0.06
423	T V AND	8	0.09	981	NO NOT AT	5	0.06
424	THAT IT WAS	8	0.09	982	NO NOT YET	5	0.06
425	THAT WOULD BE	8	0.09	983	NOT SATISFIED WITH	5	0.06
426	THAT YOU CAN	8	0.09	984	NOT THAT MUCH	5	0.06
427	THAT'S A GOOD	8	0.09	985	O'CLOCK IN THE	5	0.06
428	THE I AGE	8	0.09	986	OF COURSE YEAH	5	0.06
429	THE I ER	8	0.09	987	OF THE COUNTRY	5	0.06
430	THE I EVENING	8	0.09	988	OF THE OF	5	0.06
431	THE WAY THEY	8	0.09	989	ON HER FACE	5	0.06
432	THEN YOU HAVE	8	0.09	990	ON THE SECOND	5	0.06
433	THEY HAD A	8	0.09	991	ONE OF THEM	5	0.06
434	THEY WERE VERY	8	0.09	992	OR IN THE	5	0.06
435	TO BE VERY	8	0.09	993	OR SOMETHING AND	5	0.06
436	TO FIND A	8	0.09	994	PEOPLE AND ERM	5	0.06
437	TO TELL YOU	8	0.09	995	PICTURE TO HER	5	0.06
438	TO TO GET	8	0.09	996	PICTURE TO THE	5	0.06
439	USED TO IT	8	0.09	997	POINT OF VIEW	5	0.06
440	VERY VERY NICE	8	0.09	998	QUITE A WHILE	5	0.06
441	WANT TO BE	8	0.09	999	REALLY WANT TO	5	0.06
442	WAS A BIT	8	0.09	1000	SEE ALL THE	5	0.06
443	WAS IN THE	8	0.09	1001	SHE DOESN'T LOOK	5	0.06
444	WAS KIND OF	8	0.09	1002	SHE IS ERM	5	0.06
445	WENT TO ERM	8	0.09	1003	SHE LIKES IT	5	0.06

446	WHEN I CAME	8	0.09	1004	SHE THINKS THAT	5	0.06
447	WITH THE I	8	0.09	1005	SHE WAS REALLY	5	0.06
448	YEAH ERM I	8	0.09	1006	SHE WENT TO	5	0.06
449	YEAH ERM WELL	8	0.09	1007	SITTING IN THE	5	0.06
450	YEAH I I	8	0.09	1008	SITTING THERE AND	5	0.06
451	YEAH I MEAN	8	0.09	1009	SO ERM IT'S	5	0.06
452	YEAH YEAH OKAY	8	0.09	1010	SO HE HAS	5	0.06
453	YOU CAN SEE	8	0.09	1011	SO I THOUGHT	5	0.06
454	YOU IF YOU	8	0.09	1012	SO IT'S ERM	5	0.06
455	A LOT AND	7	0.08	1013	SO THIS WAS	5	0.06
456	A REALLY GOOD	7	0.08	1014	SOME THINGS THAT	5	0.06
457	ABOUT A COUNTRY	7	0.08	1015	SOMETHING AND I	5	0.06
458	AND ALL THE	7	0.08	1016	THAT I CAN	5	0.06
459	AND ER SHE	7	0.08	1017	THAT I I	5	0.06
460	AND ERM IT	7	0.08	1018	THAT I THINK	5	0.06
461	AND ERM WHAT	7	0.08	1019	THAT VERY MUCH	5	0.06
462	AND HE SAID	7	0.08	1020	THAT YEAH YEAH	5	0.06
463	AND I ERM	7	0.08	1021	THE FIRST DAY	5	0.06
464	AND I KNOW	7	0.08	1022	THE FIRST THING	5	0.06
465	AND I LIKE	7	0.08	1023	THE I EYES	5	0.06
466	AND I MEAN	7	0.08	1024	THE I ISLAND	5	0.06
467	AND I THOUGHT	7	0.08	1025	THE PEOPLE I	5	0.06
468	AND SO IT	7	0.08	1026	THE PEOPLE THERE	5	0.06
469	AND THAT'S WHY	7	0.08	1027	THE REST OF	5	0.06
470	AND THEN IT	7	0.08	1028	THE RING AND	5	0.06
471	AND THEN THEY	7	0.08	1029	THE SECOND ONE	5	0.06
472	AND THEY THEY	7	0.08	1030	THE U K	5	0.06
473	AND WE WENT	7	0.08	1031	THE WAY OF	5	0.06
474	AS WELL SO	7	0.08	1032	THE WHOLE THING	5	0.06
475	BE ABLE TO	7	0.08	1033	THEM AND THEY	5	0.06
476	BECAUSE I DON'T	7	0.08	1034	THEN I WAS	5	0.06
477	BECAUSE THE I	7	0.08	1035	THEN THE I	5	0.06
478	BECAUSE THEY ARE	7	0.08	1036	THERE AND ER	5	0.06
479	BUT ERM IT'S	7	0.08	1037	THERE AND THEN	5	0.06
480	BUT ERM YEAH	7	0.08	1038	THEY ARE JUST	5	0.06
481	BUT THE I	7	0.08	1039	THEY ARE WEARING	5	0.06
482	DO YOU SAY	7	0.08	1040	THEY DIDN'T HAVE	5	0.06
483	DON'T KNOW WHY	7	0.08	1041	THEY DON'T LOOK	5	0.06
484	ER I MEAN	7	0.08	1042	THEY DON'T REALLY	5	0.06
485	ERM AND I	7	0.08	1043	THEY THEY HAVE	5	0.06
486	ERM BECAUSE I	7	0.08	1044	THEY WANT TO	5	0.06
487	ERM I HAVE	7	0.08	1045	THINK IT'S IT'S	5	0.06
488	ERM I LIKE	7	0.08	1046	THINK SO I	5	0.06
489	ERM I LIKED	7	0.08	1047	THINK THIS IS	5	0.06
490	ERM I MEAN	7	0.08	1048	THIS KIND OF	5	0.06
491	ERM I THOUGHT	7	0.08	1049	THOUGHT THAT WAS	5	0.06
492	ERM IF YOU	7	0.08	1050	TO ALL HER	5	0.06
493	ERM THERE WAS	7	0.08	1051	TO BE HONEST	5	0.06
494	ERM THIS IS	7	0.08	1052	TO BE ON	5	0.06
495	ERM WE WENT	7	0.08	1053	TO BE THE	5	0.06
496	ERM YEAH WELL	7	0.08	1054	TO DO AND	5	0.06
497	FOR EXAMPLE THE	7	0.08	1055	TO DO MY	5	0.06
498	GO BACK TO	7	0.08	1056	TO DO SOMETHING	5	0.06
499	HARD TO SAY	7	0.08	1057	TO DO THE	5	0.06
500	HE HAS TO	7	0.08	1058	TO DO WITH	5	0.06
501	HER AND ERM	7	0.08	1059	TO ERM I	5	0.06
502	I AGE OF	7	0.08	1060	TO ERM THE	5	0.06

503	I DIDN'T REALLY	7	0.08	1061	TO GET INTO	5	0.06
504	I END OF	7	0.08	1062	TO KEEP ON	5	0.06
505	I ERM THE	7	0.08	1063	TO KNOW WHAT	5	0.06
506	I I DIDN'T	7	0.08	1064	TO LIVE WITH	5	0.06
507	I I HAVE	7	0.08	1065	TO SEE A	5	0.06
508	I I WANT	7	0.08	1066	TO THINK ABOUT	5	0.06
509	I THINK ABOUT	7	0.08	1067	TO WORK AND	5	0.06
510	I THINK THIS	7	0.08	1068	TO WORK WITH	5	0.06
511	I WAS ALWAYS	7	0.08	1069	TOPIC NUMBER TWO	5	0.06
512	I WAS ERM	7	0.08	1070	TWO THOUSAND AND	5	0.06
513	I WOULD BE	7	0.08	1071	TWO THOUSAND ONE	5	0.06
514	IF YOU DON'T	7	0.08	1072	VERY DIFFICULT TO	5	0.06
515	IF YOU GO	7	0.08	1073	VERY FRIENDLY AND	5	0.06
516	IN ER IN	7	0.08	1074	WANTED ME TO	5	0.06
517	IN GERMANY AND	7	0.08	1075	WANTED TO DO	5	0.06
518	IN THE CITY	7	0.08	1076	WANTED TO GO	5	0.06
519	IN THE COUNTRY	7	0.08	1077	WANTED TO HAVE	5	0.06
520	IN THE FIRST	7	0.08	1078	WAS I THINK	5	0.06
521	IN THE MORNING	7	0.08	1079	WAS LIKE OH	5	0.06
522	IN THE NORTH	7	0.08	1080	WAS NOT SO	5	0.06
523	IN THE SOUTH	7	0.08	1081	WAS QUITE GOOD	5	0.06
524	IN THE U	7	0.08	1082	WAS QUITE INTERESTING	5	0.06
525	IS THAT ERM	7	0.08	1083	WAS REALLY A	5	0.06
526	IT VERY MUCH	7	0.08	1084	WAS REALLY INTERESTING	5	0.06
527	IT WAS ALL	7	0.08	1085	WAS SUPPOSED TO	5	0.06
528	IT'S A NICE	7	0.08	1086	WAS THE I	5	0.06
529	IT'S A VERY	7	0.08	1087	WE DON'T HAVE	5	0.06
530	I'VE BEEN TO	7	0.08	1088	WE HAVE A	5	0.06
531	I'VE NEVER BEEN	7	0.08	1089	WE WE HAD	5	0.06
532	KIND OF ER	7	0.08	1090	WELL ERM I	5	0.06
533	KNOW IF YOU	7	0.08	1091	WELL I DON'T	5	0.06
534	KNOW IT WAS	7	0.08	1092	WELL I I	5	0.06
535	LIKE TO TALK	7	0.08	1093	WELL IT'S IT'S	5	0.06
536	NO I DON'T	7	0.08	1094	WELL THEY HAD	5	0.06
537	NO IT WAS	7	0.08	1095	WENT TO SCHOOL	5	0.06
538	NOW AND ERM	7	0.08	1096	WHAT IT WAS	5	0.06
539	OF COURSE YOU	7	0.08	1097	WHAT'S GOING ON	5	0.06
540	OF HER AND	7	0.08	1098	WHEN I FIRST	5	0.06
541	OKAY # SEC	7	0.08	1099	WHICH IS A	5	0.06
542	ON THE ON	7	0.08	1100	WHICH WAS QUITE	5	0.06
543	ONE OR TWO	7	0.08	1101	WITH IT AND	5	0.06
544	OR OR OR	7	0.08	1102	WOULD BE A	5	0.06
545	OR SO AND	7	0.08	1103	YEAH NO I	5	0.06
546	PART OF THE	7	0.08	1104	YEAH NO NO	5	0.06
547	SO I I	7	0.08	1105	YEAH THE I	5	0.06
548	SO I WAS	7	0.08	1106	YEAH YOU CAN	5	0.06
549	SO ON AND	7	0.08	1107	YOU CAN'T DO	5	0.06
550	SO THIS IS	7	0.08	1108	YOU HAD TO	5	0.06
551	SO TO SAY	7	0.08	1109	YOU HAVE A	5	0.06
552	SOME OF THE	7	0.08	1110	YOU KNOW THAT	5	0.06
553	TELL YOU ABOUT	7	0.08	1111	YOU KNOW THE	5	0.06
554	THAT THEY ARE	7	0.08	1112	YOU SAY THAT	5	0.06
555	THAT YOU HAVE	7	0.08	1113	YOU SEE A	5	0.06
556	THAT'S WHY I	7	0.08	1114	YOU WOULD HAVE	5	0.06
557	THE BEGINNING I	7	0.08	1115	YOU YOU YOU	5	0.06
558	THE LAST PICTURE	7	0.08				

Table A3: List of all 4-grams in LOCNEC with a minimum frequency of 5

No	4-gram	Freq.	ptw	No	4-gram	Freq.	ptw
1	I DON'T KNOW I	60	0.51	152	I WAS THERE FOR	6	0.05
2	IT WAS IT WAS	45	0.38	153	I WENT BACK TO	6	0.05
3	AND THINGS LIKE THAT	33	0.28	154	I WORKED IN A	6	0.05
4	YEAH YEAH YEAH YEAH	30	0.25	155	I'D LIKE TO DO	6	0.05
5	ERM I DON'T KNOW	27	0.23	156	IN THE FIRST PICTURE	6	0.05
6	AT THE I END OF	22	0.19	157	IT COULD HAVE BEEN	6	0.05
7	THE I END OF THE	21	0.18	158	IT WAS A REALLY	6	0.05
8	I END OF THE	21	0.18	159	IT WAS JUST A	6	0.05
9	A BIT OF A	20	0.17	160	IT WAS VERY INTERESTING	6	0.05
10	YEAH THAT'S RIGHT YEAH	20	0.17	161	KNOW I DON'T KNOW	6	0.05
11	I THINK IT WAS	19	0.16	162	LIKE TO GO BACK	6	0.05
12	I'D LIKE TO GO	19	0.16	163	LOOK AT IT AND	6	0.05
13	AND STUFF LIKE THAT	18	0.15	164	LOT OF THE TIME	6	0.05
14	I DON'T KNOW IF	17	0.14	165	NO I DON'T THINK	6	0.05
15	IT WAS REALLY GOOD	17	0.14	166	NO IT WAS JUST	6	0.05
16	OR SOMETHING LIKE THAT	17	0.14	167	NO NO I DON'T	6	0.05
17	I THINK I THINK	15	0.13	168	SHE DOESN'T LIKE THE	6	0.05
18	A LOT OF PEOPLE	14	0.12	169	SHE'S NOT VERY HAPPY	6	0.05
19	I DON'T I DON'T	14	0.12	170	SO THAT WAS QUITE	6	0.05
20	I THOUGHT IT WAS	14	0.12	171	SO YEAH I MEAN	6	0.05
21	YEAH YEAH IT WAS	14	0.12	172	SORT OF YOU KNOW	6	0.05
22	AND IT WAS REALLY	13	0.11	173	THE I END OF IT	6	0.05
23	BUT I DON'T KNOW	13	0.11	174	THINGS LIKE THAT YEAH	6	0.05
24	YEAH YEAH THAT'S RIGHT	13	0.11	175	TO SEE IT AND	6	0.05
25	AND IT WAS JUST	12	0.10	176	WAS IT WAS IT	6	0.05
26	AND IT WAS LIKE	12	0.10	177	WHEN I CAME HERE	6	0.05
27	I WANT TO DO	12	0.10	178	WHEN I WAS IN	6	0.05
28	IN THE MIDDLE OF	12	0.10	179	WHEN I WAS THERE	6	0.05
29	THAT KIND OF THING	12	0.10	180	YEAH IT WAS REALLY	6	0.05
30	THAT WAS THAT WAS	12	0.10	181	YEAH IT WAS VERY	6	0.05
31	THAT'S RIGHT YEAH YEAH	12	0.10	182	YOU HAVE TO BE	6	0.05
32	YEAH I DON'T KNOW	12	0.10	183	YOU HAVE TO DO	6	0.05
33	YOU KNOW WHAT I	12	0.10	184	YOU HAVE TO YOU	6	0.05
34	DON'T KNOW I DON'T	11	0.09	185	YOU KNOW IT'S IT'S	6	0.05
35	I WAS GOING TO	11	0.09	186	YOU KNOW SORT OF	6	0.05
36	I WENT TO SEE	11	0.09	187	YOU KNOW YOU CAN	6	0.05
37	IT WAS A BIT	11	0.09	188	A BIT SORT OF	5	0.04
38	IT WAS JUST LIKE	11	0.09	189	A COUPLE OF YEARS	5	0.04
39	KNOW WHAT I MEAN	11	0.09	190	A LOT OF WORK	5	0.04
40	THERE'S A LOT OF	11	0.09	191	A SORT OF A	5	0.04
41	YEAH YEAH YEAH I	11	0.09	192	ALL HER FRIENDS AND	5	0.04
42	YOU KNOW IT WAS	11	0.09	193	AND I DON'T KNOW	5	0.04
43	A LOT OF THE	10	0.08	194	AND I HAD TO	5	0.04
44	AND YOU HAVE TO	10	0.08	195	AND I THOUGHT OH	5	0.04
45	DON'T KNOW I THINK	10	0.08	196	AND I THOUGHT WELL	5	0.04
46	I WAS I WAS	10	0.08	197	AND I WAS LIKE	5	0.04
47	SHE DOESN'T LIKE IT	10	0.08	198	AND IT WAS IT	5	0.04
48	STUFF LIKE THAT AND	10	0.08	199	AND SORT OF LIKE	5	0.04

49	THAT SORT OF THING	10	0.08	200	AND THAT KIND OF	5	0.04
50	TO ALL HER FRIENDS	10	0.08	201	AND THAT SORT OF	5	0.04
51	TO GO AND SEE	10	0.08	202	AS WELL I THINK	5	0.04
52	AND THEN I WENT	9	0.08	203	AS WELL SO I	5	0.04
53	AS A FOREIGN LANGUAGE	9	0.08	204	AT THE PICTURE AND	5	0.04
54	I DON'T KNOW IT	9	0.08	205	BECAUSE YOU HAVE TO	5	0.04
55	I DON'T KNOW WHAT	9	0.08	206	BUT I DON'T THINK	5	0.04
56	I DON'T REALLY KNOW	9	0.08	207	BUT I'D LIKE TO	5	0.04
57	I DON'T WANT TO	9	0.08	208	DO A LOT OF	5	0.04
58	I THOUGHT THAT WAS	9	0.08	209	DON'T I DON'T KNOW	5	0.04
59	I WANTED TO DO	9	0.08	210	END OF THE YEAR	5	0.04
60	IN THE FIRST YEAR	9	0.08	211	ERM I WENT TO	5	0.04
61	IT WOULD HAVE BEEN	9	0.08	212	ERM I'D LIKE TO	5	0.04
62	IT'S NOT TOO BAD	9	0.08	213	GO AND SEE A	5	0.04
63	O'CLOCK IN THE MORNING	9	0.08	214	GOOD THING ABOUT IT	5	0.04
64	QUITE A LOT OF	9	0.08	215	HAPPY WITH IT SO	5	0.04
65	SO I DON'T KNOW	9	0.08	216	HAVE A LOOK AT	5	0.04
66	SO IT WAS QUITE	9	0.08	217	HAVE TO DO A	5	0.04
67	THINGS LIKE THAT AND	9	0.08	218	HAVE TO GO TO	5	0.04
68	YEAH YEAH I MEAN	9	0.08	219	I CAME TO UNIVERSITY	5	0.04
69	A PICTURE OF A	8	0.07	220	I CAN'T REMEMBER WHAT	5	0.04
70	ALL OVER THE PLACE	8	0.07	221	I DIDN'T KNOW WHAT	5	0.04
71	ALL THE TIME YEAH	8	0.07	222	I DIDN'T REALLY KNOW	5	0.04
72	AN AWFUL LOT OF	8	0.07	223	I DON'T KNOW HOW	5	0.04
73	ENGLISH AS A FOREIGN	8	0.07	224	I DON'T KNOW WHY	5	0.04
74	I DON'T KNOW YEAH	8	0.07	225	I DON'T LIKE IT	5	0.04
75	I HAD TO GO	8	0.07	226	I JUST WANTED TO	5	0.04
76	I'D LIKE TO TEACH	8	0.07	227	I KNOW I KNOW	5	0.04
77	IT WAS A LOT	8	0.07	228	I MEAN IT'S NOT	5	0.04
78	MOST OF THE TIME	8	0.07	229	I MEAN THAT WAS	5	0.04
79	PAINTING A PICTURE OF	8	0.07	230	I MEAN YOU YOU	5	0.04
80	SO IT WAS A	8	0.07	231	I REALLY ENJOYED IT	5	0.04
81	THINGS LIKE THAT SO	8	0.07	232	I REALLY WANTED TO	5	0.04
82	TO BE ABLE TO	8	0.07	233	I THINK IT'S JUST	5	0.04
83	WE WENT TO THE	8	0.07	234	I THINK THAT WAS	5	0.04
84	WHAT I WANT TO	8	0.07	235	I WAS A BIT	5	0.04
85	WHEN I WENT TO	8	0.07	236	I WAS INTERESTED IN	5	0.04
86	YEAH IT WAS IT	8	0.07	237	I WAS ON CAMPUS	5	0.04
87	A REALLY GOOD FILM	7	0.06	238	I WOULD LIKE TO	5	0.04
88	ALL THE TIME AND	7	0.06	239	IF I WANTED TO	5	0.04
89	AND A LOT OF	7	0.06	240	IF YOU WANT TO	5	0.04
90	AND I WAS JUST	7	0.06	241	IN THE CENTRE OF	5	0.04
91	AND SHE DOESN'T LIKE	7	0.06	242	IN THE MORNING AND	5	0.04
92	AND YOU KNOW I	7	0.06	243	IN THE SUMMER AND	5	0.04
93	AT THE SAME TIME	7	0.06	244	IT WAS SORT OF	5	0.04
94	BUT I MEAN I	7	0.06	245	IT WAS YOU KNOW	5	0.04
95	FOR A COUPLE OF	7	0.06	246	IT'S A BIT OF	5	0.04
96	G C S E	7	0.06	247	IT'S A LONG WAY	5	0.04
97	I CAN'T I CAN'T	7	0.06	248	I'VE ALWAYS WANTED TO	5	0.04
98	I DIDN'T WANT TO	7	0.06	249	LIKE THAT I DON'T	5	0.04
99	I DON'T KNOW ERM	7	0.06	250	LIKE THAT YEAH YEAH	5	0.04
100	I DON'T THINK I	7	0.06	251	LIKE TO GO TO	5	0.04
101	I DON'T THINK SO	7	0.06	252	LIVING IN A HOUSE	5	0.04
102	I MEAN I WAS	7	0.06	253	LOOKS AT THE PICTURE	5	0.04

103	I THINK IT'S A	7	0.06	254	NO I MEAN I	5	0.04
104	IN MY FIRST YEAR	7	0.06	255	NO NO NOT REALLY	5	0.04
105	IN THE SECOND YEAR	7	0.06	256	NO NOT AT ALL	5	0.04
106	IT IS IT IS	7	0.06	257	ON CAMPUS LAST YEAR	5	0.04
107	IT WAS LIKE A	7	0.06	258	ONE OF MY FRIENDS	5	0.04
108	IT WAS QUITE A	7	0.06	259	ONE OF THE BEST	5	0.04
109	IT WAS QUITE GOOD	7	0.06	260	PICTURE OF A WOMAN	5	0.04
110	SO I HAD TO	7	0.06	261	SO I'D LIKE TO	5	0.04
111	SORT OF THING AND	7	0.06	262	SO THAT WAS A	5	0.04
112	SORT OF THING SO	7	0.06	263	SOME OF THEM ARE	5	0.04
113	TEACHING ENGLISH AS A	7	0.06	264	SOMETHING LIKE THAT YEAH	5	0.04
114	THERE WAS THERE WAS	7	0.06	265	SORT OF LIKE I	5	0.04
115	WELL I DON'T KNOW	7	0.06	266	SORT OF THING BUT	5	0.04
116	WHAT I WANTED TO	7	0.06	267	THAT I DON'T LIKE	5	0.04
117	YEAH I MEAN I	7	0.06	268	THAT YEAH YEAH YEAH	5	0.04
118	YEAH I'D LIKE TO	7	0.06	269	THE MIDDLE OF THE	5	0.04
119	YEAH YEAH I DON'T	7	0.06	270	THINGS LIKE THAT ERM	5	0.04
120	YEAH YEAH YEAH THAT'S	7	0.06	271	THINK I THINK I	5	0.04
121	YOU HAVE TO GO	7	0.06	272	TO GO TO A	5	0.04
122	YOU KNOW IT'S NOT	7	0.06	273	TO GO TO THE	5	0.04
123	YOU KNOW YOU KNOW	7	0.06	274	TWO AND A HALF	5	0.04
124	A COUPLE OF WEEKS	6	0.05	275	WAS A REALLY GOOD	5	0.04
125	A GOOD FILM AND	6	0.05	276	WAS IT WAS A	5	0.04
126	A LOOK AT IT	6	0.05	277	WAS IT WAS REALLY	5	0.04
127	AND A HALF HOURS	6	0.05	278	WAS THE FIRST TIME	5	0.04
128	AND ER IT WAS	6	0.05	279	WENT TO SEE IT	5	0.04
129	AND THEN AND THEN	6	0.05	280	WHEN I WHEN I	5	0.04
130	AND THEN YOU GET	6	0.05	281	WHICH IS A BIT	5	0.04
131	AND THERE WAS A	6	0.05	282	WHILE I WAS THERE	5	0.04
132	AND WE WENT TO	6	0.05	283	YEAH I KNOW I	5	0.04
133	AS A SORT OF	6	0.05	284	YEAH I THINK I	5	0.04
134	DON'T KNOW I MEAN	6	0.05	285	YEAH I THINK IT	5	0.04
135	END OF THE DAY	6	0.05	286	YEAH I THINK IT'S	5	0.04
136	ER I DON'T KNOW	6	0.05	287	YEAH IT IS YEAH	5	0.04
137	GO AND SEE IT	6	0.05	288	YEAH YEAH ERM I	5	0.04
138	HAD A LOT OF	6	0.05	289	YEAH YEAH I I	5	0.04
139	I DON'T KNOW I'M	6	0.05	290	YEAH YEAH I THINK	5	0.04
140	I DON'T KNOW IT'S	6	0.05	291	YEAH YEAH IT IS	5	0.04
141	I END OF IT	6	0.05	292	YEAH YEAH THAT'S IT	5	0.04
142	I HAD TO DO	6	0.05	293	YEAH YEAH YEAH AND	5	0.04
143	I I DON'T THINK	6	0.05	294	YEAH YEAH YEAH ERM	5	0.04
144	I MEAN I DON'T	6	0.05	295	YEAH YEAH YEAH IT'S	5	0.04
145	I MEAN I I	6	0.05	296	YEAH YEAH YEAH OH	5	0.04
146	I MEAN IT'S A	6	0.05	297	YOU HAD TO DO	5	0.04
147	I MEAN SORT OF	6	0.05	298	YOU KNOW BUT I	5	0.04
148	I THINK SO YEAH	6	0.05	299	YOU KNOW I THINK	5	0.04
149	I WANTED TO BE	6	0.05	300	YOU KNOW IT'S JUST	5	0.04
150	I WAS LIKE OH	6	0.05	301	YOU KNOW THEY WERE	5	0.04
151	I WAS SORT OF	6	0.05				

Table A4: List of all 4-grams in LINDSEI-GE with a minimum frequency of 5

No.	Word	Freq.	ptw	No.	Word	Fre q.	ptw
1	ERM I DON'T KNOW	32	0.37	86	AND THEN YOU HAVE	6	0.07
2	I DON'T KNOW IF	29	0.34	87	AT THE PICTURE AND	6	0.07
3	I WOULD LIKE TO	27	0.31	88	DON'T KNOW IF I	6	0.07
4	IN THE I END	23	0.27	89	ERM HOW DO YOU	6	0.07
5	I DON'T KNOW I	22	0.26	90	ERM IN THE I	6	0.07
6	YEAH YEAH YEAH YEAH	19	0.22	91	I DON'T KNOW BUT	6	0.07
7	AND I DON'T KNOW	15	0.17	92	I DON'T KNOW WHETHER	6	0.07
8	I THINK IT WAS	14	0.16	93	I DON'T KNOW YET	6	0.07
9	ON THE I OTHER	14	0.16	94	I I THINK I	6	0.07
10	OR SOMETHING LIKE THAT	14	0.16	95	I REALLY LIKED THE	6	0.07
11	I DON'T KNOW THE	13	0.15	96	I THOUGHT IT WAS	6	0.07
12	IT TO HER FRIENDS	13	0.15	97	IN THE BEGINNING I	6	0.07
13	THE I OTHER HAND	13	0.15	98	IN THE I AFTERNOON	6	0.07
14	ER I DON'T KNOW	12	0.14	99	IN THE SECOND PICTURE	6	0.07
15	I DON'T KNOW IT'S	12	0.14	100	IS A PAINTER AND	6	0.07
16	I DON'T KNOW WHAT	12	0.14	101	LIKE TO TALK ABOUT	6	0.07
17	THE I THE I	12	0.14	102	MHM # SEC BREAK	6	0.07
18	TO HER FRIENDS AND	12	0.14	103	MOST OF THE TIME	6	0.07
19	AND IN THE I	11	0.13	104	ON THE ON THE	6	0.07
20	I'M NOT QUITE SURE	11	0.13	105	ON THE ONE HAND	6	0.07
21	SHE SHOWS IT TO	11	0.13	106	SHE LOOKS A BIT	6	0.07
22	IT WAS IT WAS	10	0.12	107	SHE SEEMS TO BE	6	0.07
23	SHOWS IT TO HER	10	0.12	108	SO I HAD TO	6	0.07
24	SO I DON'T KNOW	10	0.12	109	THERE WERE SO MANY	6	0.07
25	THE I OTHER ONE	10	0.12	110	TO TELL YOU ABOUT	6	0.07
26	WOULD LIKE TO GO	10	0.12	111	VERY NICE AND ERM	6	0.07
27	YEAH I DON'T KNOW	10	0.12	112	WHEN I WAS IN	6	0.07
28	# SEC BREAK OKAY	9	0.10	113	YEAH YEAH YEAH I	6	0.07
29	A LITTLE BIT MORE	9	0.10	114	A BIT OF A	5	0.06
30	BUT ON THE I	9	0.10	115	A LOT ABOUT IT	5	0.06
31	I DON'T KNOW MAYBE	9	0.10	116	A PORTRAIT OF A	5	0.06
32	IN THE UNITED STATES	9	0.10	117	AND A HALF YEARS	5	0.06
33	WHEN I WHEN I	9	0.10	118	AND ERM I DON'T	5	0.06
34	AND ERM YEAH I	8	0.09	119	AND ERM I WAS	5	0.06
35	AND I THINK THAT	8	0.09	120	AND I THINK THAT'S	5	0.06
36	AND IT WAS REALLY	8	0.09	121	AND I WAS LIKE	5	0.06
37	AND YOU HAVE TO	8	0.09	122	AND IT WAS QUITE	5	0.06
38	ERM YEAH I THINK	8	0.09	123	AND SHE IS NOT	5	0.06
39	I DON'T KNOW ERM	8	0.09	124	AND YEAH IT WAS	5	0.06
40	I DON'T REALLY KNOW	8	0.09	125	AT THE I AGE	5	0.06
41	I THINK IT'S A	8	0.09	126	DON'T KNOW IT WAS	5	0.06
42	I WANT TO GO	8	0.09	127	DON'T KNOW THE I	5	0.06
43	IN A IN A	8	0.09	128	ERM AFTER A WHILE	5	0.06
44	IN THE I EVENING	8	0.09	129	ERM I THINK I	5	0.06
45	IN THE IN THE	8	0.09	130	ERM I WOULD LIKE	5	0.06
46	IT WAS REALLY NICE	8	0.09	131	ERM WELL I THINK	5	0.06
47	LIKE I DON'T KNOW	8	0.09	132	EXPRESSION ON HER FACE	5	0.06
48	MM I DON'T KNOW	8	0.09	133	FOR A COUPLE OF	5	0.06
49	SHE WANTS TO BE	8	0.09	134	HOW DO YOU SAY	5	0.06
50	SHOWS THE PICTURE TO	8	0.09	135	I DON'T I DON'T	5	0.06
51	THE PICTURE AND SHE	8	0.09	136	I DON'T KNOW THEY	5	0.06
52	WANT TO GO TO	8	0.09	137	I DON'T KNOW YEAH	5	0.06
53	AND ERM I I	7	0.08	138	I HAVE TO ADMIT	5	0.06

54	AND I REALLY LIKED	7 0.08	139	I HAVE TO SAY	5 0.06
55	AND SO ON AND	7 0.08	140	I I THINK IT'S	5 0.06
56	AT THE I END	7 0.08	141	I THINK SO I	5 0.06
57	DON'T KNOW HOW TO	7 0.08	142	I THOUGHT THAT WAS	5 0.06
58	DON'T KNOW IF YOU	7 0.08	143	I WAS IN THE	5 0.06
59	ERM I THINK IT	7 0.08	144	I WAS LIKE OH	5 0.06
60	ERM I THINK IT'S	7 0.08	145	IF YOU GO TO	5 0.06
61	ERM WHEN I WAS	7 0.08	146	IN THE LAST PICTURE	5 0.06
62	I DON'T KNOW HOW	7 0.08	147	IT WAS A GOOD	5 0.06
63	I DON'T KNOW IT	7 0.08	148	IT WAS A REALLY	5 0.06
64	I DON'T KNOW REALLY	7 0.08	149	IT WAS ERM THE	5 0.06
65	I DON'T KNOW WHY	7 0.08	150	IT WAS KIND OF	5 0.06
66	I HAD TO GO	7 0.08	151	IT WAS NOT REALLY	5 0.06
67	I I DON'T KNOW	7 0.08	152	LIKE THAT AND ERM	5 0.06
68	I REALLY LIKED IT	7 0.08	153	LIKE TO GO TO	5 0.06
69	I THINK I WOULD	7 0.08	154	NO NOT AT ALL	5 0.06
70	IF YOU IF YOU	7 0.08	155	ON THE SECOND PICTURE	5 0.06
71	IN THE THIRD PICTURE	7 0.08	156	SEC BREAK OKAY ERM	5 0.06
72	OKAY # SEC BREAK	7 0.08	157	SHE DIDN'T LIKE IT	5 0.06
73	THE I AGE OF	7 0.08	158	SHE IS NOT SATISFIED	5 0.06
74	THE I END OF	7 0.08	159	SO HE HAS TO	5 0.06
75	THE I ERM THE	7 0.08	160	THE PICTURE TO HER	5 0.06
76	TO BECOME A TEACHER	7 0.08	161	THEN YOU HAVE TO	5 0.06
77	TO GO TO THE	7 0.08	162	THERE IS A PAINTER	5 0.06
78	WAS REALLY NICE AND	7 0.08	163	THINK IT'S A GOOD	5 0.06
79	# SEC BREAK WELL	6 0.07	164	TO BE A TEACHER	5 0.06
80	A COUPLE OF YEARS	6 0.07	165	WHEN I WAS THERE	5 0.06
81	A FRIEND OF MINE	6 0.07	166	YEAH I THINK IT'S	5 0.06
82	A LOT OF THINGS	6 0.07	167	YEAH YEAH I THINK	5 0.06
83	A LOT OF TIME	6 0.07	168	YOU DON'T HAVE TO	5 0.06
84	ALL OVER THE WORLD	6 0.07	169	YOU HAVE TO BE	5 0.06
85	AND ERM IT WAS	6 0.07			

Appendix A3:
Final Ethics Approval

Pages 269-270 removed from Open Access version as they may contain sensitive/confidential content.

