

TABLE OF CONTENTS

Creation of a Corpus for Studying Properties of Child Speech Related to Iconicity, Frequency, Length, Orthographic Neighborhood, and Concreteness – a Technical Report with some Descriptive Statistics

Velina Slavova

1. Introduction - The Aim of the Corpus Creation	203
2. Data sources	204
3. Data-Reorganization of the Child Speech Corpus.....	207
4. Adjusting of the Word-Strings from the Child Speech Corpus.....	209
5. Data Fusion for Annotation of the Child Speech with Iconicity	224
6. The POS-problem: Some Descriptive Statistics of the Data sources	227
7. POS Transition, Mapping and Filtering	231
8. The obtained corpus	239
9. Closing Remarks.....	248
Acknowledgment.....	249
References.....	249
Authors' Information	250
Appendix A. Data from CHILDES – list of the used corpora	251
Appendix B. Example - Dialogue retrieved from CHILDES, tagged in the Child Speech Corpus.....	252
Appendix C. Word adjusting of the Child speech corpus to match the Iconicity Table	258
Appendix D. Occurrences of Normal Forms of Verbs which Match verbs in the Iconicity Table	276
Appendix E. Occurrences of 3S conjugations left to Match verbs in the Iconicity Table	281
Appendix F. Protective changes of present participles.....	284
Appendix G. Protective Changes of Past Tense Forms	288
Appendix I. Past tense forms of Regular verbs left to match the normal form in the Iconicity Table	289
Appendix J. Past tense forms of Irregular verbs protected from matching the Iconicity Table	291
Appendix K. Irregular verbs mistaken by children - with examples of use.....	292
Appendix L. Protective Changes of Past Participles	293
Appendix M. POS Transition Markers	294
Appendix N. The 100 most frequently used words in the children’s speech with their Iconicity	298
Table of Contents.....	300