ITHEA

International Journal INFORMATION MODELS & ANALYSES

2018 Volume 7 Number 3

International Journal
INFORMATION MODELS AND ANALYSES
Volume 7 / 2018, Number 3

IJ IMA is official publisher of the scientific papers of the members of the ITHEA® International Scientific Society

IJ IMA rules for preparing the manuscripts are compulsory.
The **rules for the papers** for ITHEA International Journals are given on www.ithea.org .
The camera-ready copy of the paper should be received by ITHEA® Submission system http://ij.ithea.org .
Responsibility for papers published in IJ IMA belongs to authors.

International Journal "INFORMATION MODELS AND ANALYSES" Volume 7, Number 3, 2018

Edited by the **Institute of Information Theories and Applications FOI ITHEA**, Bulgaria, in collaboration with:
University of Telecommunications and Posts, Bulgaria,
V.M.Glushkov Institute of Cybernetics of NAS, Ukraine,
Universidad Politécnica de Madrid, Spain,
Hasselt University, Belgium,
University of Perugia, Italy,
Institute for Informatics and Automation Problems, NAS of the Republic of Armenia
St. Petersburg Institute of Informatics, RAS, Russia,

Publisher: **ITHEA®** Sofia, 1000, P.O.B. 775, Bulgaria. www.ithea.org , e-mail: office@ithea.org
Technical editor: **Ina Markova**

Printed in Bulgaria

ISSN 1314-6416 (printed) **ISSN 1314-6432 (Online)**

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

***Abstract**: An extensive corpus of child speech in English, annotated with words' Iconicity, Concreteness, Orthographic neighborhood, Length, and Frequency was created. About 309 000 English words were extracted from free dialogues of children aged between 9 and 62 months, and stored in a relational database. The corpus was annotated using recently published data. The general steps performed to get the corpus annotated and the solutions applied are explained. The problems which arose to obtain the corpus, as well as the solutions applied are explained. The steps of the corpus creation are illustrated with several examples based on querying the child speech corpus. The properties of the children's speech with regard to the annotated novel characteristics are illustrated with several plots. Some initial statistical results are discussed.*

***Keywords:** I.2.7 Natural Language Processing, H.2.8 Database Applications*

***ITHEA Keywords**: H.4 Information systems applications*

1. Introduction - The Aim of the Corpus Creation

The efforts in many fields such as brain imaging, cognitive experimentation, psychology, linguistics, and so forth, have tried to explain the mechanisms and the principles which lead to establishing of mental representation of the world. However, these mechanisms are still not clear. One way to discover a bit more concerning these problems is to investigate language acquisition. Important parameters for studying the process of language acquisition are words' iconicity, level of abstraction, frequency of use in the language environment, common lemmas and length. They can clarify the impact of factors like perception, phonology, abstract thinking, and memory on the concept formation. That motivated the creation of a huge corpus of child speech annotated with these additional features.

The purpose of the work presented here is to obtain a corpus of child speech in English annotated in a reliable manner with the parameters of Iconicity, Length, Frequency, Orthographic neighborhood and

Concreteness of the used words. The present short announcement gives some aspects of the procedures performed in order to create such a corpus.

The interest of creating a corpus of annotated child speech was first discussed with specialists in linguistics and psycholinguistics from the Language Processing Lab at the University of Calgary, Canada. The intensive exchange which followed led step by step to the creation of the corpus, presented here.

The stages of attaining the corpus annotated are explained further. Several data-examples are provided. The metadata obtained and used for the annotating procedure as well as some statistical results which can be useful in the fields of psycholinguistics, child language acquisition and other fields related to language, cognition, and psychology, are given as appendices.

The created corpus is based on data taken from different sources. The used data sources are three: a Corpus of child speech in English [Slavova, 2016], the data on Iconicity, Length, Orthographic Neighborhood and Frequency [Winter et al. 2016] and data on Concreteness [Brysbaert et al., 2016]. These three sources are shortly described in the next section.

2. Data sources

2.1. Dialogues from CHILDES Data Repository (the Child Speech Corpus)

Data from 30 corpora containing 630 free dialogues with child speech in English (Appendix A) were extracted from CHILDES (Child Language Data Exchange System) and used as the main source for the corpus presented here. These child speech dialogues, as annotated with part of speech and grammar in the source (see [MacWhinney, 2014]) were used to compose the child speech utterances. The transcription of POS in CHILDES is performed by the authors of the corresponding corpora using CLAN (Computerized Language ANalysis), a computerized system designed specifically for the Exchange system's standardized format. The transcripts include for each speech utterance a separate line marked with *"%mor*". This line contains the system's standardized symbols for the parts of speech (POS), based on Hausser's MORPH system ([Hausser, 1989]).

The transcripts, with the entire linguistic annotation, were first stored locally and extracted from CLAN format into a relational database (DB) where each dialogue and each utterance is tagged with a unique identifier [Slavova, 2016]. An example of a stored in the database dialogue is provided in Appendix B. The used further corpus contains 629 dialogues of children (62 girls, 66 boys, and 7 with unspecified gender) of age between 9 months and 62 months. Some children are “recorded” during several successive months and some are not. The dialogues contain in average 520 utterances of the different participants, with 202 child utterances. This data source is further called ***Child Speech Corpus*.** Table 1 contains examples of the obtained 125, 353 children's utterances as organized in a 3NF database table "*Child Speech POS and Syntax*".

Table 1. The content of the Child Speech Corpus – examples.

Months	Module Name	Sentence utterance ID	Speech	POS (the line *%mor* from the source CHILDES)
28	zek28.st	zek28.sto041	she build house .	pro:sub|she v|build n|house .
28	zek28.st	zek28.sto172	snowman !	n|+n|snow+n|man !
28	zek28.st	zek28.sto181	uhhuh .	co|uhhuh .
28	zek28.st	zek28.sto188	a hat .	det|a n|hat .
28	zek28.st	zek28.sto303	what matter with Timmy ?	adv:int|what v|matter prep|with n:prop|Timmy ?
28	zek28.st	zek28.sto408	that's all !	pro:dem|that cop|be&3S pro:indef|all !
28	she28.po	she28.pos732	I got mail .	pro:sub|I v|get&PAST n|mail .
28	she28.po	she28.pos737	where people ?	adv:wh|where n|person&PL ?
28	zek28.st	zek28.sto023	yyy .	

2.2. Data on Iconicity, Length, Frequency and Orthographic Neighborhood (The Iconicity Table)

Iconicity, as given in the data source that is used here, is defined as quantified iconicity based on native speaker ratings. This second data source contains a statistical result based on the subjects' evaluation of the iconicity of English words.

The sources originate from recently published studies by Winter and colleagues [Winter et al. 2017] which replicates and provides an expanded set of words to the results of Perry and colleagues [Perry et al. (2015)]. In these works, the *iconicity norms* were assessed to English words using ratings for iconicity, assigned by native speakers of English following a procedure proposed by Perry and colleagues. To collect the ratings, large numbers of native English speakers were recruited. After an appropriate instruction, the participants were asked to rate each *word* on a scale from −5 ("words that sound like the opposite of what they mean") to +5 ("words that sound like what they mean"). This procedure has provided iconicity norms for 3001 English words.

The data was further organized in a separate DB table, which contains the words, their POS, their Length, their Frequency, and their Orthographic neighborhood. This source is further called here ***Iconicity Table.*** Examples of the content of the Iconicity Table are given in Table 2.

The data used here on the word's Frequency, taken as in the work of Winter and colleagues, is from the American SUBTLEX subtitle corpus of movie speech, as presented in the English Lexicon Project [Balota et al., 2007]. The POS labels were taken from Brysbaert and Keuleers' (2012) annotation of the SUBTLEX corpus. It should be mentioned that the POS accorded to each word in this data-source is based on the most frequent use of the word as a given POS in SUBTLEX.

The orthographic neighborhood is a parameter analogical to distance between strings (e.g. Levenshtein distance) which expresses in a specific way the number and the length of the words that can be obtained by changing letters while preserving the identity and positions of the others (for more details see [Balota et al. 2007]). In general, this parameter provides a measure for the number of lexical items which are phonologically similar to the target.

Table 2. The data source "Iconicity Table" - examples

Iconicity Table

Word_Winter_et	No	Iconicity	Brysetal_POS	Length	Frequency (LgSUBTLWF)	OrthoNeigberho
a	1	0,461538462	functionword	1	6,02	1,45
abide	1	0,25	verb	5	2,14	1,7
able	1	0,466666667	adjective	4	3,91	1,5
about	1	-0,1	functionword	5	5,27	1,85
above	1	1,0625	functionword	5	3,4	1,95
abrasive	1	1,3125	adjective	8	1,38	2,9
absorbent	1	0,923076923	adjective	9	0,95	3,1
academy	1	0,692307692	noun	7	2,8	3,15
accident	1	1,363636364	noun	8	3,62	2,7
accordion	1	-0,454545455	noun	9	1,83	3,6
account	1	-0,428571429	noun	7	3,36	2,6
ache	1	1,153846154	noun	4	2,11	1,65
aching	1	0,25	verb	6	2,03	1,9
acid	1	0,75	noun	4	2,71	1,7
acidic	1	1	adjective	6	1,08	2,7
acrid	1	0,615384615	adjective	5	0,48	1,9

2.3. Data on Concreteness (The Concreteness Table)

The level of abstraction of the used words, important for studying the involvement of abstract concepts in the children's speech, is represented here by the parameter Concreteness. The data originates from a study by Brysbaert and colleagues [Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014)] which provides a measure for words' Concreteness, accorded using subjects' evaluation of 37,058 generally known English word lemmas. This annotated word-set is a subset of a comprehensive list of English lemmas and contains all lemmas known by at least 85% of the raters.

The data used here was also preliminarily organized and provided by David Sidhu for the annotation-procedure. The data-source was further organized in a separate DB table which contains the words and their Concreteness (the Mean, obtained after the subjects-evaluation results). This data-source is further called here ***Concreteness Table.*** Examples of the content of the Concreteness Table are given in Table 3.

When querying these sources, became clear that 2822 words from the Concreteness Table correspond to words in the Iconicity Table. It has to be noted that the words in the Concreteness Table are not annotated for POS. For the treatment described here, the POS of the coinciding words are taken as accorded in the Iconicity Table.

Table 3. Example of the content of the Concreteness Table.

Word (CONCR)	Concreteness Mean	Concetereness StDev	Percent_known	SUBTLEX
deport	3,41	1,31	1	40
deportable	2,38	1,28	0,888888888888889	0
deportation	2,42	1,33	0,962962962962963	40
deportee	3,79	1,21	0,966666666666667	0
deportment	2,67	1,44	0,964285714285714	12
deposable	2,24	1,3	0,862068965517241	0
depose	2,52	1,12	0,958333333333333	21
deposer	2,63	1,42	0,9	1
deposit	3,92	1,23	1	551

The three data-sources used for the creation of the corpus originate from recently published results. The list of references for the used sources is given in Table 4.

Table 4. Main references for the used data sources

Data Source	Reference
Child Speech Corpus	Slavova (2016) based on CHILDES (Dialogues)
Iconicity Norms	Winter et al. (2017) enlarging data from Perry et al. (2015)
Word Frequency	Balota et al. (2007)
Orthographic Neighborhood	Balota et al. (2007)
Concreteness	Brysbaert et al. (2014)
Part-of-speech labels	Winter et al (2017) based on Brysbaert & Keuleers (2012)

The used data sources are different at the level of data organization:

1. The Child Speech Corpus is organized in sentence speech-utterances (see Table 1.)
2. The Iconicity and the Concreteness Tables are organized in words (Tables 2. and 3.)

This has necessitated a data reorganization of the Child speech corpus, described in the next section.

3. Data-Reorganization of the Child Speech Corpus

The corpus of child speech has been created and used for studying gender differences in the use of nouns and proper nouns of different semantic categories [Slavova V, Atanasov D. & Andonov F. (2016)]. For this purpose, the content of the field “POS” (that is – the *%mor* line from CHILDES) of the speech utterances, as annotated by linguists in the original source, was parsed. It contains both – the words of the child and the POS, as accorded by the authors of the corpora in CHILDES.

The obtained after the parsing words-strings were organized in separate DB tables following their POS, that is – a distinct table was defined to stock each word depending on its part of speech. The stored records were tagged with markers carrying information about the sentence-utterance identifier (leading to the Dialogue to which the sentence belongs) and the position on which the word-string starts. This was done in two technical ways in order to assure the reliability of the result. The obtained relational database has the structure shown in Fig. 1:

This previously obtained structure of the corpus was further used here for organizing the available data into words. That is, each detected by the parser word-string from the children's dialogues and annotated with POS was organized in a separate line in a novel DB Table. This was done when:

1. keeping the unique marker of the sentence-utterance in order to retain the relation to the dialogue (and the child) in which the word-string appears.
2. keeping the original POS annotation, such as given in the data source CHILDES.

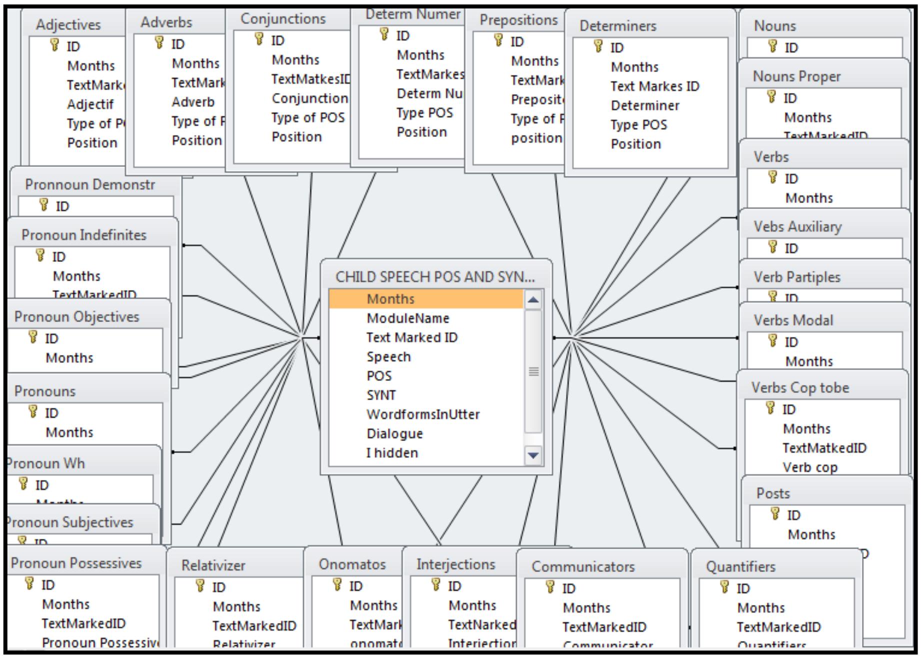


Figure 1. Structure of the database obtained in previous work [Slavova V. et al, 2016]

The novel table, named "*All Word-strings Table*", contains all the extracted by the parsing procedure word-strings and the corresponding POS. The table contained **309 117** word-strings isolated as words used by children. The dialogue, sentence and position markers were kept in order to allow reconstructing the dialogue. Example of two reconstructed sentence-utterances as detected by the parser and stored in the obtained "All Word-strings Table" is given in Table 5.

It is seen from the examples in Table 5. that the parsing of the *%mor*-string of the source CHILDES did not provide a completely correct extraction (or reconstruction) of words. This is due mainly to the fact that there are several additional markers concatenated to the end of the word in the %*mor*-string, while the parser is based on detecting the indicators for POS which are at the beginning of the word, such as *pro:sub*|, *n:prop*|, *n*|, *v*|, *det*| etc. (see the examples in Table 1. and Appendix B).

Table 5. All Word-strings Table – example of the word-strings and POS-tags, extracted from two sentence-utterances

ID	TextMarkedID	Month	Word in CHILDES	CHILDES POS	Position
12474	ad138.bro2124	38	okay	co	0
6818	ad138.bro2127	38	I	pro:sub	0
18012	ad138.bro2127	38	get&PAST	v	9
1869	ad138.bro2127	38	some	qn	20
25019	ad138.bro2127	38	sunglass-PL	n	28
5704	ad138.bro2130	38	you	pro	0
18013	ad138.bro2130	38	want	v	7
18014	ad138.bro2130	38	fit&ZERO	v	21
2328	ad138.bro2130	38	them	pro:obj	32

The extracted this way word-string could not be mapped directly to the content of the Iconicity and the Concreteness Tables. This specificity was one of the reasons which led to an extended procedure of *adjusting* the extracted word-strings. This procedure is described in the next section.

4. Adjusting of the Word-Strings from the Child Speech Corpus

The aim of this step is to extract (and "reconstruct") **words** from the word-strings parsed from CHILDES in order to match them to the words in the Iconicity Table. This match concerns uniquely the words and does not take into consideration the part of speech accorded to each word (see Fig. 3.).

The word-string, originally detected by the parser, was kept unchanged in its field in order to retain the extracted grammatical markers used in CHILDES and to use them in the next steps of the procedure, as well as to correct errors, when necessary. An additional field "*Childes word adjusted*" for the adjusted words was created in the database table All Word-strings. The adjusted words were introduced in this additional field after several steps of data observation, analysis and querying, which are shortly described below.

4.1. Dealing with annotation specificities

Initially, to reconstruct the words which have not been properly extracted from their word-strings, the diverse concatenated grammatical markers for plural, verb conjugations and so on, such as -PL , &dv-AGT , ~aux, -PAST, &3S were removed. The elimination of the concatenated grammatical markers led to considerably more cases of word-match between the words in the child speech corpus and the other data sources. The example in Fig. 2. shows the transformation from the extracted word-string to an adjusted word which matches a word in the Iconicity Table:

Child Speech
banged her on the floor like this .
CHILDES POS ANNOTATION *(%mor)*
v|bang-PAST pro:obj|her prep|on det|the n|floor prep|like pro:dem|this .

Word-string extracted from CHILDES	CHILDES POS extracted	CHILDES WORD ADJUSTED		Matches the Word in the Iconicity Table	with POS
bang-PAST	v	bang	-->	bang	verb

Figure 2. Example - adjusting of the string* bang-PAST *to match a word in the Iconicity Table

The next transformation was performed when attentively examining the content of the Iconicity Table. Strings for compound nouns like, for example, "+n|grape+n|fruit" existing in CHILDES which can be fitted to words which exist in the Iconicity Table (*grapefruit* for this example*)*, had to be discovered. Other strings of the same type do not correspond to the content of the Iconicity Table or do match it in separate components. Such strings were not adjusted. Table 6. provides some examples. This step of the adjusting procedure has been performed with queries and manually when consulting the content of both data sources.

Table 6. Adjusting of specific composite strings for further match to the words in the Iconicity Table

the extracted string	adjusted in order to match the Iconicity Table:
+n|basket+n|ball	basketball
+n|butter+n|fly	butterfly
+n|finger+n|nail	fingernail
+n|foot+n|ball	football
+n|type+n|writer	typewriter
+adj|hot+n|dog	Should not be adjusted to *hot* and *dog*
+n|bath+n|room	Should not be adjusted to *bath* and *room*
+n|birth+n|day	Should not be adjusted to *birth* and *day*
+n|fire+n|man	Should not be adjusted to *fire* and *man*

Further, several words from CHILDES are annotated as normal forms and, because of that do not match the words given in the Iconicity Table. For this reason, the content Iconicity Table and the corpus of child speech were studied attentively to discover which word-strings from CHILDES should match the Iconicity Table, but because of annotation specificities in CHILDES, they do not. This non-fit problem was regulated case by case using queries. As seen from the examples given in Table 7., *adj|good&CP*

had to be adjusted to *better, adj|sun&dn-Y* – to *sunny, adj|bug&dn-Y* – to *buggy*, *dirt&dn-Y* – to *dirty* etc. Appendix C contains all the changes performed at the level of words.

Table 7. Examples of specific cases of word-strings in CHILDES

Text Marked ID	Speech	POS annotation in CHILDES - the parsed string
law25.bwd0188	xxx to feel ***better*** ? [+ PI]	inf|to v|feel ***adj|good&CP ?***
ad931.bro4227	**many** out ?	***adj|man&dn-Y*** adv|out ?
ada41.bro2591	**buggy** .	***adj|bug&dn-Y*** .
laa21.bwd1735	**penny** [*] ice+cream .	***adj|pen&dn-Y*** n|+n|ice+n|cream .
laa29.bwd0676	use my **potty** .	v|use pro:poss:det|my ***adj|pot&dn-Y .***
nat30.sno1390	&i &i &i &i it's **sunny** out ?	pro|it cop|be&3S ***adj|sun&dn-Y*** adv|out ?
laa24.bwd0278	he 0is [*] **dirty** .	pro:sub|he 0aux|is adj|***dirt&dn-Y .***

The general purpose of this adjusting procedure was to enlarge as much as possible the number of word-matches between the data sources (Fig. 3.). All the changes were performed uniformly over the entire corpus and the changed words - stored in the field "CHILDES Word Adjusted".

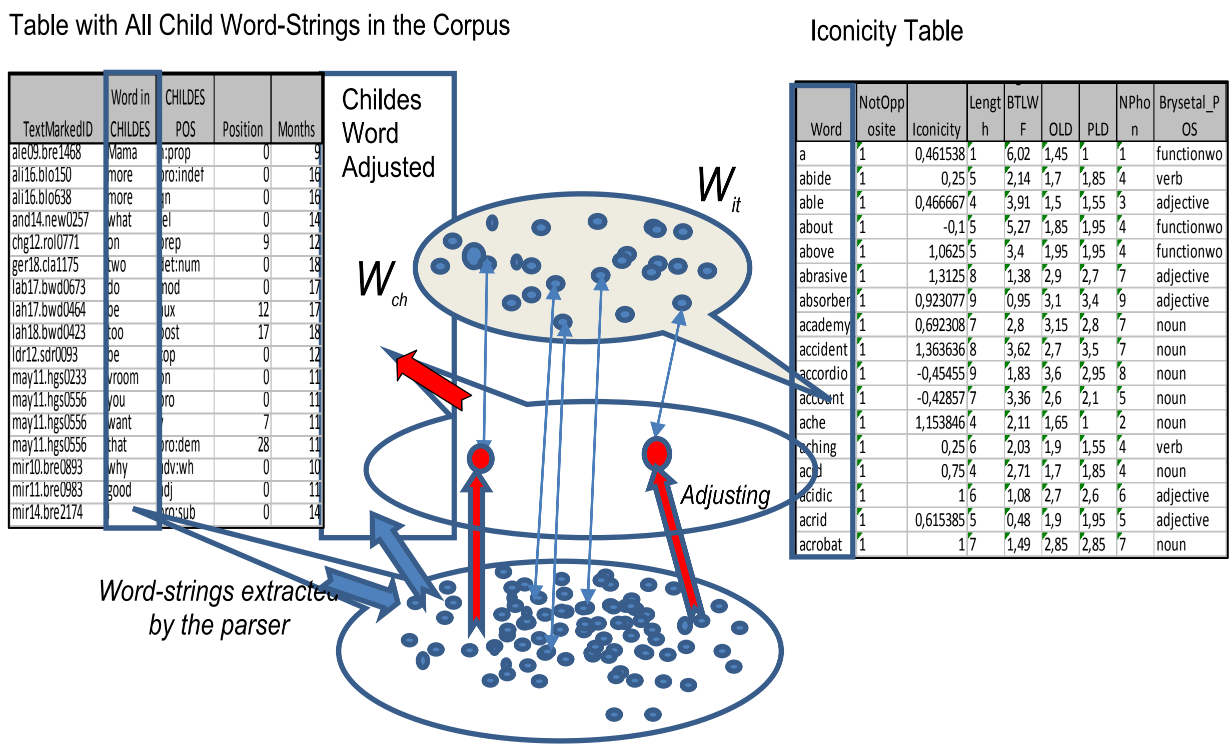


Figure 3. The Word Adjusting Procedure aims at enlarging Wch ∩ Wit .

The entire list of the adjusted word-strings is given in Appendix C. Here some examples (Table 8.).

Table 8. Examples of adjusted words.

Word-string extracted from CHILDES	CHILDES WORD ADJUSTED	To Match the Word in Iconicity Table
be (cop|be&PAST)	were	were
+n|butter+n|fly	butterfly	butterfly
animal-PL	animal	animal
blend&dv-AGT	blender	blender
call-3S	call	call
call-PAST	call	call
cock_a_doodle_doo	cockadoodledoo	cockadoodledoo
contain&dv-AGT	container	container
own&dv-AGT~poss|s	owner	owner
snow~aux|be&3S	snow	snow
sock-PL	sock	sock
work-3S	work	work
work-PL	work	work

4.2. Dealing with Verbs

The verb-forms are annotated in CHILDES exclusively with the basic form of the verb. That is, the use of the verb "*to write*", for example, with all the forms that children used (write, writes, wrote, writing, written) is annotated with the form "write" and the extracted by the parser string is "write". All these grammatical forms are distinguishable at two levels:

1. With different POS-tags (stored in the field CHILDES POS): verb, modal verb, auxiliary verb, present participle, and past participle (v, mod, aux, PRESP, cop, and PASTP), as shown in Fig. 4.
2. within the word-string, with grammatical markers concatenated to the word's basic form (that is – within the %*mor* field of the database), such as PAST for past tense and 3S for the conjugation.

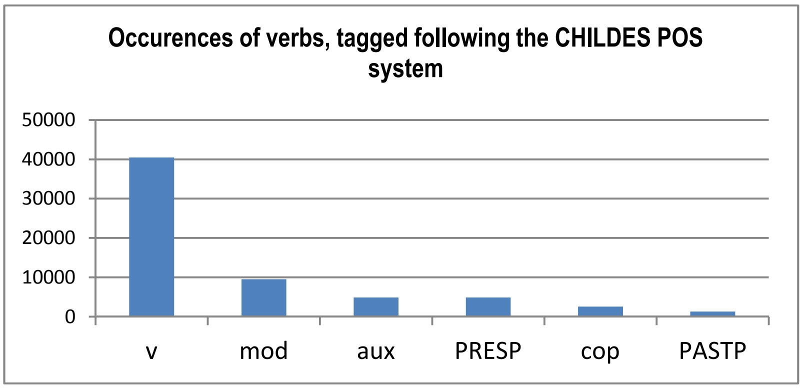


Figure 4. Verbs in CHILDES.

In the Iconicity table, the verbs are given without specifying their concrete form, as their iconicity was studied as words out of any context. After analyzing the 556 verbs given in the Iconicity Table, 73 forms of present participles, 31 forms of past tense and/or past participles and 3 forms of 3th person conjugation (*is*, *does* and *was*) were identified. The rest of the verbs in the Iconicity Table are given in the normal form.

Obviously, verbs are the POS which has more forms. The Iconicity of the verb-form occurrences had to be evaluated following some clear scheme and criteria.

The plot in Figure 5 shows how often appear the different grammatical occurrences of verbs in the child speech corpus (after having isolated the verbs tagged with - PAST or &PAST and with -3S or &3Sat the end of the word-string), The verbs *To Be* and *To Do* are not included here, they were treated separately and will be explained in the corresponding section.

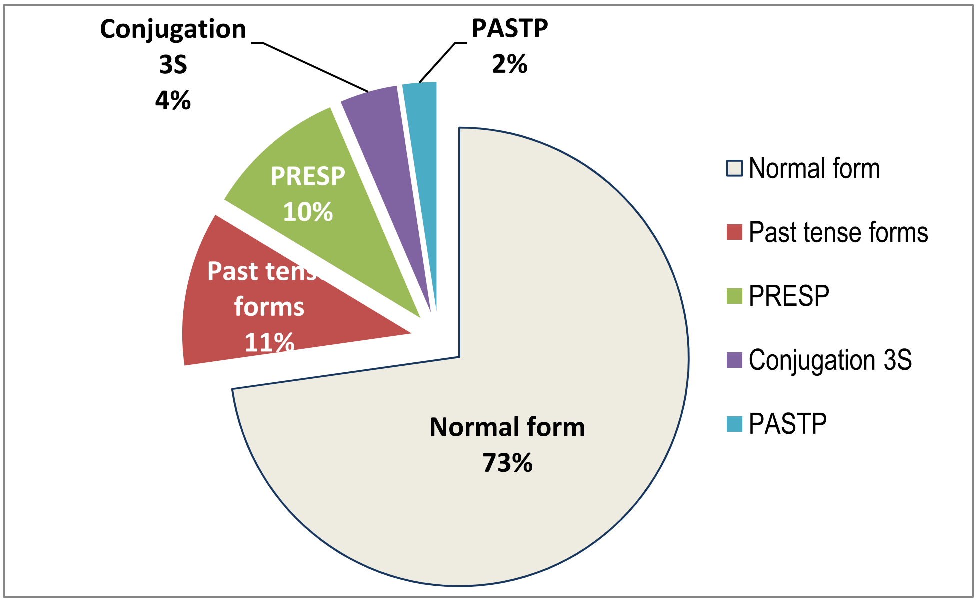


Figure 5. Occurrencies of verbs in the Child speech corpus (the verbs "to be" and "to do" are not included).

The steps undertaken to adjust the children's words in order to match appropriately as much as possible of the content of the Iconicity Table are further explained.

4.2.1. The Occurrences of Verbs in Normal Form and in 3th person conjugation

The verbs in Normal form are the case in which the match of the speech occurrences to the content of the Iconicity Table is pure. That is, the extracted by the parser normal form has to be matched to the existing normal form of the verb in the Iconicity Table.

However, the verb *to want* which exists in its normal forms in the Iconicity Table is often used by the children in a specific way. The examples of sentence-utterances in Table 9. illustrate such cases.

Table 9. Examples of Protective Changes of "wanna" from matching the normal form "want" in the Iconicity Table.

Speech	POS (%mor annotation)	Word-string extracted	CHILDES WORD ADJUSTED	To not match	Word in Iconocoty T.	Iconicity
mommy you wanna put this away ?	n|mommy pro|you v|want inf|to v|put&ZERO det|this adv|away ?	want	wanna~		want	-0,21429
you wanna <see superman> [>] ?	pro|you v|want inf|to v|see super#n|man ?	want	wanna~		want	-0,21429
(.) I wanna tell you something .	pro:sub|I v|want inf|to v|tell pro|you pro:indef|something .	want	wanna~		want	-0,21429

The problem is that the POS-tagged string (*%mor*) contains the verbs in normal form, while the child did not say a phonetic content which is similar enough to "*want*". It was suggested by the specialists in Iconicity to protect the match of these cases to the normal form of the verb. The strings "*want*", extracted from the 465 occurrences of *wanna,* were changed to "*wanna~"* in order protect their match to the verb *want* in the Iconicity Table.

The rest of the word-strings, extracted from the currencies of verbs used in normal form, were left unchanged. The number of occurrences which match verbs in the Iconicity Table is 25337, that is - 74,5% of the occurrences of normal forms of verbs in the corpus. The list of these verbs is provided in Appendix D.

It was considered that the **3th person conjugations** are close enough to the normal form of the verb. The word-strings were left as extracted from the original - in their basic form. As an example, the sentence "*and Googee sleeps right here* ." is originally tagged as "*coord|and n:prop|Googee v|sleep-3S adv|right adv|here* .", the extracted string is "*v|sleep-3S*" and the adjusted word is "*sleep*", which matches the verb "*sleep*", given in normal form in the Iconicity table, with iconicity 0,77. Appendix E provides the list of all the forms of 3th person conjugations which were left to match the normal forms of the verbs which exist in the Iconicity Table. In total 1632, that is - 86% of all the 3th person conjugations which occurred in the child speech were matched to verb normal forms, existing in the Iconicity Table.

4.2.2. The occurrences of Present Participles

All the occurrences of present participles in the Child speech corpus (tagged with *part*| at the beginning and with PRESP collated to the word) which correspond to POS - verbs (and are present participles) in the iconicity table were retrieved (as extracted in their basic form) and adjusted to the corresponding "ing" form. That ensures their mapping to the corresponding present participle in the Iconicity Table, where some of the verbs exist as both - basic form and the present participle (with small differences in the iconicity norm). Table 10 gives the list of all the adjusted word-strings.

Table 10. Present participles occurring in the Child speech, adjusted to match the Iconicity Table.

Occurrences	CHILDES POS	Word-string extracted	CHILDES WORD ADJUSTED	Word in Iconicity Table	Iconicity	POS in Iconicity Table
2	PRESP	bang	banging	banging	2,5000	verb
8	PRESP	bark	barking	barking	3,3333	verb
2	PRESP	boil	boiling	boiling	1,3846	verb
7	PRESP	burn	burning	burning	0,5333	verb
1	PRESP	crash	crashing	crashing	2,4286	verb
73	PRESP	cry	crying	crying	0,1538	verb
24	PRESP	fall	falling	falling	0,6154	verb
15	PRESP	laugh	laughing	laughing	0,5000	verb
2	PRESP	spill	spilling	spilling	1,5000	verb
1	PRESP	squeal	squealing	squealing	2,7500	verb
5	PRESP	swing	swinging	swinging	1,9231	verb
21	PRESP	write	writing	writing	0,3571	verb

There were no other occurrences of PRESP found to match present participles in the Iconicity Table. Only 161 occurrences, that is – 3% of all the PRESP occurring in the child speech, found a correspondent in the Iconicity Table.

The normal forms extracted from the rest of the occurrences of PRESP in the Child speech corpus were transformed by concatenating the string *"~ing"* to the extracted normal verb-form (see the example in Fig. 5.1.) in order to indicate in the novel data-field (CHILDES Word Adjusted) that this is not the original string.

Importantly, after this transformation, the introduced strings do not match "false" words in the Iconicity Table (e.g. build~ing does not match the noun building). All the occurrences of present participles which, after the change, do not match "wrong" words in the Iconicity Table are provided in Appendix D01. Table 11 gives some examples.

It is seen from the examples in Table 11 that the word-strings extracted from the present participles occurring in the child speech could match words which correspond to normal forms of verbs in the Iconicity Table. It was assumed that the present participles are phonetically and semantically distant from the normal form of the verb and they were transformed to not match it.

Table 11. Examples of Present participles occurring in the Child speech corpus, changed to Not match "wrong" words the Iconicity Table.

CHILDES POS ORIGINAL	Word-string extracted	CHILDES WORD ADJUSTED	**TO NOT MATCH**	Word Iconicity T.	POS Iconicity T.	Occurrences
PRESP	back	back~ing		back	adverb	4
PRESP	be	be~ing		be	verb	8
PRESP	block	block~ing		block	noun	1
PRESP	blow	blow~ing		blow	verb	7
PRESP	boss	boss~ing		boss	noun	1
PRESP	catch	catch~ing		catch	verb	1
PRESP	chop	chop~ing		chop	noun	1
PRESP	clean	clean~ing		clean	adjective	7
PRESP	close	close~ing		close	adjective	2
PRESP	color	color~ing		color	noun	9
PRESP	dream	dream~ing		dream	noun	1

4.2.3. The Occurrences of Past-Tense forms.

The question is how to evaluate the iconicity of the used by children verbs which are not in normal form.

All the verb-form occurrences (tagged with *v|* at the beginning and a marked for the conjugation and tense at the end of the word-string) had to be matched to the verbs given in the Iconicity Table. They were all extracted as POS "v".

The **past tense** occurrences of verbs (5086 in the corpus) which are originally tagged with v| at the beginning and marked with "-PAST" or "&PAST" at the end had to be adjusted. A small number of verbs are given in the Iconicity Table in past-tense forms. These verbs were first identified and next - the Child speech corpus was checked to find the corresponding occurrences of past tense forms. The existing in the child speech occurrences of these verbs were retrieved (Table 12.) and adjusted to match the word in the Iconicity Table.

Table 12. List of the adjusted word-strings of past tense forms of verbs, changed to match verbs given in past tense in the Iconicity table.

Occurr ences	CHILDES POS ORIGINAL	Word-string extracted	CHILDES WORD ADJUSTED	Word in Iconicity Table	Iconicity	POS Iconicity Table
2	v	close-PAST	cosed	closed	1,60	verb
1	v	crack-PAST	cracked	cracked	3,00	verb
260	v	fall&PAST	fell	fell	1,71	verb
12	v	fall-PAST	fell	fell	1,71	verb
2	v	feel&PAST	felt	felt	0,93	verb
1	v	feel-PAST	felt	felt	0,93	verb
8	v	paint-PAST	painted	painted	1,45	verb
3	v	roll-PAST	rolled	rolled	2,60	verb
136	v	see&PAST	saw	saw	1,73	verb

There are no other occurrences of past tense forms which match the verbs given in past tense in the Iconicity Table. In total 425 occurrences, that is – 8% of all the past tense forms occurring in the child speech, found a correspondent in the Iconicity Table.

To protect false matches, the past tense forms for which the extracted normal form matches strings of other words in the Iconicity Table were also adjusted, by concatenating the string "~d" to the extracted basic verb-form. For example, the occurrences of the past tense of "to dream" were adjusted to "dream~d" in order to not match the noun "dream" given in the Iconicity Table. Examples are provided in Table 13.

Table 13. Examples of Past Tense forms occurring in the Child speech corpus, changed in order to avoid matches to "wrong" words in the Iconicity Table.

string extracted	CHILDES WORD ADJUSTED	**Does Not Match**	Word in Iconicity Table	Brysetal_POS
back-PAST	back~d		back	adverb
clean-PAST	clean~d		clean	adjective
damage-PAST	damage~d		damage	noun
down-PAST	down~d		down	adverb
dry-PAST	dry~d		dry	adjective
glue-PAST	glue~d		glue	noun

After consulting the specialists in Iconicity from the University of Calgary, Canada, it was assumed that the past tenses of regular and irregular verbs should be treated differently. This separation is due to the fact that the change of the root morpheme plays an important role for the Iconicity index. It has to be noted that there are no indications in the data which indicates if the used by children verbs are regular and which are irregular, so this distinction was done after several consulting steps.

The questions regarding the relationship between the phonetic content, the semantic and grammatical features of the words and their Iconicity are still not fully clarified. However, it is known that such a relation exists. Assuming that the semantics (the action) and the lexical category (verb) of the forms of one and the same verb are close, the main difference within the pairs "normal form – past tense form" is the phonetic content.

The past tense forms of the **regular verbs** were assumed to correspond to the basic form of the verb. The reason is that the past tense forms of the regular verbs are phonetically and semantically close to the basic form of the verb as well as that they are derived from it following a stable grammatical rule. Having assumed this, the occurrences of the past tense forms of 72 regular verbs, which appear in the children speech and exist in normal form in the Iconicity Table, were adjusted to the normal form of the verb. So, these occurrences were accorded the Iconicity of the normal form of the verb. The list of these verbs is given in Appendix I , Table 14 provides some examples of use.

Table 14. Examples of occurrences of past tense forms of regular verbs which exist in the Iconicity table only in normal form and were matched to this existing normal form

Word in Iconicity T.	Iconicity norm	POS in Iconicity T.	CHILDES POS	Childes word ADJUSTED	Child speech with the verb-part tense occurrence - example	CHILDES POS-annotation (*%mor*)
boil	1,38	verb	v	boil	hard ***boiled*** egg .	adv|hard *v|boil-PAST* n|egg .
cry	0,87	verb	v	cry	he ***cried*** .	pro:sub|he *v|cry-PAST* .
help	1,53	verb	v	help	I ***helped*** her .	pro:sub|I *v|help-PAST* pro:obj|her .
jump	1,00	verb	v	jump	xxx ***jumped*** on the table .	*v|jump-PAST* prep|on det|the n|table .
kiss	1,25	verb	v	kiss	no girl ***kissed*** me .	qn|no n|girl *v|kiss-PAST* pro:obj|me .
pick	2,08	verb	v	pick	she ***picked*** a flower .	pro:sub|she *v|pick-PAST* det|a n|flower .
scare	0,67	verb	v	scare	***scared*** me .	*v|scare-PAST* pro:obj|me .
use	1,00	verb	v	use	I [<] ***used*** it .	pro:sub|I *v|use-PAST* pro|it .
work	1,71	verb	v	work	and (.) it ***worked*** .	coord|and pro|it *v|work-PAST* .
yell	0,58	verb	v	yell	she ***yelled*** .	pro:sub|she *v|yell-PAST* .

In total, 574 occurrences of the past tense forms of regular verbs, left in normal form, were found to match to the normal form of the corresponding verb, given in the Iconicity Table. This represents 11% of the past tense forms which exist in the Child speech corpus,

The occurrences of past tense forms of all **irregular verbs** were not allowed to fit the corresponding normal forms. The occurrences of past tense-forms for these verbs were changed to not match the normal form of the verb. Table 15 gives examples of changed strings, namely, they were transformed by concatenating the string "~‘" in order to not match the normal form of the verb, given in the Iconicity Table. Examples of the use of these verbs are given in Appendix J.

Table 15. Examples of protective changes of past forms of irregular verbs to not match the normal form in the Iconicity Table

occurences	string	Word-string extracted	CHILDES WORD ADJUSTED	to Not Match	Word in Iconicity Table	POS in Iconicity Table
77	break&PAST	break	broke~		break	verb
27	bring&PAST	bring	brought~		bring	verb
21	buy&PAST	buy	bought~		buy	verb
53	catch&PAST	catch	caught~		catch	verb
1	catch-PAST	catch	caught~		catch	verb
302	go&PAST	go	went~		go	verb
25	leave&PAST	leave	left~		leave	verb
4	steal&PAST	steal	stole~		steal	verb
111	take&PAST	take	took~		take	verb
1	take-PAST	take	took~		take	verb
1	teach&PAST	teach	taught~		teach	verb
43	think&PAST	think	thought~		think	verb

As it is seen from the description, all such adjusting steps required complex queries over the corpus, consulting the content of the data sources, checking the context of use and manual change. This adjusting operation was performed nearly manually. Several mistaken by the children past tenses were noticed, namely, children usually applied the rule for regular verbs to the irregular verbs. Examples of such cases are provided in Appendix K. Table 16 gives examples. All noticed cases of mistaken past forms were adjusted based on the form used by the child.

Table 16. Examples of occurrences of mistaken past forms adjusted to match the existing in the Iconicity Table form of the verb.

Word in Iconicity Table	Iconicity	POS in Iconicity T.	CHILDES POS	Childes word ADJUSTED	Child speech with the example	CHILDES POS-annotation (%mor)
break	2,90	verb	v	break	it ***breaked*** .	pro|it v|break-PAST .
catch	0,46	verb	v	catch	I ***catched*** another fish .	pro:sub|I v|catch-PAST qn|another n|fish .
do	0,85	verb	v	do	see what he ***doed*** ?	v|see rel|what pro:sub|he v|do-PAST ?
draw	0,20	verb	v	draw	who ***drawed*** on it ?	pro:wh|who v|draw-PAST prep|on pro|it ?
go	1,45	verb	v	go	she ***goed*** away .	pro:sub|she v|go-PAST adv|away .
hide	1,14	verb	v	hide	we ***hided*** [: hid] [* m:=ed] the ice cream .	pro:sub|we v|hide&PAST det|the n|ice n|cream .
steal	0,00	verb	v	steal	they [= bees] ***stealed*** the honey .	pro:sub|they v|steal-PAST det|the n|honey .

4.2.4. The Past Participles

All the occurrences of past participles (tagged with *part*| at the beginning and "-*PASTP*" or "&*PASTP*" at the end of the word-string and extracted as a separate POS) used in the children's speech which correspond to those which exist in the iconicity table were retrieved in the corpus and adjusted to the corresponding past given in the Iconicity Table. The existing matches are given in Table 17.

Table 17. List of the occurrences of past participles which match the Iconicity Table.

Occurrences	CHILDES POS ORIGINAL	Word-string extracted	CHILDES WORD ADJUSTED	Word in Iconicity T	Iconicity	POS in Iconicity T
137	PASTP	break	broken	broken	1,5	verb
2	PASTP	burn	burnt	burnt	0,666666667	verb
10	PASTP	close	closed	closed	1,6	verb
1	PASTP	leave	left	left	0,454545455	verb
1	PASTP	paint	painted	painted	1,454545455	verb
17	PASTP	stick	stuck	stuck	1,583333333	verb

There were no other occurrences of PASTP found to match past participles in the Iconicity Table. In total 168 occurrences of PASTP (15% of all the occurrences of PASTP) have a correspondent in the Iconicity Table.

The rest of the occurrences of PASTP in the Child speech corpus were marked by concatenating the string *"~d"* to the existing basic verb-form (or *"~"* for some irregular forms) as shown in the examples provided in Table 18, in order to indicate in the novel data-field (CHILDES Word Adjusted) that this is not the original string. Thus, the obtained "pseudo-words" do not match inconvenient words the Iconicity table (e.g. *land~d* will not match the noun *land*). The list of the past participles which are adjusted this way is given in Appendix L.

Table 18. Examples of changes of Past participles performed to avoid false matches

Occurr ences	CHILDES POS ORIGINAL	Word-string extracted	CHILDES WORD ADJUSTED	To Not Match	Word in Iconicity Table	Iconicity	POS in Iconicity T.
13	PASTP	wet	wet~d		wet	2,25	adjective
1	PASTP	like	like~d		like	-0,33	functionword
7	PASTP	name	name~d		name	0,07	noun
5	PASTP	pop	pop~d		pop	4,08	onomatopoeia

This adjusting operation was performed nearly manually. Several mistaken by the children past participles were noticed. Examples of such cases are provided in Appendix K.

Obviously, several past participles are the same as the corresponding past tense forms. The approach used to match the past participles to the normal form was the same as this applied for the past tense forms. That is, the occurrences of past participles of the regular verbs were matched to the normal form of the verb. The occurrences past participles of irregular verbs were assumed distant from the normal form of the verb and were not matched to the normal form correspondent in the Iconicity Table.

4.2.5. The verbs *to be* and *to do*.

The verb *to be* is annotated in CHILDES as "cop" or "aux" and the verb *to do* as "mod" or "v". They are annotated in CHILDES, like all the other verbs, with their normal forms. The forms of these verbs which exist separately in the Iconicity Table are: *be*, *is*, *was*, *were* and *did*. Again, the occurrences of, for example, cop|be&3S, aux|be&3S, cop|be&1S, aux|be&1S, cop|be&PAST, mod|do&3S, mod|do&PAST, v|do&PAST, etc. were to be adjusted in order to match the right word in the Iconicity table. These occurrences were retrieved case by case using queries and adjusted to fit the targets.

The retrieved occurrences of *to be* are 12835 and these of *to do* are 7001. The forms of these verbs which exist separately in the Iconicity Table are: *be*, *is*, *was*, *were* and *do*, *does* and *did*. For the matching cases:

- The occurrences of *cop|be&13S* and *aux|be&13S* were retrieved and adjusted to ***is***.
- The occurrences of *cop|be&1S,* and *aux|be&1S* were retrieved and adjusted to ***am***.
- The occurrences of *cop|be&PRES* and *aux|be&PRES* (which are not *am* or *is*) were retrieved and adjusted to ***are***.
- The occurrences of *cop|be&PAST&13S*, and *aux|be&PAST&13S* were retrieved and adjusted to ***was***.
- The occurrences *aux|be&PAST* and *cop|be&PAST* were retrieved and adjusted to ***were***.
- The occurrences of *mod|do&3S* and these of *v|do&3S* were retrieved and adjusted to ***does***.
- The occurrences *mod|do&PAST* and ***v|do&PAST*** were retrieved and adjusted to ***did***.

These occurrences were retrieved case by case using queries to find the targets "does", "did", "are" "am", "were", "was" and "is" in the string “Child speech”. The reason is that the tagging of tense and conjugation in the original do not always respect strict rules.

The occurrences of *to be* and *to do* which are in normal form were not changed, they match the normal forms in the Iconicity Table.

The forms of *to be* and *to do* which do not exist separately in the Iconicity Table are the past participles *done* (168 occurrences were retrieved) and *been* (48 occurrences were retrieved). They were adjusted to *done*~ and *been*~ to not match the normal form of the verbs. The present participles were changed as for the other verbs (see Appendix C).

In addition, in compliance with the recommendation of the specialists in psycholinguistics from the University of Calgary, 5071 occurrences of 3S conjugation of to be (cop|be&3S or aux|be&3S) which are not used by the child as a clearly pronounced “is” were changed to “is~” in order to protect them from matching the form “is”, given in the Iconicity table. Examples of such occurrences are given in Table 19.

Table 19. Examples of the cases where the used word as form of “to be” was changed to Not match the form “is” in the Iconicity Table.

Word-string extracted	CHILDES WORD ADJUSTED	Child Speech	POS annotation (the %mor string)
be	is~	**that's** okay .	pro:dem|that cop|be&3S adj|okay .
be	is~	**it's** just broken .	pro|it cop|be&3S adv|just part|break&PASTP .
be	is~	**who's** over here ?	pro:wh|who cop|be&3S prep|over pro:dem|here ?

In result, the occurrences of the forms of *to be* and *to do* which are adjusted to match the existing forms in the Iconicity Table are as shown in Table 20.

Table 20. Occurrences of "to be" and "to do" which match the forms in the Iconicity Table

occurrences	CHILDES WORD ADJUSTED	Word in Iconicity T	Iconicity
1252	am	am	0,58333333
1433	are	are	-0,90000000
769	be	be	0,38461539
1131	did	did	0,46666667
5036	do	do	0,84615385
666	does	does	-0,60000000
3193	is	is	-0,14285714
895	was	was	-0,83333333
163	were	were	0,60000000

After this adjusting procedure, 60% of the occurrences of the verb "to be" and 98% of the occurrences of the verb "to do" were matched to the content of the Iconicity Table. The rest of the occurrences did not find a correspondent or were protected for matching.

The example in Fig. 6. illustrates the result (from a database report) for the verb *to do* after adjusting the extracted by the parser string *v|do&3S* to the word *does*:

wan47.gle0103

what does this do ?		rel|what v|do&3S det|this v|do ?	
what	functionword	0,142857143	0
does	verb	-0,6	8
this	functionword	0,133333333	16
do	verb	0,846153846	25
	Sum:	0,522344322	WordCount: 4

wan47.gle0111

Figure 6. Adjusting of the verb to do – example of a sentence, a DB Report.

4.2.6. The Result of Adjusting and Protecting the Verbs

The described procedure, applied to the occurrences of verbs in order to match correctly their forms to the content of the Iconicity Table, gave the result illustrated in Fig. 7.

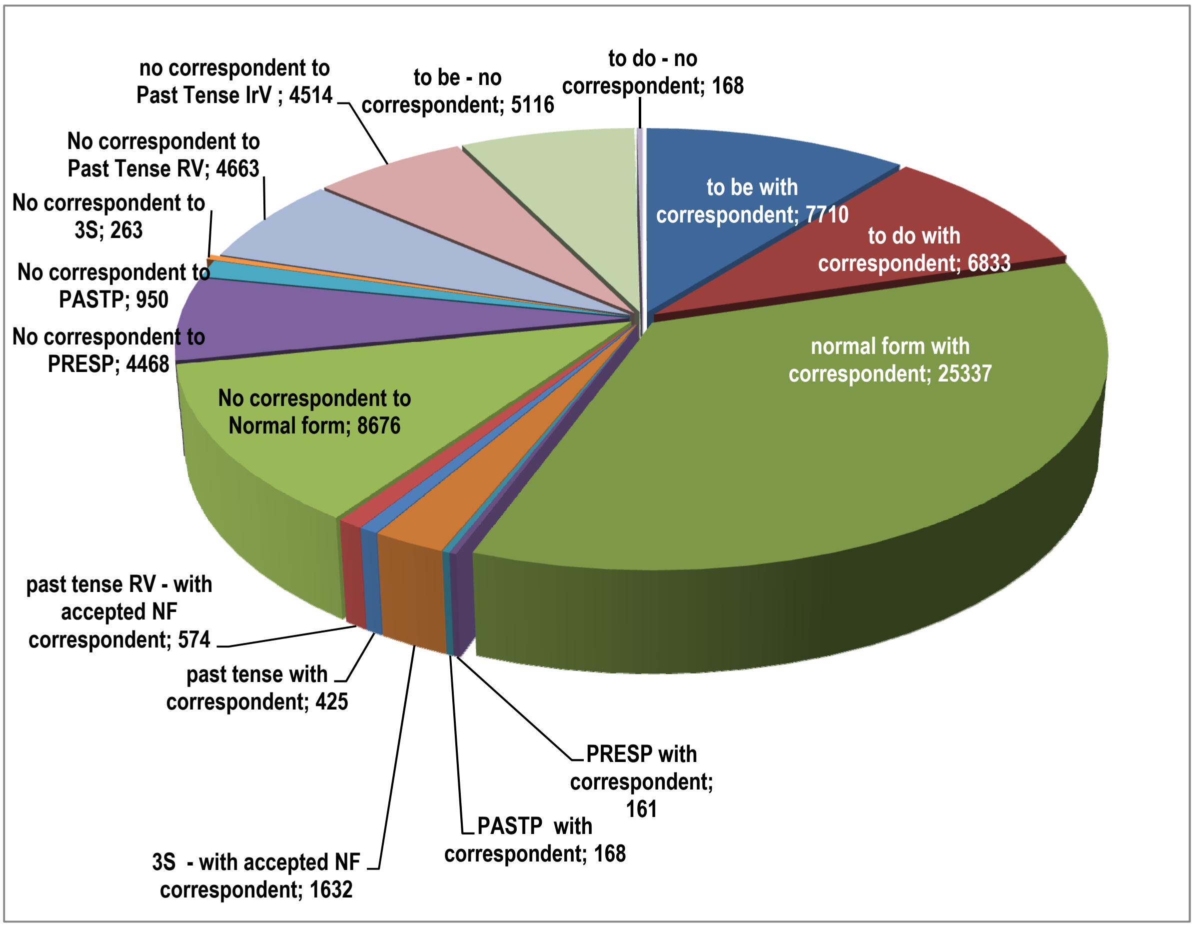


Figure 7. All the occurrencies of verbs in the Child speech corpus – these which have and these wich have not a correspondent in the Iconicity Table after the procedure of adjusting and protecting.

In total, 60% of the verbs occurring in the Child speech corpus obtained a correspondent in the Iconicity Table.

4.3 The Result of word-string adjusting

After the procedure of adjusting the word-strings, 36 940 word-string occurrences (1864 different words) were modified in the separate field “CHILDES Word Adjusted“ to match the words in the Iconicity Table. This allowed evaluating the percentage of word-occurrences which have “obtained” a correspondent word in the Iconicity Table at 78%.

However, the adjusting is made at the level of words and does not concern the parts of speech.

One and the same word can be used as different POS. This is, of course, done by children too, as seen in Table 8., where he word *work* is used by children as a verb and as a noun. This can lead to false matches, as shown with the example given in Fig. 8. :

Child Speech	CHILDES POS ANNOTATION (%mor)
(.) did it **blast** off in the air ?	mod|do&PAST pro|it v|blast adv|off prep|in det|the n|air ?

Word Parsed from CHILDES	CHILDES POS	Word ADJUSTED		Matches the Word in the Iconicity Table	with POS
blast	v	blast	-->	blast	noun

Figure 8. False match because of POS contradiction

5. Data Fusion for Annotation of the Child Speech with Iconicity

The strategy undertaken for the data fusion was to use as join fields the words in the two tables – this with the adjusted children words from the Child speech corpus and the word in the Iconicity Table (Fig. 9). The obtained intermediate result contains all the words (CHILDES Word Adjusted) used in the children dialogues and the words from the Iconicity Table, which match them, together with their POS and their Iconicity norm. This intermediate result was temporarily stored in a separate table for further treatment, namely - the POS-transition and filtering.

The result of this (Left) Join operation provides all the available matches, at the level of words:

$$WordMatch = W_{ch} \cap W_{It} \qquad (2)$$

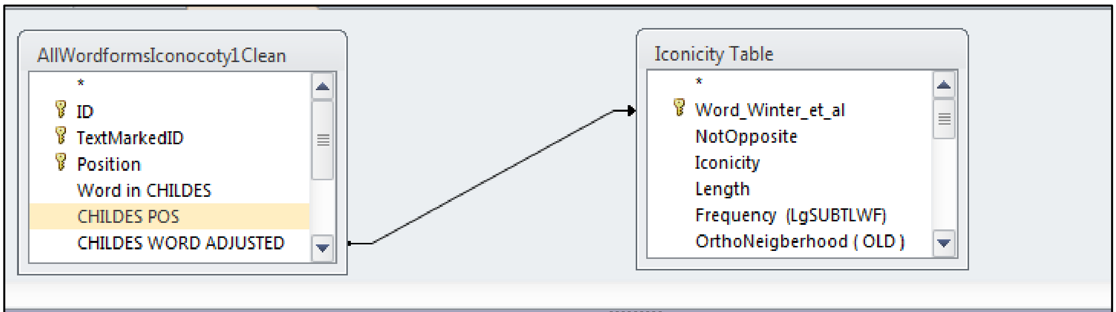
Where W_{ch} are the adjusted words from the children's dialogues, and W_{It} are the words from the Iconicity Table.

The result of the data fusion stored in the temporary Table with all Children's Words allows following the course of the dialogues using the sentence and position markers. This is shown with the example of two sentence-utterances given in Fig. 9.

The "fused" table is in 3th NF and stores the following information for each word-utterance of the child speech corpus:

- The identifier of the word from the separated POS tables
- The identifier of the sentence -utterance (for relation to the dialogue and the child)
- The position on which the word is situated in the string of the sentence–utterance.
- The word-string as initially extracted from the POS annotation in CHILDES (*mor%*)
- The adjusted Word-string (the steps described above)
- The Word in the Iconicity Table which matches the adjusted word
- The Iconicity of this word in the Iconicity Table
- The POS annotation in CHILDES

- The POS annotation in the Iconicity Table



ID	TextMarked	Position	Word in CHILDES	CHILDES POS	CHILDES WORD ADJU	Word_Winter_	Iconicity	Length	Frequenc	OrthoNei
AllWc	AllWordfor	AllWordf	AllWordformsIco	AllWordformsI	AllWordformsIconocc	Iconicity Table	Iconicity	Iconicity	Iconicity	Iconicity

ID	TextMarkedI	Po	CHILDES POS	Word in CHILI	CHILDES WORD ADJUST	Word_Winter_et_	Iconicity	Brysetal_POS
21408	add40.bro1270	0	v	cut&ZERO	cut	cut	0,705882353	verb
2558	add40.bro1270	10	pro:poss:det	your	your	your	0	functionword
28233	add40.bro1270	28	n	hair	hair	hair	-0,416666667	noun
21409	add40.bro1277	0	v	see	see	see	1,272727273	verb
888	add40.bro1277	5	pro:wh	what	what	what	0,142857143	functionword
1981	add40.bro1277	17	aux	be	is	is	-0,142857143	verb
2926	add40.bro1277	28	PRESP	happen	happen~ing			

Figure 9. Data fusion - Join of the data sources.

As said, the resulting table contains all the extracted and adjusted words from the child speech corpus (as described in the previous sections), that is – it contains 309 116 word occurrences. At this first step, the utterances which have a corresponding word in the Iconicity Table have as attributes the POS-annotation following the Iconicity Table as well as their iconicity norms. This organization of data was useful for further analysis as it allows queries which reconstruct the course of the sentences within the dialogues together with the accorded iconicity. An example (of a report) is given in Fig. 10.

			tre36.dem1486
is this a good egg .		cop|be&3S det|this det|a adj|good n|egg .	
is	verb	-0,142857143	0
this	functionword	0,133333333	9
a	functionword	0,461538462	18
good	adjective	0,928571429	24
egg	noun	1,818181818	33
	Sum:	3,198767899	WordCount: 5

Figure 10. One sentence-utterance, annotated with Iconicity after the data fusion based on Word-match.

After the described procedure of data fusion, 80% of the word occurrences in the child speech corpus were found to have a correspondent in the Iconicity Table. It is important to examine how this correspondence is distributed over the sentence-utterances, by months. The plot in Fig. 11. shows that this distribution is quite uniform.

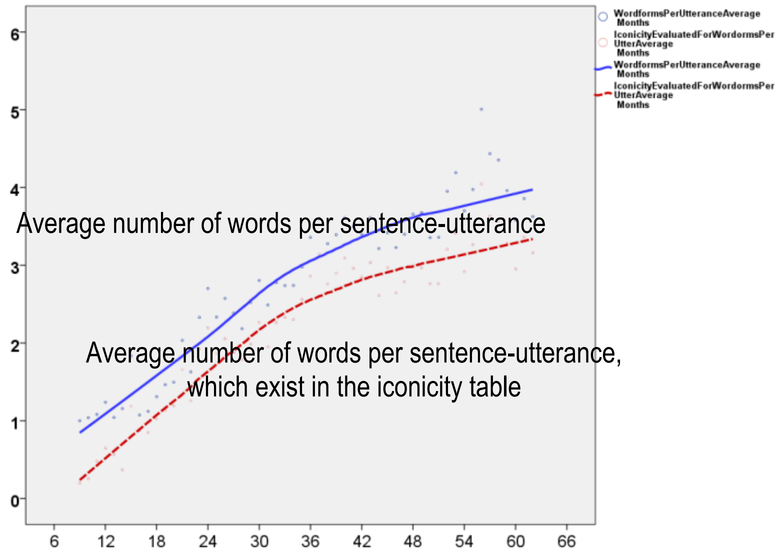


Figure 11. Use of words which exist in the Iconicity Table in the child speech

The fact of keeping both POS annotations gives the opportunity to check quickly is the word used by the child the same POS which is in the Iconicity Table or it is used as some other part of speech. This allows two manners of analysis depending on the matching condition:

- Word Match condition;
- Total Match condition.

The Word Match gives the iconicity of the words used as *words* without taking into account do they fit as POS in the two tables. The Total Match takes into account the Iconicity norm of the said ***if and only if*** the word is used by the child as the same POS as the POS given in the Iconicity Table.

The next problem to be solved concerns the filtering of errors occurring because lack of match at the POS-level. That is, the word-match has to be reduced -- all the cases $W_{uncohPOS}$ of use of words where the POS indicated in the two data sources are not coherent have to be excluded:

$$Total\ Match = WordMatch - W_{uncohPOS} \qquad (3)$$

Where: $W_{uncohPOS}$ is the set of words for which the POS in CHILDES does not correspond to the POS in the Iconicity Table.

One of the biggest difficulties to annotate the described here corpus was related to the Total Match condition (POS-match) because the two POS–annotation systems are quite different. The next section describes some statistical properties of the data sources in order to provide some details concerning the differences in the POS systems.

6. The POS-problem: Some Descriptive Statistics of the Data sources

6.1. The Iconicity Table

The short description provided here is based on the data published by Winter and colleagues [Winter et al. 2017] where the *iconicity norms* were assessed to English words using ratings for iconicity, assigned by native speakers of English. The data source contains iconicity norms for 3001 English words.

The statistical properties of the Iconicity Table are given, as calculated from the data sources used for this report, in Table 21. and in Fig. 12. and 13:

Table 21. Descriptive statistics of the content of the Iconicity Table.

POS (Iconicity Table)	Number Words	Percentage	Average Iconicity	Variance Of Iconicity
noun	1714	0,58121	0,67583	1,04927
verb	556	0,18854	1,35770	1,52315
adjective	538	0,18243	1,17631	1,19740
functionword	71	0,02408	0,45949	0,76934
adverb	40	0,01356	0,79827	0,72439
onomatopoeia	19	0,00644	3,47665	0,32279
interjection	11	0,00373	2,30673	0,67324
not accorded	52	not taken for further analysis		

In total 2949 words from the Iconicity Table match the words used in the Child speech corpus (after the adjusting procedure described above), these words are taken for the further POS-fit procedures.

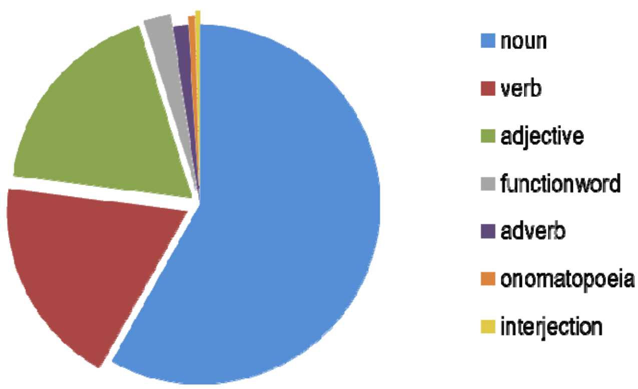


Figure 12. Percentage of words in the Iconicity Table, by POS.

As reported in the specialized studies and illustrated in Fig. 13, the seven parts of speech existing in the Iconicity Table have different Iconicity profiles, with differences in their means and standard deviations.

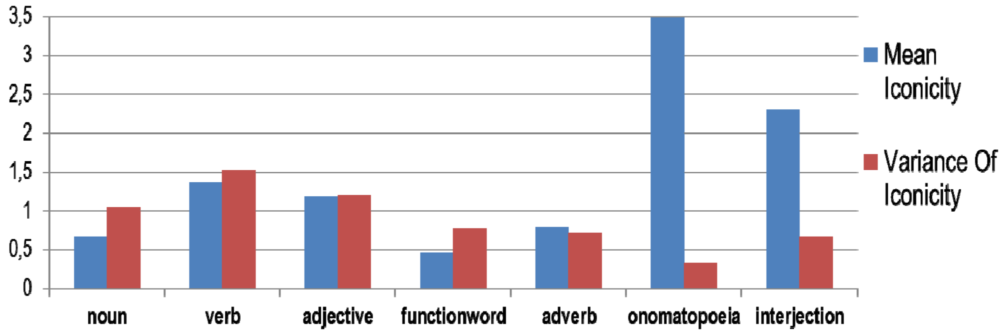


Figure 13. Profiles of the different POS, derived from the data in the Iconicity Table.

6.2. The Child Speech corpus

After the described word-adjusting procedure, the children's vocabulary consists of 5313 words. The list of the 100 more frequently used words with their iconicity (when they exist in the Iconicity Table) is given in Appendix N.

Some statistical characteristics of the children speech are given in Table 22, along with their number of occurrences in the Child speech corpus.

Table 22. POS of the Words used by children in the dialogues, ascending following their frequency of use, with their rough POS-to-POS correspondence to the POS categories in the Iconicity Table.

POS in CHILDES	occurrences in the dialogues	Percentage	Words in the Iconicity Table are usually		POS in CHILDES	occurrences in the dialogues	Percentage	Words in the Iconicity Table are usually
n	47778	0,160	noun		pro:obj	5187	0,017	functionword
v	40486	0,135	verb		aux	4868	0,016	verb
co	25168	0,084	interjection		PRESP	4853	0,016	verb
pro:sub	21184	0,071	functionword		pro:poss:det	4731	0,016	functionword
det	19167	0,064	functionword		conj	4326	0,014	functionword
adv	16664	0,056	adverb		det:num	4174	0,014	functionword
pro	14622	0,049	functionword		qn	4099	0,014	functionword
n:prop	12673	0,042			pro:indef	4070	0,014	functionword
prep	12391	0,041	functionword		adv:wh	3838	0,013	adverb
adj	9505	0,032	adjective		pro:wh	1977	0,007	functionword
mod	9501	0,032	verb		adv:int	1647	0,006	adverb
cop	8403	0,028	verb		PASTP	1286	0,004	verb
pro:dem	7373	0,025	functionword		adv:tem	869	0,003	adverb
rel	5533	0,019	functionword		on	860	0,003	onomatopoeia
pro:obj	5187	0,017	functionword		post	743	0,002	functionword
aux	4868	0,016	verb		pro:poss	445	0,001	functionword
PRESP	4853	0,016	verb		int	278	0,001	interjection
pro:poss:det	4731	0,016	functionword		pro:refl	128	0,000	functionword

The proportion of the occurrences (during the overall period) of use of the different POS, classified following the original CHILDES POS-annotation is given in Fig. 14. The plot in the figure is mostly to illustrate that the different POS deployed in CHILDES (with very poor use of some of them), defers strongly from the list of POS deployed in the Iconicity Table.

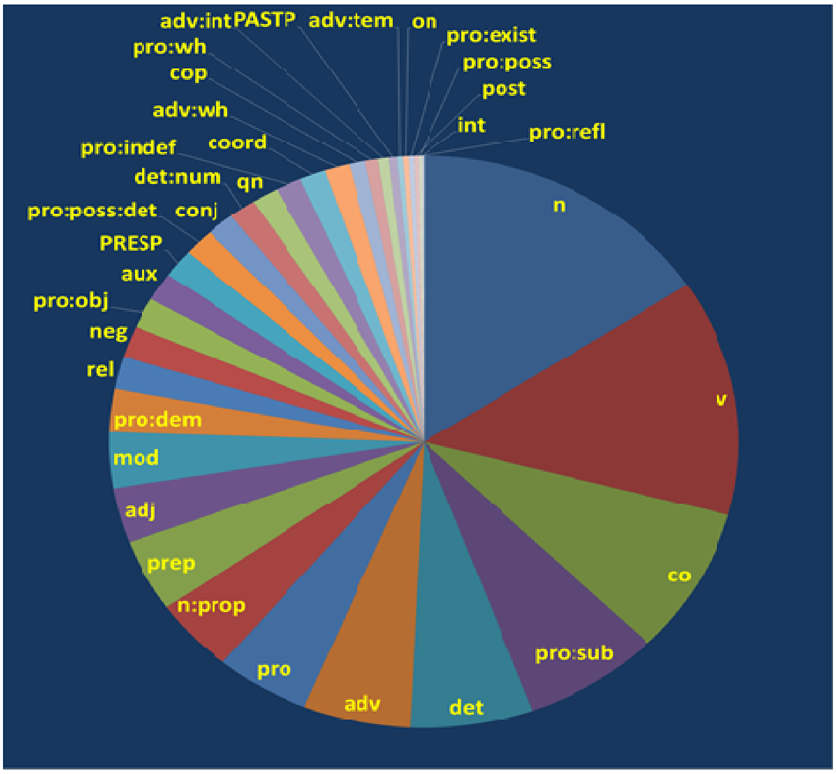


Figure 14. Word occurrences in the Child speech corpus, classified following the CHILDES POS-system

It is important to note that, naturally, the proportion of use of the different POS changes over the time of language acquisition. This is illustrated with the plot given in Fig. 15. The plot is based on ***all*** the occurrences of words in the child dialogues, independently of whether or not they exist in the Iconicity table. The annotation in CHILDES is “roughly” shifted to this in the Iconicity Table, as shown in Table 22. For example, the verbs on the plot comprise all the words which are originally annotated in CHILDES as verbs, past and present participles, modal and auxiliary verbs.

As the plot shows, the part of function words in the child speech grows significantly until 42 months. This could be explained with the content of the set “function words”, called also “grammatical words”, which comprises pronouns, prepositions, conjunctions, quantifiers, determiners, etc. lexical sub-categories (see [Winter et al. 2017]) necessary to form longer and grammatically correct expressions.

As it is seen from the plot, till 18 months, the children’s speech is dominated by nouns which part of use initially grows, together with a big, but decreasing participation of onomatopoeias, interjections, and communicators. During this period the portion of the used verb and function words grows quickly. At about 28 months, verbs and function words start representing the prevailing part of the used words. The use of adjectives stays low during the entire period under investigation. The use of adverbs is also low and displays an insignificant increase to 24 months.

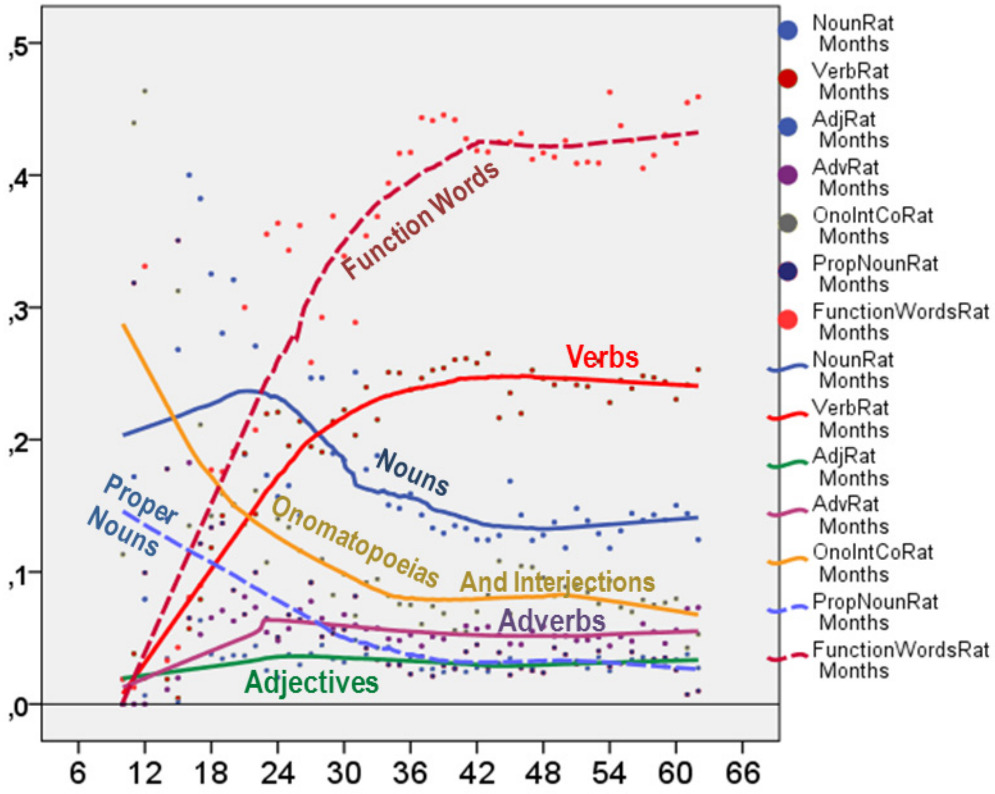


Figure 15. Development of the proportion of POS used in the child speech corpus, by month (number of each POS over the number of all the used words)

The plot in Figure 15. illustrates that, for the investigated period of language acquisition, the proportion of different POS involved in the children's speech changes significantly with the age of the children. This plot is, however, based on a roughly established correspondence of the POS system deployed in CHILDES to the POS given in the Iconicity Table.

The obvious problem which arises is how to decide that a given word was actually used by children as the POS for this word, given in the Iconicity Table. For example, the word *talk* is used by children as both – verb (21624 times) and noun (14076 times). It should not be fitted to the verb *talk* in the Iconicity Table when it is used as a noun. If it is used as a verb, it must be matched. The solution to such a problem crossed several obstacles and necessitated a separate approach which is discussed in the next section.

7. POS Transition, Mapping and Filtering

7.1. The Problem of Tying the two POS-systems

It is seen from the previous section that the two POS systems – this used in CHILDES and this postulated in the Iconicity table, are quite different. The problem, as explained, is how to decide that a word from the child speech is used as the POS, given in the Iconicity Table. To recall - the unique POS accorded to a word in the Iconicity Table is the POS as which the word is the most frequently used in the adults' speech corpus (SUBTLEX).

The first step for the POS-Transition was based on data observation. Table 23 gives the general POS-transition picture based on *Rough POS-to-POS* correspondences.

Table 23. Rough POS-to-POS correspondences of the POS-classification systems deployed in the data sources

CHILDES POS	Iconicity Table POS	CHILDES POS	Iconicity Table POS
n	noun	pro:sub	functionword
v	verb	det	functionword
mod	verb	pro	functionword
cop	verb	prep	functionword
aux	verb	pro:dem	functionword
PASTP	verb	rel	functionword
PRESP	verb	pro:obj	functionword
adj	adjective	pro:poss:det	functionword
adv	adverb	conj	functionword
adv:wh	adverb	det:num	functionword
adv:int	adverb	qn	functionword
adv:tem	adverb	pro:indef	functionword
pro:exist	adverb	pro:wh	functionword
int	interjection	post	functionword
on	onomatopoeia	pro:poss	functionword
		pro:refl	Functionword
n:prop	not accorded	co	Functionword

The transition based on such general Rough POS-to-POS correspondences (Table 23.) does not provide a clear-cut conversion of the used vocabulary and proved to be quite inaccurate. The querying of the child speech corpus showed, as illustrated in Fig. 16., the use of one and the same word as different POS happens unexpectedly often. For example, the word *laugh* (Fig. 16) is used as a noun and as a verb, and also - several uses of the word are annotated as communicators (following the context of the dialogue). And, for example, the word *last,* which is a verb in the Iconicity Table, has never been used by the children as a verb.

CHILDES WORD ADJU	AsNoun	AsVerb	AsAdjectiv	AsAdverb	AsOnoIr	AsFumc	AsPrope	AllUses	Brysetal_POS
last	120	0	1040	440	0	0	0	1600	verb
Lastname	0	0	0	0	0	0	16	16	
late	0	0	18	63	0	0	0	81	
later	0	0	0	1089	0	0	0	1089	adverb
laugh	81	513	0	0	0	135	0	729	verb
laughing	0	405	0	0	0	324	0	729	verb

Figure 16. Examples of use words as different POS - a query-result.

The cases in which a given word is, for example, a noun in the Iconicity Table and is used (and annotated) as a noun in the Child speech corpus have to be retained for the Total Match condition (see the example in Fig. 23.). One supposes that such coincidences represent the big majority of the cases. However, the statistical picture is a bit different from this expectation.

Table 24. Distribution of the word occurrences by POS in the Iconicity Table as percentage of POS of these word-occurrences in the Child Speech corpus, classified following the Rough POS-to-POS correspondence

	Appear in the Child speech corpus (% of occurrences):							
The words with POS In the Iconicity Table:	As Nouns	As Verbs	As Adjectives	As Adverbs	As FunctWords	As On, Int, Com	As Proper Nouns	AllUses
adjectives	2,6%	0,8%	82,8%	3,9%	9,3%	0,0%	0,7%	100,0%
adverbs	0,5%	0,0%	0,4%	79,1%	20,0%	0,0%	0,0%	100,0%
functionwords	0,0%	0,2%	0,0%	0,0%	99,8%	0,0%	0,0%	100,0%
interjections	0,0%	0,0%	0,0%	0,0%	100,0%	0,0%	0,0%	100,0%
nouns	59,6%	0,5%	0,6%	0,0%	39,1%	0,0%	0,2%	100,0%
onomatopoeias	47,9%	7,6%	1,1%	0,0%	3,8%	38,8%	0,8%	100,0%
verbs	0,3%	98,2%	0,0%	0,0%	1,4%	0,0%	0,0%	100,0%

Table 24. and the plot in Fig. 17. show how the words from each POS-group in the Iconicity Table are distributed (in percentage) as occurrences with different POS in the Child Speech corpus when applying the Rough POS-to-POS correspondence.

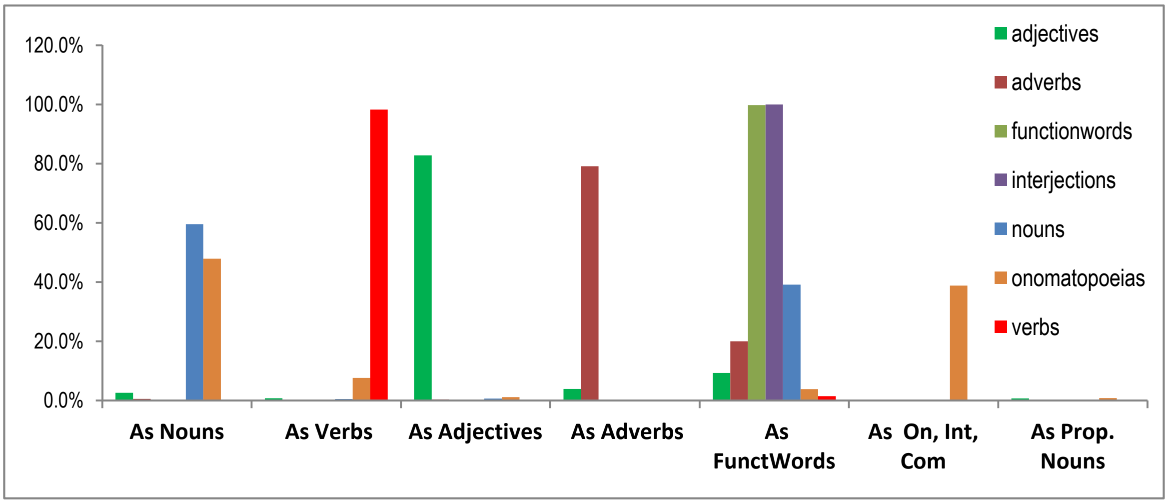


Figure 17. Distribution of the word occurrences by POS in the Iconicity Table as percentage of POS of these word-occurrences in the Child Speech corpus, classified following the Rough POS-to-POS correspondence

From this statistical picture, it can be concluded that the application of the Rough POS-to-POS correspondence leads to a result which has to be disambiguated. It is not clear at this point are the words used actually as different POS or there exist some systematic lack of correspondences between the two systems. The decision was to not modify or correct the original POS annotations in CHILDES. However, when performing the data-treatment, some errors in the original annotation were noticed. Examples are given in Table 25.

Table 25. Examples of supposed POS-annotation errors in the source

Text Marked ID	Speech	POS marked in CHILDES
dan42.t1i434	the mother has_to push him .	det|the ***v|mother*** mod:aux|has_to v|push pro:obj|him .
bba56.hal0284	musical chairs , we had .	***n|musical*** n|chair-PL cm|cm pro:sub|we v|have&PAST .
apr25.hig0561	close (.) must (.) must close eyes .	adj|close n|must n|must adj|close n|eye-PL .
laa43.bwd2017	nail polishing is really hard .	v|nail n:gerund|polish-PRESP cop|be&3S adv|real&dadj-LY adj|hard .
laa24.bwd0839	I write that name . [+ SR]	pro:sub|I v|write pro:dem|that v|name .
che26.cla0427	you name is Shem .	pro|you v|name cop|be&3S n:prop|Shem .
laa29.bwd0296	xxx <near you> [?] .	v|near pro|you .
lae24.bwd0167	No .	n:prop|No .
ad227.bro2127	take no(se) off ?	v|take v|nose adv|off ?
cou48.bfs3074	where is number one (.) where is it (.) where is it ?	adv:wh|where aux|be&3S v|number pro:indef|one rel|where cop|be&3S pro|it rel|where aux|be&3S pro|it ?
tre30.dem472	and here's your nuts ?	coord|and pro:exist|here cop|be&3S pro:poss:det|your adj|nuts ?

Such cases were stored for correction. Their number is small and can be evaluated to less than 0,1%.The general strategy undertaken is to respect the original POS annotation in CHILDES even though it can contain some small percentage of inaccuracies. In the linguistics verification procedure, however, all such cases were accounted as errors (see section 8.2.).

7.2. The POS-Transition Markers

If in a given dialogue, at a given place, a given word appears (is originally annotated) like one POS and in another places the same word appears as some other POS, only the occurrences of the word which match the POS in the Iconicity table have to be taken into consideration for the Total Match condition (see also the example in Fig. 23.). This seems clear to be solved. However, the examples in Table 26 show why the transition task has necessitated another approach.

Table 26. The problem of the POS-Transition and examples of POS-Transition Markers

Dialogues Corpus			Iconicity Table		Remark
CHILDES POS Original	CHILDES Word Adjusted	**POS Transition Marker**	POS in the IconicityTable	Word in the IconicityTable	Number of occurrences
co	sorry	**CoToAdj**	adjective	sorry	43
co	so	**coToAdv**	adverb	so	274
conj	so	**conjToAdv**	adverb	so	131
prep	about	**prepToFun**	functionword	about	216
conj	after	**conjToFun**	functionword	after	31
adv	all	**advToFun**	functionword	all	148
adv:int	all	**advToFun**	functionword	all	201
pro:indef	all	**proToFun**	functionword	all	109
adv	before	**advToFun**	functionword	before	29
prep	before	**prepToFun**	functionword	before	13
pro:dem	that	**proToFun**	functionword	that	3994
adv	that	**advToFun**	functionword	that	79
co	hello	**coToInt**	interjection	hello	74
n	hello	**nToInt**	interjection	hello	16
co	no	**coToInt**	interjection	no	4991
co	thanks	**coToNoun**	noun	thanks	33
co	cockadoodledoo	**CoToOnom**	onomatopoeia	cockadoodledoo	6
co	look	**coToVerb**	verb	look	392
co	wait	**coToVerb**	verb	wait	67
adv:wh	what	**advToFun**	functionword	what	101

The words which are systematically annotated in the original CHILDES POS annotation as one unique POS, are often annotated in the Iconicity table as another POS. For example, following CHILDES, *no* as a communicator, and following the Iconicity Table, this word is an interjection (Table 26). At the same time, as illustrated in Table 26, not all the communicators following CHILDES are interjections following the Iconicity Table (the examples in Table 26 are of the words *thanks, cockadoodledoo, look* and *wait)*.

The problem of tying up the two POS systems was solved by inserting in the Child speech corpus metadata which expresses how to link the original CHILDES POS-annotation of a given word-occurrence to the corresponding word in the Iconicity Table. This metadata is called here ***POS-Transition Markers***. The *POS-Transition Markers* were stored separately, in an additional field, individually for each word-occurrence of the children's dialogues. The steps to perform such an annotation are further shortly explained with some examples.

The analysis of the situation displayed by the examples in Table 26 showed that the POS-transition has to be made based word-representatives. 941 words (18%) out of the 5365 words isolated in the child speech vocabulary are used as more than one POS.

There are 4424 words which are used as a unique POS in the Child speech corpus. The querying of the corpus showed that all the occurrences of the word *no* (for example) are annotated as communicators, while in the iconicity table *no* is an interjection. Such systematic cases were solved based on word-select queries with a subsequent introduction of a POS-Transition Marker for the entire set of the word-occurrences in the corpus. In the case of *no*, the marker is *coToInt*.

Cases as, for example, this of the word "*so*" (Table 26), which has been used in the children dialogues as both - a communicator and as a conjunction had to be mapped to the word so, which in the iconicity table is an adverb. In this case, all the word-occurrences have to be first selected by querying the corpus and next - the POS-Transition Marker has to be accorded in correspondence of the original CHILDES annotation. For the example of the word "*so",* the POS-Transition Markers are *coToAdverb* and *conjToAdvreb*. Table 27. provides the example of the words *here* and *there* which exist (both) with three different POS-taggers in CHILDES (*adv*, *pro:dem* and *pro:exist*). These two words had to be accorded POS-Transition markers as follows – the occurrences of the word *here* - transition "to adverb" and the occurrences of the word *there* – transition "to functionword".

Table 27. Examples of Word-representative sets with their POS-subsets – a query result.

Occurances	CHILDES WORD A	CHILDES POS ORIGI	Transition Marker	POS in Iconicity Ta	Word in Iconicity Ta
1778	here	adv	adv	adverb	here
315	here	pro:dem	proToAdv	adverb	here
304	here	pro:exist	ProToAdv	adverb	here
2273	there	adv	advToFun	functionword	there
610	there	pro:dem	proToFun	functionword	there
506	there	pro:exist	ProToFun	functionword	there

This approach spreads the set of occurrences of a given word in the Child speech corpus into subsets depending on their original POS-annotation. All such cases were solved using word-select queries. The 941 words which are used as multiple POS were identified, their POS subsets – isolated and the word occurrences in each POS subset – discerned with the corresponding POS-Transition marker. This necessitated treating the corpus nearly manually. The general form of the POS Transition Markers is given in Fig. 18:

One main concern was to not accord an erroneous transition marker. For example, if in the child speech corpus a given word is used as a verb and as a noun and in the Iconicity Table this word is a noun, the transition marker “verbToNoun was not accorded in the cases in which the word is annotated in CHILDES as a verb. In fact, a big part of such word-occurrences had to be examined in the context of the sentence-utterance and even of the dialogue.

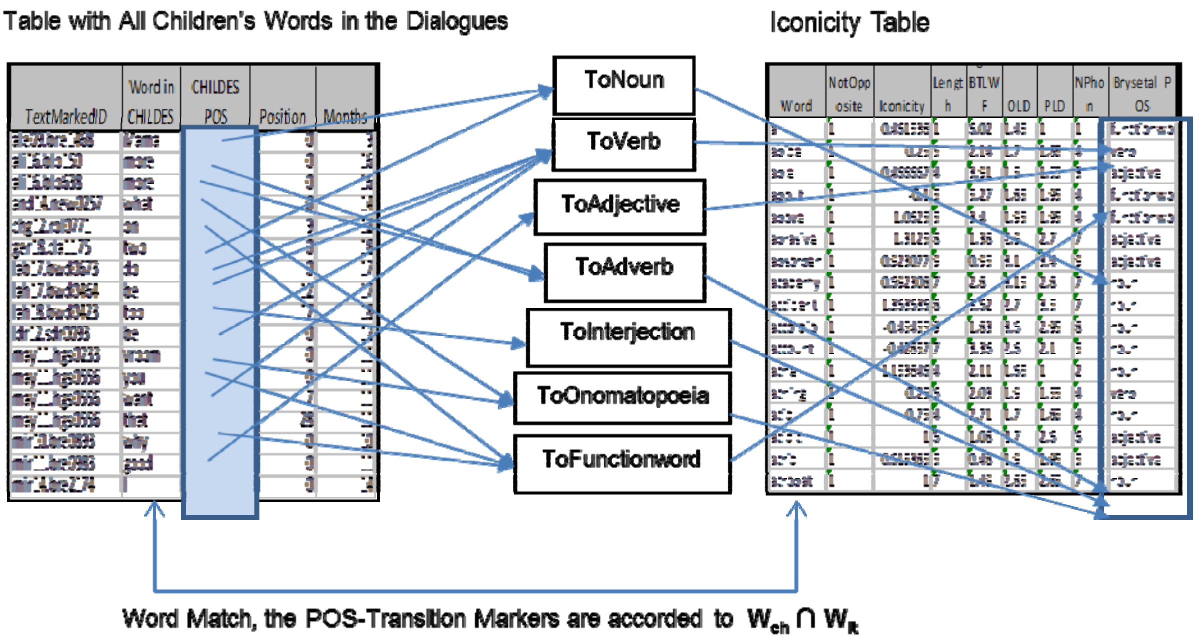


Figure 18. POS-Transition Markers.

The use of Proper Nouns in the children dialogues represents a separate case. There are no Proper Nouns in the Iconicity Table. However, words like “Mommy” and “Daddy” exist, classified as nouns. That has necessitated the retrieval of all the cases of use of such words, in the context of the dialogues, as some occurrences are to name movies, toys, etc. without the necessity of knowing the meaning of the noun itself. Table 28 gives cases where the POS Transition Markers was accorded to Proper Nouns.

Table 28. The cases of transition from Proper Nouns to Nouns.

Dialogues Corpus			Iconicity Table		Remark
CHILDES POS Original	CHILDES Word Adjusted	POS Transition Marker	POS in the Iconicity Table	Word in the IconicityTable	Number of occurrences where the Transition Marker was accorded
n:prop	Aunt	n:propToNoun	noun	aunt	1
n:prop	Baby	n:propToNoun	noun	baby	3
n:prop	Bear	n:propToNoun	noun	bear	4
n:prop	Bird	n:propToNoun	noun	bird	12
n:prop	Dad	n:propToNoun	noun	dad	219
n:prop	Daddy	n:propToNoun	noun	daddy	720
n:prop	Father	n:propToNoun	noun	father	7
n:prop	Mommy	n:propToNoun	noun	mommy	2054
n:prop	Mother	n:propToNoun	noun	mother	15
n:prop	Sister	n:propToNoun	noun	sister	2
n:prop	Uncle	n:propToNoun	noun	uncle	33

All the accorded POS-Transition Markers are given in Appendix M. As illustrated in Fig. 18, the transition was: “word annotated as several CHILDES POS - to – the same word, with one precise POS given in the Iconicity Table.

It should be noted that the procedure of linguistic verification, described in section 8.2. has accounted all the words which can be used as several POS as errors and the POS-transition was discarded because it is difficult and even impossible to evaluate are there missed words in the ~309 000 word occurrences in the Child speech corpus. However, all the statistical investigations are performed after the POS-filtering and the evaluated error is significantly reduced.

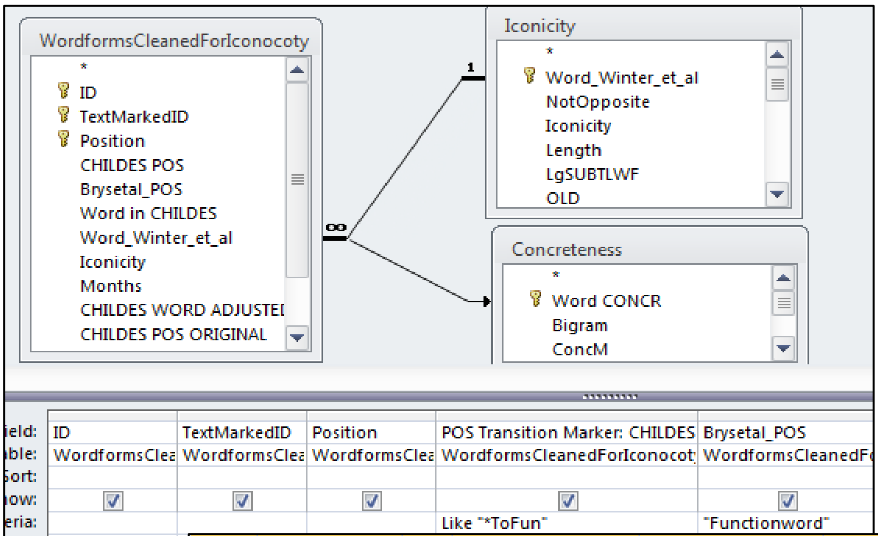
7.3. The POS-Filtering and the Two Matching Conditions - Word Match and Total Match

All the accorded POS-Transition Markers are given in Appendix M. As illustrated in Fig. 18, the transition was: “word annotated as several CHILDES POS - to – the same word, with one precise POS given in the Iconicity Table.

It should be noted that the procedure of linguistic verification, described in section 8.2. has accounted all the words which can be used as several POS as errors and the POS-transition was discarded because it is difficult and even impossible to evaluate are there missed words in the ~309 000 word occurrences in the Child speech corpus. However, all the statistical investigations are performed after the POS-filtering and the evaluated error is significantly reduced.

7.3. The POS-Filtering and the Two Matching Conditions - Word Match and Total Match

The POS-Transition Marker accorded to each child's word was used for filtering the POS correspondence. The filtering queries are based on both – the POS-Transition marker and the POS in the Iconicity Table. This is shown in Fig. 19 with the example of the filter for Functionwords. As it is seen, the query selects from the Child speech corpus the words for which the POS-Transition Marker terminates with “ToFun” and the words from the Iconicity Table, which are “function words”.



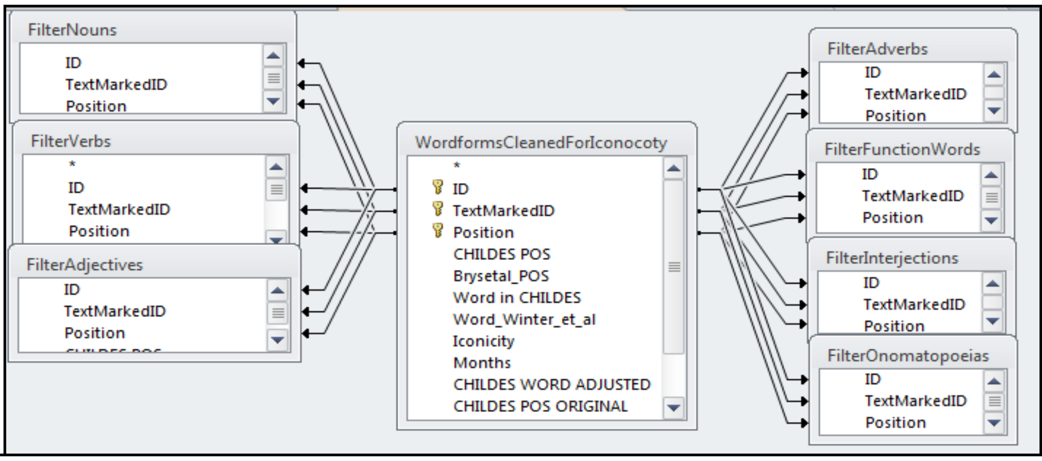
ID	TextMarked	Po	POS Transtion M	Brysetal_POS	Word	Word	FuncWordIconicity
8838	abe51.hsl0426	36	prepToFun	functionword	in	in	1,461538462
5887	abe51.hsl0426	44	proToFun	functionword	there	there	0,461538462
3952	abe51.hsl0441	31	proToFun	functionword	his	his	0,588235294
2439	abe51.hsl0441	0	conjToFun	functionword	and	and	0,5625
5888	abe51.hsl0441	8	proToFun	functionword	this	this	0,133333333
3035	abe51.hsl0443	28	proToFun	functionword	more	more	1,125
15057	abe51.hsl0447	22	proToFun	functionword	they	they	0,454545455
2440	abe51.hsl0447	0	conjToFun	functionword	but	but	0
14268	abe51.hsl0454	42	detToFun	functionword	a	a	0,461538462
15058	abe51.hsl0454	11	proToFun	functionword	I	I	3,181818182
3953	abe51.hsl0462	19	proToFun	functionword	my	my	1,5

Figure 19. Filtering query for the POS-transition to functionwords.

Such filtering queries were created for each POS in the Iconicity Table. These queries were further used to according to the children's words the parameters related to Iconicity, Length, Frequency, Orthographic neighborhood, and Concreteness. The filtered this way corpus was stored in a novel table, in a separate database (in order to not violate the 3th NF), called Iconic Corpus. It contains all the words used by the children over the course of the dialogues, annotated in two ways:

- Word Match condition: the parameters of Iconicity, Length, Frequency, Orthographic neighborhood and Concreteness of the corresponding words from the used data sources;

- Total Match (POS strict use) condition: all these parameters, separately by POS, if and only if the POS-transition filters classified the word-occurrence in its concrete use as being used as the corresponding POS in the Iconicity Table.
- Figure 20 Illustrates the result of the described double-step approach and gives an example of the annotation obtained for one sentence-utterance in the two matching-conditions.
- As seen in Fig. 20. the Iconic Corpus is ready to use in a) the Word March condition and b) the Total Match (POS-filtered) condition of the annotated child speech.



a) Word Match condition: Values of the word-parameters which are given in the Data Sources:

M	ID	TextMarkedID	Po	CHILDES	Brysetal_POS	IconicityTa	Le	Freq	OrtoTa	Concr
51	37929	abc51.hsl0123	0	daddy	noun	1,428571429	5	3,98	1,85	4,24
51	6580	abc51.hsl0123	8	will	verb	0,214285714	4	5,03	1	2,64
51	10972	abc51.hsl0123	21	you	functionword	-0,4	3	6,33	1,85	4,11
51	31194	abc51.hsl0123	29	tie	verb	0,461538462	3	3,36	1,05	4,81
51	3932	abc51.hsl0123	35	my	functionword	1,5	2	5,54	1,7	2,42
51	37930	abc51.hsl0123	51	shoe	noun	-0,3125	4	3,19	1,4	4,97

b) Total Match Condition (POS strict use): Values which crossed the POS-transition filters:

CHILDES	NoounIconic		Nou	Nou	Nc	Ver	VerbIconic	Vei	Verb	Veri	Vei	Fur	Fu	Fi	Fui	Fui	Fu
daddy	1,428571429	5	3,98	1,85	4,24												
will						will	0,214285714	4	5,03	1	2,64						
you												you	-0,4	3	6,33	1,85	4,11
tie						tie	0,461538462	3	3,36	1,05	4,81						
my												my	1,5	2	5,54	1,7	2,42
shoe	-0,3125	4	3,19	1,4	4,97												

Figure 20. The result of querying of the two steps of the peformed data fusion - Word match and POS-filtering

8. The obtained corpus

8.1. Short Description of the Corpus

The obtained corpus comprises 629 dialogues of children aged between 9 and 62 months, The corpus contains 125, 353 sentence-utterances with 309 116 word-utterances annotated with POS following the original POS system in CHILDES. After the word-adjusting procedure, 239 592 (78%) of the children word utterances correspond to words in the Iconicity Table and are annotated also with POS following the system adopted in this data source as well as with values for their Iconicity, Orthographic Neighborhood, Length and Frequency. 284,016 of the word-occurrences in the child Speech corpus were found to correspond to words in the Concreteness Table and annotated with data on Concreteness. The POS-transition and filtering “discovered” a compatible use as POS for 232 915 word-occurrences, that is - 75% of the words used by the children that correspond to the Total Match condition are annotated with Iconicity, Length, Frequency and Orthographic Neighborhood. As the words are not assigned with POS in the Concreteness Table, their POS were taken equivalent to the POS in the Iconicity Table.

The data is organized in a relational database (in 3NF) which allows investigating the dependencies between the enumerated features of the words and the characteristics of the children such as age, gender, identity, and consulting data from the dialogues and the corpora in CHILDES which are used as a source. (Figure 21.).

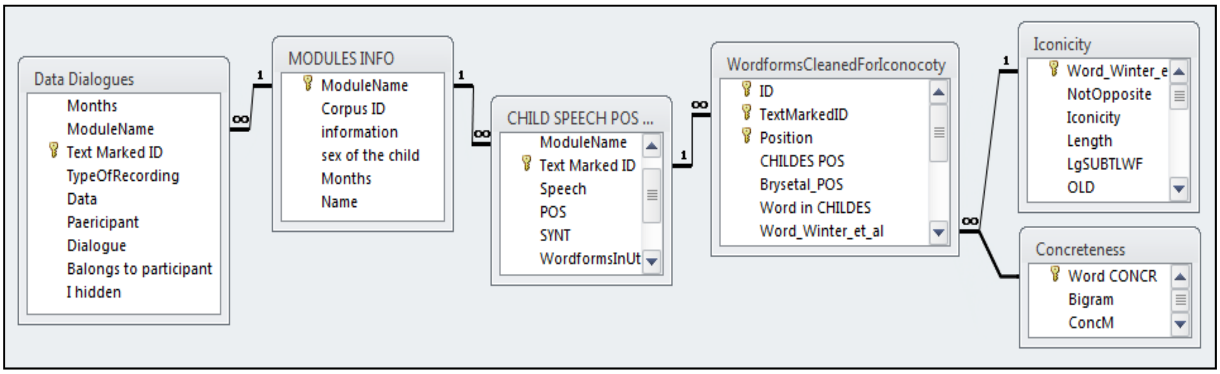


Figure 21. The relational structure of the corpus.

The steps performed to obtain the parameters of Iconicity, Length, Orthographic neighborhood, Frequency, and Concreteness, accorded to the children’s words can be tracked via the stored in separate fields intermediate metadata used to perform the fusion and the filtering.

The technical procedure described here does not prove that the result is correct. As discussed, several steps performed for the data-fusion necessitated consulting the content of all the data sources, the sentence-utterances and the context in dialogues. An additional check of the result had to be performed.

The corpus underwent a verification procedure for linguistics consistency by specialists in linguistics and psycholinguistics from the Language Processing Lab at the University of Calgary, Canada.

8.2. Linguistics verification of the corpus

The resulting corpus of annotated child speech has been checked for linguistics consistency by the collaborating specialists in linguistics and psycholinguistics from the Language Processing Lab at the University of Calgary, Canada, under the guidance of dr. Penny Pexman.

The task was to evaluate the result for linguistics consistency. As mentioned, this task can be properly performed only when taking into account the context of the word within the sentence-utterance and, often, the context of the sentence within the dialogue.

The procedure was the following: 25 randomly chosen dialogues (about 4% of the dialogues and 5% of the word occurrences in the corpus) were submitted to the evaluators in two forms. The first represents the whole texts of the dialogue, containing the original text from CHILDES together with the used in the database sentence-identifiers (as given in Appendix B). The second form presents the result for these dialogues after the described here corpus-annotating procedure. This form was generated by querying the annotated corpus with the identifiers of these precise dialogues and presented under the form of a report as shown in Fig. 22,. where the example is for one sentence from the dialogue given in Appendix B.

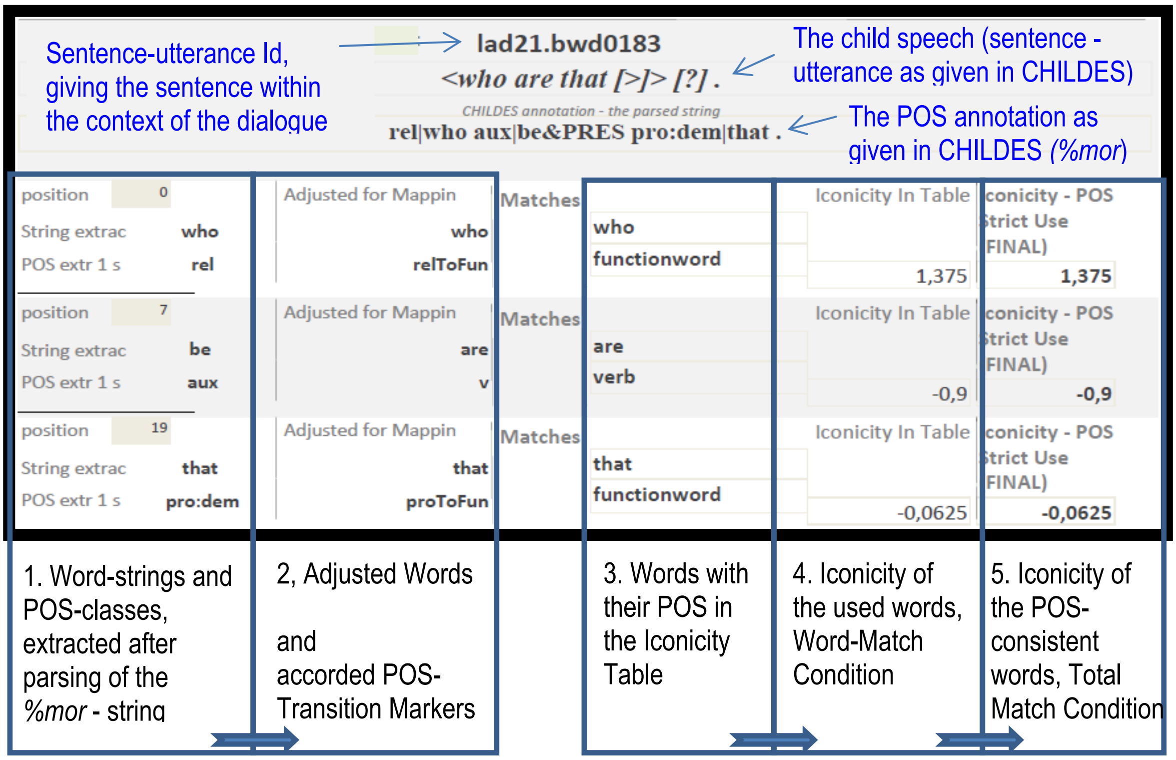


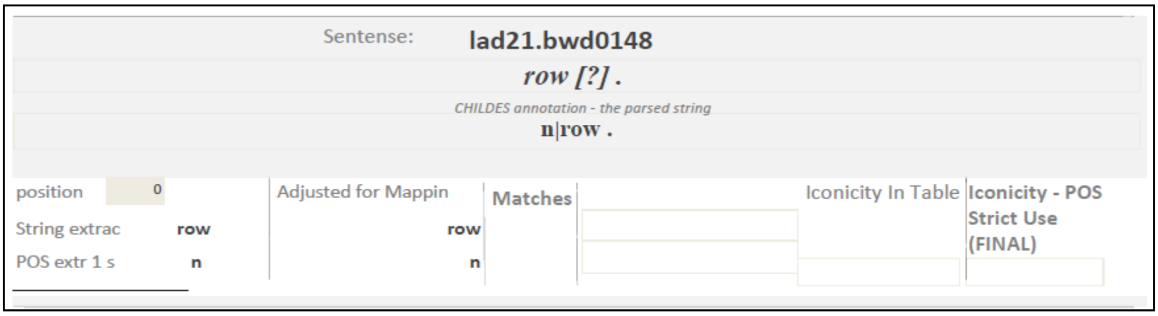
Figure 22. The format of the DB reports, submitted for evaluation of the annotated corpus.

As illustrated in Fig. 22., the content available in the initial data-sources is visualized, as well as the results of the intermediate steps of the procedure for attaining the annotation of the child speech. This visualization allows identifying the step on which a detected error has occurred.

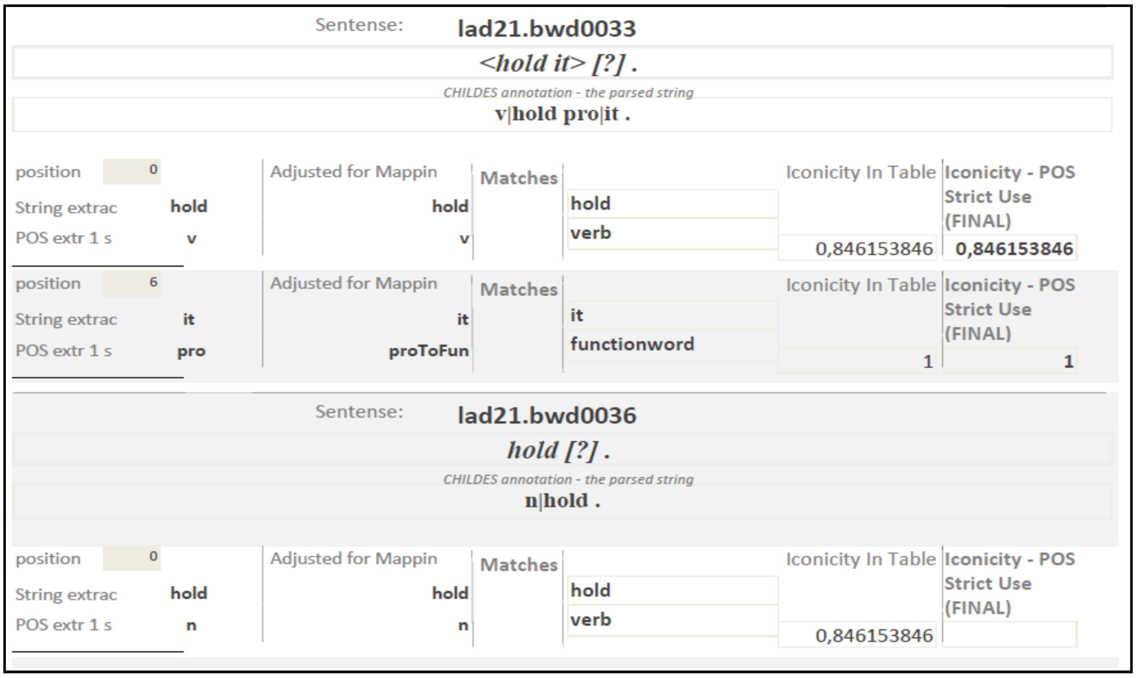
The task was to read the sentence-utterance as given in the original (the sentence-utterance as in CHILDES) and as adjusted word by word (column 2 in Fig. 22.). As also illustrated in Fig. 22., columns 3 and 4 contain the corresponding words in the Iconicity Table, with their respective POS and iconicity norms. This corresponds to the Word-Match condition. If the child's word does not match a word in the Iconicity Table, the content in columns 3 and 4 does not appear (see Fig. 23. a).

The content of column 5 appears depending on the result of the queries for POS-filtering. If a given word does not cross the POS-filtering, its iconicity norm (and all the other characteristics from the data sources) is not assigned. Column 5 in Fig. 22. corresponds to the Total Match condition.

As illustrated in Fig. 23.b., one and the same word can cross or not the POS-filtering. The examples in Fig 23. are also for sentences from the dialogue in Appendix B.



a) The used by the child word does not exist in the Iconicity Table.



b) All the used words exist in the Iconicity Table, but do not always cross the POS filtering

Figure 23. Examples of not-respecting a) the word-match condition and b) the total match condition

The verification procedure took several months, every linguistics inconsistency was taken into account and reported as an error. Here, as an example, the report for one dialogue (hel52.gl.) provided by Jennifer Williamson, a research assistant in the Language Processing Lab at the University of Calgary, Canada:

Dialogue ID: hel52.gl.

There were 564 iconicity trials and 123 errors or unusual cases:

- 14 cases involving "what" or "right" where the word was coded as a function word, but was being used as an interjection in context
- 8 cases involving irregular verbs being coded as a regular verb (eg. "did" becomes "do")
- 3 cases where "gimme" was broken down into "give me" and thus incorrectly had an iconicity value
- 10 cases involving typos in the table (eg. "ma" for "may")
- 37 cases involving contractions where one or more component words had an iconicity value
- 9 cases involving numbers where the POS was unclear
- 11 cases involving compound nouns ("potato chips", "ice cream", "Happy Birthday", various numbers)
- 5 cases involving words from the transcript missing from the table
- 6 trials involving possible POS errors ("no six", "not like it", "kick ball", "do you like...", “so alright", "not me")
- 15 indeterminate trials involving the word "rock"
- 1 trial where "there" was used as an adverb but was coded as a function word
- 1 case where the table did not match the transcript
- 1 parsing error ("thank_you")
- 1 possible parsing error ("anymore")
- 1 case involving an irregular plural ("feet" was coded as "foot")

As it can be seen, all the conflicts have been taken into account, including the parsing errors, the lack of correspondence in the POS systems, the errors in the original POS annotation as given in CHILDES, etc. This was done in order to evaluate the maximal possible error which is potentially included in the annotated corpus when having these data sources.

Further, an examination of the reports and a counting procedure over the checked dialogues was performed by David Sidhu, Ph.D. candidate in the Department of Psychology at the University of Calgary, Canada, whose research work in psycholinguistics is related to such data (see e.g., [Sidhu, D. M., & Pexman, P. M. (2017)], [Sidhu, D. M., & Pexman, P. M. (2018)], [Westbury et al., (2017)].

The result showed that the identified this way errors encompass 19,7% of the word occurrences. This percentage reflect the maximal error susceptible to arise from the data sources.

It has to be accentuated, again, that this evaluation takes into account the maximal possible value of the error without taking into consideration the word adjusting, the POS filtering and the rules respected in these procedures, such as considering the plural forms as singular forms in order match the words to the Iconicity Table.

This evaluation shows that the statistical analysis of language-related features and parameters related to psycholinguistics based on the corpus will provide a reliable result.

8.3. Some Descriptive Statistics based on the obtained corpus

The used by the children words in the Child Corpus form a vocabulary of 5313 words. This vocabulary starts with a small set of words which increases with novel members over the time of acquisition. The novel words included in the vocabulary are not equivalently distributed as POS. This was checked when comparing the overall vocabulary content (that is – all the dialogues of all the children were included) month by month. As illustrated on the plot in Figure 24, for all the POS, the rate of inclusion of novel words grows till 24 -26 months, after which point it starts slowing down. This phenomenon is very strongly manifested for the nouns. One may imagine that the 24 -26 months point corresponds to a saturation of, figuratively speaking, some “sufficient vocabulary”, which is necessary to express children’s thoughts. This result is consistent with specialized studies (see e.g., [Cairns, H. S., & Fernández, E. M. (2010)], [O'grady, W., & Archibald, J. (2015)]) reporting that after 12 months children’s expressive vocabulary grows to approximately 50 words in the following six months. The result here suggests that the period of extensive growth is even longer - one year, that is – from 12 to 24 months.

The rate of inclusion of novel Proper nouns in the vocabulary is related to the specificities of the concrete children environments and their intersections, so the dissimilar pattern for proper nouns, displayed on the plot, is not astonishing.

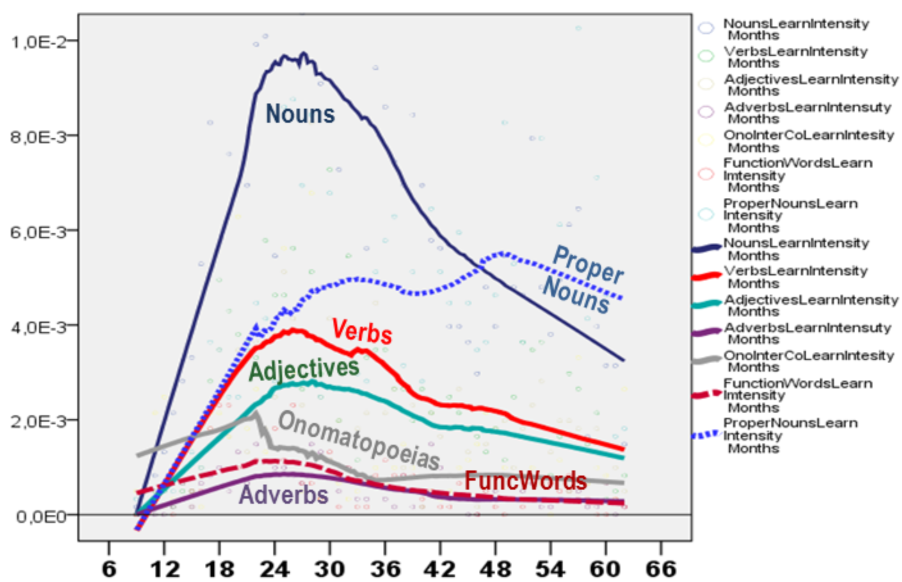


Figure 24. Rates of extension of the Vocabulary over time – inclusion of novel words by POS

In the first months of production, between 9 and 12 months, the rate of expansion of the vocabulary by novel nouns, onomatopoeias and interjections is considerably higher this of the other POS. The inclusion of novel representatives of these grammatical categories drastically drops after 24 months. It has to be noted that the inclusion of novel verbs is quite high - higher than this of novel adjectives. This corresponds to latest findings concerning the ease of gaining the meaning of verbs [Gogate, L., & Maganti, M. (2017)] reporting that long before they talk, infants acquiring a noun-dominant language, learn synchronous word-action relations.

The inclusion of novel functionwords in the vocabulary starts intensively at the beginning and further remains quite monotonic. This is not surprising either because, as discussed, functionwords are a set containing pronouns, prepositions, conjunctions, etc., which, in the proposed above interpretation of “sufficient vocabulary”, can be seen as the inclusion of novel functionwords serving to relate the other POS into longer sentence-utterances.

The behavior of the vocabulary concerning the rate of inclusion of novel adverbs is a bit surprising.

The annotated corpus provides the possibility to investigate how to develop over the time of language acquisition the parameters of the used by the children words - this development is given on the plot in Fig. 25 The values of the displayed parameters cannot be compared between them, but the tendencies are visible. It is seen that children’s word-occurrences, over time, start including more and more words which are frequent in the adult’s language.

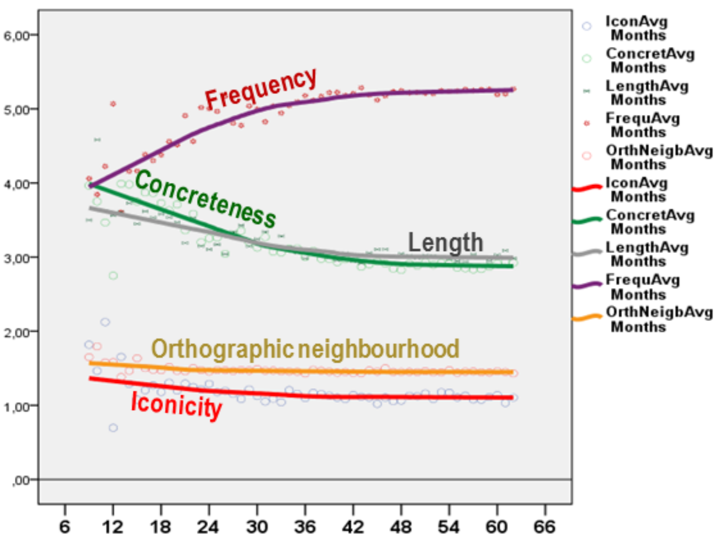


Figure 25. Development of the Average Iconicity, Length, Frequency, Orthographic Neighborhood and Concreteness over the time of language acquisition

The decrease of the average Length of the used words seems initially surprising, but when consulting the rates of use of different POS, one simple explanation of this statistical result can be proposed: children's expressions contain more and more functionwords over the time of acquisition. As mentioned, functionwords contain pronouns, conjunctions, and prepositions, etc. which are short. This influences the average length of the used words.

The average orthographic neighborhood changes slightly with age. It has to be additionally studied to what extent this statistical picture is consistent with the suggestions given the specialized studies, for example those derived from the acquisition of sign languages [Caselli, N. K., & Pyers, J. E. (2017)], namely - that the orthographic neighborhood is a factor which influences the acquisition of spoken languages.

One important observation concerns the growth of the average level of Abstraction (taken as opposite to Concreteness) of the used words and the development of the words' average Iconicity. As it is seen from the plot, the concreteness of the used words drops down in a noticeable way. This statistical picture is consistent with the expectations and with the results reported in the specialized studies.

The average iconicity also shows a noticeable decrease. This is in support of the iconicity bootstrapping hypothesis in language acquisition.

However, as shown in Fig. 15, the POS-content of the children's speech changes significantly over the time of acquisition. As explained, the recent studies (e.g., [Winter et al. 2016], [Perry et al. 2017]) report that the average iconicity of the different POS is different (see Table 21). These facts necessitate a more attentive approach to the Iconicity parameter. Figure 26 displays a plot of the development of the average Iconicity, separately for the different POS. It has to be pointed out that the data used for the plot is averaged when taking into account all the words which have an Iconicity norm accorded, independently of its value (positive, negative or zero).

The obtained statistical picture shows that children may be "rely" on the iconicity of the different POS in a different way. The iconicity norm averaged for the occurrences of Onomatopoeias and Interjections (which have a very high iconicity, see Table 21), does not display big differences over the entire period under investigation. Verbs and nouns, on the contrary, are gradually involved in children's speech with word-occurrences which average iconicity decreases. Initially the average iconicity of the involved verbs and nouns is high; further, it constantly decreases. The average iconicity of the used nouns is slightly higher than this of the used verbs. This is somewhat surprising because the mean iconicity of verbs is higher than this of nouns (see Table 21). When comparing the iconicity for nouns and verbs from the Iconicity Table (Fig. 13.) and these of the used in the child speech corpus (Fig. 26.), it ca**n** be supposed that small children use frequently a set of verbs which are less iconic compared to the entire set of verbs in the Iconicity Table.

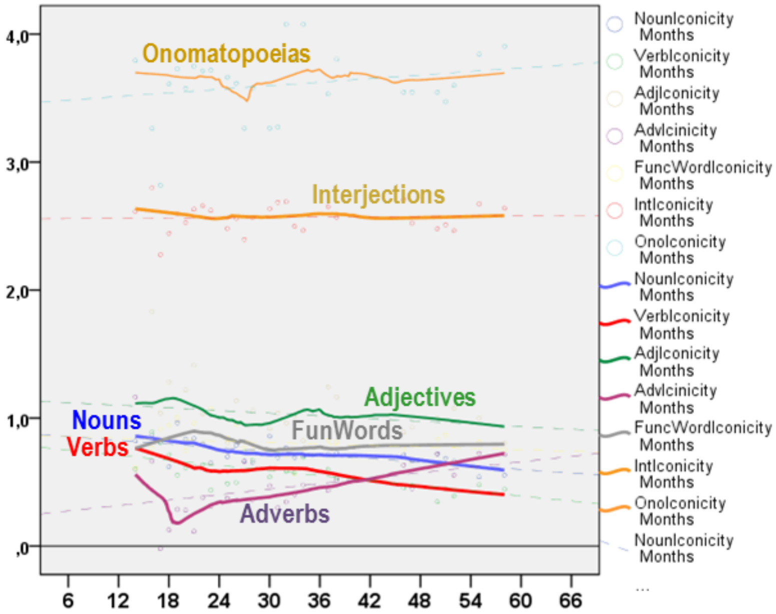


Figure 26. Development of the Average Iconicity of the used words, by POS, over the investigated period

The reliability of the result displayed for adjectives has to be attested using accurate statistical approaches as the number of adjectives in the children's speech is quite low (see Fig. 15.).

The line-pattern for functionwords displays a growth at the beginning. The reasons for that could reside in the big variance in iconicity for functionwords (see Table 21) and in the fact that this lexical category contains multiple sub-categories, some of which are maybe included in children's speech more intensively at the beginning of speech production whereas others are included later on (see Fig. 15.).

At this point, the corpus-result cannot be analyzed using unequally descriptive statistics without risk of drawing wrong conclusions. It is, however, astonishing to observe that adverbs show a constantly growing pattern of iconicity. Initially, the word-occurrences of adverbs which enter the children's speech are with a mean iconicity which is considerably lower than the mean iconicity accorded for adverbs (following the Iconicity Table). The occurring adverbs, in average, start comprising more and more iconic ones, a process that continues till the end of the period under investigation, when the mean iconicity of the occurring adverbs is significantly higher than the average iconicity for adverbs. One possible hypothesis which can be proposed at this point is that adverbs are more "difficult" to be metalized compared to nouns and verbs and their use by children necessitates an iconicity-assistance during the overall period under investigation. This is coherent with the hypothesis proposed by the author in [Slavova (2017)], namely, that the different lexical categories are language-realizations of mental categories with different cognitive loads related to their level of abstractness. The hypothesis proposed there was that the adverbs' cognitive complexity is due to their higher level of abstractness.

The comparison of the plots for the mean parameters (Fig. 27) with the changing proportions of the used POS during the acquisition (Fig. 15) makes appearing some uncertainty – perhaps the developing patterns of the annotated parameters are due simply to the changes in the POS proportion over the time of acquisition. Such and other similar hypothesis based on observation of the statistical patterns should be attentively studied using appropriate statistical models. Moreover, the displayed here descriptive patterns are based on data collected from different children and do not take into consideration individual differences. For the moment, the profiles of nouns, verbs, and adverbs displayed in Figure 27 show that the further investigation has to be performed when taking into account several of the annotated parameters, in their interdependencies.

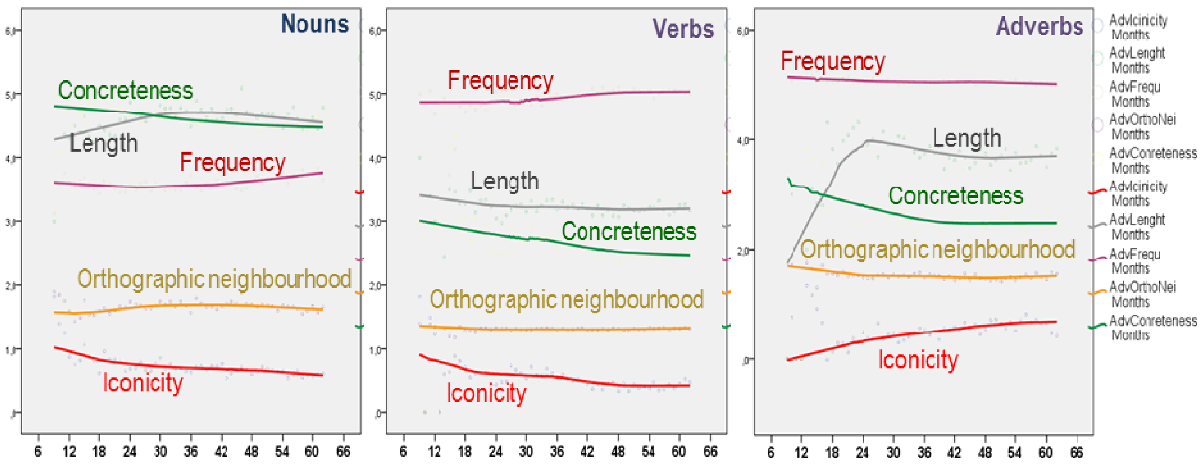


Figure 27. Plots of the averaged (per month) values for iconicity, concreteness, length, orthographic neighborhood and frequency, for three POS

9. Closing Remarks

An extensive corpus of 309 116 words extracted from free dialogues in English of children aged between 9 and 62 months was annotated with words' Iconicity, Concreteness, Orthographic neighborhood, Length, and Frequency. The obtained corpus was evaluated for linguistics consistency and proved to be reliable for statistical analysis. After the procedure of verification, the corpus was provided for detailed investigation of the iconicity parameters in the child speech to the collaborating specialists in psycholinguistics from the Department of Psychology and the Language Processing Lab at the University of Calgary, Canada.

The corpus is ready to serve as a basis for studying the parameters of child speech related to Iconicity, Concreteness, Orthographic neighborhood, Length and Frequency.

Acknowledgment

The author is thankful to Dr. Penny Pexman, head of the Language Processing Lab, to David Sidhu, Ph. D. candidate, and to Jennifer Williamson, research assistant, all of them from the University of Calgary, Canada, who confirmed the interest to create the corpus and continuously advised the linguistics and psycholinguistics aspects of the work on it. These collaborators achieved the task of evaluation of the corpus' maximal error.

References

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., ... & Treiman, R. (2007). The English lexicon project. Behavior research methods, 39(3), 445-459.

Brysbaert, M., New, B., & Keuleers, E. (2012). Adding part-of-speech information to the SUBTLEX-US word frequencies. Behavior research methods, 44(4), 991-997., 44, 991–997.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. Behavior research methods, 46(3), 904-911.

Brysbaert, N. K., & Pyers, J. E. (2017). The Road to Language Learning Is Not Entirely Iconic: Iconicity, Neighborhood Density, and Frequency Facilitate Acquisition of Sign Language. Psychological science, 28(7), 979-987

Cairns, H. S., & Fernández, E. M. (2010). The Fundamentals of Psycholinguistics. Wiley-Blackwell.

Caselli, N. K., & Pyers, J. E. (2017). The road to language learning is not entirely iconic: Iconicity, neighborhood density, and frequency facilitate acquisition of sign language. Psychological science, 28(7), 979-987.

CHILDES - the child language component of the TalkBank - system for sharing and studying conversational interactions. http://childes.psy.cmu.edu/ or http://childes.talkbank.org/

Gogate, L., & Maganti, M. (2017). The Origins of Verb Learning: Preverbal and Postverbal Infants' Learning of Word–Action Relations. Journal of Speech, Language, and Hearing Research, 60(12), 3538-3550.

Hausser, R. 1989. Principles of computational morphology. Pittsburgh, PA, Carnegie Mellon University, Laboratory for Computational Linguistic, Technical report.

MacWhinney, B. (2014). The CHILDES project: Tools for analyzing talk, Volume II: The database. Psychology Press.

O'grady, W., & Archibald, J. (2015). Contemporary linguistic analysis: An introduction. Pearson Canada.

Perry, L. K., Perlman, M., & Lupyan, G. (2015). Iconicity in English and Spanish and its relation to lexical category and age of acquisition. PloS one, 10(9), e0137147.

Perry, L. K., Perlman, M., Winter, B., Massaro, D. W., & Lupyan, G. (2017). Iconicity in the speech of children and adults. Developmental science.:

Sidhu, D. M., & Pexman, P. M. (2017). Five mechanisms of sound symbolic association. Psychonomic bulletin & review, 1-25.

Sidhu, D. M., & Pexman, P. M. (2018). Lonely sensational icons: semantic neighbourhood density, sensory experience and iconicity. Language, Cognition and Neuroscience, 33(1), 25-31.

Slavova V., Andonov F., Atanasov D, (2016, July). A Study on Noun Concept Categories Using Data from Child Language Acquisition – an Outline.In Proc. of the 12th Annual International Conference on Computer Science and Education in Computer Science (CSECS 2016)

Slavova, V. (2016, July). Data Collection for Studying Language Acquisition. In Proc. of the 12th Annual International Conference on Computer Science and Education in Computer Science (CSECS 2016)

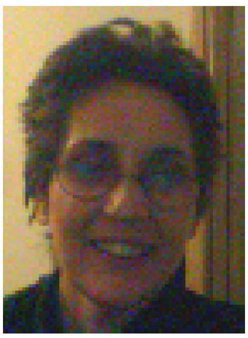
Slavova, V., (2017). On native semantic roles – comparative study based on data from child language acquisition of English and French. International Journal of Cognitive Research in Science, Engineering and Education, 5(2).

Slavova, V., Atanasov, D., & Andonov, F. (2016). Gender differences in the use of noun concept categories – a statistical study based on data from child language acquisition. International Journal" Information Content and Processing, 3(2), 103-116.

Westbury, C., Hollis, G., Sidhu, D. M., & Pexman, P. M. (2017). Weighing up the evidence for sound symbolism: Distributional properties predict cue strength. Journal of Memory and Language.

Winter, B., Perlman, M., Perry, L. K., & Lupyan, G. (2017). Which words are most iconic?. Interaction Studies, 18(3), 430-451.

Authors' Information



Velina Slavova, *Ph. D. Assoc. Prof at the Department of Computer Science, New Bulgarian University,*

e-mail: *vslavova@nbu.bg*

Major Fields of Scientific Research: *Models of language and cognition*

Appendix A. Data from CHILDES – list of the used corpora

Age of the child (Months)	dialogues	Age of the child (Months)	dialogues	Age of the child (Months)	dialogues
06	1	25	21	**43**	**12**
English Rollins Corpus	1	English Braunwald Corpus	9	English Belfast Corpus	2
09	5	English Demetras1 Corpus	1	English Braunwald Corpus	3
English Brent Corpus	4	English Feldman Corpus	5	English Gleason Corpus	2
English Rollins Corpus	1	English Gleason Corpus	2	English HSLLD Corpus	3
10	7	English Higginson Corpus	2	English Snow Corpus	1
English Brent Corpus	7	English Post Corpus	2	English VanHouten Corpus	1
11	6	26	15	**44**	**19**
English Brent Corpus	5	English Braunwald Corpus	6	English Gleason Corpus	1
English Higginson Corpus	1	English Clark Corpus	1	English HSLLD Corpus	18
12	24	English Cornell Corpus	1	**45**	**14**
English Brent Corpus	1	English Feldman Corpus	3	English Belfast Corpus	3
English Rollins Corpus	22	English Gleason Corpus	2	English HSLLD Corpus	9
Eng-USA Soderstrom Corpus	1	English Post Corpus	2	English Snow Corpus	2
13	9	27	8	**46**	**12**
English Bernstein-Ratner Corpus	1	English Braunwald Corpus	3	English Demetras1 Corpus	3
English Brent Corpus	3	English Brown Corpus	1	English HSLLD Corpus	9
English NewEngland Corpus	5	English Feldman Corpus	2	**47**	**11**
14	17	English Gleason Corpus	1	English Belfast Corpus	2
English Brent Corpus	2	English Post Corpus	1	English Braunwald Corpus	1
English Feldman Corpus	3	**28**	**79**	English Demetras1 Corpus	2
English NewEngland Corpus	11	English Bloom73 Corpus	1	English Gleason Corpus	2
English Sachs Corpus	1	English Braunwald Corpus	1	English HSLLD Corpus	4
15	15	English Brown Corpus	3	**48**	**9**
English Bernstein-Ratner Corpus	1	English Gleason Corpus	1	English Belfast Corpus	1
English Braunwald Corpus	4	English Post Corpus	1	English Gleason Corpus	4
English Brent Corpus	2	English-USA Bates Corpus	72	English HSLLD Corpus	4
English Feldman Corpus	4	**29**	**10**	**49**	**19**
English NewEngland Corpus	2	English Braunwald Corpus	4	English Belfast Corpus	1
English Peters Corpus	2	English Brown Corpus	1	English Gleason Corpus	3
16	7	English Cornell Corpus	1	English HSLLD Corpus	12
English Bloom73 Corpus	1	English Gleason Corpus	2	English MacWhinney Corpus	3
English Feldman Corpus	1	English Post Corpus	2	**50**	**13**
English Peters Corpus	4	**30**	**12**	English Braunwald Corpus	2
Phonbank English Providence Corpus	1	English Braunwald Corpus	5	English Gleason Corpus	8
17	11	English Brown Corpus	2	English HSLLD Corpus	3
English Bernstein-Ratner Corpus	1	English Demetras1 Corpus	4	**51**	**19**
English Braunwald Corpus	8	English Snow Corpus	1	English Bliss Corpus	1
English Peters Corpus	2	**31**	**4**	English Braunwald Corpus	4
18	11	English Braunwald Corpus	2	English Gleason Corpus	2
English Bernstein-Ratner Corpus	1	English Brown Corpus	2	English HSLLD Corpus	12
English Braunwald Corpus	8	**32**	**8**	**52**	**9**
English Cornell Corpus	2	English Braunwald Corpus	2	English Belfast Corpus	1
19	13	English Brown Corpus	2	English Gleason Corpus	3
English Bernstein-Ratner Corpus	1	English Demetras1 Corpus	3	English Hall Corpus	1
English Bloom73 Corpus	1	English Gleason Corpus	1	English HSLLD Corpus	2
English Braunwald Corpus	8	**33**	**3**	English MacWhinney Corpus	1
English Cornell Corpus	2	English Brown Corpus	2	English Weist Corpus	1
English Post Corpus	1	English Suppes Corpus	1	**53**	**7**
20	34	**34**	**5**	English Gleason Corpus	1
English Bloom73 Corpus	1	English Bloom73 Corpus	1	English Hall Corpus	1
English Braunwald Corpus	2	English Brown Corpus	3	English MacWhinney Corpus	4
English Cornell Corpus	1	English Demetras1 Corpus	1	English Weist Corpus	1
English Post Corpus	2	**35**	**7**	**54**	**4**
English-USA Bates Corpus	28	English Brown Corpus	3	English Bliss Corpus	1

21	7
English Braunwald Corpus	4
English NewEngland Corpus	1
English Post Corpus	2
22	9
English Bloom73 Corpus	1
English Braunwald Corpus	4
English Higginson Corpus	1
English Post Corpus	3
23	10
English Braunwald Corpus	6
English Cornell Corpus	1
English Feldman Corpus	1
English Post Corpus	2
24	10
English Braunwald Corpus	5
English Demetras1 Corpus	3
English Post Corpus	2
	21
English Braunwald Corpus	9
English Demetras1 Corpus	1
English Feldman Corpus	5
English Gleason Corpus	2
English Higginson Corpus	2
English Post Corpus	2
26	15
English Braunwald Corpus	6
English Clark Corpus	1
English Cornell Corpus	1
English Feldman Corpus	3
English Gleason Corpus	2
English Post Corpus	2
27	8
English Braunwald Corpus	3
English Brown Corpus	1
English Feldman Corpus	2
English Gleason Corpus	1
English Post Corpus	1
English Demetras1 Corpus	2
English Gleason Corpus	2
36	**10**
English Braunwald Corpus	3
English Brown Corpus	1
English Demetras1 Corpus	3
English Gleason Corpus	3
37	**6**
English Braunwald Corpus	2
English Brown Corpus	2
English Demetras1 Corpus	1
English Warren Corpus	1
38	**6**
English Braunwald Corpus	1
English Brown Corpus	2
English Gleason Corpus	3
39	**7**
English Braunwald Corpus	2
English Brown Corpus	2
English Demetras1 Corpus	2
English Gleason Corpus	1
40	**16**
English Braunwald Corpus	2
English Brown Corpus	2
English Cornell Corpus	1
English VanHouten Corpus	11
41	**6**
English Braunwald Corpus	3
English Brown Corpus	2
English VanHouten Corpus	1
42	**9**
English Braunwald Corpus	3
English Brown Corpus	1
English Gleason Corpus	2
English HSLLD Corpus	1
English VanHouten Corpus	2
English MacWhinney Corpus	1
English Warren Corpus	1
English Weist Corpus	1
55	**10**
English Braunwald Corpus	4
English HSLLD Corpus	2
English Sachs Corpus	2
English Weist Corpus	2
56	**8**
English HSLLD Corpus	2
English MacWhinney Corpus	3
English Weist Corpus	3
57	**8**
English MacWhinney Corpus	4
English Sachs Corpus	1
English Warren Corpus	1
English Weist Corpus	2
58	**5**
English Fletcher Corpus	1
English MacWhinney Corpus	3
English Weist Corpus	1
59	**9**
English Gleason Corpus	1
English HSLLD Corpus	4
English MacWhinney Corpus	3
English Warren Corpus	1
60	**7**
English Cornell Corpus	2
English Fletcher Corpus	3
English HSLLD Corpus	1
English Weist Corpus	1
61	**4**
English Fletcher Corpus	4
62	**3**
English Fletcher Corpus	2
English Gleason Corpus	1
Total	**630**

Appendix B. Example - Dialogue retrieved from CHILDES, tagged in the Child Speech Corpus

Extracted Momths	ID assigned of the Dialodue	ID assigned of the Record	Line In Childes (original)	Content In CHILDES (original)	Tag For Participant	Tag For Corce Of Dialoque	Tag for to whom it belongs
21	lad21.bw	lad21.bwd0001	@Begin			0	0
21	lad21.bw	lad21.bwd0002	@Languages:	eng		0	0
21	lad21.bw	lad21.bwd0003	@Participants:	CHI Laura Target_Child , FAT Father , MOT Mother , SIS Sister		0	0
21	lad21.bw	lad21.bwd0004	@ID:	eng|Braunwald|CHI|1;9.16|female|||Target_Child|||		0	0
21	lad21.bw	lad21.bwd0005	@ID:	eng|Braunwald|FAT|||||Father|||		0	0
21	lad21.bw	lad21.bwd0006	@ID:	eng|Braunwald|MOT|||||Mother|||		0	0
21	lad21.bw	lad21.bwd0007	@ID:	eng|Braunwald|SIS|||||Sister|||		0	0

<table>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0008</td><td>@Media:</td><td>1-09-16, audio</td><td></td><td>0</td><td>0</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0009</td><td>@Date:</td><td>06-DEC-1972</td><td></td><td>0</td><td>0</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0010</td><td>@Situation:</td><td>Eating breakfast</td><td></td><td>0</td><td>0</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0011</td><td>@Transcriber:</td><td>Lianne Heys</td><td></td><td>0</td><td>0</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0012</td><td>*MOT:</td><td>no .</td><td>MOT:</td><td>1</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0013</td><td>%mor:</td><td>co|no .</td><td></td><td>1</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0014</td><td>%gra:</td><td>1|0|INCROOT 2|1|PUNCT</td><td></td><td>1</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0015</td><td>*MOT:</td><td>don't [/] don't eat in [?] xxx . [+ PI]</td><td>MOT:</td><td>2</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0016</td><td>%mor:</td><td>mod|do~neg|not v|eat adv|in .</td><td></td><td>2</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0017</td><td>%gra:</td><td>1|3|AUX 2|1|NEG 3|0|ROOT 4|3|JCT 5|3|PUNCT</td><td></td><td>2</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0018</td><td>*FAT:</td><td>xxx .</td><td>FAT:</td><td>3</td><td>FAT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0019</td><td>*CHI:</td><td>xxx .</td><td>CHI:</td><td>4</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0020</td><td>*MOT:</td><td><over here [>]> [?] .</td><td>MOT:</td><td>5</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0021</td><td>%mor:</td><td>prep|over pro:dem|here .</td><td></td><td>5</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0022</td><td>%gra:</td><td>1|0|INCROOT 2|1|POBJ 3|1|PUNCT</td><td></td><td>5</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0023</td><td>*CHI:</td><td>Mom [?] .</td><td>CHI:</td><td>6</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0024</td><td>%mor:</td><td>n:prop|Mom .</td><td></td><td>6</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0025</td><td>%gra:</td><td>1|0|INCROOT 2|1|PUNCT</td><td></td><td>6</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0026</td><td>*CHI:</td><td>Mom [?] [>] .</td><td>CHI:</td><td>7</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0027</td><td>%mor:</td><td>n:prop|Mom .</td><td></td><td>7</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0028</td><td>%gra:</td><td>1|0|INCROOT 2|1|PUNCT</td><td></td><td>7</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0029</td><td>*SIS:</td><td>can [<] we sing a song ?</td><td>SIS:</td><td>8</td><td>SIS:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0030</td><td>%mor:</td><td>mod|can pro:sub|we v|sing det|a n|song ?</td><td></td><td>8</td><td>SIS:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0031</td><td>%gra:</td><td>1|3|LINK 2|3|SUBJ 3|0|ROOT 4|5|DET 5|3|OBJ 6|3|PUNCT</td><td></td><td>8</td><td>SIS:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0032</td><td>%com:</td><td>SIS and MOT talking</td><td></td><td>8</td><td>SIS:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0033</td><td>*CHI:</td><td><hold it> [?] .</td><td>CHI:</td><td>9</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0034</td><td>%mor:</td><td>v|hold pro|it .</td><td></td><td>9</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0035</td><td>%gra:</td><td>1|0|ROOT 2|1|OBJ 3|1|PUNCT</td><td></td><td>9</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0036</td><td>*CHI:</td><td>hold [?] .</td><td>CHI:</td><td>10</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0037</td><td>%mor:</td><td>n|hold .</td><td></td><td>10</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0038</td><td>%gra:</td><td>1|0|INCROOT 2|1|PUNCT</td><td></td><td>10</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0039</td><td>*CHI:</td><td>hold [?] pockets [= actually says packets] . [+ PI]</td><td>CHI:</td><td>11</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0040</td><td>%mor:</td><td>n|hold v|pocket-3S .</td><td></td><td>11</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0041</td><td>%gra:</td><td>1|2|SUBJ 2|0|ROOT 3|2|PUNCT</td><td></td><td>11</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0042</td><td>*MOT:</td><td>you don't have a pocket to put it in .</td><td>MOT:</td><td>12</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0043</td><td>%mor:</td><td>pro|you mod|do~neg|not v|have det|a n|pocket inf|to v|put&ZERO</td><td></td><td>12</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0044</td><td></td><td>pro|it adv|in .</td><td></td><td>12</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0045</td><td>%gra:</td><td>1|4|SUBJ 2|4|AUX 3|2|NEG 4|0|ROOT 5|6|DET 6|4|OBJ 7|8|INF
8|6|XMOD</td><td></td><td>12</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0046</td><td></td><td>9|8|OBJ 10|8|JCT 11|4|PUNCT</td><td></td><td>12</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0047</td><td>*CHI:</td><td>hold . [+ SR]</td><td>CHI:</td><td>13</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0048</td><td>%mor:</td><td>n|hold .</td><td></td><td>13</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0049</td><td>%gra:</td><td>1|0|INCROOT 2|1|PUNCT</td><td></td><td>13</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0050</td><td>*MOT:</td><td>okay .</td><td>MOT:</td><td>14</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0051</td><td>%mor:</td><td>co|okay .</td><td></td><td>14</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0052</td><td>%gra:</td><td>1|0|INCROOT 2|1|PUNCT</td><td></td><td>14</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0053</td><td>*MOT:</td><td>we'll put it right here , , alright ?</td><td>MOT:</td><td>15</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0054</td><td>%mor:</td><td>pro:sub|we~mod|will v|put&ZERO pro|it adv|right adv|here cm|cm</td><td></td><td>15</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0055</td><td></td><td>cm|cm adj|alright ?</td><td></td><td>15</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0056</td><td>%gra:</td><td>1|3|SUBJ 2|3|AUX 3|0|ROOT 4|3|OBJ 5|3|JCT 6|3|JCT 7|3|LP 8|7|LP
9|7|MOD</td><td></td><td>15</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0057</td><td></td><td>10|3|PUNCT</td><td></td><td>15</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0058</td><td>*CHI:</td><td>xxx .</td><td>CHI:</td><td>16</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0059</td><td>*MOT:</td><td>Mummy hold the bottle .</td><td>MOT:</td><td>17</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0060</td><td>%mor:</td><td>n:prop|Mummy n|hold det|the n|bottle .</td><td></td><td>17</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0061</td><td>%gra:</td><td>1|0|INCROOT 2|1|OBJ 3|4|DET 4|2|OBJ 5|1|PUNCT</td><td></td><td>17</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0062</td><td>*CHI:</td><td>huh [?] .</td><td>CHI:</td><td>18</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0063</td><td>%mor:</td><td>co|huh .</td><td></td><td>18</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0064</td><td>%gra:</td><td>1|0|INCROOT 2|1|PUNCT</td><td></td><td>18</td><td>CHI:</td></tr>
</table>

<table>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0065</td><td>*MOT:</td><td>how are you gonna eat the rest of your breakfast [>] ?</td><td>MOT:</td><td>19</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0066</td><td>%mor:</td><td>adv:wh|how aux|be&PRES pro|you part|go-PRESP~inf|to v|eat det|the</td><td></td><td>19</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0067</td><td></td><td>n|rest prep|of pro:poss:det|your n|breakfast ?</td><td></td><td>19</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0068</td><td>%gra:</td><td>1|4|LINK 2|4|AUX 3|4|SUBJ 4|0|ROOT 5|6|INF 6|4|XCOMP 7|8|DET 8|6|OBJ</td><td></td><td>19</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0069</td><td></td><td>9|8|NJCT 10|11|MOD 11|9|POBJ 12|4|PUNCT</td><td></td><td>19</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0070</td><td>*CHI:</td><td>no [<] [/] no .</td><td>CHI:</td><td>20</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0071</td><td>%mor:</td><td>co|no .</td><td></td><td>20</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0072</td><td>%gra:</td><td>1|0|INCROOT 2|1|PUNCT</td><td></td><td>20</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0073</td><td>*CHI:</td><td>don't do [?] with it .</td><td>CHI:</td><td>21</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0074</td><td>%mor:</td><td>mod|do~neg|not v|do prep|with pro|it .</td><td></td><td>21</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0075</td><td>%gra:</td><td>1|3|AUX 2|1|NEG 3|0|ROOT 4|3|JCT 5|4|POBJ 6|3|PUNCT</td><td></td><td>21</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0076</td><td>*CHI:</td><td><do it> [?] .</td><td>CHI:</td><td>22</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0077</td><td>%mor:</td><td>mod|do pro|it .</td><td></td><td>22</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0078</td><td>%gra:</td><td>1|0|INCROOT 2|1|OBJ 3|1|PUNCT</td><td></td><td>22</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0079</td><td>*CHI:</td><td>no .</td><td>CHI:</td><td>23</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0080</td><td>%mor:</td><td>co|no .</td><td></td><td>23</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0081</td><td>%gra:</td><td>1|0|INCROOT 2|1|PUNCT</td><td></td><td>23</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0082</td><td>*CHI:</td><td>I [*] do it .</td><td>CHI:</td><td>24</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0083</td><td>%mor:</td><td>pro:sub|I v|do pro|it .</td><td></td><td>24</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0084</td><td>%gra:</td><td>1|2|SUBJ 2|0|ROOT 3|2|OBJ 4|2|PUNCT</td><td></td><td>24</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0085</td><td>%err:</td><td>[?]</td><td></td><td>24</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0086</td><td>*MOT:</td><td>my do it .</td><td>MOT:</td><td>25</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0087</td><td>%mor:</td><td>pro:poss:det|my v|do pro|it .</td><td></td><td>25</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0088</td><td>%gra:</td><td>1|2|MOD 2|0|ROOT 3|2|OBJ 4|2|PUNCT</td><td></td><td>25</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0089</td><td>*FAT:</td><td>you wanna hold the bottle plus the baby [>] ?</td><td>FAT:</td><td>26</td><td>FAT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0090</td><td>%mor:</td><td>pro|you v|want~inf|to v|hold det|the v|bottle prep|plus det|the</td><td></td><td>26</td><td>FAT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0091</td><td></td><td>n|baby ?</td><td></td><td>26</td><td>FAT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0092</td><td>%gra:</td><td>1|2|SUBJ 2|0|ROOT 3|4|INF 4|2|XCOMP 5|6|DET 6|4|OBJ 7|6|JCT 8|9|DET</td><td></td><td>26</td><td>FAT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0093</td><td></td><td>9|7|POBJ 10|2|PUNCT</td><td></td><td>26</td><td>FAT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0094</td><td>*MOT:</td><td>yeah [<] .</td><td>MOT:</td><td>27</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0095</td><td>%mor:</td><td>co|yeah .</td><td></td><td>27</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0096</td><td>%gra:</td><td>1|0|INCROOT 2|1|PUNCT</td><td></td><td>27</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0097</td><td>*MOT:</td><td>and my door [?] .</td><td>MOT:</td><td>28</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0098</td><td>%mor:</td><td>coord|and pro:poss:det|my n|door .</td><td></td><td>28</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0099</td><td>%gra:</td><td>1|0|INCROOT 2|3|MOD 3|1|COORD 4|1|PUNCT</td><td></td><td>28</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0100</td><td>*MOT:</td><td>ha .</td><td>MOT:</td><td>29</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0101</td><td>%mor:</td><td>co|ha .</td><td></td><td>29</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0102</td><td>%gra:</td><td>1|0|INCROOT 2|1|PUNCT</td><td></td><td>29</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0103</td><td>%com:</td><td>FAT , SIS and MOT talking</td><td></td><td>29</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0104</td><td>*SIS:</td><td>www [>] .</td><td>SIS:</td><td>30</td><td>SIS:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0105</td><td>*CHI:</td><td>xxx [<] .</td><td>CHI:</td><td>31</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0106</td><td>%com:</td><td>FAT , SIS and MOT talking</td><td></td><td>31</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0107</td><td>*CHI:</td><td>grumpy [?] [>] .</td><td>CHI:</td><td>32</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0108</td><td>%mor:</td><td>adj|grump&dn-Y .</td><td></td><td>32</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0109</td><td>%gra:</td><td>1|0|INCROOT 2|1|PUNCT</td><td></td><td>32</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0110</td><td>*MOT:</td><td><I don't> [<] think <so , sweetie> [>] .</td><td>MOT:</td><td>33</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0111</td><td>%mor:</td><td>pro:sub|I mod|do~neg|not v|think co|so cm|cm n|sweet&dadj-DIM .</td><td></td><td>33</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0112</td><td>%gra:</td><td>1|4|SUBJ 2|4|AUX 3|2|NEG 4|0|ROOT 5|7|COM 6|5|LP 7|4|OBJ 8|4|PUNCT</td><td></td><td>33</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0113</td><td>*FAT:</td><td><I doubt it> [<] .</td><td>FAT:</td><td>34</td><td>FAT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0114</td><td>%mor:</td><td>pro:sub|I v|doubt pro|it .</td><td></td><td>34</td><td>FAT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0115</td><td>%gra:</td><td>1|2|SUBJ 2|0|ROOT 3|2|OBJ 4|2|PUNCT</td><td></td><td>34</td><td>FAT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0116</td><td>*SIS:</td><td>xxx [>] .</td><td>SIS:</td><td>35</td><td>SIS:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0117</td><td>*CHI:</td><td><happy children> [?] [=! singing] [<] .</td><td>CHI:</td><td>36</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0118</td><td>%mor:</td><td>adj|happy n|child&PL .</td><td></td><td>36</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0119</td><td>%gra:</td><td>1|2|MOD 2|0|INCROOT 3|2|PUNCT</td><td></td><td>36</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0120</td><td>%com:</td><td>SIS and MOT talking</td><td></td><td>36</td><td>CHI:</td></tr>
</table>

<table>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0121</td><td>*CHI:</td><td>xxx .</td><td>CHI:</td><td>37</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0122</td><td>*MOT:</td><td>yes .</td><td>MOT:</td><td>38</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0123</td><td>%mor:</td><td>co|yes .</td><td></td><td>38</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0124</td><td>%gra:</td><td>1|0|INCROOT 2|1|PUNCT</td><td></td><td>38</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0125</td><td>*MOT:</td><td>I know a quarter [?] xxx . [+ PI]</td><td>MOT:</td><td>39</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0126</td><td>%mor:</td><td>pro:sub|I v|know det|a n|quarter .</td><td></td><td>39</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0127</td><td>%gra:</td><td>1|2|SUBJ 2|0|ROOT 3|4|DET 4|2|OBJ 5|2|PUNCT</td><td></td><td>39</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0128</td><td>*MOT:</td><td>you're [>] not supposed to notice that .</td><td>MOT:</td><td>40</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0129</td><td>%mor:</td><td>pro|you~cop|be&PRES neg|not adj|supposed inf|to v|notice</td><td></td><td>40</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0130</td><td></td><td>pro:dem|that .</td><td></td><td>40</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0131</td><td>%gra:</td><td>1|2|SUBJ 2|0|ROOT 3|2|NEG 4|2|PRED 5|6|INF 6|2|XCOMP 7|6|OBJ 8|2|PUNCT</td><td></td><td>40</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0132</td><td>*CHI:</td><td>xxx [<] .</td><td>CHI:</td><td>41</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0133</td><td>*MOT:</td><td>0 [=! laughs] .</td><td>MOT:</td><td>42</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0134</td><td>*CHI:</td><td>xxx on . [+ PI]</td><td>CHI:</td><td>43</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0135</td><td>%mor:</td><td>adv|on .</td><td></td><td>43</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0136</td><td>%gra:</td><td>1|0|INCROOT 2|1|PUNCT</td><td></td><td>43</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0137</td><td>%com:</td><td>SIS , FAT and MOT talking</td><td></td><td>43</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0138</td><td>*CHI:</td><td>xxx <that make> [?] xxx cup [?] . [+ PI]</td><td>CHI:</td><td>44</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0139</td><td>%mor:</td><td>rel|that v|make n|cup .</td><td></td><td>44</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0140</td><td>%gra:</td><td>1|2|LINK 2|0|ROOT 3|2|OBJ 4|2|PUNCT</td><td></td><td>44</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0141</td><td>*MOT:</td><td>www .</td><td>MOT:</td><td>45</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0142</td><td>*CHI:</td><td>good .</td><td>CHI:</td><td>46</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0143</td><td>%mor:</td><td>adj|good .</td><td></td><td>46</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0144</td><td>%gra:</td><td>1|0|INCROOT 2|1|PUNCT</td><td></td><td>46</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0145</td><td>*MOT:</td><td>what's good ?</td><td>MOT:</td><td>47</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0146</td><td>%mor:</td><td>rel|what~cop|be&3S adj|good ?</td><td></td><td>47</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0147</td><td>%gra:</td><td>1|2|LINK 2|0|ROOT 3|2|PRED 4|2|PUNCT</td><td></td><td>47</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0148</td><td>*CHI:</td><td>row [?] .</td><td>CHI:</td><td>48</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0149</td><td>%mor:</td><td>n|row .</td><td></td><td>48</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0150</td><td>%gra:</td><td>1|0|INCROOT 2|1|PUNCT</td><td></td><td>48</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0151</td><td>*MOT:</td><td>row [?] ?</td><td>MOT:</td><td>49</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0152</td><td>%mor:</td><td>n|row ?</td><td></td><td>49</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0153</td><td>%gra:</td><td>1|0|INCROOT 2|1|PUNCT</td><td></td><td>49</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0154</td><td>*CHI:</td><td>Frera Jaques . [+ R]</td><td>CHI:</td><td>50</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0155</td><td>%mor:</td><td>n:prop|Frera n:prop|Jaques .</td><td></td><td>50</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0156</td><td>%gra:</td><td>1|0|INCROOT 2|1|APP 3|1|PUNCT</td><td></td><td>50</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0157</td><td>*MOT:</td><td>oh .</td><td>MOT:</td><td>51</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0158</td><td>%mor:</td><td>co|oh .</td><td></td><td>51</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0159</td><td>%gra:</td><td>1|0|INCROOT 2|1|PUNCT</td><td></td><td>51</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0160</td><td>*MOT:</td><td>you sing Frere Jaques too ?</td><td>MOT:</td><td>52</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0161</td><td>%mor:</td><td>pro|you v|sing n:prop|Frere n:prop|Jaques post|too ?</td><td></td><td>52</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0162</td><td>%gra:</td><td>1|2|SUBJ 2|0|ROOT 3|2|OBJ 4|3|APP 5|4|PQ 6|2|PUNCT</td><td></td><td>52</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0163</td><td>*FAT:</td><td>oh .</td><td>FAT:</td><td>53</td><td>FAT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0164</td><td>%mor:</td><td>co|oh .</td><td></td><td>53</td><td>FAT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0165</td><td>%gra:</td><td>1|0|INCROOT 2|1|PUNCT</td><td></td><td>53</td><td>FAT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0166</td><td>*CHI:</td><td>Frere Jaques . [+ R] [+ SR]</td><td>CHI:</td><td>54</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0167</td><td>%mor:</td><td>n:prop|Frere n:prop|Jaques .</td><td></td><td>54</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0168</td><td>%gra:</td><td>1|0|INCROOT 2|1|APP 3|1|PUNCT</td><td></td><td>54</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0169</td><td>*CHI:</td><td>0 [=! laughs] [>] .</td><td>CHI:</td><td>55</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0170</td><td>*MOT:</td><td>0 [=! laughs] [<] .</td><td>MOT:</td><td>56</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0171</td><td>*SIS:</td><td>0 [=! laughs] .</td><td>SIS:</td><td>57</td><td>SIS:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0172</td><td>*CHI:</td><td>0 [=! laughs] .</td><td>CHI:</td><td>58</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0173</td><td>*CHI:</td><td>0 [=! laughs] .</td><td>CHI:</td><td>59</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0174</td><td>*CHI:</td><td>xxx <come on> [?] . [+ PI]</td><td>CHI:</td><td>59</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0175</td><td>%mor:</td><td>v|come adv|on .</td><td></td><td>59</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0176</td><td>%gra:</td><td>1|0|ROOT 2|1|JCT 3|1|PUNCT</td><td></td><td>59</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0177</td><td>*CHI:</td><td>tig [: tick] [?] .</td><td>CHI:</td><td>60</td><td>CHI:</td></tr>
</table>

<table>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0178</td><td>%mor:</td><td>n|tick .</td><td></td><td>60</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0179</td><td>%gra:</td><td>1|0|INCROOT 2|1|PUNCT</td><td></td><td>60</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0180</td><td>*SIS:</td><td>you know what ?</td><td>SIS:</td><td>61</td><td>SIS:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0181</td><td>%mor:</td><td>pro|you v|know pro:wh|what ?</td><td></td><td>61</td><td>SIS:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0182</td><td>%gra:</td><td>1|2|SUBJ 2|0|ROOT 3|2|OBJ 4|2|PUNCT</td><td></td><td>61</td><td>SIS:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0183</td><td>*CHI:</td><td><who are that [>]> [?] .</td><td>CHI:</td><td>62</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0184</td><td>%mor:</td><td>rel|who aux|be&PRES pro:dem|that .</td><td></td><td>62</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0185</td><td>%gra:</td><td>1|3|LINK 2|3|AUX 3|0|ROOT 4|3|PUNCT</td><td></td><td>62</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0186</td><td>*SIS:</td><td>www [<] .</td><td>SIS:</td><td>63</td><td>SIS:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0187</td><td>%com:</td><td>www .</td><td></td><td>63</td><td>SIS:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0188</td><td>*CHI:</td><td>more milk .</td><td>CHI:</td><td>64</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0189</td><td>%mor:</td><td>qn|more n|milk .</td><td></td><td>64</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0190</td><td>%gra:</td><td>1|2|QUANT 2|0|INCROOT 3|2|PUNCT</td><td></td><td>64</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0191</td><td>*SIS:</td><td>www [>] .</td><td>SIS:</td><td>65</td><td>SIS:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0192</td><td>*MOT:</td><td>more milk [<] .</td><td>MOT:</td><td>66</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0193</td><td>%mor:</td><td>qn|more n|milk .</td><td></td><td>66</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0194</td><td>%gra:</td><td>1|2|QUANT 2|0|INCROOT 3|2|PUNCT</td><td></td><td>66</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0195</td><td>*CHI:</td><td>xxx .</td><td>CHI:</td><td>67</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0196</td><td>*CHI:</td><td>recorder .</td><td>CHI:</td><td>68</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0197</td><td>%mor:</td><td>n|record&dv-AGT .</td><td></td><td>68</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0198</td><td>%gra:</td><td>1|0|INCROOT 2|1|PUNCT</td><td></td><td>68</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0199</td><td>*MOT:</td><td>well .</td><td>MOT:</td><td>69</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0200</td><td>%mor:</td><td>co|well .</td><td></td><td>69</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0201</td><td>%gra:</td><td>1|0|INCROOT 2|1|PUNCT</td><td></td><td>69</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0202</td><td>*MOT:</td><td>we should just make sure it's recording , , alright ?</td><td>MOT:</td><td>70</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0203</td><td>%mor:</td><td>pro:sub|we mod|should adv|just v|make adv|sure pro|it~aux|be&3S</td><td></td><td>70</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0204</td><td></td><td>part|record-PRESP cm|cm cm|cm adj|alright ?</td><td></td><td>70</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0205</td><td>%gra:</td><td>1|4|SUBJ 2|4|AUX 3|4|JCT 4|0|ROOT 5|4|JCT 6|8|SUBJ 7|8|AUX
8|4|COMP</td><td></td><td>70</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0206</td><td></td><td>9|8|LP 10|8|LP 11|8|JCT 12|4|PUNCT</td><td></td><td>70</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0207</td><td>*FAT:</td><td>here's some milk .</td><td>FAT:</td><td>71</td><td>FAT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0208</td><td>%mor:</td><td>pro:exist|here~cop|be&3S qn|some n|milk .</td><td></td><td>71</td><td>FAT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0209</td><td>%gra:</td><td>1|2|SUBJ 2|0|ROOT 3|4|QUANT 4|2|PRED 5|2|PUNCT</td><td></td><td>71</td><td>FAT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0210</td><td>*SIS:</td><td>www .</td><td>SIS:</td><td>72</td><td>SIS:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0211</td><td>*CHI:</td><td>come on [>] .</td><td>CHI:</td><td>73</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0212</td><td>%mor:</td><td>v|come adv|on .</td><td></td><td>73</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0213</td><td>%gra:</td><td>1|0|ROOT 2|1|JCT 3|1|PUNCT</td><td></td><td>73</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0214</td><td>*SIS:</td><td>www [<] .</td><td>SIS:</td><td>74</td><td>SIS:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0215</td><td>*MOT:</td><td>www .</td><td>MOT:</td><td>75</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0216</td><td>*MOT:</td><td>oh no .</td><td>MOT:</td><td>76</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0217</td><td>%mor:</td><td>co|oh co|no .</td><td></td><td>76</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0218</td><td>%gra:</td><td>1|2|COM 2|0|INCROOT 3|2|PUNCT</td><td></td><td>76</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0219</td><td>*MOT:</td><td>you want some more egg ?</td><td>MOT:</td><td>77</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0220</td><td>%mor:</td><td>pro|you v|want qn|some qn|more n|egg ?</td><td></td><td>77</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0221</td><td>%gra:</td><td>1|2|SUBJ 2|0|ROOT 3|5|QUANT 4|5|QUANT 5|2|OBJ 6|2|PUNCT</td><td></td><td>77</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0222</td><td>*CHI:</td><td>mm mm .</td><td>CHI:</td><td>78</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0223</td><td>%mor:</td><td>co|mm co|mm .</td><td></td><td>78</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0224</td><td>%gra:</td><td>1|2|COM 2|0|INCROOT 3|2|PUNCT</td><td></td><td>78</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0225</td><td>*CHI:</td><td>mm mm .</td><td>CHI:</td><td>79</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0226</td><td>%mor:</td><td>co|mm co|mm .</td><td></td><td>79</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0227</td><td>%gra:</td><td>1|2|COM 2|0|INCROOT 3|2|PUNCT</td><td></td><td>79</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0228</td><td>*CHI:</td><td>cereal@z:sc .</td><td>CHI:</td><td>80</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0229</td><td>%mor:</td><td>unk|cereal .</td><td></td><td>80</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0230</td><td>%gra:</td><td>1|0|INCROOT 2|1|PUNCT</td><td></td><td>80</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0231</td><td>*CHI:</td><td>cereal@z:sc . [+ SR]</td><td>CHI:</td><td>81</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0232</td><td>%mor:</td><td>unk|cereal .</td><td></td><td>81</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0233</td><td>%gra:</td><td>1|0|INCROOT 2|1|PUNCT</td><td></td><td>81</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0234</td><td>*MOT:</td><td>cereal ?</td><td>MOT:</td><td>82</td><td>MOT:</td></tr>
</table>

<table>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0235</td><td>%mor:</td><td>n|cereal ?</td><td></td><td>82</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0236</td><td>%gra:</td><td>1|0|INCROOT 2|1|PUNCT</td><td></td><td>82</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0237</td><td>*CHI:</td><td>yeah .</td><td>CHI:</td><td>83</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0238</td><td>%mor:</td><td>co|yeah .</td><td></td><td>83</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0239</td><td>%gra:</td><td>1|0|INCROOT 2|1|PUNCT</td><td></td><td>83</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0240</td><td>*MOT:</td><td>no .</td><td>MOT:</td><td>84</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0241</td><td>%mor:</td><td>co|no .</td><td></td><td>84</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0242</td><td>%gra:</td><td>1|0|INCROOT 2|1|PUNCT</td><td></td><td>84</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0243</td><td>*MOT:</td><td><you won't> [?] have any cereal xxx . [+ PI]</td><td>MOT:</td><td>85</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0244</td><td>%mor:</td><td>pro|you mod|will~neg|not v|have qn|any n|cereal .</td><td></td><td>85</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0245</td><td>%gra:</td><td>1|4|SUBJ 2|4|AUX 3|2|NEG 4|0|ROOT 5|6|QUANT 6|4|OBJ 7|4|PUNCT</td><td></td><td>85</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0246</td><td>*CHI:</td><td>no . [+ I]</td><td>CHI:</td><td>86</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0247</td><td>%mor:</td><td>co|no .</td><td></td><td>86</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0248</td><td>%gra:</td><td>1|0|INCROOT 2|1|PUNCT</td><td></td><td>86</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0249</td><td>*SIS:</td><td>only for Dwww [% sister] .</td><td>SIS:</td><td>87</td><td>SIS:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0250</td><td>%mor:</td><td>adv|only prep|for n:prop|Dwww .</td><td></td><td>87</td><td>SIS:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0251</td><td>%gra:</td><td>1|0|INCROOT 2|1|JCT 3|2|POBJ 4|1|PUNCT</td><td></td><td>87</td><td>SIS:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0252</td><td>*MOT:</td><td>Dwww [% sister] didn't have an egg .</td><td>MOT:</td><td>88</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0253</td><td>%mor:</td><td>n:prop|Dwww mod|do&PAST~neg|not v|have det|a n|egg .</td><td></td><td>88</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0254</td><td>%gra:</td><td>1|4|SUBJ 2|4|AUX 3|2|NEG 4|0|ROOT 5|6|DET 6|4|OBJ 7|4|PUNCT</td><td></td><td>88</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0255</td><td>*MOT:</td><td>we chose egg .</td><td>MOT:</td><td>89</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0256</td><td>%mor:</td><td>pro:sub|we v|choose&PAST n|egg .</td><td></td><td>89</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0257</td><td>%gra:</td><td>1|2|SUBJ 2|0|ROOT 3|2|OBJ 4|2|PUNCT</td><td></td><td>89</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0258</td><td>*CHI:</td><td>no [=! moans] .</td><td>CHI:</td><td>90</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0259</td><td>%mor:</td><td>co|no .</td><td></td><td>90</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0260</td><td>%gra:</td><td>1|0|INCROOT 2|1|PUNCT</td><td></td><td>90</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0261</td><td>*MOT:</td><td>will you really need [>] cereal ?</td><td>MOT:</td><td>91</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0262</td><td>%mor:</td><td>mod|will pro|you adv|real&dadj-LY v|need n|cereal ?</td><td></td><td>91</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0263</td><td>%gra:</td><td>1|2|AUX 2|4|SUBJ 3|4|JCT 4|0|ROOT 5|4|OBJ 6|4|PUNCT</td><td></td><td>91</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0264</td><td>*CHI:</td><td>huh [=! moans] .</td><td>CHI:</td><td>92</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0265</td><td>%mor:</td><td>co|huh .</td><td></td><td>92</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0266</td><td>%gra:</td><td>1|0|INCROOT 2|1|PUNCT</td><td></td><td>92</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0267</td><td>*MOT:</td><td>well I'll have to get you a clean bowl .</td><td>MOT:</td><td>93</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0268</td><td>%mor:</td><td>co|well pro:sub|I~mod|will v|have inf|to v|get pro|you det|a</td><td></td><td>93</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0269</td><td></td><td>adj|clean n|bowl .</td><td></td><td>93</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0270</td><td>%gra:</td><td>1|4|COM 2|4|SUBJ 3|4|AUX 4|0|ROOT 5|6|INF 6|4|XCOMP 7|6|OBJ 8|10|DET</td><td></td><td>93</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0271</td><td></td><td>9|10|MOD 10|6|OBJ 11|4|PUNCT</td><td></td><td>93</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0272</td><td>*MOT:</td><td>wait a second xxx (be)cause you don't want milk in yours anyway . [+ PI]</td><td>MOT:</td><td>94</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0273</td><td>%mor:</td><td>v|wait det|a n|second conj|because pro|you mod|do~neg|not v|want</td><td></td><td>94</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0274</td><td></td><td>n|milk prep|in pro:poss|yours adv|anyway .</td><td></td><td>94</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0275</td><td>%gra:</td><td>1|0|ROOT 2|3|DET 3|1|OBJ 4|8|LINK 5|8|SUBJ 6|8|AUX 7|6|NEG 8|3|CJCT</td><td></td><td>94</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0276</td><td></td><td>9|8|OBJ 10|8|JCT 11|10|POBJ 12|11|JCT 13|1|PUNCT</td><td></td><td>94</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0277</td><td>*FAT:</td><td>here's some bread .</td><td>FAT:</td><td>95</td><td>FAT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0278</td><td>%mor:</td><td>pro:exist|here~cop|be&3S qn|some n|bread .</td><td></td><td>95</td><td>FAT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0279</td><td>%gra:</td><td>1|2|SUBJ 2|0|ROOT 3|4|QUANT 4|2|PRED 5|2|PUNCT</td><td></td><td>95</td><td>FAT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0280</td><td>*FAT:</td><td>do you want some bread ?</td><td>FAT:</td><td>96</td><td>FAT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0281</td><td>%mor:</td><td>mod|do pro|you v|want qn|some n|bread ?</td><td></td><td>96</td><td>FAT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0282</td><td>%gra:</td><td>1|3|AUX 2|3|SUBJ 3|0|ROOT 4|5|QUANT 5|3|OBJ 6|3|PUNCT</td><td></td><td>96</td><td>FAT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0283</td><td>*CHI:</td><td>&uh [/] &uh .</td><td>CHI:</td><td>97</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0284</td><td>*CHI:</td><td>milk .</td><td>CHI:</td><td>98</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0285</td><td>%mor:</td><td>n|milk .</td><td></td><td>98</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0286</td><td>%gra:</td><td>1|0|INCROOT 2|1|PUNCT</td><td></td><td>98</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0287</td><td>*MOT:</td><td>www [>] .</td><td>MOT:</td><td>99</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0288</td><td>*CHI:</td><td><xxx <in it> [?]> [<] . [+ PI]</td><td>CHI:</td><td>100</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0289</td><td>%mor:</td><td>prep|in pro|it .</td><td></td><td>100</td><td>CHI:</td></tr>
</table>

<table>
<tr><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0290</td><td>%gra:</td><td>1|0|INCROOT 2|1|POBJ 3|1|PUNCT</td><td></td><td>100</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0291</td><td>*FAT:</td><td>xxx .</td><td>FAT:</td><td>101</td><td>FAT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0292</td><td>*CHI:</td><td>xxx [>] .</td><td>CHI:</td><td>102</td><td>CHI:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0293</td><td>*MOT:</td><td>where's [<] a spoon ?</td><td>MOT:</td><td>103</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0294</td><td>%mor:</td><td>adv:wh|where~cop|be&3S det|a n|spoon ?</td><td></td><td>103</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0295</td><td>%gra:</td><td>1|2|LINK 2|0|ROOT 3|4|DET 4|2|PRED 5|2|PUNCT</td><td></td><td>103</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0296</td><td>*MOT:</td><td>anybody got an extra spoon ?</td><td>MOT:</td><td>104</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0297</td><td>%mor:</td><td>pro:indef|anybody v|get&PAST det|a n|extra n|spoon ?</td><td></td><td>104</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0298</td><td>%gra:</td><td>1|2|SUBJ 2|0|ROOT 3|5|DET 4|5|MOD 5|2|OBJ 6|2|PUNCT</td><td></td><td>104</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0299</td><td>%com:</td><td>SIS , FAT and MOT talking</td><td></td><td>104</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0300</td><td>*MOT:</td><td>okay .</td><td>MOT:</td><td>105</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0301</td><td>%mor:</td><td>co|okay .</td><td></td><td>105</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0302</td><td>%gra:</td><td>1|0|INCROOT 2|1|PUNCT</td><td></td><td>105</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0303</td><td>*MOT:</td><td>Dwww [% sister] will get you a spoon .</td><td>MOT:</td><td>106</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0304</td><td>%mor:</td><td>n:prop|Dwww mod|will v|get pro|you det|a n|spoon .</td><td></td><td>106</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0305</td><td>%gra:</td><td>1|3|SUBJ 2|3|AUX 3|0|ROOT 4|6|SUBJ 5|6|DET 6|3|OBJ 7|3|PUNCT</td><td></td><td>106</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0306</td><td>%com:</td><td>MOT , FAT and SIS talking</td><td></td><td>106</td><td>MOT:</td></tr>
<tr><td>21</td><td>lad21.bw</td><td>lad21.bwd0307</td><td>@End</td><td></td><td></td><td>106</td><td>MOT:</td></tr>
</table>

Appendix C. Word adjusting of the Child speech corpus to match the Iconicity Table

<table>
<tr><th>CHILDES POS ORIGINAL</th><th>Word-string Extracted</th><th>CHILDES WORD ADJUSTED</th><th>Word in Iconicity Table</th><th>Iconicity</th><th>Number Occurences</th></tr>
<tr><td>n</td><td>+n|basket+n|ball</td><td>basketball</td><td>basketball</td><td>0,5</td><td>17</td></tr>
<tr><td>n</td><td>+n|butter+n|fly</td><td>butterfly</td><td>butterfly</td><td>0,545455</td><td>62</td></tr>
<tr><td>n</td><td>+n|butter+n|fly-PL</td><td>butterfly</td><td>butterfly</td><td>0,545455</td><td>6</td></tr>
<tr><td>n</td><td>+n|finger+n|nail</td><td>fingernail</td><td>fingernail</td><td>-0,153846</td><td>5</td></tr>
<tr><td>n</td><td>+n|finger+n|nail-PL</td><td>fingernail</td><td>fingernail</td><td>-0,153846</td><td>2</td></tr>
<tr><td>n</td><td>+n|foot+n|ball</td><td>football</td><td>football</td><td>-1</td><td>33</td></tr>
<tr><td>n</td><td>+n|foot+n|ball-PL</td><td>football</td><td>football</td><td>-1</td><td>2</td></tr>
<tr><td>n</td><td>+n|grape+n|fruit</td><td>grapefruit</td><td>grapefruit</td><td>1,090909</td><td>16</td></tr>
<tr><td>n</td><td>+n|type+n|writer</td><td>typewriter</td><td>typewriter</td><td>1,818182</td><td>9</td></tr>
<tr><td>n</td><td>ache-PL</td><td>ache</td><td>ache</td><td>1,153846</td><td>2</td></tr>
<tr><td>v</td><td>add-PAST</td><td>add</td><td>add</td><td>-0,153846</td><td>1</td></tr>
<tr><td>n</td><td>airplane-PL</td><td>airplane</td><td>airplane</td><td>2,545455</td><td>5</td></tr>
<tr><td>n</td><td>alligator-PL</td><td>alligator</td><td>alligator</td><td>1,090909</td><td>1</td></tr>
<tr><td>n</td><td>animal-PL</td><td>animal</td><td>animal</td><td>-0,933333</td><td>38</td></tr>
<tr><td>n</td><td>ant-PL</td><td>ant</td><td>ant</td><td>0,6</td><td>1</td></tr>
<tr><td>n</td><td>apple-PL</td><td>apple</td><td>apple</td><td>0,083333</td><td>15</td></tr>
<tr><td>n</td><td>area-PL</td><td>area</td><td>area</td><td>-0,214286</td><td>1</td></tr>
<tr><td>v</td><td>arm-3S</td><td>arm</td><td>arm</td><td>0,2</td><td>2</td></tr>
<tr><td>n</td><td>arm-PL</td><td>arm</td><td>arm</td><td>0,2</td><td>14</td></tr>
<tr><td>n</td><td>ash-PL</td><td>ash</td><td>ash</td><td>0,538462</td><td>3</td></tr>
<tr><td>n</td><td>ashtray-PL</td><td>ashtray</td><td>ashtray</td><td>1,909091</td><td>1</td></tr>
<tr><td>n</td><td>aunt-DIM</td><td>aunt</td><td>aunt</td><td>0,2</td><td>3</td></tr>
<tr><td>n</td><td>aunt-PL</td><td>aunt</td><td>aunt</td><td>0,2</td><td>2</td></tr>
<tr><td>n</td><td>baby~aux|be&3S</td><td>baby</td><td>baby</td><td>2,230769</td><td>4</td></tr>
<tr><td>n</td><td>baby~poss|s</td><td>baby</td><td>baby</td><td>2,230769</td><td>12</td></tr>
<tr><td>n</td><td>baby-PL</td><td>baby</td><td>baby</td><td>2,230769</td><td>44</td></tr>
<tr><td>n</td><td>bad&dadj-DIM</td><td>bad</td><td>bad</td><td>0,5625</td><td>2</td></tr>
<tr><td>n</td><td>bad&dadj-DIM-PL</td><td>bad</td><td>bad</td><td>0,5625</td><td>1</td></tr>
<tr><td>v</td><td>bag-3S</td><td>bag</td><td>bag</td><td>1,090909</td><td>1</td></tr>
</table>

n	bag-PL	bag	bag	1,090909	17
v	ball-3S	ball	ball	0,75	4
n	ball-DIM	ball	ball	0,75	7
n	balloon~aux|be&3S	balloon	balloon	1,714286	2
v	balloon-3S	balloon	balloon	1,714286	2
n	balloon-PL	balloon	balloon	1,714286	27
n	ball-PL	ball	ball	0,75	33
n	banana-PL	banana	banana	1	5
v	band-3S	band	band	0,307692	3
n	band-PL	band	band	0,307692	3
PRESP	bang	banging	banging	2,5	2
v	bang-3S	bang	bang	3,833333	1
v	bang-PAST	bang	bang	3,833333	2
n	banjo-PL	banjo	banjo	1,363636	2
n	banner-PL	banner	banner	-0,727273	1
PRESP	bark	barking	barking	3,333333	8
n	barrel-PL	barrel	barrel	1	1
n	basket-PL	basket	basket	-0,166667	1
n	bath-DIM	bath	bath	0,3	2
n	bath-PL	bath	bath	0,3	1
n	bat-PL	bat	bat	0,153846	1
aux	be	am	am	0,583333	893
aux	be	are	are	-0,9	699
aux	be	is	is	-0,142857	2367
aux	be	was	was	-0,833333	1
aux	be	was	was	-0,833333	230
aux	be	were	were	0,6	100
cop	be	am	am	0,583333	357
cop	be	are	are	-0,9	736
cop	be	is	is	-0,142857	64
cop	be	was	was	-0,833333	664
cop	be	were	were	0,6	63
cop|be&3S	be	is	is	-0,142857	5838
v	bear-3S	bear	bear	-0,25	1
n	bear-PL	bear	bear	-0,25	12
n	beaver-PL	beaver	beaver	0,2	1
n	bed~cop|be&3S	bed	bed	0,416667	1
v	bed-3S	bed	bed	0,416667	2
n	bed-PL	bed	bed	0,416667	12
n	bee-PL	bee	bee	1,545455	14
n	beep-PL	beep	beep	4,357143	1
n	beer-PL	beer	beer	0,846154	2
v	begin&PAST	begin	begin	1,571429	4
n	bell-PL	bell	bell	1,181818	2
v	belong-3S	belong	belong	-0,166667	4
v	belt-3S	belt	belt	1,2	1
n	belt-PL	belt	belt	1,2	2
n	bench-DIM	bench	bench	-0,916667	1
n	berry-PL	berry	berry	1,1	1
n	bike&dv-AGT	bike	bike	0,9	1
n	bike-PL	bike	bike	0,9	7
n	bird-DIM	bird	bird	0,727273	69
n	bird-DIM-PL	bird	bird	0,727273	1
n	bird-PL	bird	bird	0,727273	14
n	biscuit-PL	biscuit	biscuit	-0,153846	1
v	bite&PAST	bite	bite	2	41
v	bite-3S	bite	bite	2	4
n	bite-PL	bite	bite	2	3

v	blanket-3S	blanket	blanket	0,117647	1
n	blanket-PL	blanket	blanket	0,117647	5
v	blast-PAST	blast	blast	3,545455	1
n	blend&dv-AGT	blender	blender	2,1	2
v	block-3S	block	block	2,428571	1
n	block-PL	block	block	2,428571	29
v	blow&PAST	blow	blow	1,363636	6
v	blow-3S	blow	blow	1,363636	1
v	blow-PAST	blow	blow	1,363636	5
n	blueberry-PL	blueberry	blueberry	2,6	10
n	boat-PL	boat	boat	1,142857	10
PRESP	boil	boiling	boiling	1,384615	2
v	boil-PAST	boil	boil	1,384615	1
n	bolt-PL	bolt	bolt	1,1	7
n	bone-PL	bone	bone	0,090909	6
n	book-DIM	book	book	0,090909	4
n	book-PL	book	book	0,090909	31
v	bottle-3S	bottle	bottle	0,5	2
n	bottle-PL	bottle	bottle	0,5	9
v	box-3S	box	box	0,5625	2
n	box-PL	box	box	0,5625	6
n	boy~aux|be&3S	boy	boy	1,7	2
n	boy~mod|do&3S	boy	boy	1,7	1
n	boy~poss|s	boy	boy	1,7	4
n	boy-PL	boy	boy	1,7	26
n	branch-PL	branch	branch	0,571429	2
n	bread-PL	bread	bread	0,333333	1
v	break&PAST	break	break	2,9	2
v	break-3S	break	break	2,9	3
n	breakfast-PL	breakfast	breakfast	0,916667	1
v	break-PAST	break	break	2,9	3
n	break-PL	break	break	2,9	4
v	brick-3S	brick	brick	1,25	2
n	brick-PL	brick	brick	1,25	5
n	bridge-PL	bridge	bridge	-0,166667	1
v	bring&PAST	bring	bring	0,133333	2
v	bring-3S	bring	bring	0,133333	2
v	bring-PAST	bring	bring	0,133333	3
on	brooom	broom	broom	0,571429	3
on	brooommm	broom	broom	0,571429	1
on	brooommmm	broom	broom	0,571429	1
n	brother~mod|do&3S	brother	brother	0,4	1
n	brother~poss|s	brother	brother	0,4	2
n	brother-PL	brother	brother	0,4	3
n	brown&dadj-DIM	brown	brown	0,571429	2
n	brown&dadj-DIM-PL	brown	brown	0,571429	1
v	brush-PAST	brush	brush	1,727273	2
v	bubble-3S	bubble	bubble	1,461538	3
n	bubble-PL	bubble	bubble	1,461538	23
n	bucket-PL	bucket	bucket	-0,529412	7
adj	bug	buggy	buggy	0,727273	10
n	bug-PL	bug	bug	1,9	8
v	build&PAST	build	build	1,153846	3
v	build-3S	build	build	1,153846	1
n	building-PL	building	building	1,636364	4
n	bullet-PL	bullet	bullet	0,076923	8
v	bump-3S	bump	bump	3,363636	2
v	bump-PAST	bump	bump		1

v	bump-PAST	bump	bump	3,363636	13
n	bump-PL	bump	bump	3,363636	1
n	bunny-PL	bunny	bunny	0,4	2
n	burglar-PL	burglar	burglar	1,363636	2
PRESP	burn	burning	burning	0,533333	7
v	burn&PAST	burn	burn	0,416667	3
v	burn-PAST	burn	burn	0,416667	2
v	burst&ZERO	burst	burst	2,916667	2
n	bush-PL	bush	bush	0,25	2
n	bus-PL	bus	bus	-0,928571	1
n	button-PL	button	button	0,5	18
v	buy&PAST	buy	buy	-0,083333	4
v	buy-3S	buy	buy	-0,083333	5
v	buy-PAST	buy	buy	-0,083333	1
n	buy-PL	buy	buy	-0,083333	2
co	byebye	bye	bye	1,583333	76
n	cabbage-PL	cabbage	cabbage	-0,9	1
n	cabin-PL	cabin	cabin	0,090909	1
v	cake-3S	cake	cake	0,461538	1
n	cake-PL	cake	cake	0,461538	11
v	call-3S	call	call	0,416667	7
v	call-PAST	call	call	0,416667	39
n	candle-PL	candle	candle	0,909091	10
n	candy-PL	candy	candy	1,117647	1
n	cape-PL	cape	cape	0,454545	1
n	car-DIM-PL	car	car	0,466667	2
n	car-PL	car	car	0,466667	70
n	carrot-PL	carrot	carrot	0,083333	13
v	carry-3S	carry	carry	0	3
n	cat~aux|be&3S	cat	cat	0,538462	2
v	catch&PAST	catch	catch	0,461538	2
v	catch-3S	catch	catch	0,461538	1
v	catch-PAST	catch	catch	0,461538	1
n	caterpillar-PL	caterpillar	caterpillar	-1	2
n	cat-PL	cat	cat	0,538462	28
v	cause	because	because	0,214286	42
n	cereal-PL	cereal	cereal	0,214286	1
v	chain-3S	chain	chain	-0,1	1
n	chair-PL	chair	chair	0,0625	29
n	chalk-PL	chalk	chalk	0,636364	1
v	chase-3S	chase	chase	0,692308	1
n	cheek-PL	cheek	cheek	1,3	10
n	cheese-DIM-PL	cheese	cheese	-0,363636	1
n	cherry-PL	cherry	cherry	0,8	4
n	chicken-PL	chicken	chicken	1,090909	13
n	child&PL	child	child	0,933333	51
n	child&PL-PL	child	child	0,933333	2
n	chocolate-DIM	chocolate	chocolate	0,545455	2
n	chocolate-PL	chocolate	chocolate	0,545455	2
on	chooo_chooo_do_do	cockadoodledoo	cockadoodledoo	3,733333	1
v	choose&PAST	choose	choose	-0,4	5
n	chop-PL	chop	chop	3,272727	1
n	circle-PL	circle	circle	1,307692	6
n	circus-PL	circus	circus	0,6	1
v	clamp-3S	clamp	clamp	2,818182	1
v	climb-3S	climb	climb	0,846154	4
v	climb-PAST	climb	climb	0,846154	10
n	clock~cop|be&3S	clock	clock	0,611111	1

n	clock-PL	clock	clock	0,611111	1
v	close-3S	close	close	-0,125	1
v	close-PAST	closed	closed	1,6	2
n	cloth-PL	cloth	cloth	-0,7	1
n	cloud-PL	cloud	cloud	1	12
n	coat-PL	coat	coat	-0,8	1
on	cock_a_doodle	cockadoodledoo	cockadoodledoo	3,733333	4
co	cock_a_doodle_doo	cockadoodledoo	cockadoodledoo	3,733333	6
n	cock_a_doodle_doo	cockadoodledoo	cockadoodledoo	3,733333	1
on	cock_a_doodle_doo	cockadoodledoo	cockadoodledoo	3,733333	1
on	cockacockadoo	cockadoodledoo	cockadoodledoo	3,733333	1
on	cockadoodle	cockadoodledoo	cockadoodledoo	3,733333	1
on	cockle_doodle_doo	cockadoodledoo	cockadoodledoo	3,733333	1
n	coffee-PL	coffee	coffee	1,3	4
n	coINVe	coin	coin	0,4	3
v	color-3S	color	color	0,083333	2
n	color-PL	color	color	0,083333	18
v	come&PAST	come	come	0,214286	128
v	come-3S	come	come	0,214286	93
n	contain&dv-AGT	container	container	1,181818	5
n	cookie-PL	cookie	cookie	0,454545	98
v	cook-PAST	cook	cook	0,9	1
n	cook-PL	cook	cook	0,9	1
v	cool-3S	cool	cool	2,214286	1
v	cost&ZERO	cost	cost	0,538462	5
v	cost-3S	cost	cost	0,538462	5
v	count-PAST	count	count	-0,2	1
n	course-PL	course	course	0,4	1
v	cover-PAST	cover	cover	0,428571	4
n	cover-PL	cover	cover	0,428571	2
n	cow~cop|be&3S	cow	cow	1,454545	1
n	cow~mod|do&3S	cow	cow	1,454545	1
n	cow-PL	cow	cow	1,454545	28
n	crab-PL	crab	crab	0	3
v	crack-3S	crack	crack	3,1	1
n	cracker-PL	cracker	cracker	0,583333	11
v	crack-PAST	cracked	cracked	3	1
PRESP	crash	crashing	crashing	2,428571	1
v	crawl-3S	crawl	crawl	3,272727	1
v	crawl-PAST	crawl	crawl	3,272727	1
v	crayon-3S	crayon	crayon	0,818182	2
n	crayon-PL	crayon	crayon	0,818182	19
v	creep-3S	creep	creep	2,1	1
n	cricket-PL	cricket	cricket	2,6	2
n	crisp-PL	crisp	crisp	3,7	2
v	cross-PAST	cross	cross	0,615385	1
n	cross-PL	cross	cross	0,615385	1
n	crown-PL	crown	crown	0,9	1
n	crow-PL	crow	crow	1,909091	2
adj	cry	crying	crying	0,153846	16
PRESP	cry	crying	crying	0,153846	73
v	cry-3S	cry	cry	0,866667	11
v	cry-PAST	cry	cry	0,866667	7
n	cry-PL	cry	cry	0,866667	2
n	cucumber-PL	cucumber	cucumber	0,1	4
v	cup-3S	cup	cup	1,538462	2
n	cup-DIM	cup	cup	1,538462	2
n	cup-PL	cup	cup	1,538462	30

n	cushion-PL	cushion	cushion	2,090909	1
n	customer-PL	customer	customer	0,428571	1
v	cut&ZERO	cut	cut	0,705882	135
v	cut-3S	cut	cut	0,705882	1
n	daddy~poss|s	daddy	daddy	1,428571	3
n	daddy-PL	daddy	daddy	1,428571	6
n	day-PL	day	day	-0,153846	35
n	dead-PL	dead	dead	1,307692	1
v	destroy-PAST	destroy	destroy	1,857143	1
v	die-3S	die	die	0,333333	3
v	die-PAST	die	die	0,333333	9
n	die-PL	die	die	0,333333	1
n	dime-PL	dime	dime	0,363636	2
n	dinner~cop|be&3S	dinner	dinner	0,263158	1
adj	dirt	dirty	dirty	1	57
n	dirt-DIM-PL	dirt	dirt	1	1
n	dirt-PL	dirt	dirt	1	4
n	dish-PL	dish	dish	0,75	13
n	dive&dv-AGT-PL	dive	dive	1,571429	1
v	dive&PAST	dive	dive	1,571429	1
mod	do	did	did	0,466667	650
mod	do	does	does	-0,6	438
v	do	did	did	0,466667	20
v	do&3S	does	does	-0,6	228
v	do&PAST	did	did	0,466667	460
v	do&PAST	do	do	0,846154	5
v	do-3S	do	do	0,846154	2
n	doctor~aux|be&3S	doctor	doctor	1,6	1
n	doctor~poss|s	doctor	doctor	1,6	3
n	doctor-PL	doctor	doctor	1,6	8
n	dog&dn-Y	dog	dog	1,272727	1
n	dog~poss|s	dog	dog	1,272727	3
v	dog-3S	dog	dog	1,272727	6
n	dog-DIM	dog	dog	1,272727	318
n	dog-DIM-PL	dog	dog	1,272727	6
n	dog-PL	dog	dog	1,272727	23
n	doll-DIM	doll	doll	-0,428571	15
n	doll-DIM-PL	doll	doll	-0,428571	5
n	doll-PL	doll	doll	-0,428571	4
n	door-PL	door	door	1,272727	14
v	do-PAST	do	do	0,846154	1
v	drag-PAST	drag	drag	2,083333	1
v	draw&PAST	draw	draw	0,2	6
n	drawer-PL	drawer	drawer	-0,923077	4
v	draw-PAST	draw	draw	0,2	2
v	drink-3S	drink	drink	1	2
n	drink-PL	drink	drink	1	3
v	drive&PAST	drive	drive	0,230769	2
v	drive-3S	drive	drive	0,230769	1
v	drive-PAST	drive	drive	0,230769	1
n	drive-PL	drive	drive	0,230769	2
v	drop-3S	drop	drop	1,071429	4
v	drop-PAST	drop	drop	1,071429	21
v	dry-3S	dry	dry	1,363636	7
n	duck~poss|s	duck	duck	0,3	2
v	duck-3S	duck	duck	0,3	3
n	duck-DIM	duck	duck	0,3	6
n	duck-DIM-PL	duck	duck	0,3	1

n	duck-PL	duck	duck	0,3	13
v	dump-3S	dump	dump	2,9375	3
v	dump-PAST	dump	dump	2,9375	3
n	ear-PL	ear	ear	1,1	26
v	eat&PAST	eat	eat	0,647059	119
v	eat-3S	eat	eat	0,647059	25
co	ee_yay_ee_yay_oh	aye	aye	1,727273	1
n	egg-DIM	egg	egg	1,818182	20
n	egg-PL	egg	egg	1,818182	72
n	elephant-PL	elephant	elephant	2,142857	7
n	enemy-PL	enemy	enemy	1,416667	1
n	envelope-PL	envelope	envelope	-0,1	4
v	eye-3S	eye	eye	1	3
n	eye-PL	eye	eye	1	100
v	face-3S	face	face	0,272727	1
n	face-PL	face	face	0,272727	1
adj	fall	falling	falling	0,615385	2
PRESP	fall	falling	falling	0,615385	24
v	fall	fell	fell	1,714286	3
v	fall&PAST	fell	fell	1,714286	260
v	fall-3S	fall	fall	1,533333	23
v	fall-PAST	fall	fall	1,533333	3
v	fall-PAST	fell	fell	1,714286	12
n	fall-PL	fall	fall	1,533333	1
mod	FAMld	family	family	0,416667	1
n	farm-PL	farm	farm	-0,105263	1
n	favor-PL	favor	favor	0,153846	2
n	feather-PL	feather	feather	1,153846	5
v	feed&PAST	feed	feed	0,8	1
v	feed-3S	feed	feed	0,8	1
v	feel&PAST	felt	felt	0,928571	2
v	feel-3S	feel	feel	1,25	8
v	feel-PAST	feel	feel	1,25	1
n	fee-PL	fee	fee	-0,166667	1
n	fight-PL	fight	fight	0,363636	2
v	find&PAST	find	find	0	99
v	find-3S	find	find	0	1
v	find-PAST	find	find	0	2
n	finger-3S	finger	finger	0,181818	1
n	finger-PL	finger	finger	0,181818	22
v	finish-3S	finish	finish	0,181818	1
v	finish-PAST	finish	finish	0,181818	34
v	fire-3S	fire	fire	1,461538	2
v	fish-3S	fish	fish	0,071429	3
n	fish-DIM	fish	fish	0,071429	4
n	fish-PL	fish	fish	0,071429	5
v	fit&ZERO	fit	fit	0,470588	110
v	fit-3S	fit	fit	0,470588	20
n	fit-PL	fit	fit	0,470588	12
v	fix-PAST	fix	fix	0,538462	1
v	flake-3S	flake	flake	1	1
n	flower~poss|s	flower	flower	-0,083333	1
v	flower-3S	flower	flower	-0,083333	2
n	flower-PL	flower	flower	-0,083333	53
v	fly&PAST	fly	fly	1,923077	3
v	fly-PAST	fly	fly	1,923077	2
n	fly-PL	fly	fly	1,923077	4
v	fold-PAST	fold	fold	0,636364	1

v	follow-PAST	follow	follow	0,142857	1
n	food-PL	food	food	1,071429	1
n	foot&PL	foot	foot	-0,4	73
n	foot-DIM	foot	foot	-0,4	7
v	fork-3S	fork	fork	0,818182	1
n	fork-PL	fork	fork	0,818182	13
v	freeze&PAST	freeze	freeze	1,066667	1
n	friend-PL	friend	friend	0,818182	51
n	frog-DIM	frog	frog	2,181818	4
n	frog-PL	frog	frog	2,181818	19
n	fuss-PL	fuss	fuss	-0,230769	1
v	game-3S	game	game	0,066667	6
n	game-PL	game	game	0,066667	38
n	gate-PL	gate	gate	0,25	3
v	get&PAST	get	get	-0,583333	996
v	get-3S	get	get	-0,583333	47
n	ghost-PL	ghost	ghost	0,5	2
n	giant~poss|s	giant	giant	0,538462	1
n	gift-PL	gift	gift	0,785714	1
n	girl-PL	girl	girl	0,636364	27
v	give&PAST	give	give	1,272727	56
v	give-3S	give	give	1,272727	9
n	glass-PL	glass	glass	0,8	48
v	go&PAST	go	go	1,454545	5
v	go&PAST	went~	go	1,454545	1
v	go-3S	go	go	1,454545	446
n	gold&dadj-DIM	gold	gold	0,692308	1
adv	good	better	better	-0,1	103
n	good&dadj-DIM	good	good	0,928571	3
v	go-PAST	go	go	1,454545	2
n	grape-PL	grape	grape	0,909091	6
v	green-3S	green	green	1,1	1
v	grind-3S	grind	grind	2	2
n	grit-PL	grit	grit	1,266667	2
n	gum-PL	gum	gum	1,6	1
n	gun-PL	gun	gun	0,5	9
n	hair-PL	hair	hair	-0,416667	1
n	hamburger-PL	hamburger	hamburger	-0,357143	6
n	hammer-PL	hammer	hammer	1,384615	2
v	hand-3S	hand	hand	0,388889	8
n	hand-PL	hand	hand	0,388889	69
v	hang-3S	hang	hang	0	1
v	hate-3S	hate	hate	1,4375	1
n	hate-DIM	hate	hate	1,4375	1
n	hat-PL	hat	hat	1,461538	5
v	have&3S	have	have	0,733333	1
v	have&3S	have	have	1,733333	1
v	have&3S	have	have	2,733333	1
v	have&3S	have	have	3,733333	1
v	have&3S	have	have	4,733333	1
v	have&3S	have	have	5,733333	1
v	have&3S	have	have	6,733333	1
v	have&PAST	have	have	-0,266667	230
v	have-3S	have	have	-0,266667	1
n	head-PL	head	head	0,6875	6
v	hear&PAST	hear	hear	0,066667	16
v	hear-PAST	hear	hear	0,066667	1
n	heart-PL	heart	heart	0,928571	1

n	helicopter-PL	helicopter	helicopter	0,909091	2
v	help-3S	help	help	1,533333	1
v	help-PAST	help	help	1,533333	10
n	hen-PL	hen	hen	0,454545	2
n	herb-DIM	herb	herb	0,230769	1
pro:exist	here~cop|be&3S	here	here	-0,2	11
n	hero-PL	hero	hero	-0,153846	2
v	hide&PAST	hide	hide	1,142857	5
n	hide-PL	hide	hide	1,142857	1
v	hit&ZERO	hit	hit	1	169
v	hit-3S	hit	hit	1	5
n	hit-PL	hit	hit	1	1
v	hold-3S	hold	hold	0,846154	3
v	hold-PAST	hold	hold	0,846154	3
n	hold-PL	hold	hold	0,846154	1
n	hole-PL	hole	hole	2,538462	15
n	holiday-PL	holiday	holiday	0,7	4
n	home-PL	home	home	2	3
v	hook-3S	hook	hook	1,454545	1
n	hook-PL	hook	hook	1,454545	1
v	hop-3S	hop	hop	3,3	2
v	hop-PAST	hope	hope	0,125	1
n	horn-PL	horn	horn	1,3	6
n	horse-DIM	horse	horse	-0,363636	64
n	horse-DIM-PL	horse	horse	-0,363636	6
n	horse-PL	horse	horse	-0,363636	46
n	hose-PL	hose	hose	0,642857	2
n	hour-PL	hour	hour	-0,071429	3
n	house-DIM	house	house	0,454545	1
n	house-PL	house	house	0,454545	13
v	hurt&ZERO	hurt	hurt	0,9	154
v	hurt-3S	hurt	hurt	0,9	33
n	hurt-PL	hurt	hurt	0,9	17
n	inch-PL	inch	inch	1,833333	1
v	invite-PAST	invite	invite	0,230769	2
n	jam-DIM-PL	jam	jam	0,571429	2
n	jar-PL	jar	jar	0,461538	1
n	jeep-PL	jeep	jeep	0,166667	3
n	jewel-PL	jewel	jewel	1,25	2
n	job-PL	job	job	0,307692	1
v	join-PAST	join	join	0	1
n	joke-PL	joke	joke	0,545455	3
v	jump-3S	jump	jump	1	9
v	jump-PAST	jump	jump	1	15
n	jump-PL	jump	jump	1	2
n	ketchup-PL	ketchup	ketchup	0,3	1
n	key-PL	key	key	0,3	22
v	kick-3S	kick	kick	1,95	3
v	kick-PAST	kick	kick	1,95	6
v	kid-3S	kid	kid	1,454545	6
n	kid-DIM	kid	kid	1,454545	2
n	kid-PL	kid	kid	1,454545	37
v	kill-3S	kill	kill	1,692308	2
v	kill-PAST	kill	kill	1,692308	5
n	kind-PL	kind	kind	-0,357143	12
n	king-PL	king	king	0,75	2
v	kiss-3S	kiss	kiss	1,25	1
v	kiss-PAST	kiss	kiss	1,25	3

n	kitten-PL	kitten	kitten	0,1	3
n	kitty-PL	kitty	kitty	0	6
n	knee-PL	knee	knee	0,133333	7
n	knife-PL	knife	knife	0,52381	2
n	knight-PL	knight	knight	1,4	1
v	knit&ZERO	knit	knit	0,181818	2
v	knock-3S	knock	knock	3,0625	2
v	knock-PAST	knock	knock	3,0625	17
n	knot-PL	knot	knot	-0,636364	2
v	know&PAST	know	know	0,769231	12
v	know-3S	know	know	0,769231	6
n	ladder-PL	ladder	ladder	1,384615	2
n	lady~aux|be&3S	lady	lady	-0,076923	1
n	lady-PL	lady	lady	-0,076923	13
n	lamb-PL	lamb	lamb	0,416667	1
n	lamp-PL	lamp	lamp	0,071429	1
v	land-3S	land	land	-0,8	1
adj	laugh	laughing	laughing	0,5	12
PRESP	laugh	laughing	laughing	0,5	15
n	laugh-PL	laugh	laugh	0,75	2
n	lawn-PL	lawn	lawn	0,6	1
v	leave-3S	leave	leave	0,615385	1
n	leave-PL	leave	leave	0,615385	13
n	leg-PL	leg	leg	0,315789	31
n	leopard-PL	leopard	leopard	0,4	3
n	letter-PL	letter	letter	-0,363636	18
v	lick-PAST	lick	lick	1,166667	4
v	lie-3S	lie	lie	-0,636364	1
v	lie-PAST	lie	lie	-0,636364	1
n	life-PL	life	life	0,333333	5
v	lift-PAST	lift	lift	0,571429	2
v	light-3S	light	light	0,9	1
n	light-PL	light	light	0,9	13
v	like-3S	like	like	-0,333333	43
n	lion-PL	lion	lion	-0,352941	7
n	lip-PL	lip	lip	0,714286	10
n	lipstick-PL	lipstick	lipstick	2,4	2
v	listen-PAST	listen	listen	0,75	2
v	live-3S	live	live	0,416667	19
v	live-PAST	live	live	0,416667	1
n	load-PL	load	load	0,0625	2
n	lobster-PL	lobster	lobster	-0,636364	1
v	look-3S	look	look	1,818182	64
v	look-PAST	look	look	1,818182	19
n	look-PL	look	look	1,818182	1
n	loop-PL	loop	loop	2,416667	1
v	lose&PAST	lose	lose	-0,307692	73
n	louse&PL	louse	louse	0,333333	2
v	love-3S	love	love	1,125	13
v	love-PAST	love	love	1,125	1
n	lump-PL	lump	lump	2,384615	4
v	make&PAST	make	make	0,642857	173
v	make&PAST|det|a	make	make	0,642857	1
v	make-3S	make	make	0,642857	32
adj	man	many	many	0,076923	10
n	man&PL	man	man	0,647059	46
n	man~aux|be&3S	man	man	0,647059	2
n	man~poss|s	man	man	0,647059	1

n	marble-PL	marble	marble	1,4	18
n	mat-PL	mat	mat	0,545455	1
v	meet&PAST	meet	meet	-0,615385	4
v	meet-3S	meet	meet	-0,615385	2
v	melt-PAST	melt	melt	0,615385	2
n	microscope-PL	microscope	microscope	-0,1	3
n	mine~poss|s	mine	mine	1,5	2
n	mine-PL	mine	mine	1,5	3
n	minute-PL	minute	minute	1,071429	16
v	mirror-3S	mirror	mirror	2,090909	1
v	miss-3S	miss	miss	1,923077	3
n	missile-PL	missile	missile	0,1875	1
n	miss-PL	miss	miss	1,923077	2
n	mix&dv-AGT	mixer	mixer	2,4	16
n	mix&dv-AGT-PL	mixer	mixer	2,4	2
n	mommy-PL	mommy	mommy	1,4	1
n	money-PL	money	money	-0,6	1
n	monkey-PL	monkey	monkey	0,285714	9
n	monster-PL	monster	monster	1,071429	5
v	moon-3S	moon	moon	0,1	1
n	moon-PL	moon	moon	0,1	2
n	morning-PL	morning	morning	-1,8	1
n	mother~aux|be&3S	mother	mother	0,307692	1
n	mother~poss|s	mother	mother	0,307692	2
n	mother-PL	mother	mother	0,307692	2
n	motorcycle-PL	motorcycle	motorcycle	1,2	9
n	mountain-PL	mountain	mountain	1,083333	2
n	mouse&PL	mouse	mouse	0	10
n	mouse-DIM	mouse	mouse	0	2
n	mouse-DIM-PL	mouse	mouse	0	1
n	mouth-PL	mouth	mouth	1,333333	2
v	move-3S	move	move	0,3	2
v	move-PAST	move	move	0,3	2
n	move-PL	move	move	0,3	2
n	movie-PL	movie	movie	0,285714	10
n	muffin-PL	muffin	muffin	1,076923	2
n	mum&dadj-DIM	mommy	mommy	1,4	35
n	mum&dadj-DIM~aux|be&3S	mommy	mommy	1,4	1
n	mum&dadj-DIM~cop|be&3S	mommy	mommy	1,4	3
n	mum&dadj-DIM~mod|do&3S	mommy	mommy	1,4	3
n	mum&dadj-DIM-PL	mommy	mommy	1,4	1
n	muscle-PL	muscle	muscle	-0,153846	16
n	mushroom-PL	mushroom	mushroom	-0,615385	1
n	nail-PL	nail	nail	0,181818	3
n	name~mod|do&3S	name	name	0,071429	1
n	name-PL	name	name	0,071429	8
n	nap-DIM	nap	nap	1,272727	4
n	napkin-PL	napkin	napkin	-0,545455	6
n	nap-PL	nap	nap	1,272727	1
n	neck-PL	neck	neck	-1	1
n	nickel-PL	nickel	nickel	0,727273	4
n	node-PL	node	node	-1	1
adj	noise	noisy	noisy	1,4	1
n	noise-DIM	noise	noise	0,909091	1
n	noise-PL	noise	noise	0,909091	5
n	nose-PL	nose	nose	0,25	4
n	number~mod|do&3S	number	number	-0,3	1
v	number-3S	number	number	-0,3	1

n	number-PL	number	number	-0,3	12
v	nurse-3S	nurse	nurse	1	1
n	nurse-PL	nurse	nurse	1	2
n	nut-PL	nut	nut	1,285714	2
n	olive-PL	olive	olive	0,6	1
n	onion-PL	onion	onion	0,181818	3
v	open-3S	open	open	1,1	15
v	open-PAST	open	open	1,1	12
n	orange-PL	orange	orange	0,230769	14
n	owl-PL	owl	owl	0,090909	4
n	own&dv-AGT	owner	owner	1	1
n	own&dv-AGT~poss|s	owner	owner	1	1
v	paint-PAST	painted	painted	1,454545	8
n	paint-PL	paint	paint	1,727273	3
n	pair-PL	pair	pair	-0,5	1
n	pan-PL	pan	pan	1,2	4
v	pant-3S	pant	pant	1,416667	7
n	paper-PL	paper	paper	0,416667	9
n	park-PL	park	park	0,384615	1
n	peach-DIM	peach	peach	0	3
n	peach-PL	peach	peach	0	3
n	pea-PL	pea	pea	1,6	19
n	pear-PL	pear	pear	0	8
adj	pen	penny	penny	-0,2	39
n	pencil-PL	pencil	pencil	0,625	2
n	penguin-PL	penguin	penguin	-1,583333	2
n	penny-PL	penny	penny	-0,2	33
n	pen-PL	pen	pen	0,6875	20
n	pepper-PL	pepper	pepper	1,571429	1
n	person&PL	person	person	-0,2	240
n	person&PL~aux|be&3S	person	person	-0,2	1
n	person&PL-PL	person	person	-0,2	13
n	person-PL	person	person	-0,2	1
n	piano-PL	piano	piano	-0,7	1
v	pick-3S	pick	pick	2,076923	3
n	pickle-PL	pickle	pickle	1,545455	3
v	pick-PAST	pick	pick	2,076923	11
n	picnic-PL	picnic	picnic	-0,538462	1
n	picture-PL	picture	picture	-0,071429	26
v	piece-3S	piece	piece	0,8	2
n	piece-PL	piece	piece	0,8	37
n	pie-PL	pie	pie	-0,272727	1
n	pig-DIM	pig	pig	0,5	50
n	pig-DIM-PL	pig	pig	0,5	9
n	pigeon-PL	pigeon	pigeon	0,181818	2
n	pig-PL	pig	pig	0,5	17
n	pillow-PL	pillow	pillow	0,466667	4
n	pill-PL	pill	pill	0,3	6
n	pineapple-PL	pineapple	pineapple	0,1	1
n	pine-PL	pine	pine	1,3	3
n	pink&dadj-DIM	pink	pink	-0,153846	1
n	pin-PL	pin	pin	1,909091	3
n	pipe-PL	pipe	pipe	0,333333	3
n	plane~poss|s	plane	plane	-0,461538	1
n	plane-PL	plane	plane	-0,461538	1
n	planet-PL	planet	planet	-0,615385	2
n	plan-PL	plan	plan	-0,076923	4
n	plant-PL	plant	plant	-0,2	12

n	plate-PL	plate	plate	0,928571	21
v	play-3S	play	play	1,083333	9
v	play-PAST	play	play	1,083333	42
n	play-PL	play	play	1,083333	1
n	plum-PL	plum	plum	0,727273	16
n	pole-PL	pole	pole	0,181818	3
n	pony-PL	pony	pony	-0,071429	4
v	pop-3S	pop	pop	4,076923	2
n	pop-PL	pop	pop	4,076923	5
adj	pot	potty	potty	0,7	15
n	potato-PL	potato	potato	0,857143	11
n	pot-DIM	pot	pot	1,545455	13
n	pot-PL	pot	pot	1,545455	2
v	pound-3S	pound	pound	0,25	1
n	pound-PL	pound	pound	0,25	1
v	pour-3S	pour	pour	1,3	1
v	pour-PAST	pour	pour	1,3	4
n	present-PL	present	present	0,307692	11
n	president-PL	president	president	0,583333	1
v	pretend-3S	pretend	pretend	-1	1
v	pretend-PAST	pretend	pretend	-1	3
n	prize-PL	prize	prize	-0,230769	1
v	pull-3S	pull	pull	2,083333	2
v	pull-PAST	pull	pull	2,083333	5
n	pumpkin-PL	pumpkin	pumpkin	-0,583333	1
adj	pup	puppy	puppy	0,666667	21
n	puppy~poss|s	puppy	puppy	0,666667	1
n	puppy-PL	puppy	puppy	0,666667	17
v	push-3S	push	push	2,333333	6
v	push-PAST	push	push	2,333333	12
v	put&ZERO	put	put	1,230769	1990
v	put-3S	put	put	1,230769	13
v	puzzle-3S	puzzle	puzzle	1,666667	1
n	puzzle-PL	puzzle	puzzle	1,666667	4
n	rabbit-PL	rabbit	rabbit	1,090909	8
n	raccoon-PL	raccoon	raccoon	0,727273	4
n	race-PL	race	race	-0,083333	1
n	rain-PL	rain	rain	0,333333	2
n	raisin-PL	raisin	raisin	0	22
n	rat-PL	rat	rat	1,636364	3
n	rattle-PL	rattle	rattle	3,272727	1
v	read&ZERO	read	read	0,090909	399
n	rectangle-PL	rectangle	rectangle	0,230769	1
n	red-PL	red	red	1,833333	1
n	ride-PL	ride	ride	1,166667	2
n	rifle-PL	rifle	rifle	0	1
v	ring-3S	ring	ring	3,153846	1
n	ring-PL	ring	ring	3,153846	2
v	rip-PAST	rip	rip	3,736842	2
n	robin-PL	robin	robin	0,818182	1
v	rock-3S	rock	rock	1,272727	4
n	rock-DIM	rock	rock	1,272727	1
n	rock-PL	rock	rock	1,272727	16
v	roll-3S	roll	roll	1,333333	2
v	roll-PAST	rolled	rolled	2,6	3
n	roll-PL	roll	roll	1,333333	1
n	room-PL	room	room	0,058824	2
n	rooster-PL	rooster	rooster	0,833333	1

n	rope-PL	rope	rope	0	4
n	rose-PL	rose	rose	0,545455	1
n	rubber-PL	rubber	rubber	0,8	1
v	run&PAST	run	run	0,909091	28
v	run-3S	run	run	0,909091	5
v	run-PAST	run	run	0,909091	1
n	run-PL	run	run	0,909091	2
n	sandwich-PL	sandwich	sandwich	-0,133333	3
n	sardine-PL	sardine	sardine	0,9	2
n	saucer-PL	saucer	saucer	1	4
v	saw-3S	saw	saw	1,733333	2
v	say&PAST	say	say	0,8	196
v	say-3S	say	say	0,8	75
n	scale-PL	scale	scale	-1,363636	2
adj	scare	scared	scared	1,384615	1
v	scare-3S	scare	scare	0,666667	1
v	scare-PAST	scare	scare	0,666667	2
n	school-PL	school	school	-0,214286	4
v	scramble-PAST	scramble	scramble	2,3	1
v	scratch-PAST	scratch	scratch	3,285714	11
adj	scream	screaming	screaming	2,083333	2
v	scream-3S	scream	scream	2,5	2
v	scream-PAST	scream	scream	2,5	2
n	scream-PL	scream	scream	2,5	1
n	seal-PL	seal	seal	-0,454545	6
v	seat-3S	seat	seat	1,454545	2
n	seat-PL	seat	seat	1,454545	1
n	second-PL	second	second	0,230769	1
v	see	saw	saw	1,733333	6
v	see&PAST	saw	saw	1,733333	131
v	see&PAST	saw	saw	2,733333	2
v	see&PAST	saw	saw	3,733333	2
v	see&PAST	saw	saw	4,733333	1
v	see&PAST	saw	see	1,272727	2
v	see&PAST	see	see	1,272727	1
v	see-3S	see	see	1,272727	5
v	seed-3S	seed	seed	-0,083333	3
n	seed-PL	seed	seed	-0,083333	20
adj	shag	shaggy	shaggy	1	10
v	shake-3S	shake	shake	2,615385	1
n	shake-PL	shake	shake	2,615385	1
v	shape-3S	shape	shape	0,8	1
n	shape-PL	shape	shape	0,8	1
v	share-3S	share	share	-0,444444	1
v	shell-3S	shell	shell	0,928571	2
n	shell-PL	shell	shell	0,928571	10
adj	shine	shiny	shiny	2,214286	17
n	ship-PL	ship	ship	0,8	1
n	shirt-PL	shirt	shirt	-0,375	3
n	shoe~mod|do&3S	shoe	shoe	-0,3125	1
n	shoe-PL	shoe	shoe	-0,3125	86
v	shoot&PAST	shoot	shoot	2,6	8
v	shoot-3S	shoot	shoot	2,6	9
n	shoot-PL	shoot	shoot	2,6	1
n	shot-PL	shot	shot	2,909091	2
v	shoulder-3S	shoulder	shoulder	-0,5	1
n	shoulder-PL	shoulder	shoulder	-0,5	1
n	shovel-PL	shovel	shovel	1,2	4

v	show-3S	show	show	0,846154	2
v	show-PAST	show	show	0,846154	9
n	show-PL	show	show	0,846154	6
n	sing&dv-AGT	singer	singer	1,909091	1
n	sing&dv-AGT-PL	singer	singer	1,909091	1
v	sing&PAST	sing	sing	1,7	2
v	sing-3S	sing	sing	1,7	3
n	sister-PL	sister	sister	0,533333	5
v	sit&PAST	sit	sit	0,866667	11
v	sit-3S	sit	sit	0,866667	16
n	skate-PL	skate	skate	1,416667	2
adj	skin	skinny	skinny	1,928571	7
n	skirt~poss|s	skirt	skirt	0,076923	1
n	skunk-PL	skunk	skunk	1,6	1
n	slack-PL	slack	slack	-0,3	1
adj	sleep	sleepy	sleepy	1,416667	19
v	sleep&PAST	sleep	sleep	0,769231	5
v	sleep-3S	sleep	sleep	0,769231	5
v	slide&PAST	slide	slide	1,416667	2
n	slide-PL	slide	slide	1,416667	2
v	slip-3S	slip	slip	2,615385	1
v	slip-PAST	slip	slip	2,615385	3
n	slipper-PL	slipper	slipper	1,538462	6
v	smash-3S	smash	smash	3,5	1
v	smash-PAST	smash	smash	3,5	6
adj	smell	smelly	smelly	0,923077	3
v	smell-3S	smell	smell	0,923077	3
v	smell-PAST	smell	smell	0,923077	1
n	smile-PL	smile	smile	1,727273	1
adj	smoke	smoky	smoky	1,3	1
v	snack-3S	snack	snack	0,583333	2
n	snack-PL	snack	snack	0,583333	5
v	sneak&PAST	sneak	sneak	3,4	2
v	sneak-3S	sneak	sneak	3,4	1
n	sneaker-PL	sneaker	sneaker	0	6
v	sneak-PAST	sneak	sneak	3,4	3
v	snore-3S	snore	snore	2,769231	1
n	snow~aux|be&3S	snow	snow	0,25	1
v	sock-3S	sock	sock	0,090909	5
n	sock-PL	sock	sock	0,090909	20
n	soldier-PL	soldier	soldier	0,6	2
n	song-PL	song	song	-0,3	2
v	sound-3S	sound	sound	0,833333	2
v	sound-PAST	sound	sound	0,833333	2
n	sound-PL	sound	sound	0,833333	1
v	speckle-PAST	speckle	speckle	-0,066667	1
n	spice-PL	spice	spice	0,538462	1
n	spider-PL	spider	spider	-0,545455	4
n	spike-PL	spike	spike	1,923077	1
PRESP	spill	spilling	spilling	1,5	2
v	spill&PAST	spill	spill	1,916667	4
v	spill-PAST	spill	spill	1,916667	36
v	spin-3S	spin	spin	1,454545	1
n	splash-PL	splash	splash	3,5	1
n	spoon-PL	spoon	spoon	0,3	15
v	spot-3S	spot	spot	1,285714	2
n	spot-PL	spot	spot	1,285714	9
n	square-PL	square	square	-0,214286	2

v	squeak-3S	squeak	squeak	4,230769	1
n	squeak-PL	squeak	squeak	4,230769	2
PRESP	squeal	squealing	squealing	2,75	1
v	squeeze-PAST	squeeze	squeeze	2,538462	1
n	squid-PL	squid	squid	1,727273	1
n	squirrel-PL	squirrel	squirrel	0,363636	1
v	stand&PAST	stand	stand	1,5	1
v	stand-3S	stand	stand	1,5	3
v	stand-PAST	stand	stand	1,5	4
v	star-3S	star	star	1,583333	1
n	star-PL	star	star	1,583333	9
v	start-3S	start	start	-0,181818	9
v	start-PAST	start	start	-0,181818	11
n	state-PL	state	state	0,071429	1
v	stay-3S	stay	stay	0,833333	5
v	stay-PAST	stay	stay	0,833333	12
n	stay-PL	stay	stay	0,833333	2
n	steak-PL	steak	steak	-1,4	1
v	steal-PAST	steal	steal	0	1
adj	stick	sticky	sticky	2,928571	17
v	stick&PAST	stick	stick	1,928571	33
v	stick-3S	stick	stick	1,928571	5
n	stick-PL	stick	stick	1,928571	3
v	sting-3S	sting	sting	2,615385	1
adj	stink	stinky	stinky	2,076923	12
v	stink-3S	stink	stink	1,5	1
n	stone-PL	stone	stone	1,272727	2
v	stop-3S	stop	stop	2,5	2
v	stop-PAST	stop	stop	2,5	2
n	stop-PL	stop	stop	2,5	5
adj	store	story	story	0,454545	24
n	store-PL	store	store	-0,875	2
n	storm-PL	storm	storm	0,428571	1
n	story-PL	story	story	0,454545	11
n	stranger-PL	stranger	stranger	-0,142857	1
n	strawberry-PL	strawberry	strawberry	0,375	26
n	stripe-PL	stripe	stripe	0,75	9
n	stroll&dv-AGT	stroller	stroller	0,785714	5
adj	sun	sunny	sunny	2	6
n	sweet&dadj-DIM	sweet	sweet	1,25	5
n	sweet&dadj-DIM-PL	sweet	sweet	1,25	13
n	sweet-PL	sweet	sweet	1,25	7
v	swim&PAST	swim	swim	1	6
v	swim-3S	swim	swim	1	4
n	swim-PL	swim	swim	1	1
adj	swing	swinging	swinging	1,923077	3
PRESP	swing	swinging	swinging	1,923077	5
v	swing-PAST	swing	swing	2,5	1
n	swing-PL	swing	swing	2,5	1
n	sword-PL	sword	sword	0,933333	2
n	tail-PL	tail	tail	-1,076923	1
v	take-3S	take	take	1,2	16
v	talk-3S	talk	talk	0,214286	6
v	talk-PAST	talk	talk	0,214286	5
n	talk-PL	talk	talk	0,214286	2
adj	tang	tangy	tangy	1,461538	1
n	tape-PL	tape	tape	0,5	10
n	tarantula-PL	tarantula	tarantula	1,2	1

adj	taste	tasty	tasty	1,076923	12
v	taste-3S	taste	taste	1,142857	6
n	taste-PL	taste	taste	1,142857	1
adj	tax	taxi	taxi	-0,5	5
n	teach&dv-AGT	teacher	teacher	0,9375	46
n	teach&dv-AGT~poss|s	teacher	teacher	0,9375	1
n	teach&dv-AGT-PL	teacher	teacher	0,9375	7
v	teach&PAST	teach	teach	0,153846	1
v	tear-3S	tear	tear	0,555556	1
n	tear-PL	tear	tear	0,555556	2
v	telephone-3S	telephone	telephone	-0,3	1
pro:exist	there~aux|be&3S	there	there	0,461538	2
pro:exist	there~cop|be&3S	there	there	0,461538	75
v	think-3S	think	think	0,363636	11
n	thread-PL	thread	thread	-0,272727	1
v	throw&PAST	throw	throw	0,785714	22
v	throw-PAST	throw	throw	0,785714	3
v	tickle-3S	tickle	tickle	2,142857	5
v	tickle-PAST	tickle	tickle	2,142857	1
n	tiger-PL	tiger	tiger	0,7	3
v	time-3S	time	time	-0,7	1
n	time-PL	time	time	-0,7	13
n	tissue-PL	tissue	tissue	1,692308	2
n	toast&dv-AGT-PL	toaster	toaster	1,5	2
v	toe-3S	toe	toe	0,142857	1
n	toe-PL	toe	toe	0,142857	17
n	toe-PL-DIM	toe	toe	0,142857	1
n	toilet-PL	toilet	toilet	0,8	1
n	tomato-PL	tomato	tomato	0	5
n	tool-PL	tool	tool	0,538462	11
n	tooth&PL	tooth	tooth	1	48
n	top-PL	top	top	1,090909	2
v	touch-PAST	touch	touch	0,823529	3
n	towel-PL	towel	towel	-0,235294	1
n	tower-PL	tower	tower	1,384615	1
v	toy-3S	toy	toy	0,909091	11
n	toy-PL	toy	toy	0,909091	132
v	track-3S	track	track	0,571429	1
n	track-PL	track	track	0,571429	1
n	tractor-PL	tractor	tractor	0,583333	3
n	trailer-PL	trailer	trailer	0,583333	2
n	train-PL	train	train	-0,090909	9
n	tree-PL	tree	tree	-0,923077	24
n	trip-PL	trip	trip	-0,384615	1
n	trouble-PL	trouble	trouble	-0,5	1
n	truck-PL	truck	truck	0,6	12
v	try-3S	try	try	-0,294118	5
v	try-PAST	try	try	-0,294118	10
n	tummy-PL	tummy	tummy	2	1
n	turkey-PL	turkey	turkey	0,066667	4
v	turn-3S	turn	turn	0	18
v	turn-PAST	turn	turn	0	25
n	turn-PL	turn	turn	0	14
n	turtle-PL	turtle	turtle	0	11
n	umbrella-PL	umbrella	umbrella	0,4	1
v	use-3S	use	use	1	1
v	use-PAST	use	use	1	15
n	voice-PL	voice	voice	0	5

v	wake&PAST	wake	wake	-0,181818	17
v	wake-3S	wake	wake	-0,181818	3
n	walk	walker	walker	0,8	1
n	walk&dv-AGT	walker	walker	0,8	8
v	walk-3S	walk	walk	0,727273	12
v	walk-PAST	walk	walk	0,727273	8
n	walk-PL	walk	walk	0,727273	5
n	wall-PL	wall	wall	0,4	4
v	want-3S	want	want	-0,214286	122
v	want-PAST	want	want	-0,214286	43
n	war-PL	war	war	-0,230769	5
v	wash-PAST	wash	wash	1,461538	5
v	watch-3S	watch	watch	1,153846	1
n	watch-PL	watch	watch	1,153846	2
v	weave&PAST	weave	weave	0,363636	1
n	weed-PL	weed	weed	1,7	1
n	week-PL	week	week	-0,461538	14
v	wet&ZERO	wet	wet	2,25	11
v	wet-3S	wet	wet	2,25	1
n	whale-PL	whale	whale	-1,3	4
n	wheel~poss|s	wheel	wheel	0,166667	1
v	wheel-3S	wheel	wheel	0,166667	6
n	wheel-DIM-PL	wheel	wheel	0,166667	2
n	wheel-PL	wheel	wheel	0,166667	72
v	whine-3S	whine	whine	2,666667	1
n	whistle-PL	whistle	whistle	2,153846	1
v	wiggle-3S	wiggle	wiggle	2,6	1
v	win&PAST	win	win	0,428571	7
v	win-3S	win	win	0,428571	2
n	win-DIM	win	win	0,428571	1
n	window-PL	window	window	-0,833333	7
v	wing-3S	wing	wing	1,6	1
n	wing-PL	wing	wing	1,6	10
n	witch-PL	witch	witch	0,416667	5
n	work&dv-AGT	worker	worker	1,5	4
n	work&dv-AGT-PL	worker	worker	1,5	1
v	work-3S	work	work	1,714286	22
v	work-PAST	work	work	1,714286	1
n	work-PL	work	work	1,714286	9
n	worm-PL	worm	worm	1,357143	5
PRESP	write	writing	writing	0,357143	21
v	write&PAST	write	write	1,181818	11
v	write-3S	write	write	1,181818	3
n	year-PL	year	year	-0,461538	23
v	yell-PAST	yell	yell	0,583333	4
n	zebra-PL	zebra	zebra	-0,538462	1
n	zero-PL	zero	zero	0,692308	2
v	zip-PAST	zip	zip	3,4	1
v	zipper-3S	zipper	zipper	2,923077	1

Appendix D. Occurrences of Normal Forms of Verbs which Match verbs in the Iconicity Table

Occurrences	CHILDES POS	Word-string extracted	CHILDES WORD ADJUSTED (the same)	Word in Iconicity Table	Iconicity	POS in Iconicity Table
2	v	add	add	add	-0,1538	verb
4	v	answer	answer	answer	0,5625	verb
3	v	avoid	avoid	avoid	-0,1667	verb
2	v	bang	bang	bang	3,8333	verb
4	v	bash	bash	bash	3,0833	verb
2	v	begin	begin	begin	1,5714	verb
27	v	belong	belong	belong	-0,1667	verb
57	v	bite	bite	bite	2,0000	verb
1	v	blame	blame	blame	0,4286	verb
81	v	blow	blow	blow	1,3636	verb
1	v	bob	bob	bob	1,2000	verb
2	v	bother	bother	bother	0,5000	verb
1	v	bounce	bounce	bounce	2,6429	verb
67	v	break	break	break	2,9000	verb
2	v	breathe	breathe	breathe	2,6667	verb
97	v	bring	bring	bring	0,1333	verb
26	v	brush	brush	brush	1,7273	verb
34	v	build	build	build	1,1538	verb
29	v	burn	burn	burn	0,4167	verb
2	v	burst&ZERO	burst	burst	2,9167	verb
96	v	buy	buy	buy	-0,0833	verb
90	v	call	call	call	0,4167	verb
31	v	carry	carry	carry	0,0000	verb
86	v	catch	catch	catch	0,4615	verb
2	v	cheat	cheat	cheat	-0,3636	verb
8	v	chew	chew	chew	2,2143	verb
3	v	chirp	chirp	chirp	4,1429	verb
13	v	choose	choose	choose	-0,4000	verb
2	v	clap	clap	clap	2,8333	verb
59	v	climb	climb	climb	0,8462	verb
9	v	comb	comb	comb	-0,0667	verb
734	v	come	come	come	0,2143	verb
51	v	cook	cook	cook	0,9000	verb
5	v	cost&ZERO	cost	cost	0,5385	verb
21	v	count	count	count	-0,2000	verb
4	v	cover	cover	cover	0,4286	verb
4	v	crawl	crawl	crawl	3,2727	verb
6	v	cross	cross	cross	0,6154	verb
46	v	cry	cry	cry	0,8667	verb

1	v	curl	curl	curl	0,6923	verb
135	v	cut&ZERO	cut	cut	0,7059	verb
20	v	dance	dance	dance	0,1538	verb
5	v	destroy	destroy	destroy	1,8571	verb
10	v	die	die	die	0,3333	verb
8	v	dig	dig	dig	0,6364	verb
11	v	dive	dive	dive	1,5714	verb
3	v	drag	drag	drag	2,0833	verb
105	v	draw	draw	draw	0,2000	verb
41	v	drive	drive	drive	0,2308	verb
76	v	drop	drop	drop	1,0714	verb
16	v	dump	dump	dump	2,9375	verb
690	v	eat	eat	eat	0,6471	verb
1	v	expect	expect	expect	-0,7500	verb
87	v	fall	fall	fall	1,5333	verb
3	v	fall	fell	fell	1,7143	verb
29	v	feed	feed	feed	0,8000	verb
44	v	feel	feel	feel	1,2500	verb
12	v	fight	fight	fight	0,3636	verb
176	v	find	find	find	0,0000	verb
58	v	finish	finish	finish	0,1818	verb
110	v	fit&ZERO	fit	fit	0,4706	verb
141	v	fix	fix	fix	0,5385	verb
1	v	float	float	float	0,5455	verb
47	v	fly	fly	fly	1,9231	verb
21	v	fold	fold	fold	0,6364	verb
13	v	follow	follow	follow	0,1429	verb
1	v	freeze	freeze	freeze	1,0667	verb
2	v	frown	frown	frown	1,4545	verb
1471	v	get	get	get	-0,5833	verb
404	v	give	give	give	1,2727	verb
2	v	glide	glide	glide	3,0000	verb
2583	v	go	go	go	1,4545	verb
28	v	hang	hang	hang	0,0000	verb
33	v	hate	hate	hate	1,4375	verb
1526	v	have	have	have	-0,2667	verb
141	v	hear	hear	hear	0,0667	verb
165	v	help	help	help	1,5333	verb
7	v	hide	hide	hide	1,1429	verb
169	v	hit&ZERO	hit	hit	1,0000	verb
170	v	hold	hold	hold	0,8462	verb
5	v	hop	hop	hop	3,3000	verb
14	v	hope	hope	hope	1,1250	verb
1	v	hunt	hunt	hunt	0,3571	verb

7	v	hurry	hurry	hurry	0,7500	verb
154	v	hurt&ZERO	hurt	hurt	0,9000	verb
1	v	jog	jog	jog	0,4545	verb
52	v	jump	jump	jump	1,0000	verb
14	v	kick	kick	kick	1,9500	verb
4	v	kill	kill	kill	1,6923	verb
10	v	kiss	kiss	kiss	1,2500	verb
3	v	kneel	kneel	kneel	2,0000	verb
2	v	knit&ZERO	knit	knit	0,1818	verb
70	v	knock	knock	knock	3,0625	verb
1474	v	know	know	know	0,7692	verb
19	v	laugh	laugh	laugh	0,7500	verb
56	v	leave	leave	leave	0,6154	verb
10	v	lick	lick	lick	1,1667	verb
4	v	lie	lie	lie	-0,6364	verb
18	v	lift	lift	lift	0,5714	verb
42	v	listen	listen	listen	0,7500	verb
47	v	live	live	live	0,4167	verb
744	v	look	look	look	1,8182	verb
8	v	lose	lose	lose	-0,3077	verb
69	v	love	love	love	1,1250	verb
635	v	make	make	make	0,6429	verb
2	v	manage	manage	manage	-0,1250	verb
3	v	march	march	march	0,9231	verb
1	v	measure	measure	measure	0,5455	verb
7	v	meet	meet	meet	-0,6154	verb
1	v	melt	melt	melt	0,6154	verb
3	v	meow	meow	meow	3,5000	verb
67	v	move	move	move	0,3000	verb
2	v	munch	munch	munch	3,4000	verb
2	v	nail	nail	nail	0,1818	verb
256	v	open	open	open	1,1000	verb
20	v	paint	paint	paint	1,7273	verb
2	v	pat	pat	pat	3,0000	verb
104	v	pick	pick	pick	2,0769	verb
340	v	play	play	play	1,0833	verb
32	v	pour	pour	pour	1,3000	verb
2	v	prefer	prefer	prefer	0,1667	verb
2	v	prepare	prepare	prepare	0,3846	verb
42	v	pretend	pretend	pretend	-1,0000	verb
99	v	pull	pull	pull	2,0833	verb
89	v	push	push	push	2,3333	verb
1990	v	put&ZERO	put	put	1,2308	verb
399	v	read&ZERO	read	read	0,0909	verb

2	v	recognize	recognize	recognize	0,0000	verb
2	v	reek	reek	reek	2,5333	verb
1	v	rip	rip	rip	3,7368	verb
36	v	roll	roll	roll	1,3333	verb
1	v	rub	rub	rub	1,5000	verb
54	v	run	run	run	0,9091	verb
1	v	sail	sail	sail	1,5455	verb
16	v	saw	saw	saw	1,7333	verb
334	v	say	say	say	0,8000	verb
5	v	scare	scare	scare	0,6667	verb
7	v	scoot	scoot	scoot	3,5455	verb
12	v	scratch	scratch	scratch	3,2857	verb
3	v	scream	scream	scream	2,5000	verb
4	v	seal	seal	seal	-0,4545	verb
6	v	see	saw	saw	1,7333	verb
1212	v	see	see	see	1,2727	verb
6	v	sew	sew	sew	-0,1250	verb
22	v	shake	shake	shake	2,6154	verb
20	v	share	share	share	-0,4444	verb
43	v	shoot	shoot	shoot	2,6000	verb
138	v	show	show	show	0,8462	verb
73	v	sing	sing	sing	1,7000	verb
347	v	sit	sit	sit	0,8667	verb
5	v	skedaddle	skedaddle	skedaddle	1,3000	verb
7	v	skip	skip	skip	2,0909	verb
56	v	sleep	sleep	sleep	0,7692	verb
5	v	slide	slide	slide	1,4167	verb
4	v	slip	slip	slip	2,6154	verb
7	v	smash	smash	smash	3,5000	verb
8	v	smell	smell	smell	0,9231	verb
9	v	smoke	smoke	smoke	2,3077	verb
2	v	sneak	sneak	sneak	3,4000	verb
3	v	sneeze	sneeze	sneeze	1,5000	verb
1	v	snore	snore	snore	2,7692	verb
1	v	sound	sound	sound	0,8333	verb
35	v	spill	spill	spill	1,9167	verb
5	v	spin	spin	spin	1,4545	verb
2	v	spit	spit	spit	1,7500	verb
1	v	squeak	squeak	squeak	4,2308	verb
4	v	squeeze	squeeze	squeeze	2,5385	verb
1	v	stab	stab	stab	2,2857	verb
50	v	stand	stand	stand	1,5000	verb
23	v	start	start	start	-0,1818	verb
44	v	stay	stay	stay	0,8333	verb

7	v	steal	steal	steal	0,0000	verb
24	v	stick	stick	stick	1,9286	verb
1	v	stomp	stomp	stomp	4,1000	verb
58	v	stop	stop	stop	2,5000	verb
2	v	sweep	sweep	sweep	3,2353	verb
15	v	swim	swim	swim	1,0000	verb
3	v	swing	swing	swing	2,5000	verb
851	v	take	take	take	1,2000	verb
94	v	talk	talk	talk	0,2143	verb
10	v	teach	teach	teach	0,1538	verb
13	v	tear	tear	tear	0,5556	verb
1	v	tend	tend	tend	-0,2500	verb
341	v	think	think	think	0,3636	verb
118	v	throw	throw	throw	0,7857	verb
36	v	tickle	tickle	tickle	2,1429	verb
25	v	tie	tie	tie	0,4615	verb
53	v	touch	touch	touch	0,8235	verb
2	v	trust	trust	trust	0,2000	verb
125	v	try	try	try	-0,2941	verb
215	v	turn	turn	turn	0,0000	verb
104	v	use	use	use	1,0000	verb
9	v	vote	vote	vote	-0,5714	verb
107	v	wait	wait	wait	-0,5000	verb
32	v	wake	wake	wake	-0,1818	verb
47	v	walk	walk	walk	0,7273	verb
2666	v	want	want	want	-0,2143	verb
35	v	wash	wash	wash	1,4615	verb
149	v	watch	watch	watch	1,1538	verb
1	v	weave	weave	weave	0,3636	verb
13	v	welcome	welcome	welcome	1,3846	verb
4	v	whisper	whisper	whisper	2,5385	verb
1	v	wiggle	wiggle	wiggle	2,6000	verb
7	v	win	win	win	0,4286	verb
22	v	wipe	wipe	wipe	2,0000	verb
14	v	wish	wish	wish	2,1818	verb
60	v	work	work	work	1,7143	verb
7	v	worry	worry	worry	0,4615	verb
211	v	write	write	write	1,1818	verb
1	v	yell	yell	yell	0,5833	verb
20	v	zip	zip	zip	3,4000	verb
7	v	zoom	zoom	zoom	3,6000	verb

Appendix E. Occurrences of 3S conjugations left to Match verbs in the Iconicity Table

Occurr ences	CHILDES POS	Word-string extracted	CHILDES WORD ADJUSTED	Word in Iconicity Table	Iconicity	POS in Iconicity Table
1	v	bang-3S	bang	bang	3,8333	verb
4	v	belong-3S	belong	belong	-0,1667	verb
4	v	bite-3S	bite	bite	2,0000	verb
1	v	blow-3S	blow	blow	1,3636	verb
3	v	break-3S	break	break	2,9000	verb
2	v	bring-3S	bring	bring	0,1333	verb
1	v	build-3S	build	build	1,1538	verb
5	v	buy-3S	buy	buy	-0,0833	verb
7	v	call-3S	call	call	0,4167	verb
3	v	carry-3S	carry	carry	0,0000	verb
1	v	catch-3S	catch	catch	0,4615	verb
1	v	clamp-3S	clamp	clamp	2,8182	verb
4	v	climb-3S	climb	climb	0,8462	verb
93	v	come-3S	come	come	0,2143	verb
5	v	cost-3S	cost	cost	0,5385	verb
1	v	crawl-3S	crawl	crawl	3,2727	verb
11	v	cry-3S	cry	cry	0,8667	verb
1	v	cut-3S	cut	cut	0,7059	verb
3	v	die-3S	die	die	0,3333	verb
228	v	do&3S	does	does	-0,6000	verb
2	v	do-3S	do	do	0,8462	verb
1	v	drive-3S	drive	drive	0,2308	verb
4	v	drop-3S	drop	drop	1,0714	verb
3	v	dump-3S	dump	dump	2,9375	verb
25	v	eat-3S	eat	eat	0,6471	verb
23	v	fall-3S	fall	fall	1,5333	verb
1	v	feed-3S	feed	feed	0,8000	verb
8	v	feel-3S	feel	feel	1,2500	verb
1	v	find-3S	find	find	0,0000	verb
1	v	finish-3S	finish	finish	0,1818	verb
20	v	fit-3S	fit	fit	0,4706	verb
47	v	get-3S	get	get	-0,5833	verb
9	v	give-3S	give	give	1,2727	verb
446	v	go-3S	go	go	1,4545	verb
2	v	grind-3S	grind	grind	2,0000	verb
1	v	hang-3S	hang	hang	0,0000	verb
1	v	hate-3S	hate	hate	1,4375	verb

7	v	have&3S	have	have	-0,2667	verb
1	v	have-3S	have	have	-0,2667	verb
1	v	help-3S	help	help	1,5333	verb
5	v	hit-3S	hit	hit	1,0000	verb
3	v	hold-3S	hold	hold	0,8462	verb
2	v	hop-3S	hop	hop	3,3000	verb
33	v	hurt-3S	hurt	hurt	0,9000	verb
9	v	jump-3S	jump	jump	1,0000	verb
3	v	kick-3S	kick	kick	1,9500	verb
2	v	kill-3S	kill	kill	1,6923	verb
1	v	kiss-3S	kiss	kiss	1,2500	verb
2	v	knock-3S	knock	knock	3,0625	verb
6	v	know-3S	know	know	0,7692	verb
1	v	leave-3S	leave	leave	0,6154	verb
1	v	lie-3S	lie	lie	-0,6364	verb
19	v	live-3S	live	live	0,4167	verb
64	v	look-3S	look	look	1,8182	verb
13	v	love-3S	love	love	1,1250	verb
32	v	make-3S	make	make	0,6429	verb
2	v	meet-3S	meet	meet	-0,6154	verb
2	v	move-3S	move	move	0,3000	verb
15	v	open-3S	open	open	1,1000	verb
3	v	pick-3S	pick	pick	2,0769	verb
9	v	play-3S	play	play	1,0833	verb
1	v	pour-3S	pour	pour	1,3000	verb
1	v	pretend-3S	pretend	pretend	-1,0000	verb
2	v	pull-3S	pull	pull	2,0833	verb
6	v	push-3S	push	push	2,3333	verb
13	v	put-3S	put	put	1,2308	verb
2	v	roll-3S	roll	roll	1,3333	verb
5	v	run-3S	run	run	0,9091	verb
2	v	saw-3S	saw	saw	1,7333	verb
75	v	say-3S	say	say	0,8000	verb
1	v	scare-3S	scare	scare	0,6667	verb
2	v	scream-3S	scream	scream	2,5000	verb
5	v	see-3S	see	see	1,2727	verb
1	v	shake-3S	shake	shake	2,6154	verb
1	v	share-3S	share	share	-0,4444	verb
9	v	shoot-3S	shoot	shoot	2,6000	verb
2	v	show-3S	show	show	0,8462	verb
3	v	sing-3S	sing	sing	1,7000	verb
16	v	sit-3S	sit	sit	0,8667	verb
5	v	sleep-3S	sleep	sleep	0,7692	verb
1	v	slip-3S	slip	slip	2,6154	verb

1	v	smash-3S	smash	smash	3,5000	verb
3	v	smell-3S	smell	smell	0,9231	verb
1	v	sneak-3S	sneak	sneak	3,4000	verb
1	v	snore-3S	snore	snore	2,7692	verb
2	v	sound-3S	sound	sound	0,8333	verb
1	v	spin-3S	spin	spin	1,4545	verb
1	v	squeak-3S	squeak	squeak	4,2308	verb
3	v	stand-3S	stand	stand	1,5000	verb
9	v	start-3S	start	start	-0,1818	verb
5	v	stay-3S	stay	stay	0,8333	verb
5	v	stick-3S	stick	stick	1,9286	verb
1	v	stink-3S	stink	stink	1,5000	verb
2	v	stop-3S	stop	stop	2,5000	verb
4	v	swim-3S	swim	swim	1,0000	verb
16	v	take-3S	take	take	1,2000	verb
6	v	talk-3S	talk	talk	0,2143	verb
1	v	tear-3S	tear	tear	0,5556	verb
11	v	think-3S	think	think	0,3636	verb
5	v	tickle-3S	tickle	tickle	2,1429	verb
5	v	try-3S	try	try	-0,2941	verb
18	v	turn-3S	turn	turn	0,0000	verb
1	v	use-3S	use	use	1,0000	verb
3	v	wake-3S	wake	wake	-0,1818	verb
12	v	walk-3S	walk	walk	0,7273	verb
122	v	want-3S	want	want	-0,2143	verb
1	v	watch-3S	watch	watch	1,1538	verb
1	v	whine-3S	whine	whine	2,6667	verb
1	v	wiggle-3S	wiggle	wiggle	2,6000	verb
2	v	win-3S	win	win	0,4286	verb
22	v	work-3S	work	work	1,7143	verb
3	v	write-3S	write	write	1,1818	verb

Appendix F. Protective changes of present participles

CHILDES POS ORIGINAL	Word-string extracted	CHILDES WORD ADJUSTED	**TO NOT MATCH**	Word_Winter_et_al	Brysetal_POS	Occurencies
PRESP	back	back~ing		back	adverb	4
PRESP	be	be~ing		be	verb	8
PRESP	begin	begin~ing		begin	verb	2
PRESP	bite	bite~ing		bite	verb	6
PRESP	block	block~ing		block	noun	1
PRESP	blow	blow~ing		blow	verb	7
PRESP	boss	boss~ing		boss	noun	1
PRESP	bother	bother~ing		bother	verb	2
PRESP	bounce	bounce~ing		bounce	verb	1
PRESP	bowl	bowl~ing		bowl	noun	3
PRESP	break	break~ing		break	verb	11
PRESP	bring	bring~ing		bring	verb	1
PRESP	brush	brush~ing		brush	verb	8
PRESP	build	build~ing		build	verb	3
PRESP	buy	buy~ing		buy	verb	7
PRESP	call	call~ing		call	verb	7
PRESP	camp	camp~ing		camp	noun	1
PRESP	carry	carry~ing		carry	verb	6
PRESP	catch	catch~ing		catch	verb	1
PRESP	chase	chase~ing		chase	noun	7
PRESP	chew	chew~ing		chew	verb	2
PRESP	chop	chop~ing		chop	noun	1
PRESP	clap	clap~ing		clap	verb	2
PRESP	clean	clean~ing		clean	adjective	7
PRESP	climb	climb~ing		climb	verb	22
PRESP	close	close~ing		close	adjective	2
PRESP	color	color~ing		color	noun	9
PRESP	come	come~ing		come	verb	97
PRESP	cook	cook~ing		cook	verb	29
PRESP	count	count~ing		count	verb	3
PRESP	crash	crash~ing		crash	noun	1
PRESP	crawl	crawl~ing		crawl	verb	5
PRESP	cross	cross~ing		cross	verb	4
PRESP	cut	cut~ing		cut	verb	8
PRESP	dance	dance~ing		dance	verb	5
PRESP	destroy	destroy~ing		destroy	verb	1
PRESP	die	die~ing		die	verb	1
PRESP	dig	dig~ing		dig	verb	1
PRESP	ditch	ditch~ing		ditch	verb	1
PRESP	dive	dive~ing		dive	verb	1

PRESP	do	do~ing		do	verb	224
PRESP	draw	draw~ing		draw	verb	1
PRESP	dream	dream~ing		dream	noun	1
PRESP	dress	dress~ing		dress	noun	3
PRESP	drink	drink~ing		drink	noun	20
PRESP	drive	drive~ing		drive	verb	26
PRESP	dump	dump~ing		dump	verb	1
PRESP	dust	dust~ing		dust	noun	1
PRESP	eat	eat~ing		eat	verb	53
PRESP	face	face~ing		face	noun	2
PRESP	feed	feed~ing		feed	verb	4
PRESP	feel	feel~ing		feel	verb	12
PRESP	fight	fight~ing		fight	verb	13
PRESP	find	find~ing		find	verb	1
PRESP	fire	fire~ing		fire	noun	2
PRESP	fish	fish~ing		fish	noun	12
PRESP	fix	fix~ing		fix	verb	8
PRESP	float	float~ing		float	verb	2
PRESP	fly	fly~ing		fly	verb	18
PRESP	fold	fold~ing		fold	verb	1
PRESP	follow	follow~ing		follow	verb	10
PRESP	fool	fool~ing		fool	noun	1
PRESP	freeze	freeze~ing		freeze	verb	3
PRESP	frown	frown~ing		frown	verb	1
PRESP	get	get~ing		get	verb	108
PRESP	give	give~ing		give	verb	11
PRESP	go	go~ing		go	verb	2050
PRESP	hang	hang~ing		hang	verb	14
PRESP	have	have~ing		have	verb	42
PRESP	hear	hear~ing		hear	verb	3
PRESP	help	help~ing		help	verb	15
PRESP	hide	hide~ing		hide	verb	13
PRESP	hit	hit~ing		hit	verb	7
PRESP	hold	hold~ing		hold	verb	28
PRESP	hop	hop~ing		hop	verb	10
PRESP	hope	hope~ing		hope	verb	3
PRESP	hug	hug~ing		hug	noun	2
PRESP	hurry	hurry~ing		hurry	verb	1
PRESP	hurt	hurt~ing		hurt	verb	4
PRESP	invite	invite~ing		invite	verb	1
PRESP	jog	jog~ing		jog	verb	6
PRESP	joke	joke~ing		joke	noun	1
PRESP	jump	jump~ing		jump	verb	16
PRESP	kick	kick~ing		kick	verb	11
PRESP	kid	kid~ing		kid	noun	8
PRESP	kiss	kiss~ing		kiss	verb	3
PRESP	kneel	kneel~ing		kneel	verb	16
PRESP	knock	knock~ing		knock	verb	8
PRESP	land	land~ing		land	noun	2

PRESP	leave	leave~ing		leave	verb	7
PRESP	lick	lick~ing		lick	verb	1
PRESP	lie	lie~ing		lie	verb	13
PRESP	lift	lift~ing		lift	verb	3
PRESP	listen	listen~ing		listen	verb	6
PRESP	live	live~ing		live	verb	5
PRESP	look	look~ing		look	verb	105
PRESP	lose	lose~ing		lose	verb	2
PRESP	love	love~ing		love	verb	1
PRESP	make	make~ing		make	verb	122
PRESP	march	march~ing		march	verb	3
PRESP	meet	meet~ing		meet	verb	2
PRESP	meow	meow~ing		meow	verb	1
PRESP	milk	milk~ing		milk	noun	1
PRESP	mind	mind~ing		mind	noun	2
PRESP	miss	miss~ing		miss	noun	9
PRESP	mop	mop~ing		mop	noun	1
PRESP	move	move~ing		move	verb	10
PRESP	name	name~ing		name	noun	1
PRESP	nap	nap~ing		nap	noun	1
PRESP	open	open~ing		open	verb	6
PRESP	paint	paint~ing		paint	verb	12
PRESP	park	park~ing		park	noun	3
PRESP	pat	pat~ing		pat	verb	1
PRESP	peep	peep~ing		peep	noun	1
PRESP	pick	pick~ing		pick	verb	8
PRESP	plant	plant~ing		plant	noun	3
PRESP	play	play~ing		play	verb	132
PRESP	pour	pour~ing		pour	verb	3
PRESP	pretend	pretend~ing		pretend	verb	10
PRESP	pull	pull~ing		pull	verb	16
PRESP	punch	punch~ing		punch	noun	2
PRESP	push	push~ing		push	verb	6
PRESP	put	put~ing		put	verb	37
PRESP	race	race~ing		race	noun	10
PRESP	rain	rain~ing		rain	noun	21
PRESP	read	read~ing		read	verb	16
PRESP	ride	ride~ing		ride	noun	19
PRESP	ring	ring~ing		ring	noun	1
PRESP	rip	rip~ing		rip	verb	1
PRESP	rock	rock~ing		rock	noun	3
PRESP	roll	roll~ing		roll	verb	10
PRESP	rope	rope~ing		rope	noun	1
PRESP	run	run~ing		run	verb	37
PRESP	rush	rush~ing		rush	noun	1
PRESP	sail	sail~ing		sail	verb	3
PRESP	saw	saw~ing		saw	verb	1
PRESP	say	say~ing		say	verb	21
PRESP	scare	scare~ing		scare	verb	1

PRESP	scratch	scratch~ing		scratch	verb	3
PRESP	scream	scream~ing		scream	verb	3
PRESP	see	see~ing		see	verb	5
PRESP	shake	shake~ing		shake	verb	2
PRESP	share	share~ing		share	verb	6
PRESP	shine	shine~ing		shine	verb	7
PRESP	shovel	shovel~ing		shovel	noun	1
PRESP	show	show~ing		show	verb	3
PRESP	sing	sing~ing		sing	verb	13
PRESP	sit	sit~ing		sit	verb	56
PRESP	skate	skate~ing		skate	verb	1
PRESP	sled	sled~ing		sled	adjective	1
PRESP	sleep	sleep~ing		sleep	verb	37
PRESP	slip	slip~ing		slip	verb	1
PRESP	smash	smash~ing		smash	verb	1
PRESP	smile	smile~ing		smile	noun	10
PRESP	smoke	smoke~ing		smoke	verb	2
PRESP	sneak	sneak~ing		sneak	verb	7
PRESP	sniff	sniff~ing		sniff	verb	1
PRESP	snow	snow~ing		snow	noun	3
PRESP	spin	spin~ing		spin	verb	5
PRESP	splash	splash~ing		splash	noun	2
PRESP	squeeze	squeeze~ing		squeeze	verb	1
PRESP	stand	stand~ing		stand	verb	28
PRESP	start	start~ing		start	verb	7
PRESP	stay	stay~ing		stay	verb	5
PRESP	stick	stick~ing		stick	verb	11
PRESP	sting	sting~ing		sting	noun	1
PRESP	stink	stink~ing		stink	verb	1
PRESP	stop	stop~ing		stop	verb	3
PRESP	sweep	sweep~ing		sweep	verb	11
PRESP	swim	swim~ing		swim	verb	21
PRESP	take	take~ing		take	verb	56
PRESP	talk	talk~ing		talk	verb	61
PRESP	tape	tape~ing		tape	noun	1
PRESP	taste	taste~ing		taste	noun	1
PRESP	teach	teach~ing		teach	verb	3
PRESP	tear	tear~ing		tear	verb	1
PRESP	tend	tend~ing		tend	verb	2
PRESP	that	that~ing		that	functionword	1
PRESP	think	think~ing		think	verb	6
PRESP	throw	throw~ing		throw	verb	4
PRESP	tickle	tickle~ing		tickle	verb	4
PRESP	tie	tie~ing		tie	verb	1
PRESP	time	time~ing		time	noun	1
PRESP	touch	touch~ing		touch	verb	3
PRESP	try	try~ing		try	verb	64
PRESP	tumble	tumble~ing		tumble	verb	1
PRESP	turn	turn~ing		turn	verb	18

PRESP	twist	twist~ing		twist	noun	4
PRESP	type	type~ing		type	noun	2
PRESP	use	use~ing		use	verb	14
PRESP	vacuum	vacuum~ing		vacuum	noun	1
PRESP	wait	wait~ing		wait	verb	31
PRESP	wake	wake~ing		wake	verb	11
PRESP	walk	walk~ing		walk	verb	44
PRESP	want	want~ing		want	verb	2
PRESP	wash	wash~ing		wash	verb	10
PRESP	watch	watch~ing		watch	verb	23
PRESP	water	water~ing		water	noun	4
PRESP	wheel	wheel~ing		wheel	noun	4
PRESP	whine	whine~ing		whine	verb	1
PRESP	whistle	whistle~ing		whistle	noun	2
PRESP	win	win~ing		win	verb	2
PRESP	wipe	wipe~ing		wipe	verb	2
PRESP	wobble	wobble~ing		wobble	verb	2
PRESP	work	work~ing		work	verb	20
PRESP	worry	worry~ing		worry	verb	1

Appendix G. Protective Changes of Past Tense Forms

string extracted	CHILDES WORD ADJUSTED	Changed to Not Match	Word in Iconicity Table	POS in Iconicity Table
back-PAST	back~d		back	adverb
block-PAST	block~d		block	noun
button-PAST	button~d		button	noun
charge-PAST	charge~d		charge	noun
chase-PAST	chase~d		chase	noun
clean-PAST	clean~d		clean	adjective
clear-PAST	clear~d		clear	adjective
color-PAST	color~d		color	noun
crash-PAST	crash~d		crash	noun
creep&PAST	creep~d		creep	noun
crowd-PAST	crowd~d		crowd	noun
damage-PAST	damage~d		damage	noun
down-PAST	down~d		down	adverb
dream&PAST	dream~d		dream	noun
dream-PAST	dream~d		dream	noun
dress-PAST	dress~d		dress	noun
drink&PAST	drink~d		drink	noun
dry-PAST	dry~d		dry	adjective
glue-PAST	glue~d		glue	noun
hand-PAST	hand~d		hand	noun

hook-PAST	hook~d		hook	noun
land-PAST	land~d		land	noun
leap-PAST	leap~d		leap	noun
light&PAST	light~d		light	noun
light-PAST	light~d		light	noun
like-PAST	like~d		like	functionword
load-PAST	load~d		load	noun
miss-PAST	miss~d		miss	noun
name-PAST	name~d		name	noun
own-PAST	own~d		own	functionword
park-PAST	park~d		park	noun
pop-PAST	pop~d		pop	onomatopoeia
powder-PAST	powder~d		powder	noun
press-PAST	press~d		press	noun
puff-PAST	puff~d		puff	noun
punch-PAST	punch~d		punch	noun
quack-PAST	quack~d		quack	onomatopoeia
ride&PAST	ride~d		ride	noun
ring&PAST	ring~d		ring	noun
ring-PAST	ring~d		ring	noun
rock-PAST	rock~d		rock	noun
rush-PAST	rush~d		rush	noun
splash-PAST	splash~d		splash	noun
sting&PAST	sting~d		sting	noun
sting-PAST	sting~d		sting	noun
taste-PAST	taste~d		taste	noun
tick-PAST	tick~d		tick	noun
trash-PAST	trash~d		trash	noun
trip-PAST	trip~d		trip	noun
whistle-PAST	whistle~d		whistle	noun
whistle-PAST	whistle~d		whistle	noun

Appendix I. Past tense forms of Regular verbs left to match the normal form in the Iconicity Table

Occurrences	CHILDES POS ORIGINAL	Word-string extracted	CHILDES WORD ADJUSTED	Word in Iconicity Table	Iconicity	POS Iconicity Table
2	v	bang-PAST	bang	bang	3,83333	verb
1	v	boil-PAST	boil	boil	1,38462	verb
39	v	call-PAST	call	call	0,41667	verb
10	v	climb-PAST	climb	climb	0,84615	verb
4	v	cover-PAST	cover	cover	0,42857	verb

1	v	crawl-PAST	crawl	crawl	3,27273	verb
1	v	cross-PAST	cross	cross	0,61538	verb
7	v	cry-PAST	cry	cry	0,86667	verb
1	v	destroy-PAST	destroy	destroy	1,85714	verb
9	v	die-PAST	die	die	0,33333	verb
1	v	drag-PAST	drag	drag	2,08333	verb
21	v	drop-PAST	drop	drop	1,07143	verb
3	v	dump-PAST	dump	dump	2,93750	verb
34	v	finish-PAST	finish	finish	0,18182	verb
1	v	follow-PAST	follow	follow	0,14286	verb
10	v	help-PAST	help	help	1,53333	verb
3	v	hold-PAST	hold	hold	0,84615	verb
2	v	invite-PAST	invite	invite	0,23077	verb
1	v	join-PAST	join	join	0,00000	verb
15	v	jump-PAST	jump	jump	1,00000	verb
6	v	kick-PAST	kick	kick	1,95000	verb
5	v	kill-PAST	kill	kill	1,69231	verb
3	v	kiss-PAST	kiss	kiss	1,25000	verb
17	v	knock-PAST	knock	knock	3,06250	verb
4	v	lick-PAST	lick	lick	1,16667	verb
1	v	lie-PAST	lie	lie	-0,63636	verb
2	v	lift-PAST	lift	lift	0,57143	verb
2	v	listen-PAST	listen	listen	0,75000	verb
1	v	live-PAST	live	live	0,41667	verb
19	v	look-PAST	look	look	1,81818	verb
1	v	love-PAST	love	love	1,12500	verb
2	v	melt-PAST	melt	melt	0,61538	verb
2	v	move-PAST	move	move	0,30000	verb
12	v	open-PAST	open	open	1,10000	verb
11	v	pick-PAST	pick	pick	2,07692	verb
42	v	play-PAST	play	play	1,08333	verb
4	v	pour-PAST	pour	pour	1,30000	verb
3	v	pretend-PAST	pretend	pretend	-1,00000	verb
5	v	pull-PAST	pull	pull	2,08333	verb
12	v	push-PAST	push	push	2,33333	verb
2	v	rip-PAST	rip	rip	3,73684	verb
2	v	scare-PAST	scare	scare	0,66667	verb
1	v	scramble-PAST	scramble	scramble	2,30000	verb
11	v	scratch-PAST	scratch	scratch	3,28571	verb
9	v	show-PAST	show	show	0,84615	verb
3	v	slip-PAST	slip	slip	2,61538	verb
6	v	smash-PAST	smash	smash	3,50000	verb
1	v	smell-PAST	smell	smell	0,92308	verb
2	v	sound-PAST	sound	sound	0,83333	verb
4	v	spill&PAST	spill	spill	1,91667	verb
36	v	spill-PAST	spill	spill	1,91667	verb

1	v	squeeze-PAST	squeeze	squeeze	2,53846	verb
1	v	stand&PAST	stand	stand	1,50000	verb
4	v	stand-PAST	stand	stand	1,50000	verb
11	v	start-PAST	start	start	-0,18182	verb
12	v	stay-PAST	stay	stay	0,83333	verb
2	v	stop-PAST	stop	stop	2,50000	verb
5	v	talk-PAST	talk	talk	0,21429	verb
22	v	throw&PAST	throw	throw	0,78571	verb
3	v	throw-PAST	throw	throw	0,78571	verb
1	v	tickle-PAST	tickle	tickle	2,14286	verb
3	v	touch-PAST	touch	touch	0,82353	verb
10	v	try-PAST	try	try	-0,29412	verb
25	v	turn-PAST	turn	turn	0,00000	verb
15	v	use-PAST	use	use	1,00000	verb
8	v	walk-PAST	walk	walk	0,72727	verb
43	v	want-PAST	want	want	-0,21429	verb
5	v	wash-PAST	wash	wash	1,46154	verb
1	v	work-PAST	work	work	1,71429	verb
4	v	yell-PAST	yell	yell	0,58333	verb
1	v	zip-PAST	zip	zip	3,40000	verb

NB. The past participles which are the same as the past tense forms were considered in the same way.

Appendix J. Past tense forms of Irregular verbs protected from matching the Iconicity Table

occurencies	Word Srting Extracted	CHILDES Word ADJUSTED	to not match	word in the Iconocoty Table	Brysetal_POS
4	begin&PAST	begin~d		begin	verb
38	bite&PAST	bite~d		bite	verb
6	blow&PAST	blow~d		blow	verb
5	blow-PAST	blow~d		blow	verb
77	break&PAST	broke~		break	verb
27	bring&PAST	brought~		bring	verb
3	burn&PAST	burn~d		burn	verb
2	burn-PAST	burn~d		burn	verb
21	buy&PAST	bought~		buy	verb
53	catch&PAST	caught~		catch	verb
1	catch-PAST	caught~		catch	verb
5	choose&PAST	choose~d		choose	verb
128	come&PAST	come~d		come	verb

6	draw&PAST	draw~d		draw	verb
119	eat&PAST	ate~d		eat	verb
99	find&PAST	foun~d		find	verb
3	fly&PAST	flew~d		fly	verb
996	get&PAST	got~d		get	verb
56	give&PAST	gave~d		give	verb
302	go&PAST	went~		go	verb
230	have&PAST	ha~d		have	verb
16	hear&PAST	hear~d		hear	verb
1	hear-PAST	hear~d		hear	verb
1	hide&PAST	hid~d		hide	verb
12	know&PAST	knew~d		know	verb
25	leave&PAST	left~		leave	verb
73	lose&PAST	los~d		lose	verb
173	make&PAST	mad~d		make	verb
4	meet&PAST	met~d		meet	verb
28	run&PAST	ran~d		run	verb
196	say&PAST	sai~d		say	verb
8	shoot&PAST	shot~d		shoot	verb
11	sit&PAST	sat~d		sit	verb
5	sleep&PAST	slep~d		sleep	verb
4	steal&PAST	stole~d		steal	verb
6	swim&PAST	swam ~d		swim	verb
111	take&PAST	took~		take	verb
1	take-PAST	took~		take	verb
1	teach&PAST	taught~d		teach	verb
43	think&PAST	thought~		think	verb
17	wake&PAST	woke~d		wake	verb
7	win&PAST	won~d		win	verb
11	write&PAST	wrote~d		write	verb

Appendix K. Irregular verbs mistaken by children - with examples of use.

These noticed mistaken occurrences were adjusted to match the form used by the child and existing in the Iconicity Table

Word in Iconicity Table	Iconicity norm	POS in Iconicity Table	CHILDES POS	word-string Extra cted	Child speech whith the verb- example	CHILDES POS-annotation (%mor)
break	2,90	verb	v	break	&eh eh and ***breaked*** .	co|eh coord|and v|break-PAST .
break	2,90	verb	v	break	it *breaked* .	pro|it v|break-PAST .
bring	0,13	verb	v	bring	I should have ***brung*** [: brought] my sun glasses .	pro:sub|I mod|should aux|have v|bring&PAST pro:poss:det|my n|sun n|glass-PL .
catch	0,46	verb	v	catch	I ***catched*** another fish .	pro:sub|I v|catch-PAST qn|another n|fish .
do	0,85	verb	v	do	see what he ***doed*** ?	v|see rel|what pro:sub|he v|do-PAST ?

<table>
<tr><th></th><th></th><th></th><th></th><th></th><th></th><th></th></tr>
<tr><td>draw</td><td>0,20</td><td>verb</td><td>v</td><td>draw</td><td>who drawed on it ?</td><td>pro:wh|who v|draw-PAST prep|on pro|it ?</td></tr>
<tr><td>drive</td><td>0,23</td><td>verb</td><td>v</td><td>drive</td><td>they drived on their bikes .</td><td>pro:sub|they v|drive-PAST prep|on pro:poss:det|their n|bike-PL .</td></tr>
<tr><td>fly</td><td>1,92</td><td>verb</td><td>v</td><td>fly</td><td>it flied out_of his hand .</td><td>pro|it v|fly-PAST prep|out_of pro:poss:det|his n|hand .</td></tr>
<tr><td>freeze</td><td>1,07</td><td>verb</td><td>v</td><td>freeze</td><td>you frozed [: froze] [* m:+ed] it .</td><td>pro|you v|freeze&PAST pro|it .</td></tr>
<tr><td>go</td><td>1,45</td><td>verb</td><td>v</td><td>go</td><td>she goed away .</td><td>pro:sub|she v|go-PAST adv|away .</td></tr>
<tr><td>hide</td><td>1,14</td><td>verb</td><td>v</td><td>hide</td><td>we hided [: hid] [* m:=ed] the ice cream .</td><td>pro:sub|we v|hide&PAST det|the n|ice n|cream .</td></tr>
<tr><td>run</td><td>0,91</td><td>verb</td><td>v</td><td>run</td><td>he ranned [: ran] [* +ed] away .</td><td>pro:sub|he v|run&PAST adv|away .</td></tr>
<tr><td>saw</td><td>1,73</td><td>verb</td><td>v</td><td>saw</td><td>did I saw dat [: that] [?] in my book ?</td><td>v|do&PAST pro:sub|I v|see&PAST rel|that prep|in pro:poss:det|my n|book ?</td></tr>
<tr><td>sneak</td><td>3,40</td><td>verb</td><td>v</td><td>sneak</td><td>and [/] and I sneaked over her door ?</td><td>conj|and pro:sub|I v|sneak-PAST prep|over pro:poss:det|her n|door ?</td></tr>
<tr><td>steal</td><td>0,00</td><td>verb</td><td>v</td><td>steal</td><td>they [= bees] stealed the honey .</td><td>pro:sub|they v|steal-PAST det|the n|honey .</td></tr>
<tr><td>teach</td><td>0,15</td><td>verb</td><td>v</td><td>teach</td><td><that's because I <i>teached</i> [: taught] [*] > [>] +...</td><td>pro:dem|that cop|be&3S conj|because pro:sub|I v|teach&PAST +...</td></tr>
<tr><td>fall</td><td>1,53</td><td>verb</td><td>PASTP</td><td>fall</td><td>it's felled [: fallen] [* ed-en] .</td><td>pro|it aux|be&3S part|fall&PASTP .</td></tr>
<tr><td>fly</td><td>1,92</td><td>verb</td><td>PASTP</td><td>fly</td><td>it's flied away .</td><td>pro|it~aux|be&3S part|fly-PASTP adv|away .</td></tr>
<tr><td>hide</td><td>1,14</td><td>verb</td><td>PASTP</td><td>hide</td><td>first he was back to his playing house and he begin to hide in there and then hided [*] at the door .</td><td>... adv:tem|then part|hide-PASTP prep|at det|the n|door .</td></tr>
<tr><td>run</td><td>0,91</td><td>verb</td><td>PASTP</td><td>run</td><td>he's been runned over .</td><td>pro:sub|he~aux|be&3S aux|be&PASTP part|run-PASTP adv|over .</td></tr>
<tr><td>wake</td><td>-0,18</td><td>verb</td><td>PASTP</td><td>wake</td><td>I waked evwybody [: everybody] up .</td><td>pro:sub|I part|wake-PASTP pro:indef|everybody prep|up .</td></tr>
</table>

Appendix L. Protective Changes of Past Participles

Occurencies	CHILDES POS ORIGINAL	Word-string extracted	CHILDES WORD ADJUSTED	To Not Match	Word in Iconicity Table	Iconicity	POS in Iconicity T.
1	PASTP	butter	butter~d		butter	0,5	noun
3	PASTP	chase	chase~d		chase	0,7	noun
5	PASTP	chop	chop~d		chop	3,3	noun
2	PASTP	clean	clean~d		clean	0,9	adjective
3	PASTP	color	color~d		color	0,1	noun
1	PASTP	crash	crash~d		crash	3,8	noun
3	PASTP	crowd	crowd~d		crowd	1,1	noun
5	PASTP	dress	dress~d		dress	0,0	noun
1	PASTP	drink	drink~d		drink	1,0	noun
3	PASTP	dry	dry~d		dry	1,4	adjective
1	PASTP	fire	fire~d		fire	1,5	noun
1	PASTP	flap	flap~d		flap	2,8	noun
1	PASTP	flavor	flavor~d		flavor	0,4	noun
1	PASTP	growl	growl~d		growl	2,9	noun
2	PASTP	ice	ice~d		ice	0,6	noun
3	PASTP	land	land~d		land	-0,8	noun
1	PASTP	like	like~d		like	-0,3	functionword
1	PASTP	mash	mash~d		mash	3,0	noun

2	PASTP	miss	miss~d		miss	1,9	noun
7	PASTP	name	name~d		name	0,1	noun
1	PASTP	park	park~d		park	0,4	noun
5	PASTP	pop	pop~d		pop	4,1	onomatopoeia
1	PASTP	press	press~d		press	0,3	noun
3	PASTP	puff	puff~d		puff	2,6	noun
3	PASTP	puzzle	puzzle~d		puzzle	1,7	noun
1	PASTP	ride	ride~d		ride	1,2	noun
1	PASTP	seat	seat~d		seat	1,5	noun
2	PASTP	shock	shock~d		shock	2,3	noun
2	PASTP	sting	sting~d		sting	2,6	noun
1	PASTP	stray	stray~d		stray	0,7	adjective
2	PASTP	stripe	stripe~d		stripe	0,8	noun
3	PASTP	taste	taste~d		taste	1,1	noun
1	PASTP	train	train~d		train	-0,1	noun
2	PASTP	trap	trap~d		trap	1,7	noun
1	PASTP	trip	trip~d		trip	-0,4	noun
13	PASTP	wet	wet~d		wet	2,3	adjective
1	PASTP	wind	wind~d		wind	1,1	noun

Appendix M. POS Transition Markers

CHILDES POS ORIGINAL	CHILDES WORD ADJUSTED	TRANSDITION :MARKER	Brysetal_POS	Existing_Winter_ et_al	occurrence s
co	aye	coToInt	interjection	aye	3
co	bye	coToInt	interjection	bye	237
co	hello	coToInt	interjection	hello	74
co	hi	coToInt	interjection	hi	350
co	no	coToInt	interjection	no	4991
co	ouch	coToInt	interjection	ouch	34
co	shh	coToInt	interjection	shh	3
co	yes	coToInt	interjection	yes	1783
co	yum	coToInt	interjection	yum	20
n	hello	nToInt	interjection	hello	16
n	ouch	nToInt	interjection	ouch	3
v	ouch	vToInt	interjection	ouch	6
co	baa	CoToOnom	onomatopoeia	baa	3
co	cockadoodledoo	CoToOnom	onomatopoeia	cockadoodledoo	6
co	vroom	CoToOnom	onomatopoeia	vroom	25

CHILDES POS ORIGINAL	CHILDES WORD ADJUSTED	TRANSDITION :MARKER	Brysetal_POS	Existing_Winter_ et_al	occurrence s
adv	about	advToFun	functionword	about	106
adv	all	advToFun	functionword	all	148
adv	at	advToFun	functionword	at	55
adv	before	advToFun	functionword	before	29
adv	behind	advToFun	functionword	behind	26
adv	by	advToFun	functionword	by	12
adv	in	advToFun	functionword	in	1052
adv	more	advToFun	functionword	more	145
adv	much	advToFun	functionword	much	9
adv	near	advToFun	functionword	near	11
adv	on	advToFun	functionword	on	808
adv	that	advToFun	functionword	that	79
adv	there	advToFun	functionword	there	2273
adv	under	advToFun	functionword	under	15
adv	what	advToFun	functionword	what	35
adv:int	all	advToFun	functionword	all	201
adv:int	much	advToFun	functionword	much	5
adv:int	that	advToFun	functionword	that	45
adv:int	what	advToFun	functionword	what	461
adv:tem	after	advToFun	functionword	after	9
adv:wh	what	advToFun	functionword	what	101
adv:wh	when	advToFun	functionword	when	10
conj	after	conjToFun	functionword	after	31
conj	and	conjToFun	functionword	and	1259
conj	because	conjToFun	functionword	because	833
conj	before	conjToFun	functionword	before	4
conj	but	conjToFun	functionword	but	998
conj	for	conjToFun	functionword	for	123
conj	if	conjToFun	functionword	if	304
conj	like	conjToFun	functionword	like	1
conj	when	conjToFun	functionword	when	535
coord	and	coordToFun	functionword	and	3854
det	a	detToFun	functionword	a	7223
det	that	detToFun	functionword	that	1496
det	the	detToFun	functionword	the	7373
det	these	detToFun	functionword	these	290
det	this	detToFun	functionword	this	2600
det	those	detToFun	functionword	those	182
det:num	one	detToFun	functionword	one	771
prep	a	prepToFun	functionword	a	1
prep	about	prepToFun	functionword	about	216
prep	above	prepToFun	functionword	above	1

CHILDES POS ORIGINAL	CHILDES WORD ADJUSTED	TRANSDITION :MARKER	Brysetal_POS	Existing_Winter_ et_al	occurrence s
prep	after	prepToFun	functionword	after	78
prep	at	prepToFun	functionword	at	662
prep	before	prepToFun	functionword	before	13
prep	behind	prepToFun	functionword	behind	19
prep	by	prepToFun	functionword	by	108
prep	for	prepToFun	functionword	for	1081
prep	in	prepToFun	functionword	in	2379
prep	into	prepToFun	functionword	into	128
prep	like	prepToFun	functionword	like	742
prep	near	prepToFun	functionword	near	32
prep	of	prepToFun	functionword	of	1005
prep	on	prepToFun	functionword	on	1527
prep	to	prepToFun	functionword	to	1451
prep	under	prepToFun	functionword	under	65
prep	with	prepToFun	functionword	with	1053
pro	it	proToFun	functionword	it	8144
pro	you	proToFun	functionword	you	6475
pro:dem	that	proToFun	functionword	that	3994
pro:dem	there	proToFun	functionword	there	610
pro:dem	these	proToFun	functionword	these	283
pro:dem	this	proToFun	functionword	this	2015
pro:dem	those	proToFun	functionword	those	156
pro:indef	all	proToFun	functionword	all	109
pro:indef	more	proToFun	functionword	more	382
pro:indef	none	proToFun	functionword	none	11
pro:indef	one	proToFun	functionword	one	2494
pro:indef	same	proToFun	functionword	same	2
pro:indef	some	proToFun	functionword	some	59
pro:obj	her	proToFun	functionword	her	275
pro:obj	him	proToFun	functionword	him	647
pro:obj	me	proToFun	functionword	me	2722
pro:obj	them	proToFun	functionword	them	940
pro:obj	us	proToFun	functionword	us	603
pro:poss	hers	proToFun	functionword	hers	11
pro:poss	mine	proToFun	functionword	mine	312
pro:poss:det	her	proToFun	functionword	her	288
pro:poss:det	his	proToFun	functionword	his	689
pro:poss:det	my	proToFun	functionword	my	2660
pro:poss:det	our	proToFun	functionword	our	201
pro:poss:det	their	proToFun	functionword	their	106
pro:poss:det	your	proToFun	functionword	your	750
pro:refl	yourself	proToFun	functionword	yourself	8

CHILDES POS ORIGINAL	CHILDES WORD ADJUSTED	TRANSDITION :MARKER	Brysetal_POS	Existing_Winter_et_al	occurrences
pro:sub	he	proToFun	functionword	he	3022
pro:sub	I	proToFun	functionword	I	13628
pro:sub	she	proToFun	functionword	she	1048
pro:sub	they	proToFun	functionword	they	1452
pro:sub	we	proToFun	functionword	we	2034
pro:wh	what	proToFun	functionword	what	1648
pro:wh	which	proToFun	functionword	which	6
pro:wh	who	proToFun	functionword	who	304
co	help	coToVerb	verb	help	88
co	look	coToVerb	verb	look	392
co	say	coToVerb	verb	say	89
co	see	coToVerb	verb	see	182
co	wait	coToVerb	verb	wait	67
co	boy	coToNoun	noun	boy	33
co	honey	coToNoun	noun	honey	9
co	pardon	coToNoun	noun	pardon	1
co	peekaboo	coToNoun	noun	peekaboo	45
co	thanks	coToNoun	noun	thanks	33
co	yahoo	coToNoun	noun	yahoo	10
n:prop	Aunt	n:propToNoun	noun	aunt	1
n:prop	Baby	n:propToNoun	noun	baby	3
n:prop	Bear	n:propToNoun	noun	bear	4
n:prop	Bird	n:propToNoun	noun	bird	12
n:prop	Dad	n:propToNoun	noun	dad	219
n:prop	Daddy	n:propToNoun	noun	daddy	720
n:prop	Father	n:propToNoun	noun	father	7
n:prop	Mommy	n:propToNoun	noun	mommy	2054
n:prop	Mother	n:propToNoun	noun	mother	15
n:prop	Sister	n:propToNoun	noun	sister	2
n:prop	Uncle	n:propToNoun	noun	uncle	33
co	bouncy	CoToAdj	Adjective	bouncy	1
co	fine	CoToAdj	Adjective	fine	1
co	icky	CoToAdj	Adjective	icky	1
co	sorry	CoToAdj	Adjective	sorry	43
co	sweet	CoToAdj	Adjective	sweet	3
pro:indef	other	proToAdj	adjective	other	15
qn	other	QnToAdj	Adjective	other	330
co	please	coToAdv	adverb	please	247
co	right	coToAdv	adverb	right	211
co	so	coToAdv	adverb	so	274
conj	so	conjToAdv	adverb	so	131
det	then	detToAdv	adverb	then	1

CHILDES POS ORIGINAL	CHILDES WORD ADJUSTED	TRANSDITION :MARKER	Brysetal_POS	Existing_Winter_et_al	occurrences
post	too	postToAdv	adverb	too	397
prep	around	prepToAdv	adverb	around	64
prep	down	prepToAdv	adverb	down	146
prep	inside	prepToAdv	adverb	inside	15
prep	off	prepToAdv	adverb	off	162
prep	out	prepToAdv	adverb	out	116
prep	over	prepToAdv	adverb	over	263
prep	up	prepToAdv	adverb	up	435
pro:dem	here	proToAdv	adverb	here	315

Appendix N. The 100 most frequently used words in the children's speech with their Iconicity

WORD	Frequency (over the 309 116 word occurrences in the corpus)	Iconicity
I	0,04421	3,181818182
is	0,02682	-0,142857143
it	0,02643	1
the	0,02393	0,428571429
that	0,02364	-0,0625
a	0,02345	0,461538462
you	0,02100	-0,4
yeah	0,01980	
do	0,01791	0,846153846
no	0,01752	2,8125
what	0,01677	0,142857143
and	0,01658	0,5625
this	0,01497	0,133333333
~not (neg Verb)	0,01422	
be	0,01210	0,384615385
in	0,01113	1,461538462
go	0,01088	1,454545455
want	0,01070	-0,214285714
one	0,01059	1,846153846
he	0,00981	1,058823529
there	0,00935	0,461538462
my	0,00908	1,5
me	0,00883	0,6

WORD	Frequency (over the 309 116 word occurrences in the corpus)	Iconicity
on	0,00757	0,916666667
can	0,00749	
oh	0,00726	
Mommy	0,00724	1,4
where	0,00692	-0,818181818
here	0,00679	-0,2
have	0,00668	-0,266666667
go~ing	0,00665	
we	0,00660	1,428571429
put	0,00656	1,230769231
get	0,00582	-0,583333333
yes	0,00578	2,25
will	0,00571	0,214285714
like	0,00547	-0,333333333
okay	0,00537	
some	0,00512	0,5
up	0,00512	1,333333333
see	0,00509	1,272727273
know	0,00484	0,769230769
they	0,00471	0,454545455
to	0,00471	-0,416666667
right	0,00439	0,916666667
why	0,00429	0,769230769

WORD	Frequency (over the 309 116 word occurrences in the corpus)	Iconicity
look	0,00424	1,818181818
for	0,00391	-1,4
down	0,00360	-0,769230769
baby	0,00357	2,230769231
not	0,00350	1,384615385
with	0,00342	1,166666667
out	0,00341	0,923076923
she	0,00340	0,714285714
now	0,00333	1,461538462
let	0,00333	
two	0,00329	1
take	0,00329	1,2
of	0,00326	0,230769231
but	0,00324	0
got~	0,00323	
all	0,00321	-0,1
come	0,00321	0,214285714
them	0,00305	-0,692307692
more	0,00289	1,125
Daddy	0,00283	1,428571429
play	0,00282	1,083333333
make	0,00282	0,642857143
eat	0,00278	0,647058824
because	0,00271	0,214285714
just	0,00269	
Mama	0,00255	
big	0,00255	1,090909091
your	0,00243	0

WORD	Frequency (over the 309 116 word occurrences in the corpus)	Iconicity
too	0,00240	1
how	0,00236	0,461538462
at	0,00233	-0,166666667
huh	0,00230	
back	0,00229	0,4
say	0,00227	0,8
off	0,00224	1,9
his	0,00223	0,588235294
car	0,00218	0,466666667
does	0,00216	-0,6
was	0,00215	-0,833333333
did	0,00211	0,466666667
him	0,00210	0,583333333
then	0,00206	0,1875
dog	0,00203	1,272727273
little	0,00199	1,15
us	0,00196	-0,769230769
another	0,00190	0,117647059
who	0,00190	1,375
good	0,00189	0,928571429
these	0,00186	-0,214285714
her	0,00183	0,7
when	0,00182	-1,125
turn	0,00180	0
over	0,00179	0,8
so	0,00178	0,909090909
three	0,00170	1,615384615

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