



# Approximating phonotactic input in children’s linguistic environments from orthographic transcripts

Sofia Strömbergsson<sup>1</sup>, Jens Edlund<sup>2</sup>, Jana Götze<sup>1,2</sup>, Kristina Nilsson Björkenstam<sup>3</sup>

<sup>1</sup>Division of Speech and Language Pathology, Department of Clinical Science, Intervention and Technology (CLINTEC), Karolinska Institutet (KI), Stockholm, Sweden

<sup>2</sup>Department of Speech, Music and Hearing, KTH, Stockholm, Sweden

<sup>3</sup>Department of Linguistics, Stockholm University, Stockholm, Sweden

sofia.strombergsson@ki.se, edlund@speech.kth.se, jagoetze@kth.se, kristina.nilsson@ling.su.se

## Abstract

Child-directed spoken data is the ideal source of support for claims about children’s linguistic environments. However, phonological transcriptions of child-directed speech are scarce, compared to sources like adult-directed speech or text data. Acquiring reliable descriptions of children’s phonological environments from more readily accessible sources would mean considerable savings of time and money. The first step towards this goal is to quantify the reliability of descriptions derived from such secondary sources.

We investigate how phonological distributions vary across different modalities (spoken vs. written), and across the age of the intended audience (children vs. adults). Using a previously unseen collection of Swedish adult- and child-directed spoken and written data, we combine lexicon look-up and grapheme-to-phoneme conversion to approximate phonological characteristics. The analysis shows distributional differences across datasets both for single phonemes and for longer phoneme sequences. Some of these are predictably attributed to lexical and contextual characteristics of text vs. speech.

The generated phonological transcriptions are remarkably reliable. The differences in phonological distributions between child-directed speech and secondary sources highlight a need for compensatory measures when relying on written data or on adult-directed spoken data, and/or for continued collection of actual child-directed speech in research on children’s language environments.

**Index Terms:** language acquisition, grapheme-to-phoneme conversion, phonology

## 1. Introduction

Children’s acquisition of the words and sounds in their surrounding language is the product not only of their developing cognitive and motor capacities; it is also dependent on statistical regularities in the surrounding language [1] [2] [3]. Such statistical characteristics may explain why certain phonological patterns are observed earlier in children’s acquisition of one language and later in another.

The child’s advancing linguistic skills are accompanied by increasingly complex linguistic input, for example in terms of syntactic structure [4]. In terms of lexical characteristics, the issue of adult accommodations to children’s linguistic competence is more unresolved. Lexical variability, often measured by the type/token ratio (TTR), has been shown to increase with the age of the child [5]. The observation of an

increasing TTR in adults’ speech directed to children of increasing age is robust, and is even attested in other media, such as in subtitles of TV shows directed to children at different ages [6]. However, when exploring other lexical aspects, such as whether adults introduce more infrequent words as the child grows older, similar trends have not been confirmed [5].

Closely linked to issues of lexical complexity in linguistic input to children is phonological complexity: what sounds and sound sequences, and what syllabic structures make up the words that children hear. And although research on segmental adaptation to children is still scarce, phonological differences have indeed been observed in infant-directed speech compared to adult-directed speech [7]. Hence, there is a risk that differences in child-directed speech (CDS) and adult-directed speech (ADS) may distort investigations of input frequency effects in phonological acquisition [8]; if there are differences in phonotactic distributions in adult-directed versus child-directed speech, frequencies based on adult speech corpora are not representative of children’s phonological input.

Although long-time language sampling of children’s environments is increasingly feasible, text corpora are still far more readily available than spoken corpora. In addition, child-directed text is a component in many children’s linguistic environments, and likely reflects linguistic adaptation to the age of the intended child audience. Comparisons between adult-directed and child-directed literature are scarce, but the findings in [9] suggest that, at least with respect to the usage of relative clauses, children’s literature is statistically intermediate between adult-directed written language and adult-directed speech. If adaption to children extends also across other linguistic domains, it is conceivable that child-directed written corpora are actually better representations of children’s phonological environments than what adult-directed speech is.

By necessity, statistical regularities in research on linguistic frequency effects are derived from any available data sources that are sufficiently large. This has led to an almost exclusive focus on written text corpora [10]. Today, improved infrastructures allow increased sharing of data resources, and we see rapid methodological advances for handling such resources. This opens entirely new arenas for research on language acquisition. Several researchers have proposed basing studies of explanatory factors to language acquisition on children’s lexicons derived from spontaneous speech rather than on adult data or written material [1] [10] [11]. However, exactly how critical this is when it comes to the description of phonological characteristics in children’s environments is unknown, and remains to be quantified.

## 1.1. Aim

We make use of a previously unseen collection of Swedish linguistic data acquired from various sources of child-directed and adult-directed speech, and of child-directed and adult-directed text. We use this data collection to estimate a number of phonological characteristics. Through the analysis of distributional variation across the age of the expected receiver (child/adult), and across modalities (written/spoken), we aim to answer the following research question:

*Does the phonological distribution in a data set vary across a) the age of the intended audience, and b) the modality of the linguistic data (written vs. spoken), and if so, to what extent?*

## 2. Method

### 2.1. Data

Table 1 presents the corpora included in this work. This collection of Swedish adult- and child-directed spoken and written data is selected to represent a broad set of genres and domains. The written data consists of text corpora from the Swedish Språkbanken at Gothenburg University and the Norwegian Språkbanken at the Norwegian National Library.

The child-directed spoken data, CHILDSPEECH, includes orthographic transcripts of speech directed to infants (3-12 months) and toddlers/pre-school children (1-6 years) in a

naturalistic free-play setting at home [12] or in a lab [13]. Similarly, the adult-directed speech, ADULTSPEECH, comes from three sources of orthographical transcripts of spontaneous interaction in both authentic environments [14] [15] and a lab setting [16]. The spoken data consists of around 2.3 million tokens adult-directed and 295 000 tokens child-directed speech.

The Swedish NST lexicon [17] was used in the retrieval of phonological transcriptions from orthographic words. This lexicon, designed to cover general vocabulary, consists of more than 630 000 entries, with manual phonemic SAMPA [18] transcriptions of the most frequent words in the NST Swedish corpus (see Table 1), as well as of complementary lexical resources. The lexicon covers 72% of the word tokens in the combined speech corpora, and 60% of the word types. For the text corpora, the lexicon covers 73% of the word tokens (with a frequency > 1), and 41% of the word types.

The original lexicon has been modified slightly for the purposes of this work. First, in order to reflect the description of the Swedish phonemic inventory as described in [19], the allophonic variants [e] and [E] were merged into one single symbol. Second, some frequent function words (like *det*, *på*, *någon*; Eng. *it*, *on*, *any*) were found missing from the lexicon, and were added manually. Third, as homographs with different pronunciation are listed as separate entries in the original lexicon, unordered with respect to frequency, all homographs with multiple pronunciation variants (n = 2219) were manually

Table 1: Descriptions of the included corpora.

| Corpus                                    | Description  | # Tokens           |
|---|--|--------------------|
| <b>CHILDSPEECH: Child-directed speech</b> |  |                    |
| CHILDES [12]                              | Spontaneous adult-child interaction; home environment; 5 children (age 1;0 to 6;0). Göteborg corpus (CHILDES, [20]).   | 248 163            |
| LONG-MINGLE [13]                          | Spontaneous parent-child interaction; lab environment; free play scenario, 17 children (age 0;3-2;9)                   | 46 437             |
| <b>Total number of tokens</b>             |  | <b>294 600</b>     |
| <b>ADULTSPEECH: Adult-directed speech</b> |  |                    |
| SB-GDC [14]                               | Spoken interaction; Context-governed. Gothenburg Dialogue Corpus Språkbanken, Gothenburg University (GU).              | 1 345 044          |
| Spontal [16]                              | Spontaneous speech; lab environment; 120 dyads.  | 79 332             |
| Swedia 2000 [15]                          | Interview transcripts; speakers of Swedish dialects.   | 923 911            |
| <b>Total number of tokens</b>             |  | <b>2 348 287</b>   |
| <b>CHILDTXT: Child-directed text</b>      |  |                    |
| Laesbart 6-9 [21]                         | Fiction. A subset of the LäSBarT corpus (sample criterion: fiction, target audience age: 6-8 years). Språkbanken, GU.  | 114 209            |
| Laesbart 9-12 [21]                        | Fiction. A subset of the LäSBarT corpus (sample criterion: fiction, target audience age: 9-12 years). Språkbanken, GU. | 291 562            |
| <b>Total number of tokens</b>             |  | <b>405 771</b>     |
| <b>ADULTTXT: Adult-directed text</b>      |  |                    |
| SB-GP2013 [22]                            | News text. Göteborgsposten 2013, Språkbanken, GU.  | 16 870 565         |
| SB-KLARSPRAK [22]                         | Official texts. Förvaltningsmyndigheters texter, Språkbanken, GU.  | 49 453             |
| SB-SAM [22]                               | Academic text, social sciences. Akademiska texter - samhällsvetenskap (subset), Språkbanken, GU.                       | 1 142 268          |
| SB-HUM [22]                               | Academic text, the humanities. Akademiska texter - humaniora (subset), Språkbanken, GU.                                | 1 104 899          |
| SB-LT2005 [22]                            | Medical news periodical. Läkartidningen 2005, Språkbanken, GU.   | 1 505 754          |
| SB-ROM99 [22]                             | Fiction. Norstedts romaner 1999, Språkbanken, GU.  | 2 534 302          |
| SB-ROMII [22]                             | Fiction. Bonniers romaner II, Språkbanken, GU.   | 4 304 239          |
| SB-SUC2 [23]                              | Stockholm-Umeå Corpus (SUC 2.0). A balanced corpus of Swedish.   | 1 166 528          |
| NST-SWE [24]                              | NST Text Corpus; news text (80%), fiction, periodicals. Språkbanken, Nasjonalbiblioteket, Norway.                      | 397 781 587        |
| <b>Total number of tokens</b>             |  | <b>426 459 595</b> |

checked, and the preferred pronunciation variant was marked. In the vast majority of cases, this was a trivial decision. Of the remaining undecided cases (n=126), preferred pronunciations of the 59 entries with at least 100 occurrences in the NST-SWE corpus (see Table 1) were inferred through concordance lists generated through Korp [22]. Preferred pronunciation variants were listed in a separate list, and used to override default pronunciations in the lexicon.

## 2.2. Procedure

Unigram frequency lists were built from each corpus listed in Table 1. Next, phonological transcriptions for the words in these lists were acquired through lexicon look-up. Fallback pronunciations for words not found in the lexicon were generated using the Sequitur grapheme-to-phoneme (g2p) converter [25], which is an open source tool based on the method described in [26]. Following these authors' recommendation, the g2p model was trained in 9 iterations. 1/4 of the data was used for training and 3/4 for testing. Table 2 summarizes the performance of the model trained on the Swedish NST lexicon. For reference, corresponding figures for CMUdict [27] presented in [26] were WER = 24.5% and PER = 5.9%.

Table 2: Performance of the g2p model, by word-error-rate (WER) and phone-error-rate (PER).

| Setting | Training size | Testing size | WER  | PER  |
|---------|---------------|--------------|------|------|
|         | 199 304       | 597 907      |      |      |
|         | (=1/4)        | (=3/4)       |      |      |
| Model8  |               |              | 9.1% | 1.1% |
| Model9  |               |              | 8.4% | 1.0% |

Finally, uni-, bi- and triphoneme frequency lists were generated from the automatically retrieved phonological transcriptions and the word frequencies in each corpus set.

## 2.3. Analysis

Consonant clusters of the format /sCC/ are reportedly a complex phonotactic structure, acquired relatively late in languages where they occur, e.g. in English [28] and Swedish [29]. For this reason, we chose this phonotactic pattern for more detailed analysis of distributional characteristics.

The reliability of the retrieved phonological transcriptions was estimated by comparing generated/retrieved phonemic transcriptions of orthographic transcripts of children's TV shows to manually transcribed versions of the same contents on a segmental level (disregarding syllable boundaries and stress notation). Three second-year SLP students carried out the manual phonemic transcription on 30 one-minute samples retrieved from the Swedish national broadcasting service svtplay.se. Inter-annotator agreement was estimated automatically on a 20% sample of the transcriptions (n = 6), which was re-transcribed by a different transcriber. A semi-automated preprocessing step was introduced in order to handle segmentation mismatches. For any pair of transcriptions A and B,  $word_{A,n}$  was compared to  $word_{B,n}$ . If identical, this pair was considered a match. If not,  $word_{A,n}$  was compared to  $word_{B,n\pm 1}$ . In cases of three or more consecutive mismatches, transcriptions were manually checked to ensure desirable glossing in the manual transcriptions. For example, multiword

sequences like *ska du*, manually transcribed as /skA:ru0/, were split up into /skA:/ and /ru0/.

Segmental agreement was estimated following the semi-automatic word-by-word alignment. Out of the 14 413 phoneme tokens present in the manual transcriptions, phoneme-by-phoneme agreement between the generated transcriptions and the manual transcriptions was 94%. Agreement between original and repeated transcriptions by different human transcribers was 90% (based on a total of 2 717 tokens).

## 3. Results

Table 3 presents the most frequent single phonemes in CHILDSPEECH, with their frequencies (relative to the total number of phoneme tokens). The shaded cells list the differences compared to the corresponding relative frequencies for the other three datasets. The table shows, for example, that /d/ is more frequent in CHILDSPEECH compared to, especially, text data. Conversely, /t/ and /e/ are less frequent in CHILDSPEECH than in the other datasets.

Table 3: Relative frequency (%) of the most frequent phonemes in CHILDSPEECH, and the difference between this and the corresponding frequencies in the other datasets. Larger differences are indicated by darker (-) or brighter (+) shading.

|   | CHILD-SPEECH | vs. ADULT-SPEECH | vs. CHILD-TEXT | vs. ADULT-TEXT |
|---|--------------|------------------|----------------|----------------|
| a | 7,87         | 0,93             | -0,37          | 0,18           |
| e | 7,49         | -0,78            | -1,40          | -1,78          |
| n | 7,32         | 0,15             | -0,31          | -0,12          |
| r | 7,01         | 0,01             | -0,56          | -0,61          |
| d | 6,56         | 1,07             | 1,97           | 2,13           |
| t | 5,29         | -1,08            | -1,46          | -1,84          |
| m | 5,19         | 0,70             | 1,46           | 1,52           |
| s | 4,74         | -0,27            | -0,89          | -1,28          |

Analysis of triphoneme frequencies reveals that /dOm/ is the most frequent triphoneme in CHILDSPEECH, and considerably more frequent than in the other datasets. Presumably, this reflects a relatively high usage of the pronouns *de/dem* (Eng. *they*, pronounced /dOm/) in CHILDSPEECH compared to the other datasets. Further inspection of bi- and triphoneme frequencies shows that phonological patterns occurring in deictic words (/den/, /dE:r/) and in first and second person pronouns (/jA:/, /d:/) are more frequent in the spoken datasets than in the text datasets, whereas conversely, third person pronouns are more frequent in the text data.

Figure 1 shows the relative frequency of the sequence /sCC/ across the four datasets. As seen, /str/, /skr/ and /spr/ are more frequent than the other /sCC/-clusters. In the spoken data, and particularly in CHILDSPEECH, /skr/ and /str/ are less frequent than in the other data sources, whereas the relative frequency of /spr/ is stable across all four datasets. Close inspection of what words contribute to the higher frequencies of /str/ and /skr/ in the written datasets reveals that the most frequent words containing /str/ are various forms of the words *extra* (Eng. *extra*), *ström* (Eng. *stream*, often occurring in proper nouns), *strand* (Eng. *beach*, also frequent in proper nouns), whereas /skr/ frequently occurs in different forms of the words *skriva* (Eng. *write*), *skratta* (Eng. *laugh*) and *skrika* (Eng. *scream*).

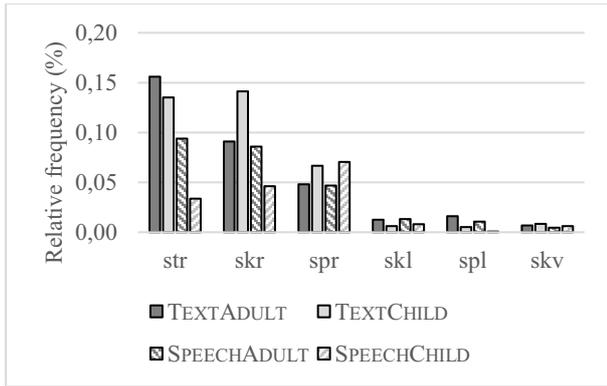


Figure 1: Relative triphoneme frequencies (%) of /sCC/-clusters across the four corpus sets.

## 4. Discussion

We have investigated Swedish phonotactic characteristics and selected aspects how these characteristics vary across different linguistic modalities (spoken/written), and across different ages of the intended audience (child/adult). Some distributional differences of uniphoneme frequencies are noted, but cannot be easily linked to, for example, lexical differences between the datasets. Differences among the most frequent bi- and triphonemes can presumably be attributed to the modality of the linguistic data: phoneme sequences typically occurring in deictic words and first and second person pronouns are more frequent in the spoken datasets than in the text datasets. Interestingly, /jA:/ (as in the first person singular pronoun *jag*, Eng. *I*) is more frequent in ADULTSPEECH than in CHILDSPEECH, where instead /d:/ (as in the second person singular pronoun *du*) is more frequent. This difference may be attributed to different contextual settings of the included spoken data, but may also reflect an expected trait in child-directed speech, with adults labelling and responding to children’s actions [30].

The examination of one particular phoneme sequence pattern – /sCC/ – revealed, on the one hand, some similarities across the different corpus sets, such that /str/, /spr/, and /skr/ are more frequent than any other /sCC/-cluster. On the other hand, distributional differences across datasets were also discovered, such that /str/ and /skr/ were less frequent in CHILDSPEECH than in the other datasets. Considering that many of the frequent words containing these sequences in the other datasets likely stem from proper nouns, and from words referring to behavioral descriptions (e.g. *writing*, *laughing*), it should not be surprising to find these less frequently in speech collected from parent-child free-play settings.

Underlying the investigation is a previously unseen collection of Swedish corpus data, selected to provide a balanced representation of contemporary spoken and written Swedish, loosely modelled on the British National Corpus [31]. The spoken data covers contemporary spoken Swedish from the two largest cities. The adult-directed data also includes regional dialects. To our knowledge, this is the first comparison of these speech resources. As far as we know, this is also the first published report of the coverage of the NST Swedish lexicon.

Although the Swedish lexicon differs in several respects from the English lexicon used in the original description of the g2p converter [26] – e.g. by including syllable boundaries, by being considerably larger, and, not least, by describing another language – the performance achieved here outperforms that

reported in the original work. Using this g2p converter to generate phonological transcriptions for words not found in the lexicon may therefore be an appealing alternative to excluding words, or to adding transcriptions by hand.

The analysis of the reliability of the generated transcriptions are encouraging, as a point-by-point agreement of 94% is well above the 85% agreement that others have suggested as a threshold of reliability between human transcribers [27]. Apart from validating the quality of the generated transcriptions, the analysis also revealed some routes of potential improvements, such as adapting the lexicon to common reductions of frequent function words like *och* (Eng. *and*) and *jag* (Eng. *I*).

### 4.1. Limitations

Using orthographic transcriptions and lexicon look-up may be considered a brute way of estimating children’s phonological environment. Certainly, there is noise in the data, for example coming from the static treatment of homographs, and of frequent suboptimal pronunciation variants of the same words (e.g. *och* and *jag*, Eng. *and* and *I*, respectively). And although accuracy may be further improved by adapting the lexicon to better reflect the most frequent pronunciation variants, the method can never generate noise free estimations of children’s phonological environment. However, as an estimation of expected pronunciation on a broad level, it may be valuable in efforts to speed up manual transcription work.

If stress and syllable notation had been included in the reliability analysis, more disagreements would have been expected. However, this follows common procedure when estimating inter-annotator agreement [32], and thus makes our figures comparable to those previously reported.

For both spoken and written data, all child-directed data was collapsed into one and the same group. Also, there is no age overlap between CHILDTXT and CHILDSPEECH. This is, of course, a simplification that may obscure linguistic adaptations to children at different ages. With larger datasets for different ages, future investigations could explore whether phonological adaptation changes with the age of the children.

Admittedly, although we have identified distributional differences, the question of whether these are impactful remains undetermined. This requires other methods of analysis (see e.g. [33]), and is an interesting venue of future research.

## 5. Conclusions

We have presented a previously unseen collection of Swedish corpus data, unique in its kind by being organized by the modality of the data (spoken/written), and by the ages of the intended audience (child/adult). Further, the approach of combining lexical look-up with a g2p converter trained for Swedish was found to generate highly reliable phonological transcriptions. Combined, this resource framework can be used to expedite large-scale investigations into lexical and phonological characteristics of Swedish. Including other languages in the same framework will open up cross-linguistic research into these phenomena.

## 6. Acknowledgements

We thank our Swedish colleagues who have generously shared their data to be included in this work. Thanks also to SLP students Karin Sproge, Hanna Bolle and Sarah Deery for performing manual transcriptions. The work was funded by the Swedish Research Council (VR 2015-01525).

## 7. References

- [1] J. A. Gierut and R. A. Dale, "Comparability of Lexical Corpora: Word frequency in phonological generalization," *Clinical Linguistics & Phonetics*, vol. 21, pp. 423-433, 2007.
- [2] P. W. Jusczyk, P. A. Luce and J. Charles-Luce, "Infants' Sensitivity to Phonotactic Patterns in the Native Language," *Journal of Memory and Language*, vol. 33, pp. 630-645, 1994.
- [3] J. Edwards, M. E. Beckman and B. Munson, "Frequency effects in phonological acquisition," *Journal of Child Language*, vol. 42, p. 306-311, Mar 2015.
- [4] J. Huttenlocher, M. Vasilyeva, H. R. Waterfall, J. L. Vevea and L. V. Hedges, "The Varieties of Speech to Young Children," *Developmental Psychology*, vol. 43, pp. 1062-1083, 2007.
- [5] D. P. Hayes and M. G. Ahrens, "Vocabulary simplification for children: a special case of "motherese"?", *Journal of Child Language*, vol. 15, pp. 395-410, 1988.
- [6] M. Kjellholm, C. Wikse Barrow and S. Strömbergsson, "Subtitles as a proxy for children's linguistic environment? Exploring the relationship between subtitles and speech in children's television programmes," in *Proceedings of Fonetik 2016*, Stockholm, Sweden.
- [7] J. Edwards and M. E. Beckman, "Some Cross-Linguistic Evidence for Modulation of Implicational Universals by Language-Specific Frequency Effects in Phonological Development," *Language Learning and Development*, vol. 4, pp. 122-156, 2008.
- [8] S. Tsuji, K. Nishikawa and R. Mazuka, "Segmental distributions and consonant-vowel association patterns in Japanese infant- and adult-directed speech," *Journal of Child Language*, vol. 41, pp. 1276-1304, 2014.
- [9] J. L. Montag and M. C. MacDonald, "Text Exposure Predicts Spoken Production of Complex Sentences in 8- and 12-Year-Old Children and Adults," *Journal of Experimental Psychology: General*, vol. 144, pp. 447-468, 2015.
- [10] J. C. Goodman, P. S. Dale and P. Li, "Does frequency count? Parental input and the acquisition of vocabulary," *Journal of Child Language*, vol. 35, pp. 515-531, 2008.
- [11] S. F. Stokes, "Neighborhood Density and Word Frequency Predict Vocabulary Size in Toddlers," *Journal of Speech, Language, and Hearing Research*, vol. 53, pp. 670-683, 2010.
- [12] S. Strömqvist, U. Richthoff and A.-B. Andersson, "Strömqvist's and Richthoff's corpora: A guide to longitudinal data from four Swedish children," *Gothenburg Papers in Theoretical Linguistics*, vol. 66, 1993.
- [13] K. N. Björkenstam, "LONG-MINGLE: A longitudinal corpus of child-directed speech," Department of Linguistics, Stockholm University, Sweden, 2014.
- [14] J. Allwood, L. Grönqvist, M. Björnberg, E. Ahlsen and C. Ottesjö, "The Spoken Language Corpus at the Linguistics Department, Göteborg University," *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, vol. 1, 2000.
- [15] A. Eriksson, "SweDia 2000: A Swedish dialect database," in *Babylonian Confusion Resolved. Proc. Nordic Symposium on the Comparison of Spoken Language, Copenhagen Working Papers in LSP*, 2004.
- [16] J. Edlund, J. Beskow, K. Elenius, K. Hellmer, S. Strömbergsson and D. House, "Spontal: a Swedish spontaneous dialogue corpus of audio, video and motion capture," in *Proc. of LREC'10*, Valetta, Malta, 2010.
- [17] G. Andersen, "Gjennomgang og evaluering av språkresurser fra NSTs konkurso," Aksis, UNIFOB, Universitetet i Bergen, Bergen, Norway, 2005.
- [18] J. C. Wells, "SAMPA - computer readable phonetic alphabet," UCL Psychology & Language Sciences, 2005. [Online]. Available: <http://www.phon.ucl.ac.uk/home/sampa/index.html>. [Accessed March 1, 2017]
- [19] O. Engstrand, *Fonetikens grunder*, Lund: Studentlitteratur, 2004.
- [20] L. Borin, M. Forsberg and J. Roxendal, "Korp - the corpus infrastructure of Språkbanken," in *Proc of LREC 2012*, Istanbul: ELRA, 2012.
- [21] B. MacWhinney, *The CHILDES Project: Tools for analyzing talk*, 3rd ed., Lawrence Erlbaum Associates, 2000.
- [22] K. Mühlhenn, "Readable, legible or plain word - Presentation of an easy-to-read Swedish corpus," in *Multilingualism: Proc of the 23rd Scandinavian Conference of Linguistics*, vol. 8 of *Acta Universitatis Upsaliensis*, Uppsala, Sweden, 2009.
- [23] G. Källgren, "Description of Stockholm-Umeå Corpus," in *Manual for SUC 2.0*, S. Gustafson-Capkova and B. Hartmann, Eds., Department of Linguistics, Stockholm University, 2006.
- [24] Språkbanken, *N-grams for Swedish (based on the NST Text Corpus)*, Oslo, Norway: Nasjonalbiblioteket, 2015.
- [25] M. Bisani, "Sequitur G2P," 2008. [Online]. <http://www-i6.informatik.rwthachen.de/web/Software/g2p.html>. [Accessed Jan 28, 2017].
- [26] M. Bisani and H. Ney, "Joint-sequence models for grapheme-to-phoneme conversion," *Speech Communication*, vol. 50, pp. 434-451, 2008.
- [27] The Carnegie Mellon Speech Group, "The CMU Pronunciation Dictionary," [Online]. <http://www.speech.cs.cmu.edu/cgi-bin/cmudict>. [Accessed Jan 1, 2017].
- [28] P. Grunwell, *Clinical Phonology*, Baltimore, MD: Williams & Wilkins, 1987.
- [29] U. Nettelbladt and E.-K. Salameh, *Språkutveckling och språkstörning hos barn*, Lund, Sweden: Studentlitteratur, 2007.
- [30] J. L. Locke, *The child's path to spoken language*, Cambridge, MA: Harvard University Press, 1995.
- [31] *The British National Corpus, version 3 (BNC XML Edition)*, Oxford University Computing Services on behalf of the BNC Consortium, 2007.
- [32] L. Shriberg and G. Lof, "Reliability studies in broad and narrow phonetic transcription," *Clinical Linguistics & Phonetics*, vol. 5, pp. 225-279, 1991.
- [33] R. Daland, "Variation in the input: a case study of manner class frequencies," *Journal of Child Language*, vol. 40, pp. 1091-1122, 2013.