

Chapter 14

Linguistic: Application of LSA to Predict Linguistic Maturity and Language Disorder in Children



Kristina Hansson, Birgitta Sahlén, Rasmus Bååth, and Sverker Sikström

- Application of LSA to quantify semantics in oral narratives
- Semantic/lexical difficulties are often part of developmental language disorder in children
- Oral narratives from 108 children with typical language development and 54 children with developmental language disorder were analysed using LSA
- Two measures were created, an index of semantic linguistic maturity (SELMA) and a semantic language impairment index (SELIMI)
- The results from SELMA and SELIMI suggest that the semantic representation of the narratives contain information on semantic maturity and suggest that semantic quality is different in children with language disorder

Introduction

In this chapter we will describe applications of latent semantic analysis to assess semantic linguistic maturity in children and how well the method can predict whether a child has developmental language disorder (DLD), based on orally

K. Hansson (✉) · B. Sahlén
Department of Logopedics, Phoniatrics and Audiology, Lund University, Lund, Sweden
e-mail: kristina.hansson@med.lu.se

R. Bååth
Department of Cognitive Science, Lund University, Lund, Sweden

S. Sikström
Department of Psychology, Lund University, Lund, Sweden

produced narratives. Assessment of narrative ability in preschool years captures important cognitive, linguistic and socio-pragmatic skills and is an important prognostic tool for later academic success. We discuss the usefulness and potential of computerized methods to quantify semantics in research on typical and disordered language development. More importantly, we also discuss if and how such methods can be applied in clinical or educational contexts for diagnosis of language disorder in different populations, as an objective measure to complement standardized and decontextualized tests and assessment methods.

Research on language development and language disorders in children most often address issues of linguistic form. Studies addressing aspects of content and semantics usually focus on the lexicon, at the word level, not on content at the text level. As a consequence, very little is known on semantic content in the language use in typical and disordered language development. One reason for this may be a lack of methods to study semantics at the text level. The latent semantic analysis paradigm has the potential to fill in the gap.

Developmental Language Disorder in Children

Children with developmental language disorder (DLD, earlier referred to as specific language impairment, SLI; see Bishop 2014; Bishop et al. 2017) are children who do not develop language as expected, given their age, in spite of normal hearing and otherwise typical development (e.g., Bishop et al. 2017; Leonard 2014). The disorder can affect different aspects of language comprehension and production (phonology, grammar, vocabulary/semantics and pragmatics). Lexical/semantic difficulties are often involved in DLD, in particular in more severe cases. DLD is persistent and does not resolve even if an individual's language profile changes over time (Conti-Ramsden and Durkin 2012; Conti-Ramsden et al. 2009; Sandgren et al. 2015). An early sign of language disorder is late onset in combination with slow vocabulary learning, both with respect to word production (Leonard 2014; Leonard et al. 1999; McGregor 2009) and word comprehension (Bishop 1997). In school age, a substantial proportion of children with DLD are described to have problems with word-finding (Dockrell et al. 1998; German and Simon 1991; Leonard et al. 1983) and to have difficulties with semantic representations (Kail and Leonard 1986; McGregor and Appel 2002; McGregor et al. 2002). In spite of this most research on DLD focuses on form (phonological and syntactic processing), rather than on content, i.e., semantics, and reliable and valid procedures for assessment and analysis are lacking.

Assessment of Lexical/Semantic Skills in Children with DLD

So far, most assessment methods related to semantic skills used in clinical or pedagogical settings focus on vocabulary, that is, the single word level. For assessment of younger children parents might be asked to report on what words their children use and understand by using for example a checklist or a questionnaire (Fenson et al. 1993, 2007). A range of standardized tests of word comprehension or word production are also in use to assess the breadth of the child's vocabulary (i.e., the number of words; for a comprehensive overview, see McGregor 2009). For the assessment of the depth of vocabulary knowledge (i.e., the knowledge of each word) different non-standardized tasks are used, like for example asking for definitions (McGregor et al. 2002) and word-to-picture matching with phonological and semantic distractors in forced choice test paradigms (e.g., Chiat and Frazier Norbury 2000). In research the ability to learn novel words has also been explored in experimental studies (e.g., Dollaghan 1987; Hansson et al. 2004; Rice et al. 1990; Sahlén and Hansson 2006).

In a study of children with DLD and age matched and vocabulary matched controls Sheng and MacGregor (2010) found that the children with DLD showed a lexical-semantic organization deficit exceeding what would be expected given their vocabulary level. The interpretation was that the main problem is not having fewer words in the vocabulary, but missing or weaker links between representations in the semantic network, where words that share semantic relations are connected. Slower and less precise word mobilisation in word fluency tasks would therefore be predicted in children with DLD, but convincing evidence is so far lacking (Weckerly et al. 2001).

Corrigan (2008) points out that most research on lexical acquisition focuses on early development. We know very little about subtler aspects of semantics in older children where the learning of new words and expressions is based on making inferences of their meaning from the spoken or written context they occur in. Children "learn about words through their co-occurrence with other words" (Corrigan 2008: 109) and the meaning of a word consists of all information from the contexts where a child has heard or read this word (Corrigan 2008). Language is learned in a cultural, social and linguistic context (Tomasello 2003). Research on language maturation and language learning difficulties must therefore necessarily take contextual influences into consideration.

We thus have to venture beyond the word level. Still, standardized measures of lexical/semantic skills are based on decontextualized tasks. For more ecologically valid measures it is urgent to find measures which are closer to the children's actual use of their vocabularies in, for example, elicited or naturalistic language samples and to study how words are used together. Language samples have been used to assess semantic development, applying measures of lexical diversity like number of different words (Klee 1992; Watkins et al. 1995), the measure D (McKee et al. 2000; Owen and Leonard 2002) or lexical density (i.e., the ratio of content words to function words) and its relation to lexical diversity (Johansson 2009). However,

these analyses still relate more to counting frequencies of specific words rather than to the semantic content that is conveyed by the text. In order to achieve a wider communicative perspective in clinical and pedagogical assessments, it is important to develop valid and reliable tools that capture content on the text level, that is, to measure semantic development in children with DLD at the text level, not only (but also) on the word and sentence levels.

Taking context dependency into account in studies of phonological and syntactic processing in children with DLD is of great importance. These basic language areas were the first to be ventured in DLD-research. Children with DLD are also at high risk for difficulties with complex language tasks, e.g., reading and writing. Vocabulary and semantics are crucial for the two factors without which reading comprehension and academic success may be severely hampered, namely correct and fluent decoding of written words, and language comprehension (Gough and Tunmer 1986).

Narratives in Children with DLD

Narrative knowledge and structure are established early in life and develop over the school years, concerning structure and cohesion (Appelbee 1978), variation in lexicon (Johansson 2009), and sentence types (e.g. Myhill 2008). Supporting narrative skills is an important content already in the pre-school curriculum (before age 6) in Sweden. In narrative tasks, even when the child is asked to tell a story from given pictures, s/he is required and given the possibility to select and create content. Narrative tasks are very useful as assessment instruments and in diagnosing language disorder (Botting 2002; Fey et al. 2004; Norbury and Bishop 2003; Miniscalco et al. 2007). They are used to provide information about grammatical skills, the ability to use cohesive devices, the ability to organise content (story grammar), and linguistic productivity as well as about pragmatic ability (adapting to the needs of the listener). Narratives also have a potential to provide information about the quality of content, thus to measure linguistic maturity from a semantic perspective.

Two Studies Applying LSA to Narratives from Children

We will summarize two studies where we used LSA to analyse picture-elicited narratives from children with typical and disordered language development. The research questions had two dimensions: a longitudinal and a cross-sectional. The purpose was to investigate if measures generated by lexical semantic analysis can reflect lexical-semantic development in children and if they can identify children with and without language disorder. Picture elicited narratives were used, because this is an easier task. It has also turned out to be a more attractive task for children, because the content is easier to control and input is similar for all children compared

to for example personal narratives. Furthermore, we have long experience of this type of data from children in different clinical groups. There is also a large body of international and cross-linguistic research using the same paradigm (Berman and Slobin 1994).

Data

The narratives were elicited using a selection of pictures from the story *One frog too many* (Mayer and Mayer 1975). The pictures were laid out, one at a time, and the child was asked to look carefully at each picture. The examiner pointed to the first picture and provided the following sentence: “This story is about a boy and his pets, who are going out on a raft”, and asked the child to continue the story. The procedure was audio- and video recorded and later transcribed orthographically.

Participants

Narratives were obtained from 108 Swedish-speaking children, who, according to parents and teachers, had typical development in all respects, including language, hearing and nonverbal IQ. Age and gender distribution are presented in Table 14.1. They were recruited from intermediate socio-economic status areas. Clinical data were collected from 54 children with DLD. All had a non-verbal IQ above 70, only few performed below IQ 80. Age and gender distribution are presented in Table 14.1 for this group too. All participants received or had received language intervention.

An Index of Semantic Linguistic Maturity (SELMA)

The first application of LSA to these narrative data was to see if the LSA measure was sensitive to quality differences due to age, i.e., whether it could reliably index typical development (Hansson et al. 2016). Thus LSA was used to generate a measure of

Table 14.1 Age and gender distribution of the participant groups

Group	N	Girls (N)	Age range (months)	Mean age
Children with typical language development	108	40	48–154	110
Children with DLD	54	22	98–153	117

semantic linguistic maturity (SELMA) based on the analysis of narratives from the 108 children with typical language development in the age range 4–11.

Applying the standard LSA algorithm (Landauer and Dumais 1997) a semantic space was created based on more than 100,000 Swedish newspaper articles from 2007 (see Chaps. 2, 3 and 4 for details on the creation of the space, Nielsen and Hansen 2020a, 2020b; Kjell et al. 2020a, 2020b). The 108 narratives were summarized in this semantic representation, by adding the semantic vectors representing each word in a narrative, so that each narrative was summarized in one vector, and then the length of the resulting vector was normalized to one. The basic assumption was that age correlates with linguistic maturity. A logistic regression analysis generated the linguistic maturity measure. We used a *leave-one-out cross validation* approach, so that the age of each child was predicted using a data set including all narratives *except* that from the child whose age was being predicted. This was repeated for all subjects (for details, see Chap. 5). The analysis was conducted in [Semanticexcel.com](https://www.semantic-excel.com) (see Chap. 6, Sikström et al. 2020).

The variable generated by LSA, i.e., SELMA, had 55% agreement with chronological age and 62% with number of words. To check validity, we also collected ratings made by humans. Raters were asked to make an overall holistic judgement about which one in a pair of narratives was the more linguistically mature. The agreement between rater judgements and SELMA values was 75%. Thus SELMA had higher agreement with the maturity ratings made by the human raters than with chronological age or number of words. This finding was not surprising. Given the great inter-individual variation in language development a 100% agreement between chronological age and quality of narratives is not to be expected. Our conclusion was that LSA was useful for measuring the semantic maturity of the children.

An Index of Language Disorder (SELIMI)

To complement SELMA, and to focus more directly on children with language disorder our next step was to see if LSA can be used for diagnostic purposes, i.e., if it can also identify children with language disorder (Bååth et al. 2019).

The 54 children with DLD in the data set originated from two different samples from studies of DLD in Swedish populations. One sample, group A, consisted of 36 children with DLD of non-specified severity attending main stream schools (Asker-Árnason et al. 2012; Hansson et al. 2004). The other sample, group B, consisted of 18 children with severe DLD attending school language units (Kalnak 2014). Due to the difference in characteristics and an age-difference (the children in group A were 1–2 years older than the children in group B) we treated them separately in the analyses.

Using the same semantic space as for the SELMA-measure, a *logistic ridge regression* (Le Cessie and Van Houwelingen 1992) was trained to discriminate between children with DLD and children with typical language development, separately for group A and group B, again applying a leave-one-out cross validation

scheme. Each child's narrative was represented by a semantic vector. Thus the *Semantic Linguistic Impairment Index (SELIMI)* was created.

In group A 68% of the participants were correctly identified as having or not having DLD, as compared to 81% in group B. Sensitivity (i.e., the ability of the measure to correctly identify a child as having DLD) was 64% and 78% respectively and specificity (i.e., the ability of the measure to correctly identify a child as not having DLD) 72% and 83% respectively. Thus, when using SELIMI to predict DLD the results for group A are below or slightly below, and results for group B are close to, what has been described as a fair level of sensitivity (Plante and Vance 1994). The conclusion was that LSA seems useful also for distinguishing between children with, or without language disorders. The fact that significant results were found in rather small data sets suggests that the proposed method has a promising potential. Typically, a larger dataset leads to better training, and higher accuracy in the predictions. Thus, having a larger number of participants is desirable to reach higher certainty.

Conclusions

We have presented two applications of LSA. One is a method to quantify semantic linguistic maturity (SELMA) and the other a method to identify narratives produced by children with language disorder (SELIMI). Both methods are built on the theoretical assumption that meaning resides in how words co-occur. The results from the comparison with other variables (chronological age and number of words) in the first study indicate that SELMA contains additional information on semantic maturity. In particular, the results suggest that the semantic representation of the narratives contains information on semantic maturity. Several other methods (Blei et al. 2003; Sahlgren 2007; Shaoul and Westbury 2010) use semantic spaces to assess text complexity, text quality or other aspects of verbal data. The unique property of SELMA is that it is directly related to semantic development. This makes it a tool that is particularly relevant to use in studies of child language development.

The findings from the first study were followed-up by creating a new LSA based method, SELIMI, to discriminate between children with DLD and typically developing children. The results suggest that SELIMI can identify children with and children without DLD with fairly good accuracy. Thus, in both studies the LSA based scores show good correspondence with outcome, in the first study regarding developmental maturity, and in the second regarding clinical diagnosis. The results indicate that semantic difficulties in children with DLD can be shown to be manifested at the text level by SELIMI. Awareness of the importance of hearing and reading words in context for learning their meaning will have theoretical implications for research in these areas and practical implications for intervention and advice to parents. Our experience from close-to-practise teacher intervention programs indicate that words are still taught in isolation (glossary, synonyms,

antonyms) and that teachers need instruction on how to expand word knowledge in sentence contexts (Dockrell et al. 2015). SELIMI has not been applied in clinical settings, but has the potential to contribute to semantic assessments to incorporate a measure of the quality of semantic networks in addition to traditional measures focusing on the word level in the assessment (McGregor 2009; Sheng and MacGregor 2010).

Further studies to replicate the results on larger samples are needed, with the aim to create a clinical application. Also, further studies based on other types of data (e.g., word fluency or word association tasks; Roll et al. 2011) and with more specific semantic questions would be very interesting. Applications in educational contexts, to measure written text quality have been made (e.g., Landauer et al. 1997; Landauer et al. 2003), but more studies are needed on clinical samples in the written modality. Furthermore, individual variation is large within the group of children with DLD. For this study the only criterion was that the children had a diagnosis of language disorder, there were no criteria as to difficulties within specific aspects of language. Investigating children with documented difficulties in lexical-semantic skills specifically might also add important insights.

This implies that LSA offers a range of future possibilities for children with immature and vulnerable language. A more accurate judgement of the need for special services to children with semantic immaturity would be within reach. It is well-known that receptive language immaturity or semantic difficulties are often hard to detect, in particular if expressive language and speech production is accurate. Research also shows that receptive and semantic language difficulties are associated with more severe difficulties and worse prognosis (e.g., Bishop and Edmundson 1987) compared with isolated expressive difficulties, and are thus more detrimental to children's academic and social success as well as mental health (e.g., Conti-Ramsden et al. 2016). Methodological advances in research on semantic maturity and vulnerability in school-age children are few and there is a need for frameworks representing 'out of the box' thinking, as in the work reported here.

LSA also offers new opportunities to assess semantic maturity in multilingual children, with or without language disorder. Currently we are collecting longitudinal data from schools with a majority of multilingual children in a community of Southern Sweden. In future studies tools like SELMA, will make it possible to explore whether second language learners' semantic maturation differs from first language learners with or without vulnerable or disordered language.

To conclude, the analysis of narrative skills and the analysis of text complexity must include several domains (Graesser et al. 2011; Manhart and Rescorla 2002) with respect to form as well as to content. With computer-generated measures of semantic-linguistic quality in texts we hope to contribute a new dimension to the analysis of content focussing on the meaning relations between words and see great opportunities for the future.

Take-Home Messages

- Semantic maturity at the text level in children's narratives can be fairly reliably assessed using computer based methods/LSA.

- A computer-based method can with fair accuracy distinguish between children with and without language disorder.
- The method has a potential to be a valuable tool to assess objectively an ability which has been largely neglected because it is difficult to catch.
- The method needs to be used on larger data sets and to be validated against other measured of language skills.

References

- Appelbee, A. (1978). *The child's concept of story*. Chicago: University of Chicago Press.
- Asker-Árnason, L., Ákerlund, V., Skoglund, C., Ek-Lagergren, I., Wengelin, Å., & Sahlén, B. (2012). Spoken and written narratives in children/adolescents with hearing impairment – Lexical measures and story grammar. *Communication Disorders Quarterly*, 3, 131–145.
- Bååth, R., Sikström, S., Kalnak, N., Hansson, K., & Sahlén, B. (2019). Latent semantic analysis discriminates children with developmental language disorder (DLD) from children with typical language development. *Journal of Psycholinguistic Research*, 48, 683–697.
- Berman, R., & Slobin, D. (Eds.). (1994). *Relating events in narrative: A crosslinguistic developmental study*. Hillsdale, NJ: Lawrence Erlbaum.
- Bishop, D. (1997). *Uncommon understanding: Development and disorders of language comprehension in children*. Hove: Psychology Press.
- Bishop, D. (2014). Ten questions about terminology for children with unexplained language problems. *International Journal of Language and Communication Disorders*, 49, 381–415.
- Bishop, D., & Edmundson, A. (1987). Language-impaired 4-year-olds: Distinguishing transient from persistent impairment. *Journal of Speech and Hearing Disorders*, 52(2), 156–173.
- Bishop, D., Snowling, M., Thompson, P., Greenhalgh, T., & The Catalise-2 Consortium. (2017). Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry*, 58, 1068–1080. <https://doi.org/10.1111/jcpp.12721>
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022. <https://doi.org/10.1162/jmlr.2003.3.4-5.993>
- Botting, N. (2002). Narrative as a clinical tool for the assessment of linguistic and pragmatic impairments. *Child Language Teaching and Therapy*, 18, 1–22.
- Chiat, S., & Frazier Norbury, C. (2000). Semantic intervention to support word recognition: A single-case study. *Child Language Teaching and Therapy*, 16, 141–163.
- Conti-Ramsden, G., & Durkin, K. (2012). Postschool educational and employment experiences of young people with specific language impairment. *Language, Speech, and Hearing Services in Schools*, 43, 507–520.
- Conti-Ramsden, G., Durkin, K., Simkin, Z., & Knox, E. (2009). Specific language impairment and school outcomes. I: Identifying and explaining variability at the end of compulsory education. *International Journal of Language and Communication Disorders*, 44, 15–35.
- Conti-Ramsden, G., Durkin, K., Mok, P. L., Toseeb, U., & Botting, N. (2016). Health, employment and relationships: Correlates of personal wellbeing in young adults with and without a history of childhood language impairment. *Social Science & Medicine*, 160, 20–28.
- Corrigan, R. (2008). Beyond the obvious: Constructing meaning from subtle patterns in the language environment. *Communication Disorders Quarterly*, 29, 109–124.
- Dockrell, J., Messer, D., George, R., & Wilson, G. (1998). Children with word finding difficulties: Prevalence, presentation, and naming problems. *International Journal of Language and Communication Disorders*, 33, 445–454.

- Dockrell, J., Bakopoulou, I., Law, J., Spencer, S., & Lindsay, G. (2015). Capturing communication supporting classrooms: The development of a tool and feasibility study. *Child Language Teaching and Therapy*, *31*, 271–286.
- Dollaghan, C. (1987). Fast mapping of normal and language-impaired children. *Journal of Speech and Hearing Disorders*, *52*, 218–222.
- Fenson, L., Dale, P., Reznick, J., Bates, E., Hartung, J., Pethick, S., & Reilly, J. (1993). *MacArthur communicative development inventories: User's guide and technical manual*. San Diego, CA: Singular Publishing Group.
- Fenson, L., Marchman, V., Thal, D., Dale, P., Bates, E., & Reznick, J. (2007). *The MacArthur-Bates communicative development inventories: User's guide and technical manual* (2nd ed.). Baltimore: Paul H. Brookes.
- Fey, M. E., Catts, H. W., Proctor-Williams, K., Tomblin, B., & Zhang, X. (2004). Oral and written story composition skills of children with language impairment. *Journal of Speech, Language and Hearing Research*, *47*, 1301–1318.
- German, D., & Simon, E. (1991). Analysis of children's word-finding skills in discourse. *Journal of Speech and Hearing Research*, *34*, 309–316.
- Gough, P., & Tunmer, W. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, *7*, 6–10.
- Graesser, A., McNamara, D., & Kulikowich, J. (2011). Och-Matrix: Providing multilevel analysis of text characteristics. *Educational Researcher*, *40*, 223–234.
- Hansson, K., Forsberg, J., Löfqvist, A., Mäki-Torkko, E., & Sahlén, B. (2004). Working memory and novel word learning in children with hearing impairment and children with specific language impairment. *International Journal of Language and Communication Disorders*, *39*, 401–422.
- Hansson, K., Bååth, R., Löhndorf, S., Sahlén, B., & Sikström, S. (2016). Quantifying semantic linguistic maturity in children. *Journal of Psycholinguistic Research*, *45*, 1183–1189.
- Johansson, V. (2009). *Developmental aspects of text production in writing and speech*. Doctoral dissertation. Department of Linguistics, Lund University, Lund.
- Kail, R., & Leonard, L. (1986). Word-finding abilities in language-impaired children. *ASHA Monographs*, *25*, 1–39.
- Kalnak, N. (2014). *Family history, clinical marker and reading skills in children with specific language impairment*. Thesis for doctoral degree. Stockholm: Department of Women's and Children's Health, Karolinska Institutet.
- Kjell, O., Kjell, K., Garcia, D., & Sikström, S. (2020a). Semantic similarity scales: Using semantic similarity scale to measure depression and worry. In S. Sikström & D. Garcia (Eds.), *Statistical semantics: Methods and applications*. Heidelberg: Springer.
- Kjell, O., Kjell, K., Garcia, D., & Sikström, S. (2020b). Prediction and semantic trained scales: Examining the relationship between semantic responses to depression and worry and the corresponding rating scales. In S. Sikström & D. Garcia (Eds.), *Statistical semantics: Methods and applications*. Heidelberg: Springer.
- Klee, T. (1992). Developmental and diagnostic characteristics of quantitative measures of children's language production. *Topics in Language Disorders*, *12*, 28–41.
- Landauer, T., & Dumais, S. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*, 211–240.
- Landauer, T., Laham, D., Rehder, B., & Scheriner, M. (1997). How well can passage meaning be derived without using word order? A comparison of latent semantic analysis and humans. In *Proceedings of the 19th annual conference of the cognitive science society* (pp. 412–417). Mahwah, NJ: Erlbaum.
- Landauer, T., Laham, D., & Foltz, P. (2003). Automatic essay assessment. *Assessment in Education*, *10*, 295–308.
- Le Cessie, S., & Van Houwelingen, J. C. (1992). Ridge estimators in logistic regression. *Applied Statistics*, *41*, 191–201.

- Leonard, L. (2014). *Children with specific language impairment* (2nd ed.). Cambridge: MIT.
- Leonard, L., Nippold, M., Kail, R., & Hale, C. (1983). Picture naming in language impaired children. *Journal of Speech and Hearing Research, 26*, 609–615.
- Leonard, L., Miller, C., & Gerber, E. (1999). Grammatical morphology and the lexicon in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 42*, 678–689.
- Manhart, J., & Rescorla, L. (2002). Oral narrative skills of late talkers at ages 8 and 9. *Applied Psycholinguistics, 23*, 1–21.
- Mayer, M., & Mayer, M. (1975). *One frog too many*. New York, NY: Penguin.
- McGregor, K. (2009). Semantics in child language disorders. In R. Schwartz (Ed.), *Handbook of child language disorders* (pp. 365–387). New York, NY: Psychology Press.
- McGregor, K., & Appel, A. (2002). On the relation between mental representation and naming in a child with specific language impairment. *Clinical Linguistics & Phonetics, 16*, 1–20.
- McGregor, K., Newman, R., Reilly, R., & Capone, N. (2002). Semantic representation and naming in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 45*, 998–1014.
- McKee, G., Malvern, D., & Richards, B. (2000). Measuring vocabulary diversity using dedicated software. *Literary and Linguistics Computing, 15*, 323–337.
- Miniscalco, C., Hagberg, B., Kadesjö, B., & Westerlund, M. (2007). Narrative skills, cognitive profiles and neuropsychiatric disorders in 7 and 8-year old children with late developing language. *International Journal of Language and Communication Disorders, 42*, 665–681.
- Myhill, D. (2008). Towards a linguistic model of sentence development in writing. *Language and Education, 22*, 271–288.
- Nielsen, F. Å., & Hansen, L. K. (2020a). Creating semantic representations. In S. Sikström & D. Garcia (Eds.), *Statistical semantics: Methods and applications*. Heidelberg: Springer.
- Nielsen, F. Å., & Hansen, L. K. (2020b). Software for creating and analyzing semantic representations. In S. Sikström & D. Garcia (Eds.), *Statistical semantics: Methods and applications*. Heidelberg: Springer.
- Norbury, F. C., & Bishop, D. V. M. (2003). Narrative skills of children with communication impairments. *International Journal of Language and Communication Disorders, 38*, 287–313.
- Owen, A., & Leonard, L. (2002). Lexical diversity in the spontaneous speech of children with specific language impairment: Application of D. *Journal of Speech, Language, and Hearing Research, 45*, 927–937.
- Plante, E., & Vance, R. (1994). Selection of preschool language tests: A data-based approach. *Language, Speech, and Hearing Services in Schools, 25*, 15–24.
- Rice, M., Buhr, J., & Nemeth, M. (1990). Fast mapping word-learning abilities of language delayed preschoolers. *Journal of Speech and Hearing Disorders, 55*, 33–42.
- Roll, M., Mårtensson, F., Sikström, S., Apt, P., Arling-Bååth, R., & Horne, M. (2011). Atypical associations to abstract words in Broca's aphasia. *Cortex, 48*, 1068–1072.
- Sahlén, B., & Hansson, K. (2006). Novel word learning and its relation to cognitive and linguistic skills. A comparison between children with SLI and children with mild-to-moderate hearing impairment. *Journal of Multilingual Communication Disorders, 4*, 95–107.
- Sahlgren, M. (2007). *An introduction to random indexing. SICS2007* (pp. 1–9). Stockholm: Stockholm University.
- Sandgren, O., Hansson, K., & Sahlén, B. (2015). Working memory and referential communication – Multimodal aspects of interaction between children with sensorineural hearing impairment and normal hearing peers. *Frontiers in Psychology, 6*, 242. <https://doi.org/10.3389/fpsyg.2015.00242>
- Shaoul, C., & Westbury, C. (2010). Exploring lexical co-occurrence space using HiDEx. *Behavior Research Methods, 42*, 393–413.
- Sheng, L., & MacGregor, K. (2010). Lexical-semantic organization in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 53*, 146–159.
- Sikström, S., Kjell, O., & Kjell, K. (2020). SemanticExcel.com: An online software for statistical analyses of text data based on natural language processing. In S. Sikström & D. Garcia (Eds.), *Statistical semantics: Methods and applications*. Heidelberg: Springer.

- Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
- Watkins, R., Kelly, D., Harbers, H., & Hollis, W. (1995). Measuring children's lexical diversity: Differentiating typical and impaired language learners. *Journal of Speech, Language, and Hearing Research*, 38, 1349–1355.
- Weckerly, J., Wulfeck, B., & Reilly, J. (2001). Verbal fluency deficits in children with specific language impairment: Slow rapid naming or slow to name? *Child Neuropsychology*, 7, 142–152.