Disentangling typical and atypical Russian acquisition patterns: an automated approach

Katsiaryna Aharodnik[1], Isabelle Barrière[1,3], Irina A. Sekerina[1,2], Valerie L. Shafer[1]

[1]CUNY Graduate Center, US
[2]CUNY College of Staten Island, US
[3]St Elizabeth University, US

https://doi.org/10.36505/ExLing-2024/15/0002/000627

Abstract

The present study investigated a machine learning approach for disentangling Developmental Language Disorder (DLD) from typical development in Russian-speaking children. We explored the BiSLI Corpora from CHILDES and conducted two studies: 1) with monolingual Russian children with typical development (TD) versus DLD; 2) with bilingual Russian-Dutch TD children and Russian monolingual children with DLD. The efficacy of using computational features was compared to that of traditional clinical measures for diagnosis of DLD. The results revealed that computational features, such as tf-idf vectors and n-gram models, performed at or above the clinical threshold for acceptable and good performance (using the "F1" measure) and demonstrated significantly better results than the traditional language sample analysis models.

Keywords: Developmental Language Disorder, machine learning, narrative skills, Russian children, bilingualism

Introduction

Reliable identification of children with Developmental Language Disorder (DLD) remains a significant challenge for researchers and clinicians. DLD is a language disorder with a high prevalence rate of approximately 7% among English-speaking preschoolers (Tomblin et al., 1997), and its language deficits often persist beyond early childhood. The pervasive nature of this disorder is evident in its impact on various linguistic domains, affecting both comprehension and production, despite the absence of generalized cognitive impairment. DLD profiles vary across languages, with highly inflectional languages posing particular challenges for clinical identification due to the subtle nature of errors in these languages (e.g, Armon-Lotem, 2014; Kornev & Balciuniene, 2021; Leonard, 2014).

Identifying DLD among bilingual children presents an additional and distinct challenge. Previous research has highlighted significant overlaps between the error profiles of typically developing (TD) bilingual children and those with DLD. Distinguishing between differences in first or second language production due to lower proficiency and the deficits indicative of DLD can be particularly difficult. Currently, no assessment tools are available that can accurately differentiate DLD from typical bilingual language development. To address this challenge, new approaches to bilingual assessment must be explored, taking into account the heterogeneity within the bilingual population (e.g. Blume et al., 2019) and the overlaps with clinical population.

The primary goal of the current study is to evaluate the effectiveness of using a machine learning technique for DLD identification in Russian-speaking children. A secondary goal is to identify computational linguistics features that outperform traditional Language Sample Analysis (LSA) approaches (Thordardottir, 2015). To achieve these objectives, two binary classification studies were conducted: 1) distinguishing between narratives produced by Russian-speaking 4-to-9-year-old monolingual children with typical development (TD) and those with DLD, and 2) distinguishing between 4-to-9-year-old Russian-Dutch bilingual TD children and Russian monolingual children with DLD.

The study tested two hypotheses: 1) computational features (e.g., tf-idf vectors and n-grams) used alone or in combination with additional narrative microstructure features would outperform the traditional LSA approach; 2) n-gram features (POS n-grams and word n-grams) would demonstrate superior performance compared to other features.

Methodology

The BiSLI Bilingual Corpus (Gülzow & Gagarina, 2007), available on CHIDLES (MacWhinney, 2000), served as the data source for this study. For Study 1, we analyzed data from the Monolingual (ML) Russian and ML Russian DLD corpora. For Study 2, we examined the data from the ML Russian DLD and Bilingual (BL) Russian corpora, focusing on simultaneous BLs with Russian-speaking mothers residing in the Netherlands. Two age groups were considered: a younger group (children aged 3;11 to 6;11 years), and an older group (children aged 7;0 to 9;11 years). Descriptive statistics for both studies are provided in Table 1.

We presented narrative microstructure features and computational features to the machine learning classifiers. The narrative microstructure feature set included language productivity and lexical diversity measures. The features traditionally employed for LSA and identified as markers of DLD in prior research (Botting, 2002) were defined as a reference point. These features were the mean length of utterance (MLU) in words, the number of different words (NDW), the total number of words (TNW), and the total number of utterances (TNU). In contrast, the computational features included tf-idf vectors (word-based and part-of-speech) and n-gram (word-based and part-of-speech) features. Part-of-speech (POS) n-grams represented extended POS tags with grammatical categories (e.g., case, gender, number) mentioned within tags.

Two supervised machine learning classifiers, the Logistic Regression and Support Vector Machines classifiers were employed in the study. We used a training-validation-test split for training and testing phases. Additionally, we implemented a k-fold (k=6) cross-validation to check the stability and generalizability of the data. Leave-one-out-cross-validation (LOOCV) was also performed on the entire training set. The LOOCV output was used to calculate statistical difference between the classifier models with McNemar Test.

Table 1. Descriptive statistics for Study 1 and Study 2.

Study	Age	N participants		Mean age		Mean number of tokens per child	
		TD	DLD	TD	DLD	TD	DLD
ML	4-6	78	64	5;5	5;6	120.3	124.7
ML	7-9	60	58	8;2	8;2	170.7	160.7
BL	4-6	77	64	5;6	5;5	62.8	61.6
BL	7-9	77	58	8;1	8;2	71.3	80.8

Results

The results of both studies were interpreted in relation to diagnostic criteria employed in speech pathology research. The performance of a measure ≥ 0.8 was considered acceptable, and ≥ 0.9 as good (Plante & Vance, 1994; Spaulding et al., 2013). The results of both studies supported the initial hypotheses. For the ML study, computational features reached the threshold for clinically acceptable performance for the older age group, with word tf-idf vectors (p = 0.03389) and POS n-grams (p = 0.000355) performing significantly better than the LSA models. In the BL study, all computational features, word tf-idf vectors (p = 0.00142), word n-grams (p = 0.00865) and POS n-grams (p = 0.011117) performed significantly better than the LSA models for the younger age group.

The results of the first study demonstrated that the machine learning classifiers were able to disentangle between ML TD and DLD based on word-based features, which could be attributable to differences in pronunciation, frequency of mazes, vocabulary choices, and lexical errors among the monolingual groups. The POS n-grams demonstrated the best and most stable result for the older age group, suggesting that it was easier for the algorithms to classify the groups with more mature syntactic patterns. This is in line with the research showing differences in complex syntax for the older DLD and TD children (Reilly et al., 2004).

The high classification performance for both word and POS n-grams in the BL study could be attributed to BLs' lower proficiency in Russian. Specifically, the differences were found in the use of Subject-Verb agreement, and in morphological and morphosyntactic markers. Further error analyses should reveal whether there are additional differences in error profiles of the groups.

Conclusion

The current study validated the machine learning approach for the DLD identification for monolingual and bilingual children and highlighted its promise as a valid, and potentially automated, tool for DLD diagnosis in a highly inflectional language. The computational features outperformed a traditional LSA approach and reached clinically acceptable thresholds.

Future follow-up studies should consider including errors as feature-predictors and explore classifiers other than traditional statistical machine learning algorithms. In addition, future studies should determine whether the current approach can be applied to the diagnoses of DLD in children speaking languages other than Russian.

References

Armon-Lotem, S. 2014. Between L2 and SLI: Inflections and prepositions in the Hebrew of bilingual children with TLD and monolingual children with SLI. Journal of Child Language, 41(1), 3-33.

Botting, N. 2002. Narrative as a tool for the assessment of linguistic and pragmatic impairments. Child language teaching and therapy, 18(1), 1-21.

Blume, M.; Barrière, I.; Dye, C.; Kang, C. 2019. Challenges for the Development of Linked Open Data for Research in Multilingualism. A. Pareja, M. Blume, B.C. Lust, M.R. Melanson (Eds.) Development of Linguistic Linked Open Data Resources for Collaborative Data-Intensive Research in the Language Sciences. (185-200) Boston: MIT press.

Kornev, A.N., Balčiūnienė, I. 2021. Lexical and grammatical errors in developmentally language disordered and typically developed children: The impact of age and discourse genre. Children, 8(12), 1114.

Reilly, J., Losh, M., Bellugi, U., Wulfeck, B. 2004. "Frog, where are you?" Narratives in children with specific language impairment, early focal brain injury, and Williams syndrome. Brain and language, 88(2), 229-247.

Spaulding, T.J., Hosmer, S., Schechtman, C. 2013. Investigating the interchangeability and diagnostic utility of the PPVT-III and PPVT-IV for children with and without SLI. International Journal of Speech-Language Pathology, 15(5), 453-462.

Thordardottir, E. 2015. The relationship between bilingual exposure and morphosyntactic development. International journal of speech-language pathology, 17(2), 97-114.