A Multilayer Annotated Corpus of Contrastive Focus in High-Functioning Autism

Authors: Valeria D'Aloia, University "G. D'Annunzio" of Chieti-Pescara
Marisa Filipe, University of Lisbon, Centre of Linguistics
Sónia Frota, University of Lisbon, Centre of Linguistics
Abstract ID: 417
Event: Language Disorders 2026
Subject: 25. DELAD Special Session: Language Disorder Corpora
Presenter Name: Valeria D\'aloia

Abstract

Objectives: The expression of contrastive focus is frequently impaired in Autism Spectrum Disorder (ASD), yet existing resources provide limited evidence about where and how prominence is assigned when it is not realized on the intended focus target, and available descriptions remain predominantly acoustic. To address these limitations, we developed a publicly available corpus by enriching an existing clinical speech dataset with a multilayer annotation scheme that captures both prominence location and its phonological realization.

Methodology: Speech data from the European Portuguese PEPS-C Focus Expression task included recordings from 18 Portuguese-speaking children with High-Functioning Autism (HFA: 5-15 years, M = 8.06, SD = 2.31, IQ ≥ 80) and 18 typically developing (TD) peers matched for age, gender, and non-verbal IQ. The task comprised 16 experimental items per child, resulting in a final corpus of 559 utterances after quality control. The recordings were re-annotated in Praat to identify: i) the intended focus target; ii) the prominence configuration obtained, including the constituent(s) receiving prominence and the corresponding error patterns (e.g., wrong single word, multiple prominence, broad focus, no prominence); and iii) the prosodic realization of each prominent constituent through the annotation of pitch accents and boundary tones following the Portuguese Tones and Break Indices (P-ToBI) framework.

Results: The multilayer annotation revealed distinct prominence-placement patterns across the two groups, allowing error types to be linked directly to their phonological realization. Error patterns showed that children with HFA more frequently distributed prominence across multiple words, whereas TD children more often assigned prominence to a single unintended word. The P-ToBI annotation further showed that both groups predominantly used the target-like pitch accent (H*+L) when prominence occurred on the intended constituent, whereas greater variability in pitch accents and boundary tones emerged when prominence was assigned elsewhere. Finally, linking prominence location with pitch-accent annotation revealed that, in children with HFA, target-like accentual patterns became increasingly frequent as prominence occurred closer to the right edge of the utterance.

Discussion: By jointly encoding intended focus, the actual location of prominence, and its phonological realization, the annotation captures aspects of contrastive focus that go beyond acoustic measures and perceptual judgments alone. This corpus provides a more detailed account of atypical prosodic marking of contrastive focus in children with HFA, revealing that reduced contrastive salience is associated not only with greater variability in intonational categories, but also with the frequent distribution of prominence across multiple constituents rather than its assignment to the intended focus target. The re-annotation of an existing clinical speech dataset transformed it into a reusable public linguistic resource, enabling analyses of prosodic strategies that were not possible with the original data and providing one of the few detailed phonological accounts of contrastive focus in children with ASD.