Automated Analysis of Natural Language Samples: Comparison of Children with ASD, DLD, and TD

Richard Sproat[1,2], Lois Black[1], Emily Prud'hommeaux[1], Jan van Santen[1], Brian Roark[1]

[1]Center for Spoken Language Understanding, Oregon Health & Science University; [2]BioSpeech, Inc.

{rws,lmblack,emily,vansanten,roark}@cslu.ogi.edu

BioSpeech — BIOMEDICAL APPLICATIONS OF SPEECH TECHNOLOGY

Background

There are areas of overlap between autism spectrum disorders (ASD) and developmental language disorders (DLD) that pose challenges for **differential diagnosis.** One such area is **language impairment.** There have been varying reports on the types of language impairments in ASD, their severity, and their incidence. These studies generally use structured, de-contextualized instruments; yet, **Language Sampling and Analysis** (LSA) methods may provide information that critically complements structured instruments. Since the paucity of LSA-based studies is likely due to the labor-intensiveness of LSA, automated methods are urgently needed.

Objectives

- Demonstrate the feasibility of **automating the analysis of natural language samples**, focusing on the IPSyn (Scarborough, 1991).
- Apply these methods to **document morphology and syntax** in high functioning verbal children with ASD, children with DLD, and typically developing children (TD).

Subjects

- Children ages 4-8 given battery of language & neurocognitive measures.
- **DLD** group (N=20): diagnosis via Tomblin's EpiSLI criteria or a CELF index score at -1 SD plus a spontaneous language measure at -1 SD, and DSM-IV-TR-based diagnosis.
- **ASD** group (N=36): diagnosis via ADOS, the Social Communication Questionnaire, and DSM-IV-TR-based clinical consensus diagnosis.
    - ➡ Divided into 2 subgroups: **ASD+DLD** group meeting DLD criteria (N=25) and **ASD-DLD** group not meeting DLD criteria (N=11).
- Groups were well-matched on age, but only the DLD and ASD+DLD groups were matched on VIQ and PIQ.
- Stringent exclusion criteria applied to all groups: children with neurodevelopmental disorders, neuropsychiatric disorders, or a sibling with ASD or DLD were excluded from the TD group.

	N	age	PIQ
TD	24	5.82	117
DLD	20	6.85	101
ASD-DLD	11	6.64	109
ASD+DLD	25	6.77	102

Index of Productive Syntax (IPSyn)

- Requires a corpus of **100 spontaneous sentences**.
- Proposes a set of **60 (mostly) English-specific constructions** based on noun phrase syntax and morphology, verb phrase syntax and morphology, sentence structure, questions and negations.
- Count **the number of occurrences of each structure** found in the transcript, up to a count of 2.

Manual IPSyn Annotation

- Trained speech-language pathology graduate students transcribed all child utterances from the following ADOS activities: *Play, Conversation, Picture Description, Wordless Picture Book.*
- Select the first 100 utterances from each child's transcript.
- Five labelers (all undergraduate linguistics majors) scored the conversations according to the IPSyn guidelines.
- Two labelers per conversation adjudicated if the difference for any of the major categories (Noun, Verb, Question, Sentence) exceeded 4 points.

Automated IPSyn Annotation

- Probabilistic parsing algorithm applied to each utterance to produce a syntactic parse tree.
- Hand-crafted rules used to extract the 60 IPSyn categories from these parse trees.

Example:

Watch this.

Parse Structure:

(S (VP (VBP *watch*) (NP (DT *this*))))

IPSyn Structures Extracted:
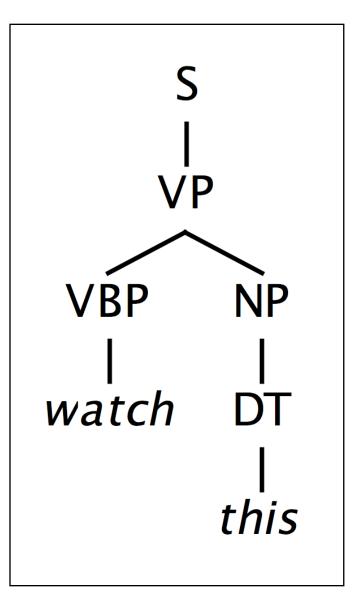
V01: Verb

(VBP *watch*)

N02: Pronoun or prolocative excluding modifiers

(NP (DT *this*))
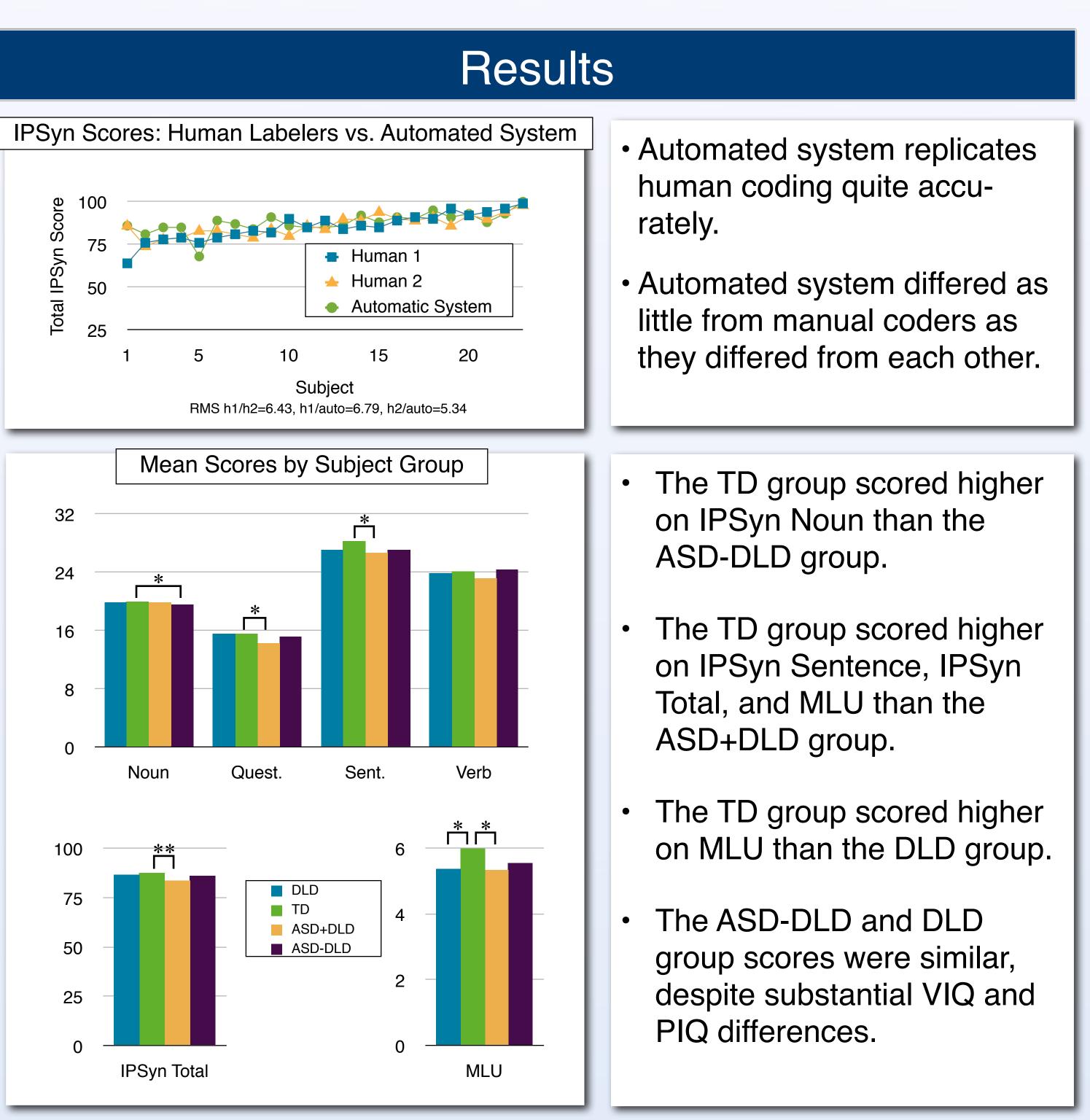
S01: Two word combination
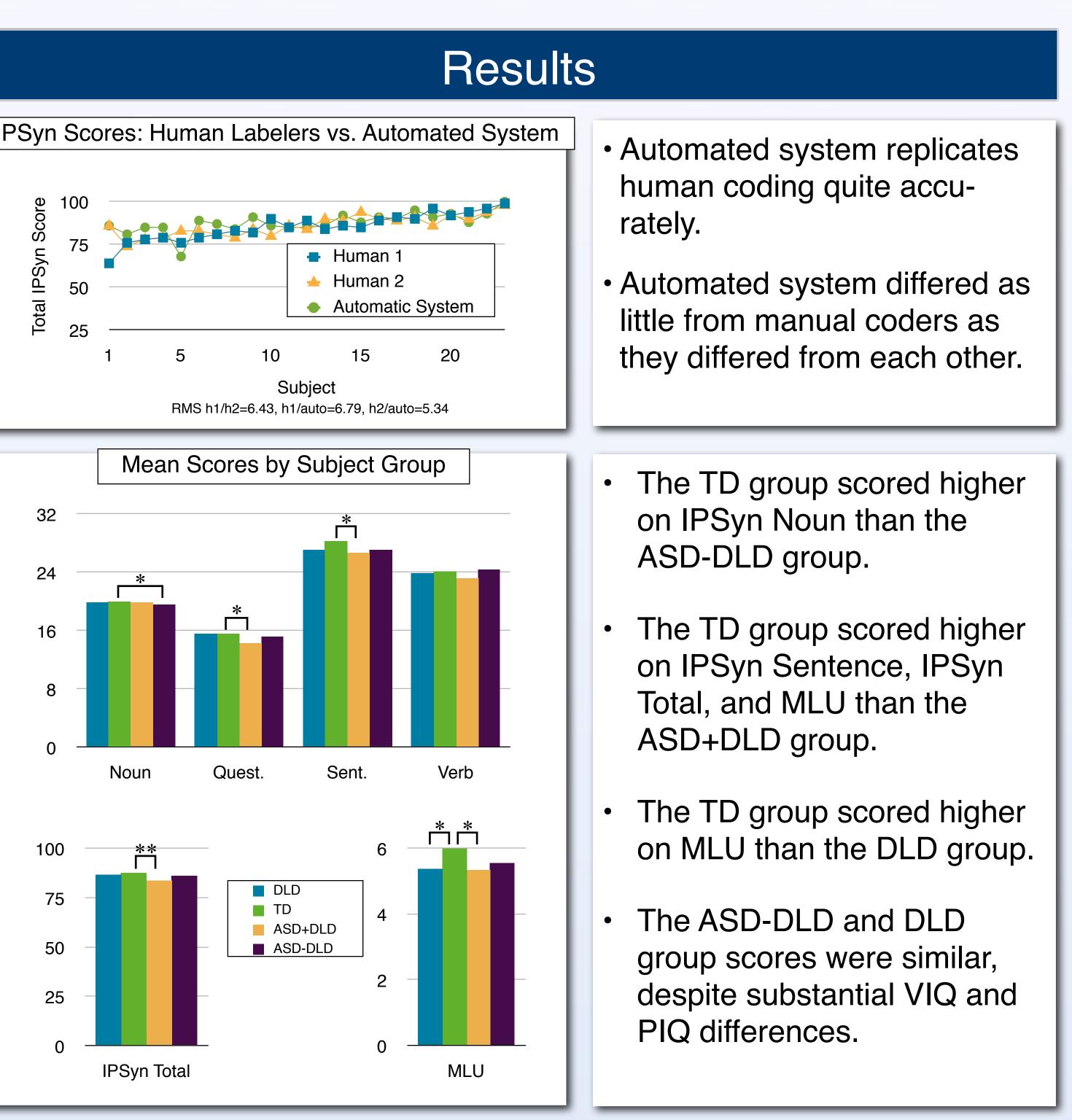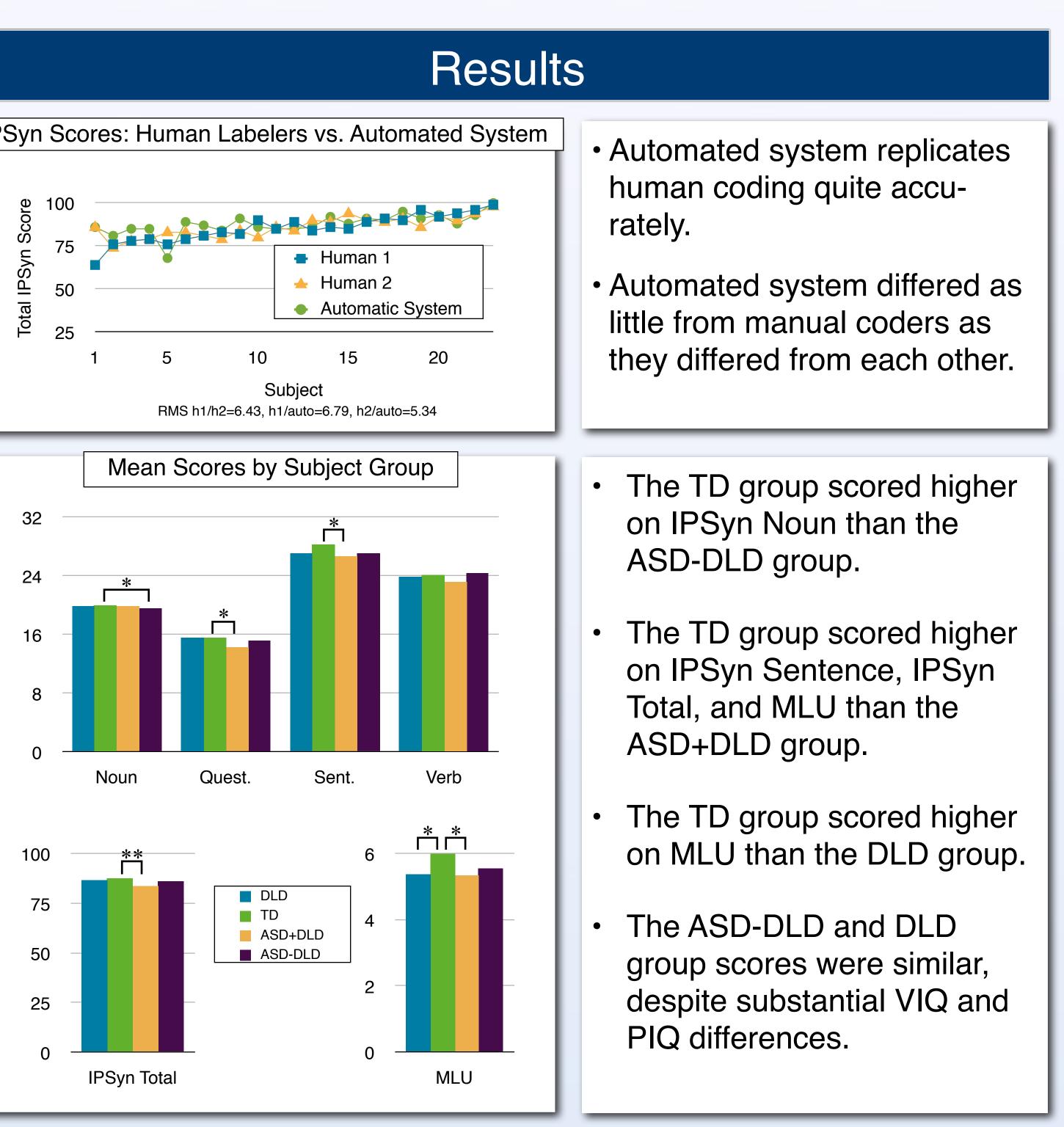
(S (VP (VBP *watch*) (NP (DT *this*))))

S03: Verb-object sequence

(VP (VBP *watch*) (NP (DT *this*)))

Results

IPSyn Scores: Human Labelers vs. Automated System
RMS h1/h2=6.43, h1/auto=6.79, h2/auto=5.34

Mean Scores by Subject Group

- Automated system replicates human coding quite accurately.
- Automated system differed as little from manual coders as they differed from each other.
- The TD group scored higher on IPSyn Noun than the ASD-DLD group.
- The TD group scored higher on IPSyn Sentence, IPSyn Total, and MLU than the ASD+DLD group.
- The TD group scored higher on MLU than the DLD group.
- The ASD-DLD and DLD group scores were similar, despite substantial VIQ and PIQ differences.

Conclusions

- The data show that the method is as accurate as human coding.
- Surprisingly, the methods showed that both ASD groups performed more poorly on IPSyn measures than one would expect based on their IQ characteristics. Specifically, the VIQ of the ASD-DLD group is 1.7 SD above that of the DLD group, yet their IPSyn scores are roughly equal.
- These results argue for the importance of LSA methods to complement structured measures.
- Moreover, given the extreme labor intensiveness of LSA, these results argue for developing additional automated LSA methods.

References & Acknowledgements

Gotham, K., S. Risi, A. Pickles, and C. Lord. 2007. The autism diagnostic observation schedule (ADOS): Revised algorithms for improved diagnostic validity. *Journal of Autism and Developmental Disorders,* 37:400.

Sagae, K., A. Lavie, and B. MacWhinney. 2005. Automatic measurement of syntactic development in child language. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics,* 197-204.

Scarborough, H. 1990. Index of productive syntax. *Applied Psycholinguistics,* 11:1-22.