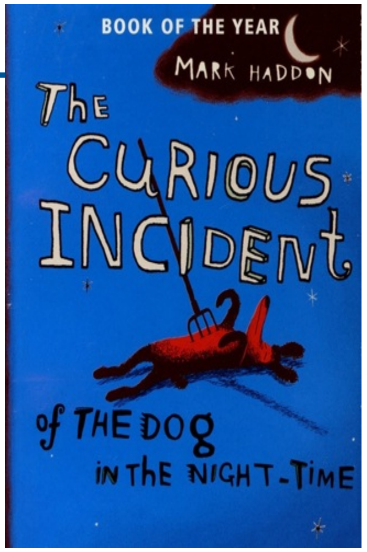
Automatically Evaluating Atypical Language in Narratives by Children with Autistic Spectrum Disorder

Michaela Regneri & Diane King

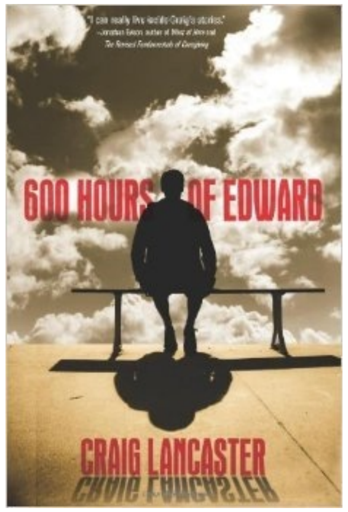
NLPCS 2014

Autism

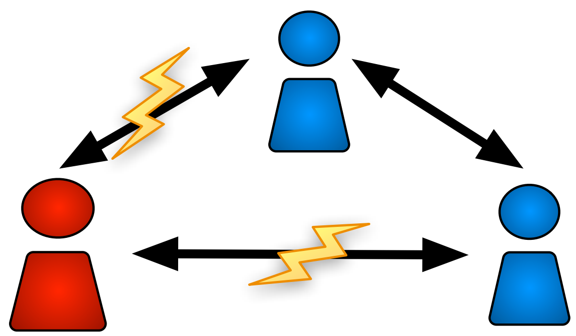

HOFFMAN
CRUISE
RAIN MAN


BOOK OF THE YEAR
MARK HADDON
The CURIOUS INCIDENT
of THE DOG
IN THE NIGHT-TIME

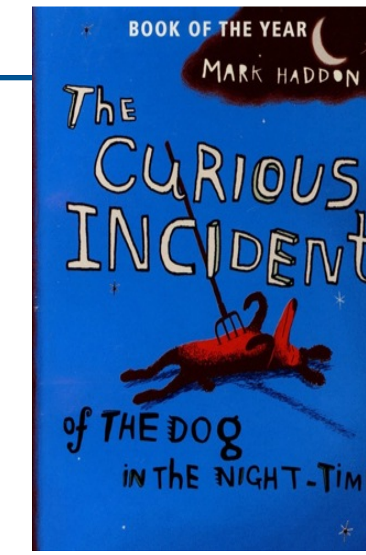

Mary and Max.


600 HOURS OF EDWARD
CRAIG LANCASTER

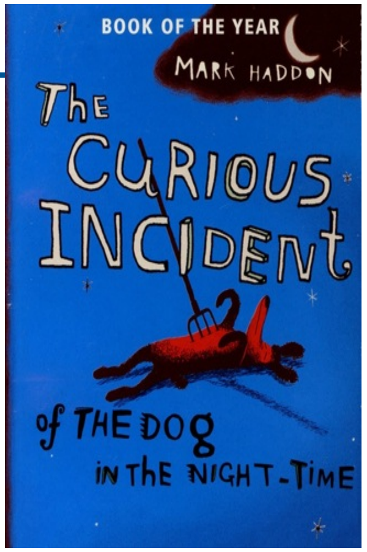
Autism

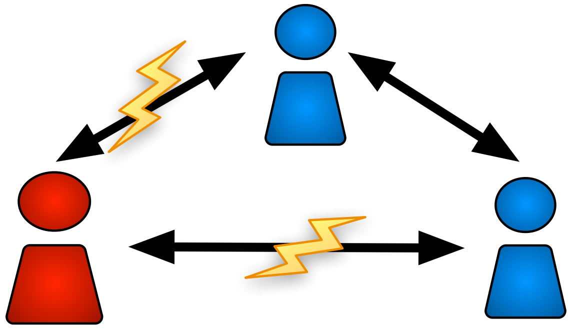


Impairment in reciprocal social interaction

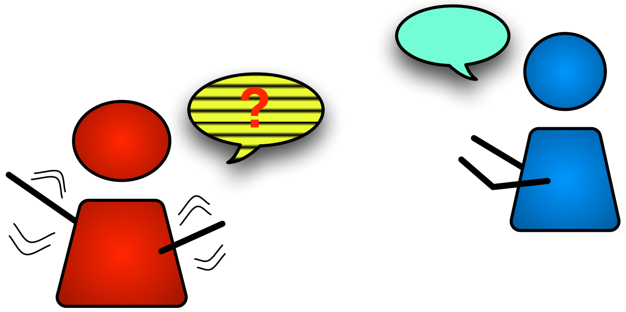


Autism

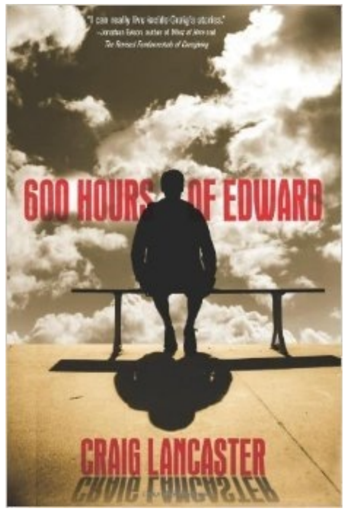




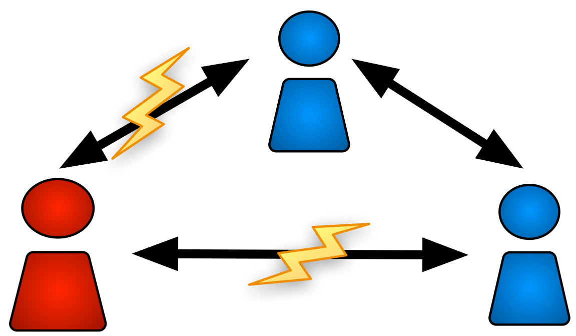
Impairment in reciprocal social interaction



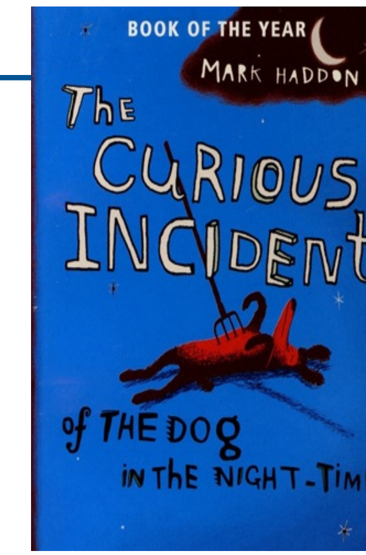
Impairment in verbal and non-verbal communication

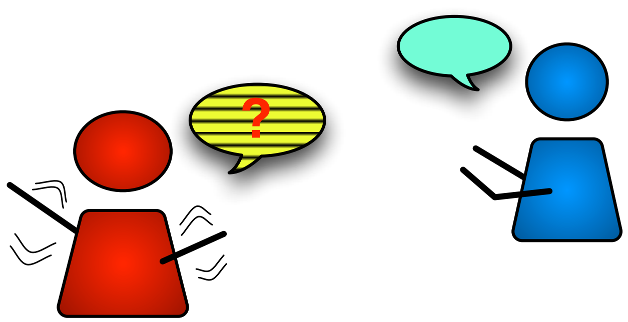


Autism

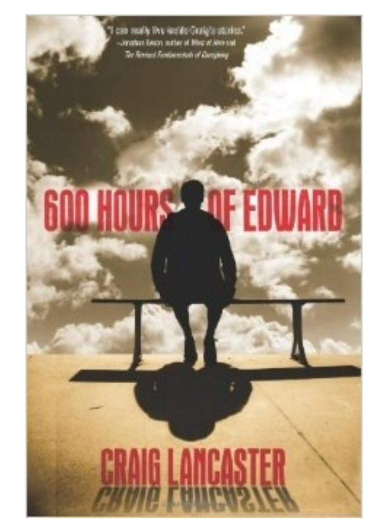


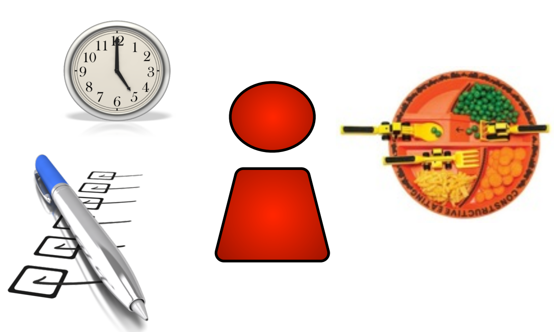
Impairment in reciprocal social interaction





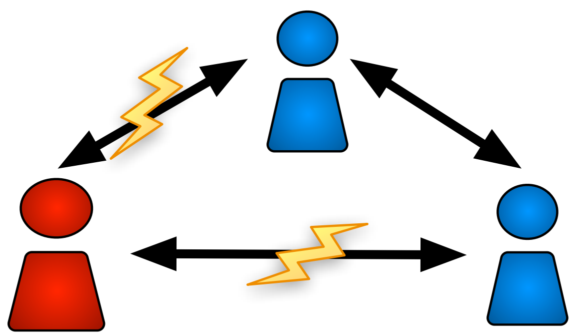
Impairment in verbal and non-verbal communication





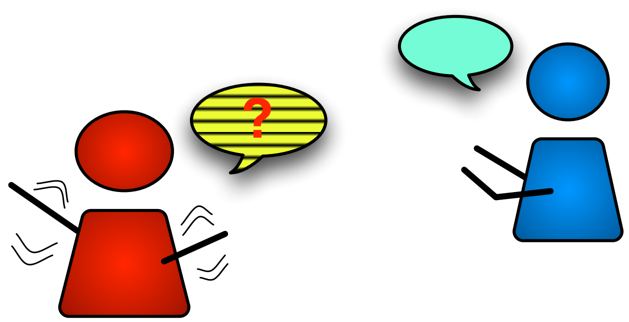
Restricted repetitive and stereotyped patterns of behaviour, interests and activities

Autism

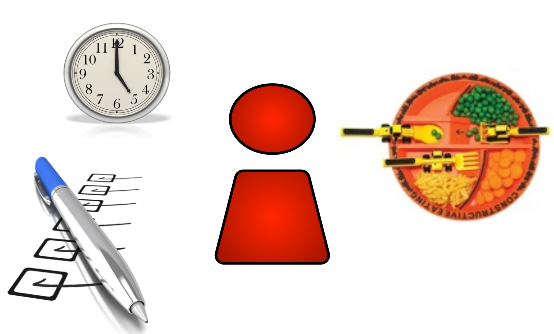


Impairment in reciprocal social interaction

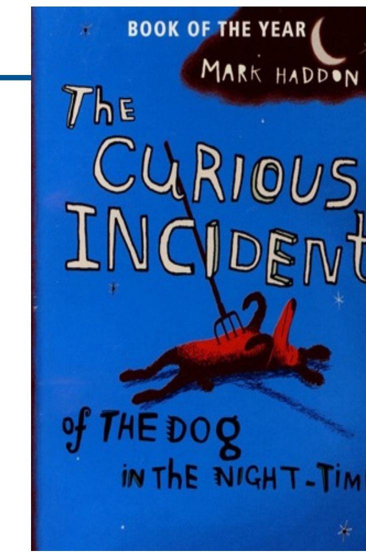
this paper



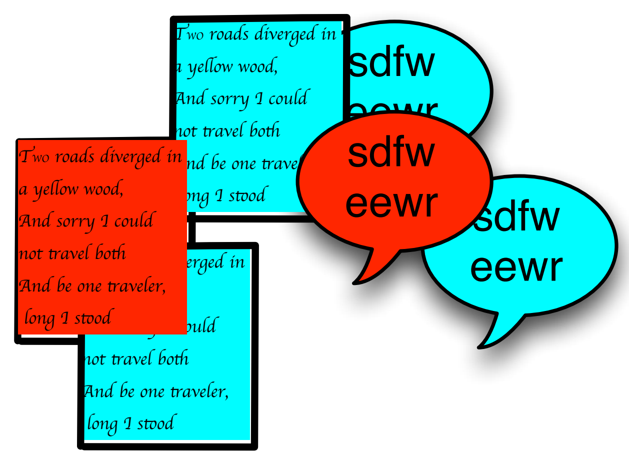
Impairment in verbal and non-verbal communication



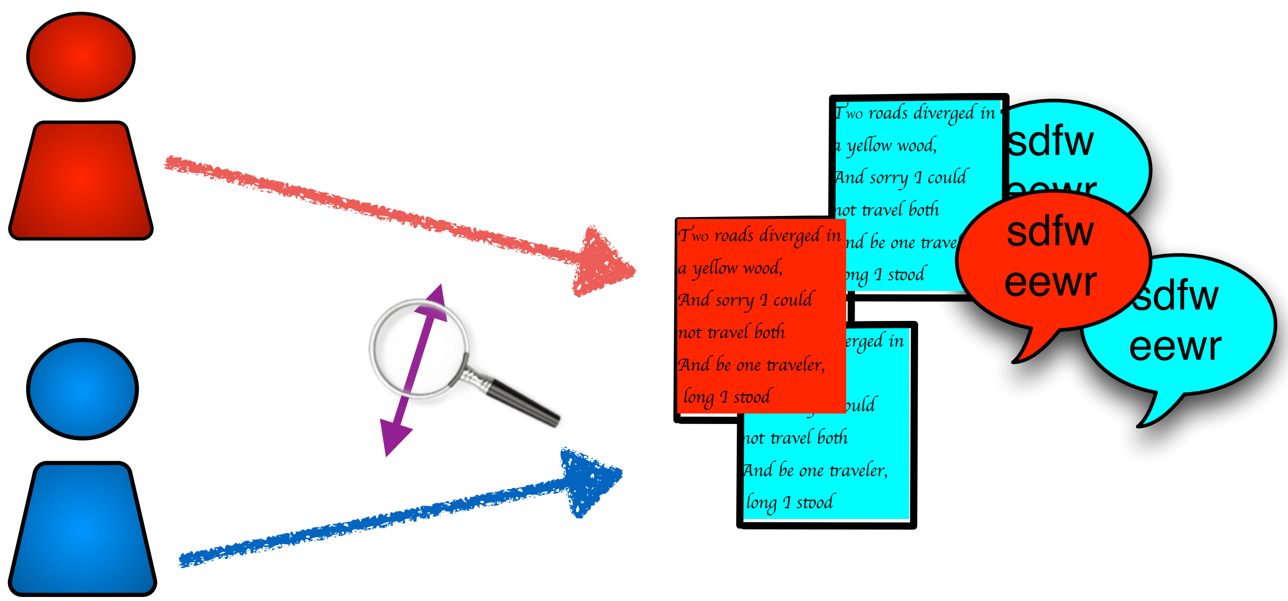
Restricted repetitive and stereotyped patterns of behaviour, interests and activities



Computational Stylometry for Diagnostic Analysis

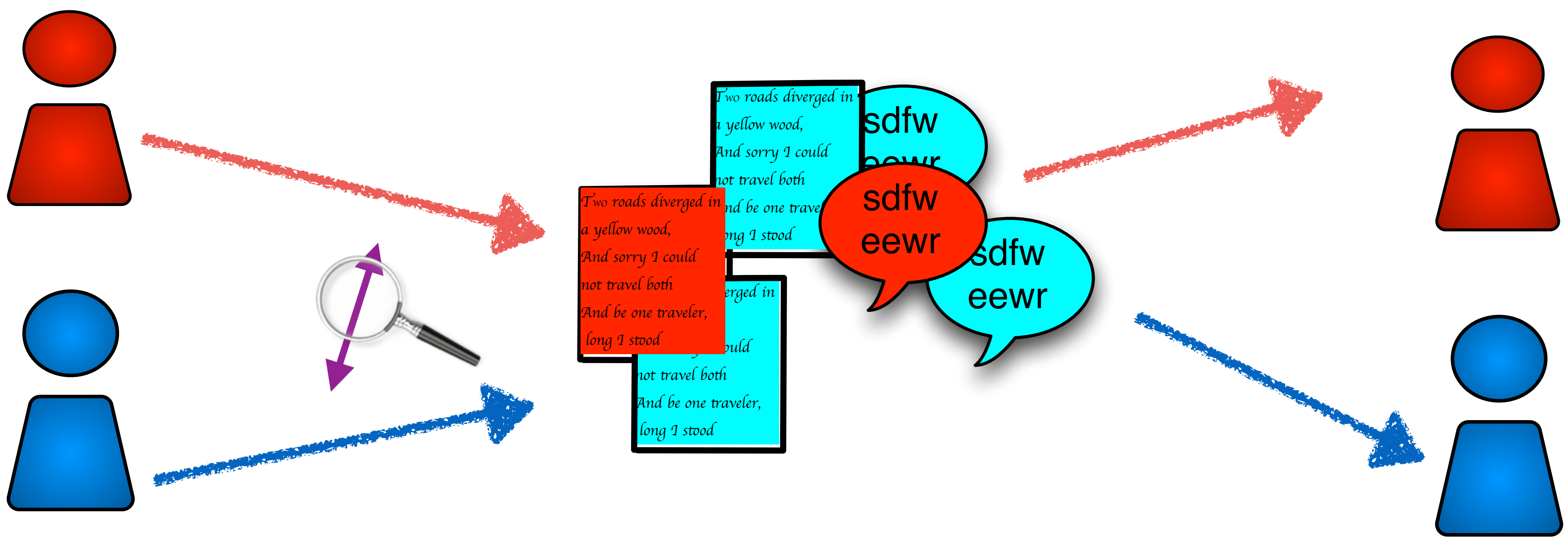


Computational Stylometry for Diagnostic Analysis



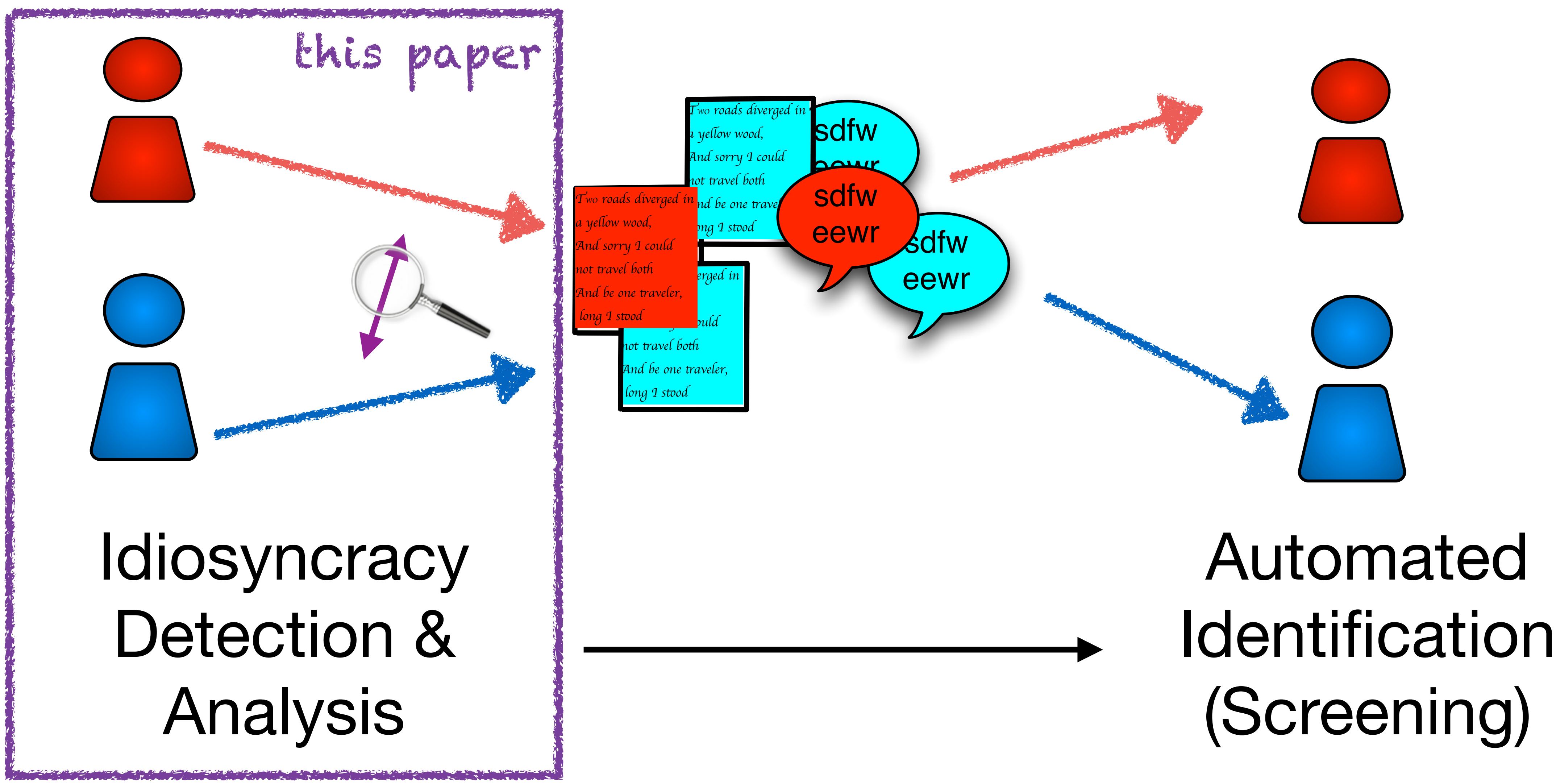
Idiosyncracy Detection & Analysis

Computational Stylometry for Diagnostic Analysis



Idiosyncracy Detection & Analysis → Automated Identification (Screening)

Computational Stylometry for Diagnostic Analysis



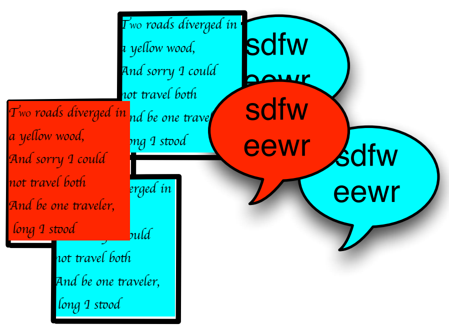
Related Work

- Work on detecting Alzheimer’s desease & other forms of dementia (Hirst and Wei Feng, 2012; Le et al., 2011, Baldas et al., 2011, Riley et al., 2005)
- Automated analysis of idiosyncrasies in language of adults with schizophrenia (Hong et al. 2012)
- ASD-related analysis: lots of work from psychology / cognitive science (Tager-Flusberg and Sullivan, 1995, Loveland and Tunali, 1993, Boucher and Bowler, 2010)
- Automated analysis on ASD data by Prud’Hommeaux et. al. (2011, 2014) / Rouhizadeh et al. (2013, 2014)

Our Experiments

- objectives: testing new features & validating features of previous work
- more authoritative (& more) data:
  - age 11-14
  - two control groups: one matched by age, one by language level
  - thematically diverse (more later)

Outline



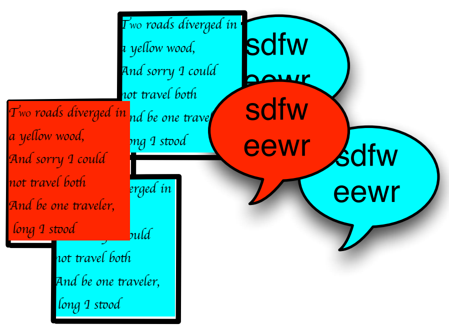
Dataset
(King et al. 2013)

Linguistic
Features

Automated
Evaluation

Feature
Analysis

Outline



Dataset
(King et al. 2013)

Linguistic
Features

Automated
Evaluation

Feature
Analysis

Data Collection

(King et al. 2013)

- short narratives about everyday scenarios:
  spending free time, going on holiday, birthday, halloween
  being scared, being angry

Data Collection

(King et al. 2013)

- short narratives about everyday scenarios:
  spending free time, going on holiday, birthday, halloween
  being scared, being angry

more episodic

more emotional

Data Collection

(King et al. 2013)

- short narratives about everyday scenarios:
  spending free time, going on holiday, birthday, halloween
  being scared, being angry

more episodic

more emotional

- 3 subject groups:
  - ASD group
  - control group matched by language-level (LM)
  - control group matched by chronological age (AM)

Data Collection

(King et al. 2013)

- short narratives about everyday scenarios:
  spending free time, going on holiday, birthday, halloween
  being scared, being angry

  more episodic

  more emotional

- 3 subject groups:
  - ASD group
  - control group matched by language-level (LM)
  - control group matched by chronological age (AM)
- narratives = oral answers to questions

Data Collection

(King et al. 2013)

- two types of prompts:
  - general narratives (GEN): "*What usually happens when someone goes on holiday?*"
  - specific narratives (SEN): "*Can you tell me about a time when you went on holiday?*"
- 3 (groups) * 2 (GEN/SEN) * 6 (scenarios) narratives (=**927**)
- each narrative was transcribed & annotated (language corrections, references to mental states, causal statements, negative comments, hedges, emphatic remarks, direct speech)

Example from Dataset

(King et al. 2013)

E	What usually happens to a thirteen year old on their birthday?
C	Get/s a cake and present/s.
E	Uh huh, anything else [P]?
C	Not[NC] necessarily[H] a cake.
E	No?
C	(um) : (um) go to see their family/s and stuff.
E	Hmm.
C	If they/'re not[NC] living with them[C].
E	Uh huh, anything else [P]?
C	Have a good[AS] time [E].

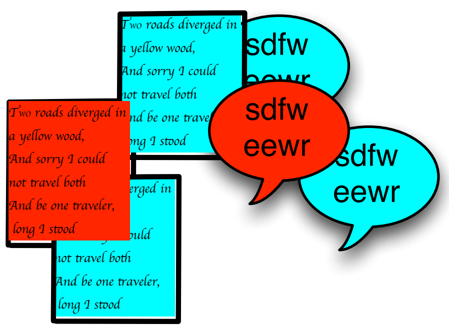
Examiner	(additional) **P**rompts		
Child	**N**egative **C**lauses	**C**onditions	**H**edging

Example from Dataset

(King et al. 2013)

E Can you tell me about what you did at halloween one time?
C Sat at home.
E Anywhere, home, school, doesn't have to be this year, it could be like, you know, five years ago.
E Just one time, what you did at halloween[P].
C I sat at home.
E You just sat at home?
C Yeah.
E Anything else [P]?
C Nope.
E So have you never done anything at halloween[P]?
C No.
E Never ever[P]?
C No, never.

Outline



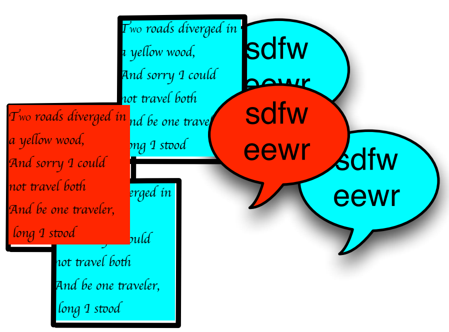
Dataset
(King et al. 2013)

Linguistic
Features

Automated
Evaluation

Feature
Analysis

Outline



Dataset
(King et al. 2013)

Linguistic
Features

Automated
Evaluation

Feature
Analysis

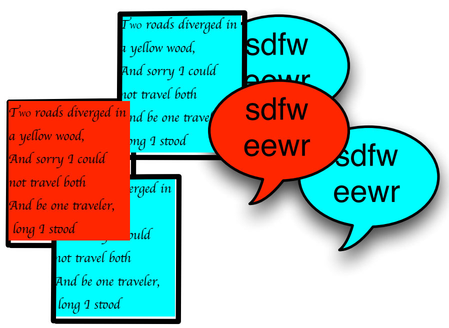
Idiosyncrasy Detection for ASD

- **General language difficulties**
  (special focus: comparison with LM group)
- **Difficulties with topic coherence**
  (theory: less coherence in ASD group;
  comparison with previous approaches)
- **Difficulties with communicating sentiments**
  (ASD: fewer references to mental states;
  question: what kind of references are there?)
- Difficulties with both **episodic memory** (SEN) and generalizing to **common sense knowledge** (GEN)

Feature overview

Feature	Objective	Prospective Outcome
Low-Frequency Words	assess **general language competence**	= LM Group (?)
Pronoun Use		Generally: like LM (~related work) (Q: for each pronoun type?)
TF-IDF scores (unusual words)	assess **topic coherence**	previous work *assumes* much higher tf-idf for ASD (?)
Sentiment Analysis	analyze **references to mental states**	Fewer references to mental states (Q: but for which sentiments?)

Outline



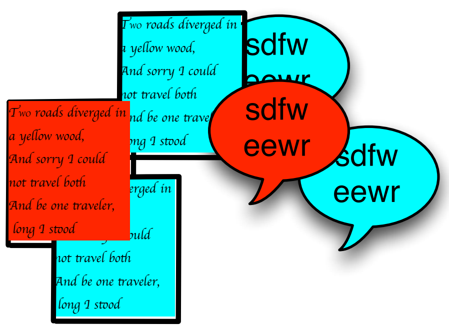
Dataset
(King et al. 2013)

Linguistic
Features

Automated
Evaluation

Feature
Analysis

Outline



Dataset
(King et al. 2013)

Linguistic
Features

Automated
Evaluation

&

Feature
Analysis

Data Preprocessing

- extract subject utterances (separate from examiner prompts)
- sentence splitting
- count word frequencies
- part-of-speech-tagging, parsing and sentiment analysis with Stanford NLP

General Language Competence: Low-Frequency Words

- words are low-frequency words if they occur rarely in a large reference corpus (BNC)
- set different thresholds and determine the proportion of low-frequency words for each group
- low-frequency words = either “difficult” (➠ language development) or idiosyncratic (➠ coherence)

General Language Competence: Low-Frequency Words

- words are low-frequency words if they occur rarely in a large reference corpus (BNC)
- set different thresholds and determine the proportion of low-frequency words for each group
- low-frequency words = either “difficult” (➠ language development) or idiosyncratic (➠ coherence)

Group / threshold	< 0.00003%	< 0.0001%	< 0.001%	< 0.01%
ASD	**0.18**	**0.26**	**0.47**	**0.57**
LM	0.23	0.31	0.52	0.65
AM	0.22	0.31	0.52	0.66

General Language Competence: Pronoun Use

- determine fraction of pronouns in all word tokens (*rises* with language competence)
- determine fraction of first person singular pronouns in all pronouns (*decreases* with language competence)
- compare generic / specific narratives

General Language Competence: Pronoun Use

- determine fraction of pronouns in all word tokens (*rises* with language competence)
- determine fraction of first person singular pronouns in all pronouns (*decreases* with language competence)
- compare generic / specific narratives

Group	pronouns / words	1 ps sg / pronouns	1 ps sg SEN	1 ps sg GEN
ASD	0.08	0.49	0.75	**0.23**
LM	0.10	0.47	0.76	0.18
AM	0.10	0.46	0.75	0.17

General Language Competence: Pronoun Use

- determine fraction of pronouns in all word tokens (*rises* with language competence)
- determine fraction of first person singular pronouns in all pronouns (*decreases* with language competence)
- compare generic / specific narratives

Group	pronouns / words	1 ps sg / pronouns	1 ps sg SEN	1 ps sg GEN
ASD	0.08	0.49	0.75	**0.23**
LM	0.10	0.47	0.76	0.18
AM	0.10	0.46	0.75	0.17

→ details...

General Language Competence: Pronoun Use

- over-proportional use of 1st ps. sg. pronouns in *general* narratives ➠ generalization issues?
- scenario-based analysis: prevalence in episodical / social scenarios

General Language Competence: Pronoun Use

- over-proportional use of 1st ps. sg. pronouns in *general* narratives ➠ generalization issues?
- scenario-based analysis: prevalence in episodical / social scenarios

Group / Scenario	spending free time	being scared	having a birthday	going on holidays	hallo-ween	being angry
ASD	**0.51**	0.22	0.20	0.18	**0.18**	0.09
LM	0.33	0.23	0.32	0.11	0.02	0.09
AM	0.36	0.13	0.21	0.13	0.09	0.10

Topic Coherence: TF * IDF

- **T**erm **F**requency * **I**nverse **D**ocument **F**requency
- scores how distinctive a word for a certain document is (within a document collection)
- high score for words that
  - occur frequently in the respective document
  - occur rarely in other documents
- normalization over document length necessary

Topic Coherence:
TF * IDF

- for us: document = narrative; document collection = all narratives from the same scenario

- hypotheses (also from previous work):
  1) more unusual words in the ASD group
  ➠ higher tf*idf on average
  2) the words from the ASD group are unusual to a higher degree ➠ higher maximum tf*idf scores

Topic Coherence: TF * IDF

Group	average tf*idf	maximum tf*idf
ASD	**1.53**	5.16
LM	1.46	5.16
AM	1.46	4.73

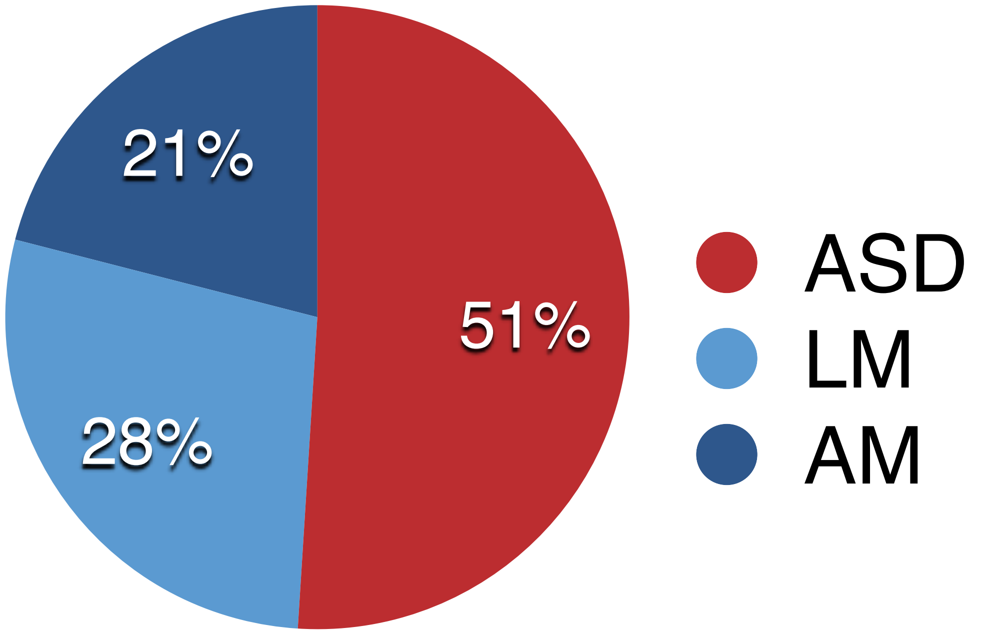
average / maximum tf*idf
(averaged over all scenarios)

Topic Coherence: TF * IDF

Group	average tf*idf	maximum tf*idf
ASD	**1.53**	5.16
LM	1.46	5.16
AM	1.46	4.73

average / maximum tf*idf (averaged over all scenarios)

division of word types with 1% highest tf*idf ("most unusual words")



References to Mental States: Automated Sentiment Analysis

- Q1: Is automated sentiment analysis a suitable tool?
- Q2: What kind of references of mental states do we find? (Particular focus: emotional scenarios)
- @ Q1: proportion of negative / positive / neutral sentences ➠ sentiment analyzer yields expected results

References to Mental States: Automated Sentiment Analysis

- Q1: Is automated sentiment analysis a suitable tool?
- Q2: What kind of references of mental states do we find? (Particular focus: emotional scenarios)
- @ Q1: proportion of negative / positive / neutral sentences ➠ sentiment analyzer yields expected results

Group	negative	positive	neutral
ASD	*0.37*	0.11	**0.52**
LM	0.54	0.10	0.36
AM	**0.60**	0.11	*0.29*

References to Mental States: Automated Sentiment Analysis

- Q1: Is automated sentiment analysis a suitable tool?
- Q2: What kind of references of mental states do we find? (Particular focus: emotional scenarios)
- @ Q1: proportion of negative / positive / neutral sentences ➠ sentiment analyzer yields expected results

Group	negative	positive	neutral
ASD	*0.37*	0.11	**0.52**
LM	0.54	0.10	0.36
AM	**0.60**	0.11	*0.29*

details (@ Q2)...

References to Mental States: Automated Sentiment Analysis

- proportion of neutral sentences:

Scenario	ASD		LM		AM		Average	
	GEN	SEN	GEN	SEN	GEN	SEN	GEN	SEN
spending free time	**0.63**	0.48	0.41	0.25	0.38	0.28	0.47	0.34
being scared	**0.65**	**0.51**	0.41	0.32	0.30	0.26	0.45	0.36
having a birthday	0.54	0.44	0.43	0.31	0.24	0.30	0.40	0.35
going on holidays	0.56	0.33	0.41	0.34	0.33	0.35	0.43	0.34
halloween	0.51	0.45	0.44	0.34	0.29	0.26	0.41	0.35
being angry	**0.64**	**0.48**	0.39	0.26	0.32	0.20	0.45	0.31
Average	0.59	0.45	0.42	0.30	0.31	0.28	0.44	0.34

marked:
+/- 0.15
vs. average
(or more)

Summarized Results

Summarized Results

Feature	Objective	Expected Outcome	Final Outcome
Low-Frequency Words	assess general language competence	= LM Group (?)	Roughly as expected
Pronoun Use		Generally: like LM (~related work) (Q: for each	Over-proportional use of 1 ps. sg. in *general* narratives (GEN)

Summarized Results

Feature	Objective	Expected Outcome	Final Outcome
Low-Frequency Words	assess general language competence	= LM Group (?)	Roughly as expected
Pronoun Use		Generally: like LM (~related work) (Q: for each	Over-proportional use of 1 ps. sg. in *general* narratives (GEN)
TF-IDF scores	assess topic coherence	previous work *assumes* much higher tf-idf for	Average / Maximum similar to control groups (!); however: words with highest tf*idf mostly in ASD group

Summarized Results

Feature	Objective	Expected Outcome	Final Outcome
Low-Frequency Words	assess general language competence	= LM Group (?)	Roughly as expected
Pronoun Use		Generally: like LM (~related work) (Q: for each	Over-proportional use of 1 ps. sg. in *general* narratives (GEN)
TF-IDF scores	assess topic coherence	previous work *assumes* much higher tf-idf for	Average / Maximum similar to control groups (!); however: words with highest tf*idf mostly in ASD group
Sentiment Analysis	assess references to mental states	Fewer references to mental states (Q: but for which sentiments?)	1) sentiment analyzer works 2) no differences for positive utterances 3) difficulties with verbalizing emotion-centric topics (especially for GEN)

Summarized Results

Feature	Objective	Expected Outcome	Final Outcome
Low-Frequency Words	assess general language competence	= LM Group (?)	Roughly as expected
Pronoun Use		Generally: like LM (~related work) (Q: for each	Over-proportional use of 1 ps. sg. in *general* narratives (GEN)
TF-IDF scores	assess topic coherence	previous work *assumes* much higher tf-idf for	Average / Maximum similar to control groups (!); however: words with highest tf*idf mostly in ASD group
Sentiment Analysis	assess references to mental states	Fewer references to mental states (Q: but for which sentiments?)	1) sentiment analyzer works 2) no differences for positive utterances 3) difficulties with verbalizing emotion-centric topics (especially for GEN)

ASD-specific!

Conclusion

- finding telling ASD-specific features is *hard*
- checking for language development is essential
- best evidence for ASD-specific features:
  - sentiment in emotional topics
  - first-person-pronouns in general narratives
- tf*idf is not necessarily a good feature for identifying text by people (children) with ASD

Future Work

- other text types (fictional stories? tweets?)
- analyze which features carry over to adult language (related work by Prud'hommeaux et al. 2014)
- show (dis-)similarities of features to other cognitive syndrome

Questions?

Thanks!

(More recent) References

Diane King, Julie E Dockrell, and Morag Stuart (2013): *Event narratives in 11-14 year olds with autistic spectrum disorder.* International Journal of Language Communication Disorders, 48(5):522–533.

Emily Prud'hommeaux, Eric Morley, Masoud Rouhizadeh, Laura Silverman, Jan van Santen, Brian Roark, Richard Sproat, Sarah Kauper, and Rachel DeLaHunta (2014): *Computational analysis of trajectories of linguistic development in autism.* Proceedings of the IEEE Spoken Language Technology Workshop (SLT), Lake Tahoe.

Masoud Rouhizadeh, Emily Prud'hommeaux, Jan van Santen, Richard Sproat (2014): *Detecting linguistic idiosyncratic interests in autism using distributional semantic models.* Proceedings of the ACL Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality.

Hiroki Tanaka, Sakriani Sakti, Graham Neubig, Tomoki Toda, Satoshi Nakamura (2014): *Linguistic and Acoustic Features for Automatic Identification of Autism Spectrum Disorders in Children's Narrative.* Proceedings of the ACL Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality.