

Attribuierung direkter Reden in deutschen Romanen des 18.-20. Jahrhunderts. Methoden zur Bestimmung des Sprechers und des Angesprochenen

Krug, Markus

markus.krug@uni-wuerzburg.de
Universität Würzburg, Deutschland

Jannidis, Fotis

fotis.jannidis@uni-wuerzburg.de
Universität Würzburg, Deutschland

Reger, Isabella

isabella.reger@uni-wuerzburg.de
Universität Würzburg, Deutschland

Macharowsky, Luisa

luisa.macharowsky@stud-mail.uni-wuerzburg.de
Universität Würzburg, Deutschland

Weimer, Lukas

lukas.weimer@stud-mail.uni-wuerzburg.de
Universität Würzburg, Deutschland

Puppe, Frank

frank.puppe@uni-wuerzburg.de
Universität Würzburg, Deutschland

Problembeschreibung

Im Folgenden wird ein Verfahren vorgestellt, das die automatische Zuordnung von direkter Rede in Erzähltexten sowohl zur Sprechenden als auch zur Angesprochenen Figur ermöglicht. Kann man eine solche automatische Zuordnung vornehmen, ermöglicht dies die Extraktion eines sozialen Netzwerks aus einem Text, wobei die Figuren als Knoten und die direkte Rede als Kanten modelliert werden (Elson / Dames 2010), aber sie kann auch eine wichtige Informationsquelle für andere analytische Schritte sein, z.B. zur Verbesserung der Koreferenzresolution oder zur Analyse der Quelle der Zuschreibung von Figurenattributen.

Stand der Forschung

Eine der ersten Arbeiten auf diesem Gebiet ist das System ESPER (Zhang et al. 2003), das direkte Reden innerhalb von Kindergeschichten erkennen soll. Das System extrahiert zunächst die direkten Reden im Text und klassifiziert diese mit einem Entscheidungsbaum in zwei Kategorien, Sprecherwechsel bzw. kein Sprecherwechsel. Evaluieren werden die Ergebnisse mit zwei manuell annotierten, sehr unterschiedlichen Geschichten. Sie berichten eine Genauigkeit, gemeint ist hier die Anzahl der korrekt bestimmten Sprecher für alle direkten Reden, von 47.6% und 86.7%. Glass und Bangay (2006), ebenfalls regelbasiert, bestimmen zunächst für eine direkte Rede das Kommunikationsverb und anschließend eine Menge von Akteuren, woraus letztendlich der Sprecher bestimmt wird. Sie evaluieren ihre Techniken auf 13 englischsprachigen fiktionalen Werken und berichten eine Genauigkeit von 79.4% (Glass / Bangay 2007). Iosif und Mishra (2014) folgen im Prinzip dem Schema von Glass und Bangay (2007), ergänzen es aber durch eine aufwendigere Vorverarbeitung einschließlich Koreferenzresolution. Sie erreichen eine Genauigkeit von ca. 84.5% und zählen damit zu den besten bisher veröffentlichten Ergebnissen. Ruppenhofer und andere (Ruppenhofer et al. 2010) berichten einen F-Score von 79% in der Zuordnung von Politikern zu ihren Aussagen in deutschsprachigen Kabinettsprotokollen aus den Jahren 1949-1960.

Neben diesen regelbasierten Ansätzen werden auch maschinelle Lernverfahren eingesetzt. Zu den ersten erfolgreichen Systemen zählt das von Elson und McKeown (2010). Ihre Daten für die Sprecherzuordnung ließen sie über Amazons *Mechanical Turk* System bearbeiten. Ihr System klassifiziert zunächst regelbasiert eine direkte Rede in eine von fünf syntaktischen Kategorien. Für jede dieser Kategorien wurden anschließend eigenständige maschinelle Lernverfahren trainiert. Insgesamt erreichen sie eine Genauigkeit von etwa 83%, ausgewertet anhand von englischen Romanen des 19. Jahrhunderts. O'Keefe und andere (O'Keefe et al. 2012), die an Elson und McKeowns Ansatz die Erstellung des Goldstandards und auch die praxisferne Verwendung von Informationen aus dem Goldstandard kritisieren, betrachten die Zuordnung als Sequenzproblem. Sie nutzen die Klassifikationsangaben von vorhergehenden direkten Reden als Features für die gesamte Sequenz. In ihrer Evaluation vergleichen Sie drei Verfahren mit einer sehr einfachen regelbasierten Baseline. Ihre Ergebnisse bei der Anwendung des Systems auf zwei Zeitungskorpora - Wall Street Journal und Sydney Morning Herald - sowie die Sammlung literarischer Texte aus der Arbeit von Elson und McKeown zeigen einen großen Unterschied zwischen den Domänen. Sie erreichen auf den beiden Zeitungskorpora 84.1% (WSJ) bzw. 91.7% (SMH) Genauigkeit. Auf dem literarischen Korpus erreichen sie dagegen lediglich eine maximale Genauigkeit von 49%. (He et al. 2013) erreichen mit einem auf Ranking basierten maschinellen Lernverfahren unter

der Ausnutzung von Features des Actor-Topic Modells (Celikyilmaz et al. 2010) auf dem Elson und McKeown-Korpus eine Genauigkeit zwischen 74.8% und 80.3%. Almeida und andere gehen von einer engen Verflechtung von Koreferenzresolution und Sprecherattribution aus und integrieren dabei beide Verfahren in ihrem Ansatz; die Ergebnisse der beiden einzelnen Lernverfahren werden in einem dritten Schritt verbunden. Sie erreichen damit 88.1% Genauigkeit (Almeida et al. 2014). Neuere Versuche mit Deep Learning-Verfahren aufgrund der Sprache der Figuren haben nur Genauigkeiten von unter 50% erreicht (Chaganty / Muzny 2014).

Die Zuordnung einer angesprochenen Figur wurde unserer Wissens noch in keiner anderen Arbeit untersucht.

Daten und Annotation

Für diese Arbeit verwenden wir Abschnitte des frei zugänglichen Korpus DROC. DROC besteht aus 89 Romanausschnitten, jeweils 130 Sätze lang, in denen alle Figurenreferenzen (mit und ohne Namen) und Koreferenzen annotiert sind. Aus dem Korpus wurden 77 Ausschnitte ausgewählt und mit einem eigens entwickelten Tool alle direkten Reden sowie die zugehörigen Sprecher und angesprochenen Figuren eingetragen. Jeder Text wurde von einem Annotator bearbeitet; eine zweite Annotation ist vorgesehen. Insgesamt wurden so 2264 direkte Reden mit Sprecher und Angesprochenen annotiert. Für die in Abschnitt 5 diskutierten Experimente wurde das Korpus in drei zufällige Mengen aufgeteilt:

Korpus	# Tokens	# Direkte Reden	# Romanfragmente
Trainingskorpus	107141	1185	37
Entwicklungskorpus	58709	615	20
Testkorpus	64330	464	20

Tab. 1: Überblick über die Auftrennung des in dieser Arbeit verwendeten Korpus.

Methoden

Wir verwenden regelbasierte Verfahren und maschinelle Lernverfahren, aber anders als in (He et al. 2013) oder (O’Keefe et al. 2012) dienen erstere nicht nur als Baseline-Verfahren, sondern wurden soweit wie möglich optimiert.

Wir verwenden die Techniken 2-Way Klassifikation und N-Way Klassifikation wie in (O’Keefe et al. 2012) vorgeschlagen. Zusätzlich evaluieren wir MaxEnt2WayToMatch, bei dem Kandidaten nur bis zum ersten tatsächlichen Sprecherkandidaten erzeugt werden.

Für die Sprecherzuordnung und Zuordnung eines Angesprochenen sind die in dieser Arbeit verwendeten Features in Tabelle A1 im Anhang zusammengefasst.

Für diese Aufgabe haben sich regelbasierte Verfahren als konkurrenzfähig mit den aktuellen ML-Verfahren

erwiesen. Sie besitzen außerdem den Vorteil, dass sie nicht so viele Trainingsbeispiele benötigen. Die Grundstruktur des Algorithmus ist der Idee des regelbasierten Koreferenzsystems von Stanford (Lee et al. 2011) angelehnt. Es werden eine Reihe von Regelpässen nacheinander ausgeführt. Die Regelpässe sind gemäß ihrer *Precision* geordnet, d. h. Regeln mit einer hohen *Precision* werden zuerst ausgeführt. Eine spätere Regel kann eine Entscheidung einer früheren Regel nicht revidieren. Tabelle A2 im Anhang zeigt die in dieser Arbeit verwendeten Regelpässe.

Mit Hilfe der Trainingsdaten konnte eine optimale Reihenfolge der Ausführung der Regeln empirisch ermittelt werden, bei der einige Regeln auch mehrfach angewendet werden.

(1)#(2)#(3)#(4)#(5)#(6)#(7)#(5)#(6)#(8)
#(9)#(5)#(6)#(7)#(10).

Evaluation

Die Parameter für die ML-Verfahren wurden auf dem Development-Anteil der Daten optimiert und anschließend gegen die Testmenge evaluiert. Für die regelbasierten Verfahren gibt es keine Unterscheidung zwischen Trainings- und Development-Korpus. Ein Sprecher gilt als korrekt bestimmt, wenn sich der vom System bestimmte Kandidat in der selben Koreferenzkette befindet, wie die Entität, die von unserem Annotator als korrekt markiert wurde. Tabelle 2 beschreibt die Ergebnisse bei der Anwendung der Verfahren auf das Testkorpus.

Verfahren	Sprecher Precision	Sprecher Recall	Sprecher Genauigkeit	Angesprochener Precision	Angesprochener Recall	Angesprochener Genauigkeit
MaxEnt2Way	72.4%	86.0%	62.3%	37.3%	86.2%	32.2%
MaxEntNWay	50.3%	97.6%	49.1%	38.5%	94.0%	36.2%
MaxEnt2WayToMatch	67.1%	81.3%	54.6%	32.3%	57.3%	18.5%
SVM2Way	58.8%	88.4%	52.0%	39.4%	37.1%	14.6%
CRFNWay	50.0%	96.6%	48.3%	35.8%	92.7%	33.2%
Regelbasiert (vorläufig)	79.6%	98.5%	78.4%	63.0%	94.2%	59.3%

Tab. 2: Ergebnisse der einzelnen Verfahren auf dem Testkorpus, bestehend aus 20 zufällig gewählten Romanfragmenten.

Unsere Experimente bestätigen die Aussagen von O’Keefe (O’Keefe et al. 2012), dass 2Way ML-Verfahren bessere Ergebnisse in der Sprechererkennung liefern, als korrespondierende NWay Verfahren. Analoges gilt für die Evaluation der CRFs, die sogar beinahe den selben Wert für die Sprechererkennung liefern wie in (O’Keefe et al. 2012). Sowohl auf dem Developmentkorpus, als auch auf dem Testkorpus zeigen regelbasierte Ansätze deutliche Vorteile gegenüber den in dieser Arbeit verwendeten ML-Verfahren. Es ist weiterhin ersichtlich, dass die Bestimmung des Sprechers einfacher ist, als die Bestimmung des Angesprochenen. Wahrscheinlich liegt das daran, dass im Fall der Sprecherzuschreibung mehr Information vorliegt, nämlich die direkte Rede und der Kontext, während bei der Ermittlung des Angesprochenen

die direkte Rede selbst nur hilfreich ist, wenn ein Angesprochener direkt darin vermerkt ist.

Ein direkter Vergleich mit dem besten in der Literatur zu findenden Verfahren (Almeida et al. 2014) kann direkt nicht durchgeführt werden. Berücksichtigt man den Unterschied, der Verfahren von O'Keefe auf den Texten des WSJ und den literarischen Texten, könnte eine Qualität von 90% Genauigkeit erreicht werden und damit ein mit der state of the art vergleichbares, sogar möglicherweise besseres Ergebnis. Im Gegensatz zu ihrem Verfahren ermitteln wir zudem auch noch eine angesprochene Entität.

Diskussion und Ausblick

Die Ergebnisse zeigen, dass das regelbasierte Verfahren für diese Aufgabe deutlich bessere Ergebnisse erzielen kann als alle ML-Verfahren, die in dieser Arbeit getestet wurden. Es ist geplant, die hier erstellte Zuordnung in die regelbasierte Koreferenzauflösung von (Krug et al. 2015) einzuarbeiten, um diese damit zu verbessern. Weil unsere Hauptmotivation die Verbesserung der Koreferenzresolution ist, diese aber im Ansatz von Almeida nicht wirksam verbessert werden konnte, haben wir darauf verzichtet, deren komplexes Lernverfahren nachzuvollziehen. Gerade die Ergebnisse, die in Tabelle 2 zu sehen sind, zeigen, dass mögliche Dialogsequenzen genauer untersucht werden müssen, um diese zuverlässig erkennen und auflösen zu können. Eine genaue Dialoganalyse vereinfacht wiederum die Koreferenzauflösung, so dass eine Extraktion von Beziehungen zwischen Personen und Attributen zu Entitäten innerhalb der Romane möglicher erscheint.

Anhang

Featurebeschreibung (zwischen Kandidat und direkten Rede)	Verwendung für Sprecherzuordnung	Zuordnung des Angesprochenen
1. Ist der Kandidat Subjekt	+	-
2. Das Verb in der Dependenzstruktur, auf das sich der Kandidat bezieht	+	+
3. Das POS-Tag des Kandidaten	-	-
4. Ist der Kandidat ein Pronomen	-	-
5-6. Befindet sich der Kandidat im Akkusativ/Dativ	+/+	+/+
7. Kandidat befindet sich in einer direkten Rede	+	-
8. Kandidat erscheint in der aktuellen direkten Rede	-	-
9. Kandidat befindet sich im selben Satz wie die direkte Rede	+	-
10. Die Direkte Rede beginnt mit einem kleingeschriebenem Wort	+	-
11. Zwischen Kandidat und direkter Rede befindet sich ein Doppelpunkt	-	-
12-14. Distanz zw. Kandidat und direkter Rede in Sätze/Wörter/Entitäten	-/-/+	-/-/-
15-16. Wort an Position +1/-1	-/-	-/-
17-18. Wort an Position +1/-1 ist Satzzeichen	+/+	-/-
19-20. Wort an Position +1/-1 ist in direkter Rede	+/+	-/-
19-20. Kandidat ist Sprecher der direkten Rede an Position -1/-2	-/-	+/+
21-22. Kandidat ist Angesprochener der	-/-	-/-

direkten Rede an Position -1/-2		
------------------------------------	--	--

Tab. A1: Ein Überblick über die in dieser Arbeit verwendeten Features. Durch + und - ist angegeben, ob dieses Feature gewinnbringend eingesetzt werden konnte. Zur Wahl der Features vgl. auch (Elson / McKeown 2010) und (He et al. 2013).

Regelbezeichnung	Regelbeschreibung
(1) Explizite Sprechererkennung	Nutzt Pattern-Matching und grammatikalische Regeln um explizite Erwähnungen eines Sprechers im direkten Umfeld einer direkten Rede zu erkennen.
(2) Explizite Erkennung des Angesprochenen	Nutzt Pattern-Matching und grammatikalische Regeln um explizite Erwähnungen eines Angesprochenen innerhalb der direkten Rede zu erkennen.
(3) Explizite Erkennung des Angesprochenen II	Wie (1) nur für den Angesprochenen
(4) Explizite Sprechererkennung II	Wie (1), nur der Kontext wird um 1 Satz außerhalb der direkten Rede erweitert.
(5) Vorwärtspropagierung	Zwei direkt aufeinanderfolgenden direkten Reden wird der Sprecher/ Angesprochener der ersten direkten Rede zugeordnet, wenn beide direkte Reden innerhalb des selben Satzes liegen
(6) Rückwärtspropagierung	wie (5) nur mit entgegengesetzter Richtung der Ausführung
(7) Nachbarschaftspropagierung	Direkten Reden, die keinen eingeschobenen Kontext aufzeigen, wechseln den Sprecher/Angesprochenen (falls vorhanden)
(8) Fragenpropagierung	Nach einer Frage wechseln Sprecher/Angesprochener
(9) Dialogpropagierung	Direkte Rede mit maximal einem zwischenliegenden Satz wechseln ihren Sprecher/ Angesprochenen
(10) Default-Sprecher/ Angesprochener	Als Sprecher wird das letzte Subjekt außerhalb direkter Reden gesetzt, als Angesprochener das letzte Subjekt, das nicht Sprecher der aktuellen direkten Rede ist.

Tab. A2: Überblick über die Regelpäse für das in dieser Arbeit vorgestellte regelbasierte Verfahren zur Sprecherzuordnung bzw. Zuordnung eines

Angesprochenen. Optimale Reihenfolge der Ausführung der Regeln aufgrund der Auswertung des Trainingsatzes:

(1)#(2)#(3)#(4)#(5)#(6)#(7)#(5)#(6)#(8)
#(9)#(5)#(6)#(7)#(10).

Das Diagramm zeigt fünf nummerierte Beispiele (1-5) aus dem Text 'Lydia'. Über dem Text sind Sprecher- und Angesprochenen-Attributionen (Sprecher: 1, 2, 3, 4; Angesprochenen: 1, 2, 3, 4) in verschiedenen Farben markiert. Die Beispiele sind:

- «Glaubst Du denn, daß es Landsfeld nicht angenehm sein würde, wenn Du allein hinführest? fragte die Forsträthin.»
- «Du kannst ja Gertrud mitnehmen.»
- «O, ich fürchte mich nicht, liebe Mutter. Und Richard hat mich ja selbst zu Besuchen aufgefordert. Auch schreibt mir Therese, daß am Bahnhofe ihr Wagen mich erwarten werde, und daß sie darauf rechne, daß ich die Nacht bei ihr bleiben werde. Aber gerade das möchte ich nicht gern.»
- «Ich sehe wirklich keinen Grund, warum Du diese freundliche Bitte ablehnen willst, liebes Kind. Ich bin ganz wohl, wie Du siehst, Landsfeld kommt auch erst morgen.»
- «Wahrscheinlich – verbesserte Lydia.»

Abb. A3: Auszug aus Aston Louise "Lydia": Beispiel für die Erkennung von Sprecher und Angesprochenem in direkten Reden gemäß den Regeln in Tabelle A2. Im ersten Durchlauf wird mit der Regel (1) die Sprecherin für die direkten Reden 1 und 5 erkannt. Anschließend erkennt Regel (7) in Rückwärtsrichtung jeweils abwechselnd Sprecherin 4 und 2 und Angesprochene in 3. Schließlich erkennt Regel (7) in Vorwärtsrichtung die Sprecherin in 3 und die Angesprochene in 4 und 2.

Bibliographie

Almeida, Mariana S.C. / Almeida, Miguel B. / Martins, André F.T. (2014): "A joint model for quotation attribution and coreference resolution", in: *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics, Gothenburg, Sweden* 39-48.

Bohnet, Bernd / Kuhn, Jonas (2012): "The best of both worlds: a graph-based completion model for transition-based parsers." In: *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*. Avignon, France: 77-87.

Chaganty, Arun / Muzny, Grace (2015): *Quote Attribution for Literary Text with Neural Networks* <https://cs224d.stanford.edu/reports/ChagantyArun.pdf> [letzter Zugriff 08. Februar 2016].

Celikyilmaz, Asli / Hakkani-Tur, Dilek / He, Hua / Kondrak, Greg / Barbosa, Denilson (2010): "The actortopic model for extracting social networks in literary narrative.", in: *Proceedings of the NIPS 2010 Workshop Machine Learning for Social Computing* <https://webdocs.cs.ualberta.ca/~denilson/files/publications/nips2010.pdf> [letzter Zugriff 08. Februar 2016].

Elson, David K. / Dames, Nicholas / McKeown, Kathleen R. (2010a): "Extracting social networks from literary fiction", in: *Proceedings of the 48th annual meeting of the association for computational linguistics*. Association for Computational Linguistics <http://www1.cs.columbia.edu/~delson/pubs/>

ACL2010-ElsonDamesMcKeown.pdf [letzter Zugriff 08. Februar 2016].

Elson, David K. / McKeown, Kathleen R. (2010b): "Automatic Attribution of Quoted Speech in Literary Narrative", in: *Proceedings of AAAI* 1013-1019.

Glass, Kevin / Bangay, Shaun (2006): "Hierarchical rule generalisation for speaker identification in fiction books", in: *Proceedings of the 2006 annual research conference of the South African institute of computer scientists and information technologists on IT research in developing countries*. South African Institute for Computer Scientists and Information Technologists: 31-40.

Glass, Kevin / Bangay, Shaun (2007): "A naive salience-based method for speaker identification in fiction books", in: *Proceedings of the 18th Annual Symposium of the Pattern Recognition Association of South Africa (PRASA'07)* <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.494.3729&rep=rep1&type=pdf> [letzter Zugriff 16. Februar 2016].

He, Hua / Barbosa, Denilson / Kondrak, Grzegorz (2013): "Identification of Speakers in Novels", in: *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*. Sofia, Bulgaria: 1312-1320.

Iosif, Elias / Mishra, Taniya (2014): "From Speaker Identification to Affective Analysis: A Multi-Step System for Analyzing Children's Stories", in: *EACL* 2014: 40-49.

Jannidis, Fotis / Krug, Markus / Reger, Isabella / Toepfer, Martin / Weimer, Lukas / Puppe, Frank (2015): "Automatische Erkennung von Figuren in deutschsprachigen Romanen", in: *Digital Humanities im deutschsprachigen Raum (Dhd 2015)*, Graz, Austria.

Joachims, Thorsten (2002): *Learning to classify text using support vector machines*. Methods, theory and algorithms (= The Springer International Series in Engineering and Computer Science 668). New York: Springer.

Krug, Markus / Puppe, Frank / Jannidis, Fotis / Macharowsky, Luisa / Reger, Isabella / Weimer, Lukas (2015): "Rule-based Coreference Resolution in German Historic Novels", in: *Proceedings of the Fourth Workshop on Computational Linguistics for Literature* 98-104.

Lee, Heeyoung / Peirsman, Yves / Chang, Angel / Chambers, Nathanael / Surdeanu, Mihai / Jurafsky, Dan (2011): "Stanford's multi-pass sieve coreference resolution system at the CoNLL-2011 shared task", in: *Proceedings of the Fifteenth Conference on Computational Natural Language Learning: Shared Task*. Association for Computational Linguistics <http://nlp.stanford.edu/pubs/conllst2011-coref.pdf> [letzter Zugriff 08. Februar 2016].

McCallum, Andrew Kachites (2002): *MALLET: A Machine Learning for Language Toolkit* <http://mallet.cs.umass.edu> [letzter Zugriff 08. Februar 2016].

Mikolov, Tomas / Sutskever, Ilya / Chen, Kai / Corrado, Greg / Dean Jeffrey (2013): "Distributed representations of words and phrases and their compositionality", in: *Advances in neural information processing systems* 26 <http://papers.nips.cc/paper/5021-distributed->

[representations-of-words-and-phrases-and-their-compositionality.pdf](#) [letzter Zugriff 08. Februar 2016].

O'Keefe, Tim / Pareti, Silvia / Curran, James R. / Koprinska, Irena / Honnibal, Matthew (2012): "A sequence labelling approach to quote attribution", in: *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*. Association for Computational Linguistics, 2012: 790-799.

Rahman, Altaf / Ng, Vincent (2011): "Narrowing the modeling gap: a cluster-ranking approach to coreference resolution", in: *Journal of Artificial Intelligence Research* 40: 469-521.

Ruppenhofer, Josef / Sporleder, Caroline / Shirokov, Fabian (2010): "Speaker Attribution in Cabinet Protocols", in: *The seventh international conference on Language Resources and Evaluation (LREC)* 2510-2515.

Schmid, Helmut (1999): "Improvements in part-of-speech tagging with an application to German", in: Armstrong, Susan / Church, Kenneth / Isabelle, Pierre / Manzi, Sandra / Tzoukermann, Evelyne / Yarowsky, David (eds.): *Natural language processing using very large corpora* (= Text, Speech and Language Technology 11). New York: Springer 13-25.

Schmid, Helmut / Laws, Florian (2008): "Estimation of conditional probabilities with decision trees and an application to fine-grained POS tagging", in: *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)* 777-784.

Sutton, Charles / McCallum, Andrew (2006): "An introduction to conditional random fields for relational learning", in: Getoor, Lise / Taskar, Ben (eds.): *Introduction to statistical relational learning*. Cambridge, MA / London: The MIT Press 93-128.

Zhang, Jason Y. / Black Alan W. / Sproat, Richard (2003): "Identifying speakers in children's stories for speech synthesis", in: *EUROSPEECH* 2041-2044.