

# Strahlt die negative Reputation von Häsitationsmarkern auf ihre Sprecher aus?

Kerstin Fischer<sup>1</sup>, Oliver Niebuhr<sup>2</sup>, Eszter Novák-Tót, Lars C. Jensen<sup>1</sup>

<sup>1</sup> Institut für Design und Kommunikation, Syddansk Universitet; kerstin@sdu.dk, larscj@sdu.dk

<sup>2</sup> Mads-Clausen-Institut, Syddansk Universitet; olni@sdu.dk

## Einleitung

Häsitationsmarker, d.h. Vokalisierungen wie *äh* und *ähm*, auch ‚gefüllte Pausen‘ genannt, werden oft als etwas betrachtet, das es zu vermeiden gilt (z.B. O’Connell & Kowall 1983). Sie gehören zu den Signalen, die in der Interaktion eine große Rolle spielen können, die wir aber selten direkt bzw. bewusst steuern und die eher als Nebeneffekte von anderen Prozessen auftreten. Im folgenden argumentieren wir dafür, dass Häsitationsmarker trotz ihrer negativen Reputation in der Wahrnehmung des Sprechers eher positive Effekte haben.

In einer Befragung von 105 Studierenden zeigt Fox Tree (2002), dass zwar die meisten Sprecher angeben, Häsitationsmarker entweder manchmal oder oft zu verwenden, dass aber 36% davon ausgehen, dass sie die Kommunikation stören, und nur 4% ihnen unterstützende Funktionen zuschreiben; 73% geben an, zu versuchen, sie zu vermeiden. Ihr Funktionsspektrum wird mit Planungsprozessen (z.B. Zeit zum Nachdenken), mit Nervosität und sogar mit Inkompetenz in Verbindung gebracht. Wenn der Wahrheitsgehalt einer Äußerung in Frage steht, werden Häsitationsmarker auch als Indikatoren für Unehrlichkeit interpretiert (Fox Tree 2002; Corley & Steward 2008).

Die Interpretation von Häsitationsmarkern als Nebeneffekte von Sprachproduktionsproblemen wird auch vielfach in der psychologischen und psycholinguistischen Literatur als Ausgangspunkt genommen. Zum Beispiel gibt es messbare Korrelationen zwischen kognitiver Komplexität und der Häufigkeit von Häsitationsmarkern (z.B. Christenfeld 1994) sowie zwischen kognitiver Belastung und Stress und Häsitationsmarkern (Berthold & Jameson 1999). Lounsbury (1954) argumentiert, dass sie gehäuft an Punkten statistischer Unsicherheit auftreten. In frühen Studien zu Häsitationsmarkern werden sie als Fenster in Sprachplanungsprozesse verstanden; so finden sie Reynolds & Paivio (1968) besonders vor abstrakten Begriffen, Goldman-Eisler (1958) häufiger vor Inhaltswörtern, Howell & Vetter (1969) vor Inhalts- und Funktionswörtern, was auf Funktionen in der Wortsuche hindeuten würde, während andere Autoren sie vor allem an grammatischen Grenzen finden (Boomer 1965; Hawkins 1971), was für grammatische Planungsprozesse sprechen würde.

Gleichzeitig gibt es aber auch Studien, die die kommunikativen Funktionen von Häsitationsmarkern hervorheben. Zum Beispiel argumentieren Maclay & Osgood (1959), dass Häsitationsmarker Funktionen in Hinblick auf das Turn-Taking-System erfüllen, und Todd (1981) zeigt, dass initiale Vorkommen den Sprecher darin unterstützen, einen sprachlichen Rhythmus beizubehalten. Dazu kommen Funktionen in der Konversation, zum

Beispiel in dispräferierten Antworten (Levinson 1983; Schegloff 2010), wo sie zur Imagearbeit beitragen und Gesichtsverlust vermeiden helfen.

In der vorliegenden Arbeit stellen wir die Frage, wie SprecherInnen wahrgenommen werden, wenn sie Häsitationsmarker verwenden. Angesichts der schlechten Reputation von Häsitationsmarkern müssten HörerInnen solche SprecherInnen, die Häsitationsmarker verwenden, als weniger kompetent, ehrlich und konzentriert ansehen. Demgegenüber stehen die Studien zu den kommunikativen Funktionen von Häsitationsmarkern, die letztere in ein positives Licht rücken und folglich erwarten lassen, dass SprecherInnen mit Häsitationsmarkern als höflicher und freundlicher empfunden werden.

Da menschliche Sprecher nicht so manipuliert werden können, dass sie exakt dieselben Äußerungen in der Interaktion produzieren, nur einmal mit und einmal ohne Häsitationsmarker, verwenden wir in unserer ersten Studie einen Roboter als Interaktionspartner. In der zweiten Studie untersuchen wir den Zusammenhang zwischen der Einschätzung von Sprechern als charismatisch und der Häufigkeit von Häsitationsmarkern in ihrer Sprache.

## Studie 1

In dieser Studie verwenden wir einen Online-Fragebogen mit einem *between-subject*-Design, so dass die Hälfte der TeilnehmerInnen Videos von Interaktionen zwischen Menschen und einem Roboter sah, die Häsitationsmarker enthielten, während die andere Hälfte dieselben Videos sah, aber in der Sprache des Roboters keine Häsitationsmarker vorkamen. Die TeilnehmerInnen wurden den beiden Videobedingungen vollständig zufällig maschinell zugewiesen.

Die Interaktionen, die die TeilnehmerInnen der Studie sahen, zeigen Szenarios, in denen der Roboter eine Person in einer Aktion anleitet, zum Beispiel in den Regeln eines Brettspiels. Die Roboteräußerungen wurden mit einem frei verfügbaren *Text-to-Speech* System synthetisiert, basierend auf "Sidney", "Peter", und "Eddie", den Microsoft-SAPI-4-Stimmen männlicher Sprecher des amerikanischen English mittleren Alters. Ein Beispiel für einen Text für den Roboter ist der folgende:

okay, now we only need to finish the roof. *Um* you put the round red block above the doors.

*uh* no, one further right.

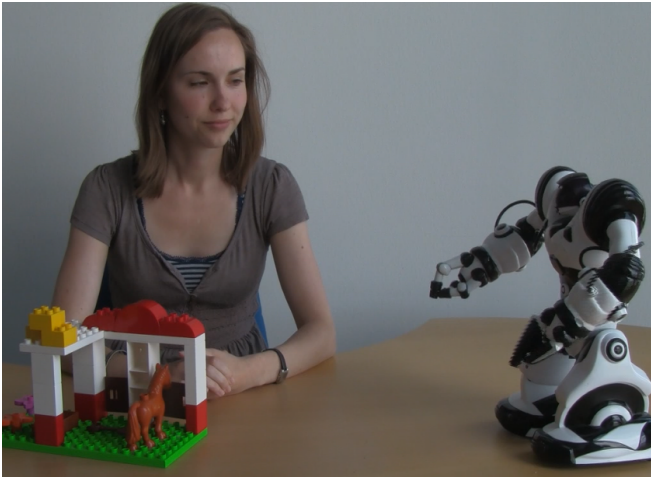
and you attach one of the red square blocks to the *uh* left corner.

okay, now we only need to finish the roof. You put the round red block above the doors.

no, one further right.

and you attach one of the red square blocks to the left corner.

897 TeilnehmerInnen, die über die Online-Plattform Crowdfunder (<https://www.crowdfunder.com/>) rekrutiert wurden, nahmen an der Studie teil. Allerdings mussten viele TeilnehmerInnen nachträglich aus der Analyse ausgeschlossen werden, da ihre Antwortzeit unter der Abspieldauer der Videos lag. Das heißt, die ausgeschlossenen Probanden hatten sich die Videostimuli nicht vollständig angesehen, sondern schon vorzeitig geantwortet. Die Ergebnisse, die wir unten berichten, beziehen sich daher auf 448 TeilnehmerInnen in der Videobedingung mit Roboterhäsitationen und 459 TeilnehmerInnen in der Videobedingung ohne Roboterhäsitationen.



**Abbildung 1:** Beispielsszenario für die Mensch-Roboter-Interaktionen: Der Roboter erklärt, wie der Stall aus Lego zusammengebaut wird.

Zwei Versionen des Fragebogens wurden hergestellt, so dass jede TeilnehmerIn nur eine Version jedes Videos zu sehen bekam. Nach jedem Video wurden die TeilnehmerInnen um eine Bewertung des Roboters auf mehreren semantischen Likert-Skalen gebeten. Konkret wurden folgende Eigenschaften abgefragt:

- How competent do you think is the robot?
- How certain do you think is it?
- How pleasant do you find it?
- How polite do you think is it?
- How important do you think was it for the robot to get the recipe/the instructions across well?
- How honest do you think it is?
- Would you like to have such a robot?

Der Bewertungsbereich erstreckte sich von 1 (trifft nicht zu) bis 7 (trifft voll zu). T-tests für unabhängige Stichproben wurden verwendet, um die Bewertungen (d.h. Skalenwerte) der TeilnehmerInnen auf Unterschiede zwischen den beiden Videobedingungen hin zu überprüfen. Die Resultate dieser t-Tests zeigen keinerlei Effekte für Kompetenz und Unsicherheit ( $p > .29$ ), siehe Tabelle 1 für Dialog 1. Stattdessen hatte die Verwendung von Häsitationen einen signifikanten Einfluss darauf, wie angenehm und höflich der Roboter wahrgenommen wurde. Benutzte der Roboter Häsitationen in der Interaktion mit dem Menschen,

empfanden Betrachter dieser Roboter-Mensch-Interaktion den Roboter als angenehmer und höflicher. Außerdem gab es einen signifikanten Trend ( $p < 0.1$ ) dahingehend, dass die Betrachter den Roboter mit Häsitationen lieber besitzen wollten als den ohne Häsitationen.

**Tabelle 1:** Resultate für Dialog 1 (Mittelwert, Standardabweichung, Signifikanzniveau)

	TTS voice: "Sidney", American English				
	n=448		n=459		p
	MW mit H.	$\sigma$	MW ohne H.	$\sigma$	
kompetent	4.72	1.36	4.79	1.50	0.492
sicher	4.75	1.37	4.85	1.48	0.291
angenehm	4.70	1.38	4.51	1.50	0.044
höflich	4.91	1.33	4.71	1.46	0.027
adressantenorientiert	5.22	1.35	5.14	1.42	0.417
ehrlich	4.95	1.47	4.90	1.54	0.609
Wunsch zu besitzen	4.67	1.84	4.46	1.85	0.086

In ihrer Generalisierbarkeit eingeschränkt werden diese Befunde dadurch, dass die besagten signifikanten Effekte und Trends nicht durchgängig für alle unsere Dialoge nachgewiesen werden konnten. Ein Grund dafür kann in der Schwierigkeit liegen, Häsitationsmarker zu synthetisieren (siehe Betz et al. 2015). Auffallend ist nämlich, dass die signifikanten Effekte und Trends nicht für die Dialoge mit der Stimme "Eddie" auftraten. Ein auditiver Vergleich der Stimulusgruppen beider Stimmen durch die Versuchsleiter hat in der Tat ergeben, dass die synthetischen Häsitationen von "Eddie" sowohl in ihren lautlichen als auch in ihren prosodischen Eigenschaften deutlich künstlicher klangen und sich schlechter in die angrenzenden Äußerungen einfügten als die Häsitationen von "Sidney" und "Peter". Weitere Studien sind nötig, um dieses erste auditive Urteil empirisch zu objektivieren und zu untermauern.

Davon abgesehen zeigt unsere Studie 1 in Übereinstimmung mit den Studien, die die pragmatische Funktion in den Vordergrund stellen, dass Häsitationsmarker einen positiven Einfluss auf die Wahrnehmung des Sprechers und dessen interpersonale Beziehung zum Gesprächspartner haben.

## Studie 2

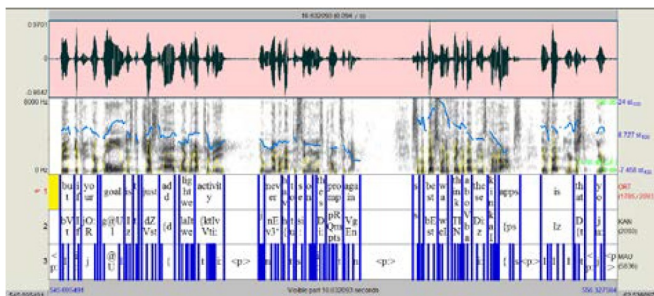
In der zweiten Studie analysieren wir die Sprache von fünf bekannte Persönlichkeiten aus der Wirtschaft in Hinblick auf ihre Verwendung von Häsitationsmarkern. Das Ziel dieser Untersuchung ist es, die Vorkommen von Häsitationsmarkern mit einer unabhängigen Einschätzung der SprecherInnen als mehr oder weniger charismatisch zu korrelieren.

Die ausgewählten SprecherInnen sind:

- Meg Whitman (H&P), die für Ihren Stil kritisiert wurde ("brash and abrasive", Samson, 2011). Sie sagt über sich selbst, sie sei "frumpy, but delivers" (Nadler 2011).

- Oprah Winfrey, die oft als charismatisch beschrieben wird (z.B. Restauri 2012)
- Ginni Rometty (IBM), der als moderner CEO und als charismatischer Führer beschrieben wird (bringing "charisma and raw leadership" to the table, siehe, z.B. Langley 2015).
- Steve Jobs (Apple), angeblich einer der charismatischsten Sprecher unserer Zeit ("One of the most charismatic speakers of the last decades", siehe Conger 1989, Niebuhr et al. 2016).
- Mark Zuckerberg (Facebook), über den gesagt wird: "Zuckerberg does not have that [= Jobs'] charisma, and the presentation skills are rough enough to impact Facebook's perception in a negative way." (siehe Niebuhr et al. 2016)

Von jedem dieser SprecherInnen wurden ca. 20 Minuten aus öffentlichen Reden, meist Produktpräsentationen, auf Vorkommen von Häsitationsmarkern untersucht. Die Untersuchung basierte auf vorhandenen, für eine vergleichende akustische Analyse der SprecherInnen angefertigten Annotationen der Sprachsignale in PRAAT-TextGrid-Dateien. Abbildung 2 zeigt einen Ausschnitt aus der Signaldatei von Mark Zuckerberg. In den Annotationen, die auf auditiven Analysen ausgebildeter PhonetikerInnen basieren, wurde zwischen primär vokoiden Häsitationen (z.B. "eeh"), primär kontoiden Häsitationen (z.B. "ehmm" oder "mmh") und "stillen Häsitationen" differenziert. Still bedeutet hier, dass eine Pause von mindestens 200 Millisekunden Länge innerhalb einer syntaktisch unvollständigen, ansonsten aber intakten Konstruktion vorhanden ist, die entsprechend nicht repariert wird (also keinen Abbruch darstellt) und die zudem nicht an den Rändern eines Turns steht.

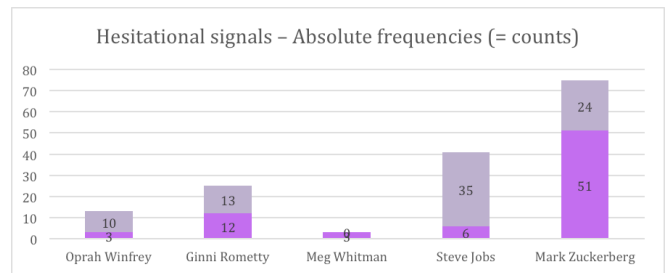


**Abbildung 2:** Ausschnitt aus einer Kombination von Sprachsignal (Oszillogramm und Spektrogramm, oben) und Annotationsdatei (unten) des Sprechers Mark Zuckerberg.

Die auditiv identifizierten und differenzierten Häsitationen wurden pro Sprecher ausgezählt. Abbildung 3 zeigt die absoluten Häufigkeiten der Häsitationen pro Sprecher, wobei vokoide und kontoide Häsitationen in einer Gruppe zusammengefasst sind (da sich für diese beiden Unterarten keine auffälligen Befunde ergeben haben). Absolute Häufigkeiten sind in diesem Fall ebenso aussagekräftig wie relative Häufigkeiten, weil pro Sprecher eine zeitlich annähernd gleich große Menge Sprachmaterial ausgewertet wurde.

Ein Chi-Quadrat-Test zeigt deutlich, dass sich die Häsitationen nicht gleichmäßig über die SprecherInnen verteilen, sondern signifikante Abweichungen in Typ und Frequenz

der Häsitationen zwischen den SprecherInnen bestehen ( $\chi^2[4]=43.037$ ,  $p<0.0001$ ).



**Abbildung 3:** Absolute Vorkommen von Häsitationsmarkern bei den fünf analysierten SprecherInnen (grau: stille Häsitationen, blau: gefüllte Pausen in Gestalt von Häsitationsmarkern primär vokoiden oder kontoiden Typs)

Die signifikanten Abweichungen bestehen erstens in einem Geschlechtsunterschied. Die drei weiblichen Sprecher haben allesamt weniger Häsitationsmarker benutzt als ihre männlichen Kollegen. Darüber hinaus zeigt sich zweitens eine unterschiedliche Verteilung in Abhängigkeit des der betreffenden Person zugeschriebenen Charismas: diejenigen SprecherInnen, die als besonders charismatisch eingeschätzt werden, zeichnen sich durch eine moderate Verwendung von Häsitationsmarkern aus. Moderat heißt für unsere Daten, etwa 2 Häsitationen pro Minute zu produzieren (Steve Jobs), mindestens aber ca. alle 90 Sekunden eine (Oprah Winfrey). Im Unterschied dazu fallen die beiden als nicht besonders charismatisch eingeschätzten SprecherInnen, Meg Whitman und Mark Zuckerberg, dadurch auf, dass sie entweder fast keine oder sehr viele Häsitationsmarker verwenden.

Neben der reinen Häufigkeit der Häsitationen gibt es auch Unterschiede in deren Dauer zwischen den SprecherInnen. So produzierte Steve Jobs innerhalb der Dreiergruppe charismatischerer SprecherInnen zwar die meisten Häsitationen, allerdings waren diese gemäß t-Tests für unabhängige Stichproben auch signifikant kürzer als die von Oprah Winfrey ( $t[52]=3.625$ ,  $p<0.001$ ) und Ginni Rometty ( $t[64]=2.991$ ,  $p<0.001$ ). Mark Zuckerberg hebt sich von den drei charismatischeren SprecherInnen nicht nur durch die meisten, sondern auch durch signifikant längere Häsitationen ab ( $t[114]=5.124$ ,  $p<0.001$ ). Mit durchschnittlich knapp 400 Millisekunden sind sie fast 50% länger als die von Steve Jobs.

## Schlussfolgerungen

Die beiden Studien haben gezeigt, dass Häsitationsmarker ihre extrem schlechte Reputation nicht verdienen. Stattdessen können sie kommunikative Funktionen erfüllen und zu Höflichkeit und Imagearbeit beitragen. Darüber hinaus müssen sie trotz ihrer schlechten Reputation die Wahrnehmung ihres Sprechers nicht unbedingt negativ beeinträchtigen. Ganz im Gegenteil legen die Ergebnisse beider Studien eindeutig nahe, dass SprecherInnen auf Grund von Häsitationsmarkervorkommen nicht als weniger höflich, angenehm oder charismatisch eingeschätzt werden. Selbst der vielfach als einer der charismatischsten Sprecher unserer Zeit betitelte Steve Jobs hat in seinen vielbeachteten Vorträgen jede Minute annähernd 2 Häsitationen produziert. Allerdings scheinen sowohl übermäßig viele als auch

außerordentlich wenige Häsitationen einen negativen Effekt auf die Wahrnehmung des Sprechers haben zu können. Das heißt zusammengefasst: Die Dosis macht das Gift. In richtiger Dosierung können Häsitationen das Salz in der Suppe sein.

An diesem Punkt überschneiden sich die Implikationen unserer beider Studien zudem darin, dass die Dosierung sich auch darin ausdrückt, wie gut sich die Häsitationen in die Rede des Sprechers einfügen. Werden sie aufgrund ungewöhnlicher Lautqualitäten, ihrer schieren Häufigkeit, großer Dauern oder anderer (unnatürlicher) Eigenschaften zu salient, können sie negative Konsequenzen für den Sprecher haben. Näher zu untersuchen, inwieweit auch noch andere Parameter eine Rolle spielen und wenn ja, welche, sollte in unseren Augen eine entscheidende Rolle in zukünftigen Studien und nicht zuletzt auch in der Synthese künstlicher Häsitationen in der Mensch-Maschine-Interaktion spielen.

## Literatur

- [1] Beattie, Geoffrey W. (1979): Planning Units in Spontaneous Speech: Some Evidence from Hesitation in Speech and Speaker Gaze Direction in Conversation. *Linguistics* 17: 61-78.
- [2] Berthold, André & Jameson, Anthony (1999): Interpreting Symptoms of Cognitive Load in Speech Input. In: Kay, J. (ed.): *User Modeling: Proceedings of the Seventh International Conference UM99*. Vienna, New York: Springer.
- [3] Betz, Simon, Wagner, Petra & Schlangen, David (2015): Micro-Structure of Disfluencies: Basics for Conversational Speech Synthesis. *Interspeech* 2015.
- [4] Boomer, Donald (1965): Hesitation and Grammatical Encoding. *Language and Speech* 8: 148-158.
- [5] Christenfeld, Nicholas (1994): Options and Ums. *Journal of Language and Social Psychology* 13, 2: 192-199.
- [6] Clark, Herbert H. & Fox Tree, Jean E. (2002): Using *uh* and *um* in Spontaneous Speaking. *Cognition* 84: 73-111.
- [7] Conger, J. A. (1989). The charismatic leader: Behind the mystique of exceptional leadership. Jossey-Bass.
- [8] Corley, M., & Stewart, O. W. (2008). Hesitation Disfluencies in Spontaneous Speech: The Meaning of *um*. *Language and Linguistics Compass*, 2(4), 589-602.
- [9] Fox Tree, Jean E. (2002): Interpreting Pauses and Ums at Turn Exchanges. *Discourse Processes* 34,1: 37-55.
- [10] Goldmann-Eisler, Frieda (1958): Speech Production and the Predictability of Words in Context. *Quarterly Journal of Experimental Psychology* 10, 96.
- [11] Hawkins, P.R. (1971): The Syntactic Location of Hesitation Pauses. *Language and Speech* 14: 277-288.
- [12] Howell, Richard W. & Vetter, Harold J. (1969): Hesitation in the Production of Speech. *The Journal of General Psychology* 81: 261-276.
- [13] Kraut, Robert E. (1978): Verbal and Nonverbal Cues in the Perception of Lying. *Journal of Personality and Social Psychology* 36, 4: 380-391.
- [14] Lalljee, Mansur & Cook, Mark (1974): Filled Pauses and Floor Holding: The Final Test? *Semiotica* 12: 219-225.
- [15] Langley, M. (2015, April 20). Behind Ginni Rometty's Plan to Reboot IBM. The Wall Street Journal. <http://www.wsj.com/articles/behind-ginni-romettys-plan-to-reboot-ibm-1429577076>
- [16] Levinson, Stephen (1983): *Pragmatics*. Cambridge: Cambridge University Press.
- [17] Lounsbury, F.G. (1954): Transitional probability, Linguistic Structure, and Systems of Habit-family Hierarchies. In: Osgood, C.E. & Sebeok, T.A. (eds.): *Psycholinguistics: A Survey of Theory and Research Problems*. Baltimore: Williams & Williams, S.93-101.
- [18] Maclay, Howard and Osgood, Charles E. (1959): Hesitation Phenomena in Spontaneous English Speech. *Word* 1: 19-44.
- [19] Martin, James G. (1967): Hesitations in the Speaker's Production and Listener's Reproduction of Utterances. *Journal of Verbal Learning and Verbal Behavior* 6: 603-609.
- [20] Martin, James G. (1971): Some Acoustic and Grammatical Features of Spontaneous Speech. In: Horton, D. and Jenkins, J. (eds.): *Perception of Language*. Columbus, OH: Charles E. Merrill.
- [21] Nadler, R. (2011, September). Will Meg be Magic for HP? Meg Whitman's Emotional Intelligence leadership under the Spotlight. Psychology Today. <https://www.psychologytoday.com/blog/leading-emotional-intelligence/201109/will-megbe-magic-hp-0>
- [22] Niebuhr, O., Brem, A., Novák-Tót, E., & Voße, J. (2016). Charisma in business speeches: A contrastive acoustic-prosodic analysis of Steve Jobs and Mark Zuckerberg. Proc. 8th International Conference of Speech Prosody, Boston, USA, 79-82.
- [23] O'Connell, Daniel C. & Kowal, Sabine (1983): Pausology. In: Sedelow, Walter A. & Sedelow, Sally Yeates (eds.): *Computers in Language Research 2*. Trends in Linguistics - Studies and Monographs 19, Berlin, New York: de Gruyter, pp. 221-301.
- [24] Schegloff, Emanuel A. (2010): Some Other "Uh(m)"s. *Discourse Processes*, 47(2), 130-174.
- [25] Restauri, D. (2012, May 3) 5 Qualities of Charismatic People. How Many Do You Have?. *Forbes*. <http://www.forbes.com/sites/deniserestauri/2012/05/03/5-qualities-of-charismatic-people-how-many-do-you-have/#3f853d42372d>
- [26] Swerts, Mark (1998): Filled Pauses as Markers of Discourse Structure. *Journal of Pragmatics* 30: 485-496.
- [27] Todd, Dietmar (1981): Zum Auftreten von Fülllauten in spontan gesprochenen Berichten. *Nova Acta Leopoldina N.F.* 54, 245: 597-611.