

Seminar Sommersemester 2010: Sprachgenerierung

Michael Strube
Heidelberger Institut für Theoretische Studien
Schloss-Wolfsbrunnenweg 35
69118 Heidelberg
<http://michael.kimstrube.de>

5. Juli 2010

Termine, Themenvorschläge

22.04.2010

Überblick Sprachgenerierung: Architektur, Diskursplanung, Mikroplanung, Oberflächenrealisierung, Terminologie: Reiter & Dale (2000), Vander Linden (2000), Melish et al. (2004), Horacek (2004a)

29.04.2010

Diskursplanung: Diskursstruktur: McKeown (1982), McKeown (1985), optional: Hovy (1993), Marcu (1997)
zur Vorbereitung: McKeown (1982)

06.05.2010

Diskursplanung: Inhaltsbestimmung: Duboue & McKeown (2003), Barzilay & Lee (2004), optional: Dale (1990), Reiter et al. (1995), Barzilay & Lapata (2005a)
zur Vorbereitung: Duboue & McKeown (2003), Barzilay & Lee (2004)

20.05.2010

Diskursplanung: Inhaltsbestimmung: Barzilay & Lee (2004) (2. Versuch)

Diskursplanung: *Content Ordering, Information Ordering*

Referat: Hartmut Senska Barzilay & Lapata (2008)

zur Vorbereitung: Barzilay & Lapata (2005b)

optional: Barzilay et al. (2002), Lapata (2003), Karamanis et al. (2004), Althaus et al. (2004), Lapata (2006), Madnani et al. (2007)

27.05.2010

Prosodie-Generierung

Referat: Thomas Kropf & Robert Schumann: Pan & McKeown (1998), Marsi et al. (2003)

zur Vorbereitung: Pan & McKeown (1998) oder Marsi et al. (2003)

optional: Hirschberg (1993), Prevost (1996), McKeown et al. (1997), Nakatani (1998), Pan & McKeown (1999), Pan & Hirschberg (2000)

Mikroplanung: Lexikalisierung: Hovy (1990), Reiter (1991), Elhadad et al. (1997), Barzilay & Lee (2002)

Mikroplanung: Paraphrasierung: Barzilay & McKeown (2001), Barzilay & Lee (2003), Wan et al. (2005b)

10.06.2010

Mikroplanung: Aggregation

Referat: Andreas Dörr, Michael Knaf: Dalianis (1999), Barzilay & Lapata (2006)

zur Vorbereitung: Dalianis (1999) oder Barzilay & Lapata (2006)

optional: Scott & Souza (1990), Dalianis & Hovy (1996), Shaw (1998a), Shaw (1998b), Reape & Mellish (1999), Dalianis (1999), Cheng & Mellish (2000), Walker et al. (2001), Shaw (2002), Di Eugenio et al. (2005)

17.06.2010

Mikroplanung: Generierung von Referenzausdrücken

Referat: Vanessa Reubold, Jonas Placzek: Dale & Reiter (1995), Krahmer et al. (2003)

zur Vorbereitung: Dale & Reiter (1995) oder Krahmer et al. (2003)

optional: Appelt (1982), Appelt (1985), Dale (1992), Horacek (1997), Poesio et al. (1999), Kibble (1999), Henschel et al. (2000), van Deemter (2002), Kelleher & van Genabith (2004), Gatt (2006), Viethen & Dale (2006), Kelleher & Kruijff (2006), Belz & Varges (2007), Viethen & Dale (2008)

24.06.2010

Mikroplanung: Sentence Fusion

Referat: Eugen Ruppert: Barzilay & McKeown (2005), Filippova & Strube (2008)

zur Vorbereitung: Barzilay & McKeown (2005) oder s.u.

optional: Daumé III & Marcu (2004), Marsi & Krahmer (2005), Wan et al. (2005a)

Anwendungen: Generierung von Zeitreihenanalysen

Referat: Alexander Henn: Belz (2007), Reiter (2007)

zur Vorbereitung: Belz (2007) oder s.o.

optional: Reiter & Dale (2000), Yu et al. (2007)

01.07.2010

Oberflächenrealisierung: Bestimmung von Wortstellung, Konstituentenreihenfolge

Referat: Dia Oumar, Johannes Nichell: Uchimoto et al. (2000), Filippova & Strube (2007a)

zur Vorbereitung: Uchimoto et al. (2000) oder Filippova & Strube (2007a)

optional: Kruijff-Korbayová et al. (2002), Ringger et al. (2004), Filippova & Strube (2006), Filippova & Strube (2007b)d

08.07.2010

Anwendungen: Generierung von Wegbeschreibungen

Referat: Julio Cesar Rodrigues, Hans-Martin Ramsel: Dale et al. (2003), Dale et al. (2005), Furlan et al. (2007)

zur Vorbereitung: Dale et al. (2003) oder Furlan et al. (2007)
optional: Denis (1997), Horacek (2004b), Tom & Denis (2004), Marciniak & Strube (2004), Marciniak & Strube (2005b), Marciniak & Strube (2005a), Marciniak & Strube (2005c)

15.07.2010

fällt aus – Michael Strube bei AAAI '10

22.07.2010

Evaluierung

Referat: Johannes Reichel: Reiter & Belz (2009)

zur Vorbereitung: Reiter & Belz (2009)

optional: Mellish & Dale (1998), Bangalore et al. (2000), Reiter & Sripada (2002), Sripada et al. (2005), Belz & Reiter (2006), Mutton et al. (2007), Viethen & Dale (2007), Reiter & Belz (2009)

Oberflächenrealisierung: Statistische Methoden

Referat: Eric Hildebrand: Langkilde & Knight (1998), Langkilde (2000), Bangalore & Rambow (2000), Varges & Mellish (2001), Paiva & Evans (2005), Belz (2005)

Diskussion, Zusammenfassung, Ausblick

Weitere Themenvorschläge:

Oberflächenrealisierung: KPML, Surge, RealPro, SPUD – Systemic Functional Grammar, Text Meaning Theory, Tree Adjoining Grammar: Bateman et al. (1991), Bateman (1997), Elhadad et al. (1997), Lavoie & Rambow (1997), Stone & Doran (1997)

Anwendungen: Generierung für Dialogsysteme: Becker et al. (2000), Rambow et al. (2001), Oh & Rudnicky (2002), Pan & Weng (2002), Chambers & Allen (2004)

Bidirektionelle Grammatiken: Appelt (1987), Shieber (1988), Strzalkowski (1991), Kay (1996), Stone & Doran (1997), Espinosa et al. (2008)

Multimedia-Generierung; Generierung für automatische Zusammenfassung und Übersetzung; Generierung von Inhaltsverzeichnissen, Bild- und Tabellenunterschriften; Generierung von Typographie, Layout (Absatzgrenzen)

Bemerkungen

Leistungsnachweise: Lektüre und aktive Teilnahme (1/3), Referat (1/3), Hausarbeit (1/3). Hausarbeit: 8-10 Seiten (Proseminar), 12-15 Seiten (Hauptseminar) inkl. Bibliographie. Die Hausarbeit kann auch per Email an mich geschickt werden, aber *nicht* als Word-Datei sondern nur als PDF-Datei. – Ich empfehle, wissenschaftliche Texte mit Latex und Bibtex zu verfassen.

Regelmäßige Teilnahme (d.i. nicht mehr als einmal unentschuldigtes Fehlen) ist Voraussetzung für Scheinerwerb. Zu jeder Sitzung müssen jeweils zwei Fragen (!) zu einem Papier abgegeben werden, das in der aktuellen Sitzung vorgestellt wird. Abgabe entweder per Email bis spätestens 15 Uhr am Tag der Sitzung oder schriftlich direkt vor der Sitzung. Dies geht in die Bewertung für aktive Teilnahme am Seminar ein.

Literatur: Viele Papiere können direkt aus der *ACL Anthology* kopiert werden (<http://acl.ldc.upenn.edu/>), insbesondere alle Papiere der (*E/NA*)*ACL*-, *Coling*- und *EMNLP*-Konferenzen, alle Workshops, die im Rahmen dieser Konferenzen veranstaltet wurden und die Zeitschrift *Computational Linguistics*. Papiere, die von der *AAAI* publiziert wurden (*AAAI*-Konferenz, *AAAI*-Workshops, *AAAI*-Symposia, etc.) sind in der *AAAI Digital Library* verfügbar (<http://www.aaai.org/Library>). – Die meisten weiteren Zeitschriften sind elektronisch verfügbar über die UB (<http://rzblx1.uni-regensburg.de/ezeit/search.phtml?bibid=UBHE>) – oder stehen dort im Regal.

Sprechstunde: Auf Vereinbarung (Email, Telefon) bei mir im Büro, ggf. auch im Anschluß an das Seminar.

Hausarbeiten:

Maximal 8-10 Seiten (Proseminar), 12-15 Seiten (Hauptseminar) inkl. Abbildungen, inkl. Literaturverzeichnis.

Inhalt: Fokus auf das vorgestellte Papier; NICHT *Related Work*-Kapitel referieren, wenn die entsprechenden Papiere nicht gelesen wurden; Evaluierung berichten; WICHTIG: mit eigener Meinung oder Bewertung abschließen.

Stil: Wissenschaftlichkeit drückt sich nicht durch lange, komplizierte Sätze und exzessiven Gebrauch von Fremdwörtern aus – deshalb bitte kurze Sätze, einfache Sprache; Hausarbeiten vor der Abgabe Korrektur lesen oder Korrektur lesen lassen (s. auch *Dos and donts: Hinweise zur Abfassung wissenschaftlicher Arbeiten* von Prof. Frank – http://www.cl.uni-heidelberg.de/~frank/materials/dos_and_donts.pdf).

Seminararbeit (d.i. eine praktische Arbeit) ist auch möglich. Sollte durch 5-6 Seiten Bericht begleitet werden.

Abgabetermin: bis spätestens 15. August 2010; per Email als PDF-Datei (kein Microsoft Word!) oder ausgedruckt per Post – Matrikelnummer nicht vergessen!

Literatur

- Althaus, Ernst, Nikiforos Karamanis & Alexander Koller (2004). Computing locally coherent discourses. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics*, Barcelona, Spain, 21–26 July 2004, pp. 400–407.
- Appelt, Douglas E. (1982). Planning natural language referring expressions. In *Proceedings of the 20th Annual Meeting of the Association for Computational Linguistics*, Toronto, Ontario, Canada, 16–18 June 1982, pp. 108–112.
- Appelt, Douglas E. (1985). *Planning English Referring Expressions*. Cambridge U.K.: Cambridge University Press.
- Appelt, Douglas E. (1987). Bidirectional grammars and the design of natural language generation systems. In *Proceedings of the Conference on Theoretical Issues in Natural Language Processing*, Las Cruces, New Mexico, USA, 7–9 January 1987, pp. 206–212.
- Bangalore, Srinivas & Owen Rambow (2000). Exploiting a probabilistic hierarchical model for generation. In *Proceedings of the 18th International Conference on Computational Linguistics*, Saarbrücken, Germany, 31 July – 4 August 2000, pp. 42–48.
- Bangalore, Srinivas, Owen Rambow & Steve Whittaker (2000). Evaluation metrics for generation. In *Proceedings of the 1st International Conference on Natural Language Generation*, Mitze Ramon, Israel, 12–16 June 2000, pp. 1–8.
- Barzilay, Regina, Noemie Elhadad & Kathleen R. McKeown (2002). Inferring strategies for sentence ordering. *Journal of Artificial Intelligence Research*, 17:35–55.
- Barzilay, Regina & Mirella Lapata (2005a). Collective content selection for concept-to-text-generation. In *Proceedings of the Human Language Technology Conference and the 2005 Conference on Empirical Methods in Natural Language Processing*, Vancouver, B.C., Canada, 6–8 October 2005, pp. 331–338.
- Barzilay, Regina & Mirella Lapata (2005b). Modeling local coherence: An entity-based approach. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*, Ann Arbor, Mich., 25–30 June 2005, pp. 141–148.
- Barzilay, Regina & Mirella Lapata (2006). Aggregation via set partitioning for natural language generation. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, New York, N.Y., 4–9 June 2006, pp. 359–366.
- Barzilay, Regina & Mirella Lapata (2008). Modeling local coherence: An entity-based approach. *Computational Linguistics*, 34(1):1–34.
- Barzilay, Regina & Lillian Lee (2002). Bootstrapping lexical choice via multiple-sequence alignment. In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing*, Philadelphia, Penn., 6–7 July 2002, pp. 164–171.
- Barzilay, Regina & Lillian Lee (2003). Learning to paraphrase: An unsupervised approach using multi-sequence alignment. In *Proc. of HLT-NAACL-03*, pp. 16–23.
- Barzilay, Regina & Lillian Lee (2004). Catching the drift: Probabilistic content models, with applications to generation and summarization. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, Boston, Mass., 2–7 May 2004, pp. 113–120.
- Barzilay, Regina & Kathleen R. McKeown (2001). Extracting paraphrases from a parallel corpus. In *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics*, Toulouse, pp. 50–57.
- Barzilay, Regina & Kathleen R. McKeown (2005). Sentence fusion for multidocument news summarization. *Computational Linguistics*, 31(3):297–327.
- Bateman, John, Christian Matthiesen, Keizo Nanri & Licheng Zeng (1991). The re-use of linguistic resources across languages in multilingual generation. In *Proceedings of the 12th International Joint Conference on Artificial Intelligence*, Sydney, Australia, 24–30 August 1991, pp. 966–971.
- Bateman, John A. (1997). Enabling technology for multilingual natural language generation. *Natural Language Engineering*, 3:15–55.

- Becker, Tilman, Anne Kilger, Patrice Lopez & Peter Poller (2000). The Verbmobil generation component VM-GECO. In W. Wahlster (Ed.), *Verbmobil: Foundations of Speech-to-Speech Translation*, pp. 481–496. Berlin: Springer-Verlag.
- Belz, Anja (2005). Statistical generation: Three methods compared and evaluated. In *Proceedings of the 10th European Workshop on Natural Language Generation*, Aberdeen, U.K., 8–10 August 2005, pp. 15–23.
- Belz, Anja (2007). Probabilistic generation of weather forecast texts. In *Proceedings of Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics*, Rochester, N.Y., 22–27 April 2007, pp. 164–171.
- Belz, Anja & Ehud Reiter (2006). Comparing automatic and human evaluation of nlg systems. In *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics*, Trento, Italy, 3–7 April 2006, pp. 313–320.
- Belz, Anja & Sebastian Varges (2007). Generation of repeated references to discourse entities. In *Proceedings of the 11th European Workshop on Natural Language Generation*, Schloss Dagstuhl, Germany, 17–20 June 2007, pp. 9–16.
- Chambers, Nathanael & James F. Allen (2004). Stochastic language generation in a dialogue system: Toward a domain independent generator. In *Proceedings of the 5th SIGdial Workshop on Discourse and Dialogue*, Cambridge, Mass., 30 April – 1 May 2004, pp. 9–18.
- Cheng, Hua & Chris Mellish (2000). An empirical analysis of constructing non-restrictive NP modifiers to express semantic relations. In *Proceedings of the 1st International Conference on Natural Language Generation*, Mitze Ramon, Israel, 12–16 June 2000, pp. 186–193.
- Dale, Robert (1990). Generating recipes: An overview of Epicure. In R. Dale, C. Mellish & M. Zock (Eds.), *Current Research in Natural Language Generation*, pp. 229–255. New York, N.Y.: Academic Press.
- Dale, Robert (1992). *Generating Referring Expressions: Constructing Descriptions in a Domain of Objects and Processes*. Cambridge, Mass.: MIT Press.
- Dale, Robert, Sabine Geldof & Jean-Philippe Prost (2003). CORAL: Using natural language generation for navigational assistance. In *Proceedings of the 26th Australasian Computer Science Conference*, Adelaide, South Australia, 4–7 February 2003, p. ??
- Dale, Robert, Sabine Geldof & Jean-Philippe Prost (2005). Using natural language generation in automatic route description. *Journal of Research and Practice in Information Technology*, 37(1):89–106.
- Dale, Robert & Ehud Reiter (1995). Computational interpretations of the Gricean maxims in the generation of referring expressions. *Cognitive Science*, 18:233–263.
- Dalianis, Hercules (1999). Aggregation in Natural Language Generation. *Journal of Computational Intelligence*, 15(4):384–414.
- Dalianis, Hercules & Eduard H. Hovy (1996). Aggregation in natural language generation. In G. Adorni & M. Zock (Eds.), *Trends in Natural Language Generation: An Artificial Intelligence Perspective*, pp. 88–105. Berlin, Germany: Springer.
- Daumé III, Hal & Daniel Marcu (2004). Generic sentence fusion is an ill-defined summarization task. In *Proceedings of the Text Summarization Branches Out Workshop at ACL '04*, Madrid, Spain, 25–26 July 2004, pp. 96–103.
- Denis, Michel (1997). The description of routes: A cognitive approach to the production of spatial discourse. *Current Psychology of Cognition*, 16:409–458.
- Di Eugenio, Barbara, Davide Fossati, Dan Yu, Susan Haller & Michael Glass (2005). Aggregation improves learning: Experiments in natural language generation for intelligent tutoring systems. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*, Ann Arbor, Mich., 25–30 June 2005, pp. 50–57.
- Duboue, Pablo A. & Kathleen R. McKeown (2003). Statistical acquisition of content selection rules. In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*, Sapporo, Japan, 11–12 July 2003, pp. 121–128.
- Elhadad, Michael, Kathleen R. McKeown & Jaques Robin (1997). Floating constraints in lexical choice. *Computational Linguistics*, 23(2):195–239.
- Espinosa, Dominic, Michael White & Dennis Mehay (2008). Hypertagging: Supertagging for

- surface realization with CCG. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, Columbus, Ohio, 15–20 June 2008, pp. 183–191.
- Filippova, Katja & Michael Strube (2006). Improving text fluency by reordering of constituents. In *Proceedings of the ESSLI Workshop on Modelling Coherence for Generation and Dialogue Systems*, Málaga, Spain, 7–11 August 2006, pp. 9–16.
- Filippova, Katja & Michael Strube (2007a). Generating constituent order in German clauses. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, Prague, Czech Republic, 23–30 June 2007, pp. 320–327.
- Filippova, Katja & Michael Strube (2007b). The German Vorfeld and local coherence. *Journal of Logic, Language and Information*, 16(4):465–485.
- Filippova, Katja & Michael Strube (2008). Sentence fusion via dependency graph compression. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, Waikiki, Honolulu, Hawaii, 25–27 October 2008, pp. 177–185.
- Furlan, Aidan, Timothy Baldwin & Alexander Klippel (2007). Landmark classification for route directions. In *Proceedings of the 4th ACL-SIGSEM Workshop on Prepositions*, Prague, Czech Republic, 28 June 2007, pp. 9–16.
- Gatt, Albert (2006). Structuring knowledge for reference generation: A clustering algorithm. In *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics*, Trento, Italy, 3–7 April 2006, pp. 321–328.
- Henschel, Renate, Hua Cheng & Massimo Poesio (2000). Pronominalization revisited. In *Proceedings of the 18th International Conference on Computational Linguistics*, Saarbrücken, Germany, 31 July – 4 August 2000, pp. 306–312.
- Hirschberg, Julia (1993). Pitch accent in context: Predicting intonational prominence from text. *Artificial Intelligence*, 63:305–340.
- Horacek, Helmut (1997). An algorithm for generating referential descriptions with flexible interfaces. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and of the 8th Conference of the European Chapter of the Association for Computational Linguistics*, Madrid, Spain, 7–12 July 1997, pp. 206–213.
- Horacek, Helmut (2004a). Textgenerierung. In U. et al. Carstensen (Ed.), *Computerlinguistik und Sprachtechnologie. Eine Einführung*, Chp. 3.7, pp. 436–465. Heidelberg, Germany: Spektrum Akademischer Verlag.
- Horacek, Helmut (2004b). Textgenerierung. In U. et al. Carstensen (Ed.), *Computerlinguistik und Sprachtechnologie*, Chp. 3.7, pp. 375–403. München, Germany: Spektrum Akademischer Verlag.
- Hovy, Eduard H. (1990). Pragmatics and natural language generation. *Artificial Intelligence*, 43(2):153–197.
- Hovy, Eduard H. (1993). Automated discourse generation using discourse structure relations. *Artificial Intelligence*, 63:341–385.
- Karamanis, Nikiforos, Massimo Poesio, Chris Mellish & Jon Oberlander (2004). Evaluating centering-based metrics of coherence for text structuring using a reliably annotated corpus. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics*, Barcelona, Spain, 21–26 July 2004, pp. 392–393.
- Kay, Martin (1996). Chart generation. In *Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics*, Santa Cruz, Cal., 24–27 June 1996, pp. 200–204.
- Kelleher, John D. & Geert-Jan Kruijff (2006). Incremental generation of spatial referring expressions in situated dialogue. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the ACL*, Sydney, Australia, July 2006, pp. 1041–1048.
- Kelleher, John D. & Josef van Genabith (2004). Exploiting visual salience for the generation of referring expressions. In *Proceedings of the 17th International Florida Artificial Intelligence Research Society Conference*, Miami Beach, Florida, 17–19 May 2004.
- Kibble, Rodger (1999). Cb or not Cb? Centering theory applied to NLG. In *ACL '99 Workshop on the Relationship between Discourse/Dialogue Structure and Reference*, University of

- Maryland, Maryland, 21 June, 1999, pp. 72–81.
- Krahmer, Emiel, Sebastiaan van Erk & André Verleg (2003). Graph-based generation of referring expressions. *Computational Linguistics*, 29(1):53–72.
- Kruijff-Korbayová, Ivana, Geert-Jan Kruijff & John Bateman (2002). Generation of appropriate word order. In K. van Deemter & R. Kibble (Eds.), *Information Sharing: Reference and Pre-supposition in Language Generation and Interpretation*, pp. 193–222. Stanford, Cal.: CSLI.
- Langkilde, Irene (2000). Forest-based statistical sentence generation. In *Proceedings of the 1st Conference of the North American Chapter of the Association for Computational Linguistics*, Seattle, Wash., 29 April – 3 May 2000, pp. 170–177.
- Langkilde, Irene & Kevin Knight (1998). Generation that exploits corpus-based statistical knowledge. In *Proceedings of the 17th International Conference on Computational Linguistics and 36th Annual Meeting of the Association for Computational Linguistics*, Montréal, Québec, Canada, 10–14 August 1998, pp. 704–710.
- Lapata, Maria (2003). Probabilistic text structuring: Experiments with sentence ordering. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, Sapporo, Japan, 7–12 July 2003, pp. 545–552.
- Lapata, Mirella (2006). Automatic evaluation of information ordering: Kendall’s tau. *Computational Linguistics*, 32(4):471–484.
- Lavoie, Benoit & Owen Rambow (1997). A fast and portable realizer for the text generation. In *Proceedings of the 5th Conference on Applied Natural Language Processing*, Washington, D.C., 31 March – 3 April 1997, pp. 265–268.
- Madnani, Nitin, Rebecca Passonneau, Necip Fazil Ayan, John Conroy, Bonnie Dorr, Judith Klavans, Dianne O’Leary & Judith Schlesinger (2007). Measuring variability in sentence ordering for news summarization. In *Proceedings of the 11th European Workshop on Natural Language Generation*, Schloss Dagstuhl, Germany, 17–20 June 2007, pp. 81–88.
- Marciniak, Tomasz & Michael Strube (2005a). Beyond the pipeline: Discrete optimization in NLP. In *Proceedings of the 9th Conference on Computational Natural Language Learning*, Ann Arbor, Mich., USA, 29–30 June 2005, pp. 136–145.
- Marciniak, Tomasz & Michael Strube (2005b). Modelling and annotating the semantics of route directions. In *Proceedings of the 6th International Workshop on Computational Semantics (IWCS-6)*, Tilburg, The Netherlands, 12–14 January, 2005, pp. 151–162.
- Marciniak, Tomasz & Michael Strube (2005c). Using an annotated corpus as a knowledge source for language generation. In *Proceedings of the Workshop on Using Corpora for Natural Language Generation*, Birmingham, U.K., 14 July, 2005, pp. 19–24.
- Marciniak, Tomasz & Michael Strube (2004). Classification-based generation using TAG. In *Proceedings of the 3rd International Conference on Natural Language Generation*, Brockenhurst, U.K., 14–16 July 2004, pp. 100–109.
- Marcu, Daniel (1997). From local to global coherence: A bottom-up approach to text planning. In *Proceedings of the 14th National Conference on Artificial Intelligence*, Providence, R.I., 27–31 July 1997, pp. 629–635.
- Marsi, Erwin & Emiel Krahmer (2005). Explorations in sentence fusion. In *Proceedings of the European Workshop on Natural Language Generation*, Aberdeen, Scotland, 8–10 August, 2005, pp. 109–117.
- Marsi, Erwin, Martin Reynaert, Antal van den Bosch, Walter Daelemans & Véronique Hoste (2003). Learning to predict pitch accents and prosodic boundaries in Dutch. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, Sapporo, Japan, 7–12 July 2003, pp. 489–496.
- McKeown, Kathleen R. (1982). The TEXT system for natural language generation: An overview. In *Proceedings of the 20th Annual Meeting of the Association for Computational Linguistics*, Toronto, Ontario, Canada, 16–18 June 1982, pp. 113–120.
- McKeown, Kathleen R. (1985). Discourse strategies for generating natural-language text. *Artificial Intelligence*, 27:1–42.
- McKeown, Kathleen R., Shimei Pan, James Shaw, Desmond Jordon & Barry Allen (1997). Language generation for multimedia healthcare briefings. In *Proceedings of the 5th Conference*

- on *Applied Natural Language Processing*, Washington, D.C., 31 March – 3 April 1997, pp. 277–282.
- Mellish, Chris & Robert Dale (1998). Evaluation in the context of natural language generation. *Computer Speech and Language*, 12:349–373.
- Mellish, Chris, Mike Reape, Donia Scott, Lynne Cahill, Roger Evans & Daniel Paiva (2004). A reference architecture for generation systems. *Natural Language Engineering*, 10:227–260.
- Mutton, Andrew, Mark Dras, Stephen Wan & Robert Dale (2007). GLEU: Automatic evaluation of sentence-level fluency. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, Prague, Czech Republic, 23–30 June 2007, pp. 344–351.
- Nakatani, Christine H. (1998). Constituent-based accent prediction. In *Proceedings of the 17th International Conference on Computational Linguistics and 36th Annual Meeting of the Association for Computational Linguistics*, Montréal, Québec, Canada, 10–14 August 1998, pp. 939–945.
- Oh, Alice & Alex Rudnicky (2002). Stochastic natural language generation for spoken dialog systems. *Computer Speech and Language*, 16(3):387–407.
- Paiva, Daniel S. & Roger Evans (2005). Empirically-based control of natural language generation. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*, Ann Arbor, Mich., 25–30 June 2005, pp. 58–65.
- Pan, Shimei & Julia Hirschberg (2000). Modeling local context for pitch accent prediction. In *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics*, Hong Kong, China, 1–8 August 2000, pp. 233–240.
- Pan, Shimei & Kathleen R. McKeown (1998). Learning intonation rules for concept to speech generation. In *Proceedings of the 17th International Conference on Computational Linguistics and 36th Annual Meeting of the Association for Computational Linguistics*, Montréal, Québec, Canada, 10–14 August 1998, pp. 1003–1009.
- Pan, Shimei & Kathleen R. McKeown (1999). Word informativeness and automatic pitch accent modeling. In *Proceedings of the 1999 SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora*, College Park, Md., 21–22 June 1999, pp. 148–157.
- Pan, Shimei & Wubei Weng (2002). Designing a speech corpus for instance-based spoken language generation. In *Proceedings of the 2nd International Conference on Natural Language Generation*, New York, N.Y., 1–3 July 2002, pp. 49–56.
- Poesio, Massimo, Renate Henschel, Janet Hitzeman & Rodger Kibble (1999). Statistical NP generation: A first report. In R. Kibble & K. van Deemter (Eds.), *Proceedings of the Workshop on The Generation of Nominal Expressions, 11th European Summer School on Logic, Language, and Information, Utrecht, 9-13 August 1999*.
- Prevost, Scott (1996). An information structural approach to spoken language generation. In *Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics*, Santa Cruz, Cal., 24–27 June 1996, pp. 294–301.
- Rambow, Owen, Srinivas Bangalore & Marilyn Walker (2001). Natural language generation in dialog systems. In *Proceedings of the 1st International Conference on Human Language Technology Research*, San Diego, Cal., 18–21 March 2001, p. ??
- Reape, Mike & Chris Mellish (1999). Just what is aggregation anyway? In *Proceedings of the 7th European Workshop on Natural Language Generation*, pp. 20–29.
- Reiter, Ehud (1991). A new model of lexical choice for nouns. *Computational Intelligence*, 7(4):240–251.
- Reiter, Ehud (2007). An architecture for data-to-text systems. In *Proceedings of the 11th European Workshop on Natural Language Generation*, Schloss Dagstuhl, Germany, 17–20 June 2007, pp. 97–104.
- Reiter, Ehud & Anja Belz (2009). An investigation into the validity of some metrics for automatically evaluating NLG systems. *Computational Linguistics*, 35(4):529–558.
- Reiter, Ehud & Robert Dale (2000). *Building Natural Language Generation Systems*. Cambridge, U.K.: Cambridge University Press.
- Reiter, Ehud, Chris Mellish & John Levine (1995). Automatic generation of technical documen-

- tation. *Applied Artificial Intelligence*, 9(3):259–287.
- Reiter, Ehud & Somayajulu Sripada (2002). Should corpora texts be gold standards for nlg? In *Proceedings of the 2nd International Conference on Natural Language Generation*, New York, N.Y., 1–3 July 2002, pp. 97–104.
- Ringger, Eric, Michael Gamon, Robert C. Moore, David Rojas, Martine Smets & Simon Corston-Oliver (2004). Linguistically informed statistical models of constituent structure for ordering in sentence realization. In *Proceedings of the 20th International Conference on Computational Linguistics*, Geneva, Switzerland, 23–27 August 2004, pp. 673–679.
- Scott, Donia R. & Clarisse Sieckenius de Souza (1990). Getting the message across in RST-based text generation. In R. Dale, C. Mellish & M. Zock (Eds.), *Current Research in Natural Language Generation*, pp. 47–73. Academic Press.
- Shaw, James (1998a). Clause aggregation using linguistic knowledge. In *Proceedings of the 9th International Workshop on Natural Language Generation*, Niagara-on-the-Lake, Ontario, Canada, August 1998, pp. 138–147.
- Shaw, James (1998b). Segregatory coordination and ellipsis in text generation. In *Proceedings of the 17th International Conference on Computational Linguistics and 36th Annual Meeting of the Association for Computational Linguistics*, Montréal, Québec, Canada, 10–14 August 1998, pp. 1220–1226.
- Shaw, James (2002). Corpus-based analysis for the ordering of clause aggregation operators. In *Proceedings of the 19th International Conference on Computational Linguistics*, Taipei, Taiwan, 24 August – 1 September 2002, p. ??
- Shieber, Stuart M. (1988). A uniform architecture for parsing and generation. In *Proceedings of the 12th International Conference on Computational Linguistics*, Budapest, Hungary, 22–27 August 1988, pp. 614–619.
- Sripada, Somayajulu, Ehud Reiter & Lezan Hawizy (2005). Evaluation of an nlg system using post-edit data. In *Proceedings of the 10th European Workshop on Natural Language Generation*, Aberdeen, U.K., 8–10 August 2005, pp. 133–139.
- Stone, Matthew & Christine Doran (1997). Sentence planning as description using tree adjoining grammar. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and of the 8th Conference of the European Chapter of the Association for Computational Linguistics*, Madrid, Spain, 7–12 July 1997, pp. 198–205.
- Strzalkowski, Tomek (Ed.) (1991). *Proceedings of a Workshop on Reversible Grammar in Natural Language Processing*, Berkeley, Cal., USA, 17 June 1991.
- Tom, Ariane & Michel Denis (2004). Language and spatial cognition: Comparing the roles of landmarks and street names in route instructions. *Applied Cognitive Psychology*, 18:1213–1230.
- Uchimoto, Kiyotaka, Masaki Murata, Qing Ma, Satoshi Sekine & Hitoshi Isahara (2000). Word order acquisition from corpora. In *Proceedings of the 18th International Conference on Computational Linguistics*, Saarbrücken, Germany, 31 July – 4 August 2000, pp. 871–877.
- van Deemter, Kees (2002). Generating referring expressions: Boolean extensions of the incremental algorithm. *Computational Linguistics*, 28(1):37–52.
- Vander Linden, Keith (2000). Natural language generation. In D. Jurafsky & J.H. Martin (Eds.), *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, Chp. 20, pp. 763–798. Upper Saddle River, N.J.: Prentice Hall.
- Varges, Sebastian & Chris Mellish (2001). Instance-based natural language generation. In *Proceedings of the 2nd Conference of the North American Chapter of the Association for Computational Linguistics*, Pittsburgh, Penn., 2–7 June 2001, pp. 1–8.
- Viethen, Jette & Robert Dale (2006). Algorithms for generating referring expressions: Do they do what people do? In *Proceedings of the 4th International Conference on Natural Language Generation*, Sidney, Australia, 15–16 July 2006, pp. 63–70.
- Viethen, Jette & Robert Dale (2007). Evaluation in natural language generation: Lessons from referring expression generation. *Traitement Automatique des Langues*, 48(1):141–160.
- Viethen, Jette & Robert Dale (2008). The use of spatial relations in referring expression gene-

- ration. In *Proceedings of the 5th International Conference on Natural Language Generation*, Salt Fork, Ohio, 12–14 June 2008, pp. 59–67.
- Walker, Marilyn A., Owen Rambow & Monica Rogati (2001). Spot: A trainable sentence planner. In *Proceedings of the 2nd Conference of the North American Chapter of the Association for Computational Linguistics*, Pittsburgh, Penn., 2–7 June 2001, pp. 17–24.
- Wan, Stephen, Robert Dale, Marc Dras & Cecile Paris (2005a). Searching for grammaticality and consistency: Propagating dependencies in the Viterbi algorithm. In *Proceedings of the 10th European Workshop on Natural Language Generation*, Aberdeen, Scotland, 8–10 August, 2005, pp. 211–216.
- Wan, Stephen, Mark Dras, Robert Dale & Cécile Paris (2005b). Towards statistical paraphrase generation: Preliminary evaluations of grammaticality. In *Proceedings of the 2nd International Joint Conference on Natural Language Processing*, Jeju Island, South Korea, 11–13 October 2005, p. ??
- Yu, Jin, Ehud Reiter, Jim Hunter & Chris Mellish (2007). Choosing the content of textual summaries of large time-series data sets. *Natural Language Engineering*, 13(1):25–50.