

Nizar Y. Habash

850 Interchurch Center, Mail Code 7717
475 Riverside Drive, New York, NY 10115
Office:(212) 870-1289, Cell:(646) 388-2315, Fax:(212) 870-1285
<http://www.NizarHabash.com>
Nizar@NizarHabash.com

EDUCATION

- Ph.D. Computer Science**, 2003. University of Maryland College Park.
Dissertation: "Generation-Heavy Hybrid Machine Translation" ([URL](#))
- M.S. Computer Science**, 2000. University of Maryland College Park.
- B.S. Computer Engineering**, 1997, *summa cum laude*. Old Dominion University.
- B.A. Linguistics and Languages**, 1997, *summa cum laude*. Old Dominion University.

RESEARCH EXPERIENCE

Columbia University January/2009 -
Center for Computational Learning Systems
Research Scientist

- PI of a Columbia University sub-contract (to IBM) under DARPA's GALE Program to develop the Columbia Arabic Tree Bank (CATiB)
- Co-PI of a Columbia University sub-contract (to IBM) under DARPA's GALE Program to develop tools for Arabic processing (machine translation group)
- Co-PI of the Columbia University sub-contract (to BBN) under a DARPA-funded Multilingual Automatic Document Classification Analysis and Translation (MADCAT) Program.
- Co-PI of a Columbia University sub-contract (to Acxiom) to work on Arabic Blog Text Processing.
- PI of a Google-awarded project on Arabic Generation in Statistical Machine Translation.

Columbia University July/2005 – January/2009
Center for Computational Learning Systems
Associate Research Scientist

- PI of a Columbia University sub-contract (to IBM) under DARPA's GALE Program to develop the Columbia Arabic Tree Bank (CATiB)
- Co-PI of a Columbia University sub-contract (to IBM) under DARPA's GALE Program to develop tools for Arabic processing (machine translation group)

- Co-PI of the Columbia University sub-contract (to BBN) under a DARPA-funded Multilingual Automatic Document Classification Analysis and Translation (MADCAT) Program.
- Co-founder of the Columbia Arabic Dialect Modeling (CADIM) group. (<http://www.ccls.columbia.edu/oldweb/cadim/>)
- Senior member in an NSF funded Johns Hopkins Summer Workshop on Arabic Dialect Parsing
- Member of the DARPA-funded "Novel Information Gathering and Harvesting Techniques in a Global Autonomous Environment (NIGHTINGALE)" project working on Arabic modeling and Machine Translation.
- Collaborated with Textwise on their Arabic CINDOR (Conceptual INterlingua DOcument Retrieval) Project.
- Collaborated with other Columbia University researcher on a KDD project on email and newsgroup summarization.

Columbia University

July/2004 – July/2005

Post-doctoral Researcher

- Worked in a NSF ITR on Arabic Dialect Modeling for Speech and Natural Language (PIs Kathy McKeown and Owen Rambow).

University of Maryland College Park

July/2003 – July/2004

Post-doctoral Researcher

- An active member of the six-site project, Interlingual Annotation for Multilingual Text Corpora (IAMTC).
- Managed the Divergence Unraveling for Statistical Translation (DUSTer) project.
- Managed the Generation-heavy Hybrid Machine Translation (GHMT) effort at University of Maryland College Park.
- Worked with a graduate student to extend the GHMT approach to handle English headline generation from foreign text (in collaboration with Bonnie Dorr).

Computer Skills

- Hardware: Strong background in digital design and microcontrollers.
- Software: Extensive software experience under Windows and Unix and internet web design.
- Resources: WordNet, CMU Toolkit, LCS Lexicons, Nitrogen/Halogen, Yamcha machine learning system, Ripper, Buckwalter Arabic analyzer, AT&T FSM toolkit, SRILM toolkit, Lextools.
- Languages: High fluency: Perl, Java, C, C++, Visual Basic, JavaScript, and Lisp. Comfortable fluency: Pascal, Basic, Prolog, MatLab, SubL, Php, SQL, Flash

WORK EXPERIENCE

Nuun Labs, Inc.

1999-2000

Chief Technical Officer

Designed and implemented the company's main products: NuunPad, NuunBuilder, NuunConvert, NuunForms, and NuunEncoding. Constructed the initial web site for Nuun technology. Provided technical support to most of NuunLabs clients. Led presentations on Nuun in an academic conference and for possible investors.

CYCORP, Inc.

Summer 1998

Intern / Ontological Engineer

The Natural Language Processing Group

Designed and implemented PreGen (Preprocessor Generator) and PIL (Preprocessing Instruction Language). Extensive work with the CYC knowledge base including writing rules and queries for CYC.

TEACHING EXPERIENCE

Lecturer

Co-instructor with Professor Bonnie Dorr. Designed and Presented half of the class lectures to a class of 35 students Designed assignments and projects and graded exams. CMSC 723/LING 645: Introduction to Computational Linguistics *University of Maryland College Park , Spring 2004* [URL](#)

Habash, Nizar. Introduction to Arabic Natural Language Processing. A two-hour tutorial

- (to be presented) at MEDAR: The 2nd International Conference on Arabic Language Resources and Tools, Cairo, Egypt, 2009.

Diab, Mona and **Nizar Habash.** Arabic Dialect Processing. A two-hour tutorial

- (to be presented) at MEDAR: The 2nd International Conference on Arabic Language Resources and Tools, Cairo, Egypt, 2009.

Habash, Nizar. Arabic Natural Language Processing for Machine Translation. A three-hour tutorial

- (12 attendees) at the meeting of the Association for Machine Translation in the Americas conference (AMTA'08), Waikiki, Hawai'i, 2008.

Diab, Mona and **Nizar Habash.** Arabic Dialect Processing. A three-hour tutorial

- (30 attendees) at the International Conference on Language Resources and Evaluation, Marrakech, Morocco, 2008.
- (15 attendees) at the Conference of the North American Association for Computational Linguistics (NAACL'07), 2007.
- (8 attendees) at the Association for Machine Translation in the Americas (AMTA'06), Boston, MA, 2006.

Habash, Nizar. Introduction to Arabic Natural Language Processing. A three-hour tutorial

- (35 attendees). Johns Hopkins University Summer School on Human Language Technology, Baltimore, 2006.
- (20 attendees) at the International Conference on Language Resources and Evaluation, Genoa, Italy, 2006.
- (40 attendees) at BBN Technologies, Boston, Massachusetts, August 29, 2005.
- (20 attendees) at the National Cryptologic Muesum. August 2005.
- (50 attendees). Johns Hopkins University Summer Workshop, Baltimore, Maryland, July 2005.
- (35 attendees) at the Association for Computational Linguistics Conference (ACL'05), Ann Arbor, Michigan, June 25, 2005.

- (17 attendees) at the Association for Machine Translation in the Americas conference (AMTA'04), Georgetown University, Washington DC, September 28, 2004.

Guest Lecturer

Habash, Nizar. Machine Translation: Challenges, Approaches and Evaluation, CS 4705: Introduction to Natural Language Processing, *Columbia University*

- *Fall 2008, Fall 2007, Fall 2006, Fall 2005, Fall 2004* ([Syllabus](#)) ([Slides](#))

Habash, Nizar. Machine Translation: Challenges and Approaches, CMSC 421: Introduction to Artificial Intelligence, *University of Maryland College Park, Fall 2003* ([Syllabus](#)) ([Slides](#))

Habash, Nizar. Generation Heavy Hybrid Machine Translation, CMSC 723/LING 645: Introduction to Computational Linguistics, *University of Maryland College Park, Spring 2003*

Habash, Nizar. Computers and Writing Systems, HONR 279I: History of the Alphabets, 2000 BCE to 2000 CE: Languages and Their Scripts, *University of Maryland College Park, Spring 2003*

Habash, Nizar. Natural Language Generation, CMSC 828: Advanced Natural Language Processing: Theory and Practice, *University of Maryland College Park, Spring 2002*

Teaching Material Design and Implementation

Currently writing a Synthesis Volume on Arabic Natural Language Processing as part of a series to be published by Morgan&Claypool. (Summer 2009)

Student Advising/Supervision

Over the last four years, I have been heavily involved in the advising of two PhD students at the University of Maryland, one PhD student at Copenhagen Business School (Denmark), one PhD student from Cairo University, one student at Universitat Politècnica de Catalunya (UPC), one Masters Student at Ben-Gurion University (Israel) and one Masters student from Georgetown University. Additionally, I supervised the work of twelve students working on eight different projects in University of Maryland and Columbia University. I have served on the doctoral committee of Otakar Smrz (2007, Charles University in Prague), Josep Crego (Universitat Politècnica de Catalunya, 2008) and Jakob Elming (Copenhagen Business School, 2008). I mentored a high school student, Hayden Metsky, doing research in the area of machine translation. The student became the regional winner in the Siemens Competition in Math, Science and Technology (top 6 out of 1893) and a finalist in the Intel Science Talent Search (top 40 out of 1600). I currently have two PhD students: Ahmed El-Kholy and Sarah Al-Kuhlani.

PUBLICATIONS AND PRESENTATIONS

Publications

In Preparation

Habash, Nizar. Introduction to Arabic Natural Language Processing, *book in preparation.*

Habash, Nizar and Fatiha Sadat. Arabic Preprocessing Schemes and Combinations for Statistical Machine Translation, *in preparation.*

Habash, Nizar, Bonnie Dorr and Christof Monz. Symbolic to Statistical Hybrid Machine Translation: The Case of Generation-Heavy MT, to appear in *MT Journal*.

2009

Biadisy, Fadi, **Nizar Habash** and Julia Hirschberg. Improving the Arabic Pronunciation Dictionary for Phone and Word Recognition with Linguistically-Based Pronunciation Rules. In *Proceedings of the North American Association for Computational Linguistics (NAACL)*, Boulder, Colorado, USA, 2009.

Habash, Nizar, Owen Rambow and Ryan Roth. MADA+TOKAN: A Toolkit for Arabic Tokenization, Diacritization, Morphological Disambiguation, POS Tagging, Stemming and Lemmatization. In *Proceedings of the 2nd International Conference on Arabic Language Resources and Tools (MEDAR)*, Cairo, Egypt, 2009.

Habash, Nizar. REMOOV: A Tool for Online Handling of Out-of-Vocabulary Words in Machine Translation. In *Proceedings of the 2nd International Conference on Arabic Language Resources and Tools (MEDAR)*, Cairo, Egypt, 2009.

Habash, Nizar, Reem Faraj and Ryan Roth. Syntactic Annotation in the Columbia Arabic Treebank. In *Proceedings of the 2nd International Conference on Arabic Language Resources and Tools (MEDAR)*, Cairo, Egypt, 2009.

Habash, Nizar and Jun Hu. Improving Arabic-Chinese Statistical Machine Translation using English as Pivot Language. In *Proceedings of the Workshop on Statistical Machine Translation at the meeting of the European Association for Computational Linguistics (EACL)*, Athens, Greece, 2009.

Biadisy, Fadi, Julia Hirschberg and **Nizar Habash**. Spoken Arabic Dialect Identification Using Phonotactic Modeling. In *Proceedings of the Workshop on Computational Approaches to Semitic Languages at the meeting of the European Association for Computational Linguistics (EACL)*, Athens, Greece, 2009.

Elming, Jakob and **Nizar Habash**. Syntactic Reordering for English-Arabic Phrase-Based Machine Translation. In *Proceedings of the Workshop on Computational Approaches to Semitic Languages at the meeting of the European Association for Computational Linguistics (EACL)*, Athens, Greece, 2009.

Buckwalter, Tim and **Nizar Habash**. Buckwalter Arabic Morphological Analyzer (BAMA) in Arabic Presentation Form. A participating system in the Arab League Educational, Cultural and Scientific Organization (ALECSO) workshop on Arabic Morphological Analyzers, Damascus, Syria, 2009.

2008

Habash, Nizar and Ahmed Elkholy. SEPIA: Surface Span Extension to Syntactic Dependency Precision-based MT Evaluation. In *Proceedings of the workshop on Metrics for Machine Translation at the meeting of Association for Machine Translation in the Americas (AMTA-2008)*, Waikiki, Hawai'i.

Habash, Nizar and Hayden Metsky. Automatic Learning of Morphological Variations for Handling Out-of-Vocabulary Terms in Urdu-English Machine Translation. In *Proceedings of the Association for Machine Translation in the Americas (AMTA-2008)*, Waikiki, Hawai'i, 2008.

Habash, Nizar. Four Techniques for Online Handling of Out-of-Vocabulary Words in Arabic-English Statistical Machine Translation. In *Proceedings of Association for Computational Linguistics (ACL)*, Columbus, Ohio. 2008.

Roth, Ryan, Owen Rambow, **Nizar Habash**, Mona Diab, and Cynthia Rudin. Arabic Morphological Tagging, Diacritization, and Lemmatization Using Lexeme Models and Feature Ranking. In *Proceedings of Association for Computational Linguistics (ACL)*, Columbus, Ohio. 2008.

Crego, Josep M. and **Nizar Habash**. Using Shallow Syntax Information to Improve Word Alignment and Reordering for SMT. In *Proceedings of the Statistical Machine Translation Workshop at the Conference of Association for Computational Linguistics (ACL)*, Columbus, Ohio. 2008.

Habash, Nizar and Ryan Roth. Identification of Naturally Occurring Numerical Expressions in Arabic. In *Proceedings of the Language Resources and Evaluation Conference (LREC)*, Marrakech, Morocco, 2008.

Farber, Benjamin, Dayne Freitag, **Nizar Habash** and Owen Rambow. Improving NER in Arabic Using a Morphological Tagger. In *Proceedings of the Language Resources and Evaluation Conference (LREC)*, Marrakech, Morocco, 2008.

Habash, Nizar, Owen Rambow, Mona Diab and Reem Farraj. Guidelines for Annotating Arabic Dialect. In *Proceedings of Workshop on Arabic and its local languages, LREC*, Marrakech, Morocco. 2008.

Elming, Jakob, **Nizar Habash** and Josep Crego. Combination of Statistical Word Alignments Based on Multiple Preprocessing Schemes. Book chapter in "Learning for Machine Translation." Editors Cyril Goutte, Nicola Cancedda, Marc Dymetman, and George Foster. MIT Press, 2008.

2007

Habash, Nizar. Syntactic Preprocessing for Statistical Machine Translation, In *Proceedings of the Machine Translation Summit (MT-Summit)*, Copenhagen, Denmark, 2007.

Diab, Mona, Mahmoud Ghoneim and **Nizar Habash**. Arabic Diacritization in the Context of Statistical Machine Translation, In *Proceedings of the Machine Translation Summit (MT-Summit)*, Copenhagen, Denmark, 2007.

Kirchhoff, Katrin, Owen Rambow, **Nizar Habash**, Mona Diab. Semi-Automatic Error Analysis for Large-Scale Statistical Machine Translation Systems, In *Proceedings of the Machine Translation Summit (MT-Summit)*, Copenhagen, Denmark, 2007.

Habash, Nizar, Ryan Gabbard, Owen Rambow, Seth Kulick and Mitch Marcus. Determining Case in Arabic: Learning Complex Linguistic Behavior Requires Complex Linguistic

Features, In *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (EMNLP), Prague, Czech Republic, 2007.

Habash, Nizar and Owen Rambow. Arabic Diacritization through Full Morphological Tagging, In *Proceedings of the North American chapter of the Association for Computational Linguistics (NAACL)*, Rochester, New York, 2007.

Elming, Jakob and **Nizar Habash**. Combination of Statistical Word Alignments Based on Multiple Preprocessing Schemes, In *Proceedings of the North American chapter of the Association for Computational Linguistics (NAACL)*, Rochester, New York, 2007.

Habash, Nizar and Owen Rambow. Morphophonemic and Orthographic Rules in a Multi-Dialectal Morphological Analyzer and Generator for Arabic Verbs, International Symposium on Computer and Arabic Language (ISCAL), Riyadh, Saudi Arabia, 2007.

Habash, Nizar. “Arabic Morphological Representations for Machine Translation.” Book Chapter. In *Arabic Computational Morphology: Knowledge-based and Empirical Methods*. Editors Antal van den Bosch and Abdelhadi Souidi. 2007.

Habash, Nizar, Abdelhadi Souidi, and Tim Buckwalter. “On Arabic Transliteration.” Book Chapter. In *Arabic Computational Morphology: Knowledge-based and Empirical Methods*. Editors Antal van den Bosch and Abdelhadi Souidi. 2007.

2006

Biadisy, Fadi, Jihad El-Sana and **Nizar Habash**. Arabic Online Handwriting Recognition. *International Workshop on Handwriting and Optical Character Recognition, Paris, France, 2006*.

Habash, Nizar, Bonnie Dorr and Christof Monz. Challenges in Building an Arabic Generation-heavy Machine Translation System and Extending it with Statistical Components. In *Proceedings of the Association for Machine Translation in the Americas (AMTA-2006)*, Boston, MA, 2006.

Habash, Nizar. “On Arabic and its Dialects,” *Multilingual Magazine*. #81 Volume 17 Issue 5, 2006.

Habash, Nizar and Owen Rambow. Morphological Analysis for Arabic Dialects. In *Proceedings of COLING-ACL*, Sydney, Australia, 2006.

Sadat, Fatiha and **Nizar Habash**. Morphological Preprocessing Scheme Combination for Statistical MT. In *Proceedings of COLING-ACL*, Sydney, Australia, 2006.

Habash, Nizar and Fatiha Sadat. Arabic Preprocessing Schemes for Statistical Machine Translation, In *Proceedings of the North American chapter of the Association for Computational Linguistics (NAACL)*, New York, 2006.

Chiang, David, Mona Diab, **Nizar Habash**, Owen Rambow, and Safi Shareef. Arabic Dialect Parsing. In *Proceedings of the European chapter of the Association of Computational Linguistics (EACL)*. 2006.

Habash, Nizar, Clinton Mah, Randy Calistri-Yeh, Sabiha Imran and Paraic Sheridan. The Design and Validation of an Arabic WordNet for Information Retrieval. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*. 2006.

Rambow, Owen, Bonnie Dorr, David Farwell, Rebecca Green, **Nizar Habash**, Stephen Helmreich, Eduard Hovy, Lori Levin, Carnegie Keith J. Miller, Teruko Mitamura, Florence Reeder, Advait Siddharthan. Parallel Syntactic Annotation of Multiple Languages. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*. 2006.

Passonneau, Rebecca, **Nizar Habash** and Owen Rambow. Interannotator Agreement on a Multilingual Semantic Annotation Task. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*. 2006.

Maamouri, Mohamed, Ann Bies, Tim Buckwalter, Mona Diab, **Nizar Habash**, Owen Rambow, Dalila Tabessi. Developing and Using a Pilot Dialectal Arabic Treebank. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*. 2006.

2005

Habash, Nizar, Owen Rambow and George Kiraz. Morphological Analysis and Generation for Arabic Dialects. In *Proceedings of the Workshop on Computational Approaches to Semitic Languages at the Conference of American Association for Computational Linguistics (ACL'05)*.

Habash, Nizar and Owen Rambow. Arabic Tokenization, Morphological Analysis, and Part-of-Speech Tagging in One Fell Swoop. In *Proceedings of the Conference of American Association for Computational Linguistics (ACL'05)*.

Darwish, Kareem, Mona Diab and **Nizar Habash**, Eds. Computational Approaches to Semitic Languages. Workshop Proceedings. Association for Computational Linguistics, Ann Arbor, Michigan, 2005.

2004

Habash, Nizar. The Use of a Structural N-gram Language Model in Generation-Heavy Hybrid Machine Translation. In *Proceedings of the Third International Conference of Natural Language Generation (INLG-04)*. Careys Manor, UK, July 2004.

Habash, Nizar. Large Scale Lexeme Based Arabic Morphological Generation. In *Proceedings of Traitement Automatique du Langage Naturel (TALN-04)*. Fez, Morocco, 2004.

Habash, Nizar and Owen Rambow. Extracting a Tree Adjoining Grammar from the Penn Arabic Treebank. In *Proceedings of Traitement Automatique du Langage Naturel (TALN-04)*. Fez, Morocco, 2004.

Habash, Nizar, Bonnie Dorr, Eduard Hovy, Florence Reeder. Eds. Determining Interlingua Utility for Machine Translation. Seventh Interlingua Workshop. Sixth Biennial Conference of the Association for Machine Translation in the Americas (AMTA-04). Georgetown, Washington DC, 2004.

Ayan, Fazil, Bonnie J. Dorr, and **Nizar Habash**, Application of Alignment to Real-World Data: Combining Linguistic and Statistical Techniques for Adaptable MT. In *Proceedings of*

the 6th Conference of the Association for Machine Translation in the Americas (AMTA-2004), Georgetown University, Washington DC, 2004.

Reeder, Florence, Bonnie Dorr, David Farwell, **Nizar Habash**, Stephen Helmreich, Eduard Hovy, Lori Levin, Teruko Mitamura, Keith Miller, Owen Rambow, Advaith Siddharthan. Interlingual Annotation for MT Development. In *Proceedings of the 6th Conference of the Association for Machine Translation in the Americas (AMTA-2004)*, Georgetown University, Washington DC, 2004.

Farwell, David, Stephen Helmreich, Bonnie J. Dorr, **Nizar Habash**, Florence Reeder, Keith Miller, Lori Levin, Teruko Mitamura, Eduard Hovy, Owen Rambow, and Advaith Siddharthan. Interlingual Annotation of Multilingual Text Corpora. In *Proceedings of the North American Chapter of the Association for Computational Linguistics Workshop on Frontiers in Corpus Annotation*, Boston, MA, pp. 55--62, 2004.

Mitamura, Teruko, Keith J. Miller, Bonnie J. Dorr, David Farwell, **Nizar Habash**, Lori Levin, Stephen Helmreich, Eduard Hovy, Lori Levin, Owen Rambow, Reeder, Florence, and Advaith Siddharthan. Semantic Annotation of Multilingual Text Corpora. In *Proceedings of the Workshop on Beyond Named Entity Recognition: Semantic Labeling for NLP Tasks*, LREC, Portugal, 2004.

Dorr, Bonnie J., Rebecca Green, Lori Levin, Owen Rambow, David Farwell, **Nizar Habash**, Stephen Helmreich, Eduard Hovy, Keith J. Miller, Teruko Mitamura, Florence Reeder, and Advaith Siddharthan. Semantic Annotation and Lexico-Syntactic Paraphrase. In *Proceedings of the Workshop on Building Lexical Resources from Semantically Annotated Corpora*, LREC, Portugal, 2004.

2003

Dorr, Bonnie J., Necip Fazil Ayan, **Nizar Habash**, Nitin Madnani, and Rebecca Hwa. Rapid Porting of DUSTER to Hindi. *ACM Transactions on Asian Language Information Processing (TALIP)*, 2:3, 2003.

Habash, Nizar. Matador: A Large Scale Spanish-English GHMT System. In *Proceedings of the MT Summit*, New Orleans, LA, pp. 149--156, 2003.

Cavalli-Sforza, Violetta, Alon Lavie and **Nizar Habash**. Eds. Proceedings of the MT Summit IX Workshop on Machine Translation for Semitic Languages: Issues and Approaches. September 23, 2003, New Orleans, LA, USA. [URL](#)

Habash, Nizar. Generation-Heavy Hybrid Machine Translation. Doctoral Dissertation. Computer Science Department, University of Maryland College Park, 2003.

Habash, Nizar and Bonnie Dorr, A Categorical Variation Database for English, *Proceedings of North American Association for Computational Linguistics*, Edmonton, Canada, pp. 96--102, 2003.

Habash, Nizar, Bonnie Dorr, and David Traum. Hybrid Natural Language Generation from Lexical Conceptual Structures. *MT Journal* volume 18 (2): 81-128, 2003.

2002

Habash, Nizar and Bonnie Dorr. Handling Translation Divergences: Combining Statistical and Symbolic Techniques in Generation-Heavy Machine Translation. In *Proceedings of the Fifth Conference of the Association for Machine Translation in the Americas, AMTA-2002*, Tiburon, CA, 2002.

Dorr, Bonnie and **Nizar Habash**. Interlingua Approximation: A Generation-Heavy Approach. In *Proceedings of Workshop on Interlingua Reliability, Fifth Conference of the Association for Machine Translation in the Americas, AMTA-2002*, Tiburon, CA, 2002.

Dorr, Bonnie, Lisa Pearl, Rebecca Hwa and **Nizar Habash**. DUSTER: A Method for Unraveling Cross-Language Divergences for Statistical Word-Level Alignment. In *Proceedings of the Fifth Conference of the Association for Machine Translation in the Americas, AMTA-2002*, Tiburon, CA, 2002.

Habash, Nizar. Generation-Heavy Hybrid Machine Translation. In *Proceedings of the International Natural Language Generation Conference (NLG-02)*. New York, 2002.

2001

Habash, Nizar and Bonnie Dorr. Large Scale Language Independent Generation Using Thematic Hierarchies. In *Proceedings of the MT Summit VIII*. Santiago de Compostella, Spain. 2001.

2000

Habash, Nizar. oxyGen: A Language Independent Language Realization Engine. In *Proceedings of the Fourth Conference of the Association for Machine Translation in the Americas, AMTA-2000*. Cuernavaca, Mexico.

Traum, David and **Nizar Habash**. Generation from Lexical Conceptual Structures. Workshop on Applied Interlinguas, ANLP-2000. Seattle, WA.

1999

Habash, Nizar. Nuun: A System for Developing Platform and Browser Independent Arabic Web Applications. In *Proceedings of the Arabic Translation and Localization Conference (ATLAS-99)*. Tunis, Tunisia, 1999. Republished in Arabic in the Arab Journal of Science, 33, June 1999.

Habash, Nizar. Issues in Palestinian Arabic Spelling Standardization. NACAL 27, 1999. Baltimore, MD.

1998

Dorr, Bonnie, **Nizar Habash** and David Traum. A Thematic Hierarchy for Efficient Generation from Lexical-Conceptual Structures. In *Proceedings of the Association of Machine Translation in the Americas, AMTA-98*. Longhorne, PA.

Habash, Nizar. Introduction to Delason: The Complete Guide to the Artificial Language. *Unpublished Manuscript*, 1998.

Invited Talks

2009

Habash, Nizar. Arabic Morphological Analysis and Disambiguation: Problem, Solution, Applications. Invited Talk. Linguistics Speaker Series. Georgetown University, Washington D.C., 2009.

2008

Habash, Nizar. Treebanking for Machine Transaltion. Invited Panel Talk, *GALE PI Meeting*, Tampa, Florida, 2008.

Habash, Nizar. Techniques for Online OOV Handling in Phrase-based Statistical Machine Translation. *Columbia University*, April 2008

Habash, Nizar. Five Techniques for OOV Reduction in Arabic-English and Urdu-English SMT. *NIST MTEval Workshop*. March 2008.

2007

Habash, Nizar. Arabic Diacritization through Full Morphological Tagging. Invited Talk, Language Technology Institute Seminar Series. Carnegie Mellon University, 2007.

2006

Habash, Nizar. Arabic Dialect Modeling: from Morphological Analysis to Parsing. Invited Talk, NSF Funded US-Morocco Workshop on Language Technology Research and Education. Ecole Nationale de l'Industrie Minérale (Rabat, Morocco) May 29 - June 2, 2006.

Habash, Nizar. Disambiguation of Rich Arabic Morphological Analyses. Invited Talk, NYU Natural Language Processing Colloquium. April 14, 2006.

2005

Mona Diab, **Nizar Habash**, and Owen Rambow. Arabic Dialect Parsing. Invited Talk, Linguistic Data Consortium, University of Pennsylvania, December 9, 2005

Habash, Nizar. Combining Symbolic and Statistical Techniques for Machine Translation and Morphological Disambiguation. Invited Talk, AT&T Labs, Florham Park, NJ, October 7, 2005.

Habash, Nizar. Generation-Heavy Hybrid Machine Translation. Invited Talk. *IBM TJ Watson Research Center*, March 9, 2005.

Habash, Nizar. Generation-Heavy Hybrid Machine Translation. Invited Talk. Computer Science Department Seminar, *City College of New York*, March 8, 2005.

2004

Habash, Nizar. Generation-Heavy Hybrid Machine Translation. Invited Talk. Language Technologies Institute Seminar. *Carnegie Mellon University*, November 19, 2004.

2003

Habash, Nizar. Semitic Linguistic Phenomena. Invited Talk. *Workshop on Machine Translation for Semitic Languages, MT Summit*, New Orleans, LA, 2003. [Announcement](#)

Software and Resources

- CATiB v1.0: Columbia Arabic Treebank (2009)
- REMOOV: Tool for Reducing Out-Of-Vocabulary in MT (2007)
- TOKAN: General Arabic Tokenizer. (2006)
- MADA Morphological Analysis and Disambiguation for Arabic . (2005)
- Qumqum A Generation-Heavy Arabic-English Machine Translation System. (2005)
- Aragen Arabic Lexeme-based Morphological Generator/Analyzer. (2004)
- K-Span: Kimmo-based Spanish Morphological Analyzer. Online Demo. 2003. (Dorr and Habash)
- Matador A Generation-Heavy Spanish-English Machine Translation System. (2003)
- Exerge Expansive Rich Generation for English System. (2003)
- CatVar A Categorical Variation Database for English (2002)

- [Sebawai Online](#) Nuun-based interface to the Arabic Morphological Analyzer Sebawai (by Kareem Darwish) (2001)
- [English Lexical Conceptual Structure Lexicon](#) web-browsable interface (2001)
- [Chinese Lexical Conceptual Structure Lexicon](#) web-browsable interface (2001)
- [Spanish Lexical Conceptual Structure Lexicon](#) web-browsable interface (2001)
- [Multi-Faceted Tree Viewer \(MFTV\)](#) A zoomable interface for viewing ambiguous meaning representation of sentences. (1999)
- [ChinMT](#) An LCS-based Chinese-English MT system. (1999)
- [Arabic NuunForm](#) Java Arabic/English input method (1999)
- [Hebrew Nuun Form](#) Java Hebrew/English input method (1999)
- [Phoenician NuunForm](#) Java Phoenician/English input method (1999)
- [NuunPad](#) Arabic notepad for non-Arabic Windows (1998)
- [Nuun Web Site](#) The original Nuun website which I fully designed and implemented (1998)
- [Arabix Page](#) A Java-based solution for Arabic display on the internet (1996)

SERVICE

Professional

- Doctoral Committees
 - Otakar Smrz, Charles University, Czech Republic, 2007.
 - Josep Crego, Universitat Politècnica de Catalunya, 2008.
 - Jakob Elming, Copenhagen Business School, 2009.
- Conference/Workshop organization
 - Workshop program co-chair for NAACL 2009
 - Senior Program Committee member for NAACL 2007 (machine translation track).
 - Program co-chair for the conference of the Association for Machine Translation in the Americas (AMTA'06)
 - One of three organizers for a workshop on Computational Approaches to Semitic Languages ACL 2005.
 - Chair of the *Workshop on Determining Interlingua Utility for Machine Translation*. AMTA-2004, Washington DC, 2004.
 - One of three organizers of the *Workshop on Machine Translation for Semitic Languages, MT Summit*, New Orleans, LA, 2003.
- Professional Organizations
 - Secretary of the board of the International Association for Machine Translation (IAMT) (2009-)
 - Secretary of the board of the Association for Machine Translation in the Americas (AMTA) (2008-)
 - Research community representative in the board of the Association for Machine Translation in the Americas (AMTA) (2006-2008)
 - Vice president of the ACL Special Interest Group on Semitic Languages (2006-2009)
- Advisory Committees
 - Member of an advisory committee to NIST MT Eval program
 - Co-chair of the advisory committee to improving the quality of the Arabic Treebank, 2006-2008.
 - Member of the machine translation team under the NSF funded MINDS workshop, 2006.

- Reviewing/Program Committee
 - NSF review panel (2009).
 - Member of the editorial board of the Journal for Machine Translation (2007-)
 - NSF proposals (2005, 2006).
 - Journal of Computational Linguistics, (2003, 2006).
 - IEEE Transactions on Audio, Speech and Language Processing (2004, 2005, 2006, 2007).
 - Journal of Natural Language Engineering, 2004.
 - ACM Transaction on Asian Languages Information Processing (TALIP), 2004.
 - Conferences: EMNLP 2009 (multiple tracks), AMTA 2008, EMNLP 2008, COLING 2008, ACL 2008 (Machine Translation track and Morphology/Phonology track), LREC 2008, ACL 2007 (Main and Semitic Languages Workshop), MT Summit 2007, EMNLP 2007, TMI 2007, CIATALA 2006, LREC 2006, EAACL 2006, NAACL 2006, ACL 2006 (Main and Student Workshop), AMTA 2006, *COLING* 2004 (*Workshop on Arabic-script based languages*), FGVienna 2003 (The Conference on Formal Grammar), INLG 2002 (main and student workshop), ACL 2002, NAACL 2001 (Student Research Workshop)

University

- Part of the team proposing the Columbia Arabic Language Institute for Modeling and Technology (CALIMAT)
- Heavily involved in the design and implementation of the website for the Center for Computational Learning Systems, Columbia University, 2005.
- Assisted with technical organizations for several Interim Progress Reviews and open houses at the Computational Linguistics and Information Processing Lab in the University of Maryland College Park. (1997-2003)
- Student host for several invited speakers at the University of Maryland College Park. (2001-2002)
- Coordinator of renovation and equipment/space allocation in CLIP laboratory (University of Maryland College Park, 2001- 2002)
- Assisted the Arabic language professor at Old Dominion University by reading texts aloud in her class on regular basis. (1993-1994)

Community

- Interviewed for an article on machine translation in Wired Magazine ([Issue 14.12](#) - December 2006)
- One of two invited guests to discuss the Middle East Conflict in a first year Hebrew language class (University of Maryland College Park, October 2002 and March 2004)
- One of five international students featured in [an article in Old Dominion University's Alumni Magazine](#). Spring 2002.
- Mediator among three invited panelists discussing views on the Middle East conflict as part of Palestine Day events (Organization of Arab Students, University of Maryland College Park, March 2002).
- Active member of several campus groups in the University of Maryland College Park including Organization of Arab Students, Graduate Lambda Coalition, and Students for Justice in Palestine (1997-2004).

- One of five graduating students featured in [a newspaper article "Pomp & Circumstance"](#) The Virginian-Pilot, May 9, 1997.

MISCELLANEOUS

- ***Citizenship***: Jordanian. United States resident (green card holder).
- ***Languages***: Palestinian Arabic (native), Modern Standard Arabic, English (near-native fluency), French (basic), Hebrew (basic), Spanish (basic).
- ***Interests, Hobbies***
 - Handcrafts, sculpture, quilting, dancing, poetry, opera, artistic experimentation, languages.
 - Designed and wrote a book on a constructed language called Delason (<http://www.delason.com>).
 - Currently working on an Arabic-Hebrew esperanto called Semitish as part of an ambitious constructed culture art project exploring a peaceful vision of the Middle East (<http://www.palisra.com/>).

REFERENCES

Available upon request