Question Answering and Machine Comprehension

1. Pranav Rajpurkar et al.: SQuAD: 100,000+ Questions for Machine Comprehension of Text. EMNLP 2015.
2. Minjoon Soo et al.: Bi-Directional Attention Flow for Machine Comprehension. ICLR 2015.

Automatic Text Summarization

1. Kevin Knight and Daniel Marcu: Summarization beyond sentence extraction. Artificial Intelligence 139, 2002.
2. James Clarke and Mirella Lapata: Modeling Compression with Discourse Constraints. EMNLP-CONLL 2007.
3. Wen-tau Yih et al.: Multi-Document Summarization by Maximizing Informative Content-Words. IJCAI 2007.
4. Alexander M Rush, et al.: A Neural Attention Model for Sentence Summarization.
5. Abigail See et al.: Get To The Point: Summarization with Pointer-Generator Networks. ACL 2017.

Coreference Resolution

1. Vincent Ng: Supervised Noun Phrase Coreference Research: The First Fifteen Years, ACL 2010.
2. Kenton Lee at al.: End-to-end Neural Coreference Resolution, EMNLP 2017.

Machine Translation & Transliteration, Sequence-to-Sequence Models

1. Peter F. Brown et al.: A Statistical Approach to Machine Translation, Computational Linguistics, 1990.
2. Kevin Knight, Graehl Jonathan. Machine Transliteration. Computational Linguistics, 1992.
3. Dekai Wu: Inversion Transduction Grammars and the Bilingual Parsing of Parallel Corpora, Computational Linguistics, 1997.
4. Kevin Knight: A Statistical MT Tutorial Workbook, 1999.
5. Kishore Papineni, et al.: BLEU: a Method for Automatic Evaluation of Machine Translation, ACL 2002.
6. Philipp Koehn, Franz J Och, and Daniel Marcu: Statistical Phrase-Based Translation, NAACL 2003.
7. Philip Resnik and Noah A. Smith: The Web as a Parallel Corpus, Computational Linguistics, 2003.
8. Franz J Och and Hermann Ney: The Alignment-Template Approach to Statistical Machine Translation, Computational Linguistics, 2004.
9. David Chiang. A Hierarchical Phrase-Based Model for Statistical Machine Translation, ACL 2005.
10. Ilya Sutskever, Oriol Vinyals, and Quoc V. Le: Sequence to Sequence Learning with Neural Networks, NIPS 2014.
11. Oriol Vinyals, Quoc Le: A Neural Conversation Model, 2015.
12. Dzmitry Bahdanau, et al.: Neural Machine Translation by Jointly Learning to Align and Translate, 2014.
13. Minh-Thang Luong, et al.: Effective Approaches to Attention-based Neural Machine Translation, 2015.
14. Rico Sennrich et al.: Neural Machine Translation of Rare Words with Subword Units. ACL 2016.

15. Yonghui Wu, et al.: Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation, 2016.
16. Melvin Johnson, et al.: Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation, 2016.
17. Jonas Gehring, et al.: Convolutional Sequence to Sequence Learning, 2017.
18. Ashish Vaswani, et al.: Attention Is All You Need, 2017.

Sequential Labeling & Information Extraction

1. Marti A. Hearst: Automatic Acquisition of Hyponyms from Large Text Corpora, COLING 1992.
2. Collins and Singer: Unsupervised Models for Named Entity Classification, EMNLP 1999.
3. Patrick Pantel and Dekang Lin, Discovering Word Senses from Text, SIGKDD, 2002.
4. Mike Mintz et al.: Distant supervision for relation extraction without labeled data, ACL 2009.
5. Zhiheng Huang et al.: Bidirectional LSTM-CRF Models for Sequence Tagging, 2015.
6. Xuezhe Ma and Eduard Hovy: End-to-end Sequence Labeling via Bi-directional LSTM-CNNs-CRF, ACL 2016.

Segmentation, Tagging, Parsing

1. Donald Hindle and Mats Rooth. Structural Ambiguity and Lexical Relations, Computational Linguistics, 1993.
2. Adwait Ratnaparkhi: A Maximum Entropy Model for Part-Of-Speech Tagging, EMNLP 1996.
3. Eugene Charniak: A Maximum-Entropy-Inspired Parser, NAACL 2000.
4. Michael Collins: Discriminative Training Methods for Hidden Markov Models: Theory and Experiments with Perceptron Algorithms, EMNLP 2002.
5. Dan Klein and Christopher Manning: Accurate Unlexicalized Parsing, ACL 2003.
6. Joakim Nivre and Mario Scholz: Deterministic Dependency Parsing of English Text, COLING 2004.
7. Ryan McDonald et al.: Non-Projective Dependency Parsing using Spanning-Tree Algorithms, EMNLP 2005.
8. Daniel Andor et al.: Globally Normalized Transition-Based Neural Networks, 2016.
9. Oriol Vinyals, et al.: Grammar as a Foreign Language, 2015.

Language Modeling

1. Joshua Goodman: A bit of progress in language modeling, MSR Technical Report, 2001.
2. Stanley F. Chen and Joshua Goodman: An Empirical Study of Smoothing Techniques for Language Modeling, ACL 2006.
3. Yee Whye Teh: A Bayesian interpretation of Interpolated Kneser-Ney, 2006.
4. Yoshua Bengio, et al.: A Neural Probabilistic Language Model, J. of Machine Learning Research, 2003.
5. Andrej Karpathy: The Unreasonable Effectiveness of Recurrent Neural Networks, 2015.
6. Yoon Kim, et al.: Character-Aware Neural Language Models, 2015.
7. Alec Radford, et al.: Language Models are Unsupervised Multitask Learners, 2018.

Topic Models

1. Thomas Hofmann: Probabilistic Latent Semantic Indexing, SIGIR 1999.
2. David Blei, Andrew Y. Ng, and Michael I. Jordan: Latent Dirichlet Allocation, J. Machine Learning Research, 2003.

Clustering & Word/Sentence Embeddings

1. Peter F Brown, et al.: Class-Based n-gram Models of Natural Language, 1992.
2. Tomas Mikolov, et al.: Efficient Estimation of Word Representations in Vector Space, 2013.
3. Tomas Mikolov, et al.: Distributed Representations of Words and Phrases and their Compositionality, NIPS 2013.
4. Quoc V. Le and Tomas Mikolov: Distributed Representations of Sentences and Documents, 2014.
5. Jeffrey Pennington, et al.: GloVe: Global Vectors for Word Representation, 2014.
6. Ryan Kiros, et al.: Skip-Thought Vectors, 2015.
7. Piotr Bojanowski, et al.: Enriching Word Vectors with Subword Information, 2017.
8. Daniel Cer et al.: Universal Sentence Encoder, 2018.

Neural Models

1. Richard Socher, et al.: Dynamic Pooling and Unfolding Recursive Autoencoders for Paraphrase Detection, NIPS 2011.
2. Ronan Collobert et al.: Natural Language Processing (almost) from Scratch, J. of Machine Learning Research, 2011.
3. Richard Socher, et al.: Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank, EMNLP 2013.
4. Xiang Zhang, Junbo Zhao, and Yann LeCun: Character-level Convolutional Networks for Text Classification, NIPS 2015.
5. Yoon Kim: Convolutional Neural Networks for Sentence Classification, 2014.
6. Christopher Olah: Understanding LSTM Networks, 2015.
7. Matthew E. Peters, et al.: Deep contextualized word representations, 2018.
8. Jacob Devlin, et al.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, 2018.
9. Yihan Liu et al. RoBERTa: A Robustly Optimized BERT Pretraining Approach, 2020.

Machine Learning

1. Avrim Blum and Tom Mitchell: Combining Labeled and Unlabeled Data with Co-Training, 1998.
2. John Lafferty, Andrew McCallum, Fernando C.N. Pereira: Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data, ICML 2001.
3. Charles Sutton, Andrew McCallum. An Introduction to Conditional Random Fields for Relational Learning.
4. Kamal Nigam, et al.: Text Classification from Labeled and Unlabeled Documents using EM. Machine Learning, 1999.
5. Kevin Knight: Bayesian Inference with Tears, 2009.

6. Marco Tulio Ribeiro et al.: "Why Should I Trust You?": Explaining the Predictions of Any Classifier, KDD 2016.
7. Marco Tulio Ribeiro et al.: Beyond Accuracy: Behavioral Testing of NLP Models with CheckList, ACL 2020.

Generation, Reinforcement Learning

1. Jiwei Li, et al.: Deep Reinforcement Learning for Dialogue Generation, EMNLP 2016.
2. Marc'Aurelio Ranzato et al.: Sequence Level Training with Recurrent Neural Networks. ICLR 2016.
3. Samuel R Bowman et al.: Generating sentences from a continuous space, CoNLL 2016.
4. Lantao Yu, et al.: SeqGAN: Sequence Generative Adversarial Nets with Policy Gradient, AAAI 2017.