15 Other Sequence-to-sequence Applications

Up until now, we have largely used machine translation as an example of sequence-to-sequence learning tasks. However, as mentioned at the beginning of the course, sequence-to-sequence models are quite general, and can be used for a large number of tasks. This chapter provides a brief survey of some other sequence-to-sequence tasks, and describes some of the unique features that make these tasks difficult or different from machine translation.

15.1 Dialog

Another interesting application area of sequence-to-sequence models is **dialog systems**. In this case, the input to the system F is an utterance by a user, and the output E is a reply generated by the system. Models for such reply generation have been created using both phrase-based [47] and neural [55, 49] translation systems.

One interesting aspect of dialog that makes it far from a straightforward application of sequence-to-sequence models is that we can expect a great **diversity** in the responses that we will expect for a certain input utterance. For example, the input "what is your name?" would only have one or a few correct translations in an MT setting, but it would have a myriad of correct conversational responses in a dialog setting. There have been a few methods proposed to resolve this problem, including the introduction of an alternative objective function for decoding [35]:

$$\hat{E} = \underset{E}{\operatorname{argmax}} \ \log P(E \mid F) - \lambda \log P(E) \tag{151}$$

The first term here is the standard one used in decoding for sequence-to-sequence models, while the second term promotes diversity (to the extent suggested by parameter λ) by penalizing results that are likely regardless of the input (such as "i don't know"). There are also a number of different methods related to diversity, including returning responses that belong to different clusters [28], or adding information about the speaker to ensure that a response reflects the trait of that speaker [36].

This diversity also poses a problem for evaluation of such systems, and [39] show that standard evaluation metrics such as BLEU are of little use in the context of dialog. There are ways to ameliorate this problem, such as using a large number of references weighted by human qualitative judgements [18]. However, this is limited to the cases where these annotations are available, and the fundamental problem of automatic evaluation is far from solved.

Another feature of dialog, is that it is heavily reliant on context, and access of external knowledge that the dialog system may be expected to have. This is particularly true for dialogs in which the user is expecting to perform a task such as making a restaurant reservation, called **task-based dialog**. Within this context, it is common to have an underlying dialog manager that handles this long-term context, then use a sequence-to-sequence model to perform only the language generation step based on the context provided by this dialog manager [58]. However, there are also some models for end-to-end trainable task-based dialog systems that also take into account context [60].

15.2 Monolingual Translation Tasks

There are also a number of sequence-to-sequence transduction tasks that are performed within a single language, translating, for example, English into English.

15.2.1 Summarization

One typical example of this is **text summarization**. In the summarization task, one is required to take a larger body of text and convert it into a smaller amount of text containing the same information for browsing purposes. This can be done at a number of levels:

Sentence Compression: The problem of compressing a single sentence into a shorter single sentence [31].

Single-document Summarization: The problem of compressing a single document into a shorter summary [10].

Multi-document Summarization: the problem of reducing the information in multiple documents into a single summary [7].

There are also typically two types of summarization: **extractive summarization** and **abstractive summarization**. In extractive summarization, we simply choose some content (usually one sentence at a time), and add these to the summary. In contrast, in abstractive summarization we actually generate a new summary, and systems using this approach have been created using the sequence-to-sequence models introduced in this course.

One unique element of summarization is that it is largely concerned with removing irrelevant content. Thus, many attempts, both using symbolic systems and neural systems, focus on simply deleting words [31, 16]. In particular, tree-based methods that explicitly use syntax have found some favor, as this is a natural way to model that fact that we can "chop off" irrelevant phrases without a major change in the main content [44]. It is common to frame these problems as a constrained optimization problem, we want to delete words to achieve a summary with a certain length while maximizing the amount of relevant content that remains in the summary.

There have also been a number of methods that move beyond only deletion, and frame the problem as a sequence-to-sequence transduction problem. Successful methods have used tree substitution grammars [14], and attentional neural networks [48]. These models can be equipped with special mechanisms to copy words [25], or control the length of the summary [30].

Summarization systems are generally evaluated based on the amount of recall of important information that can be achieved within the limited summary length. The standard measure is ROUGE, which measures recall over n-grams [37], and it is also common to perform manual human evaluation as well.

Interested readers can find a more complete survey in [19].

15.2.2 Paraphrase Generation

Another example of translation between two sentences in the same language is paraphrasing: re-wording sentences into other sentences with the same content but different surface features.

This technology has a number of applications including query expansion for information retrieval [51], improving robustness of machine translation to lexical variations [9], converting the style or register of text [45, 62], or sentence simplification for reading assistance [66, **?**]. These works take approaches that are based on phrase-based machine translation [9, 45] tree-based machine translation [66], or neural methods [57].

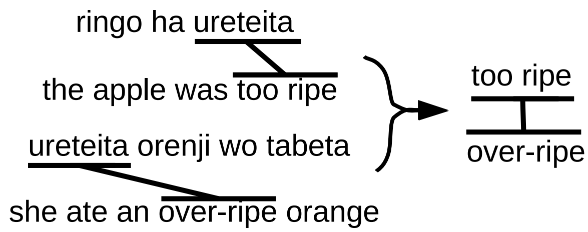

Figure 53: An example of extracting monolingual paraphrases from bilingual phrases.

One interesting aspect of paraphrasing is that it is possible to create paraphrasing models without explicitly aligned parallel data. [6] describe a simple method to extract phrasal paraphrase candidates from bilingual machine translation training data using pivoting, as shown in Figure 53. Basically, the idea is that we can calculate the probability of a paraphrase between English phrases $P(\boldsymbol{e}_2 \mid \boldsymbol{e}_1)$ by marginalizing over the probability of phrases in the source language:

$$P(\boldsymbol{e}_2 \mid \boldsymbol{e}_1) = \sum_{\boldsymbol{f}} P(\boldsymbol{e}_2 \mid \boldsymbol{f}) P(\boldsymbol{f} \mid \boldsymbol{e}_1). \tag{152}$$

This means that if we can extract a phrase table from a parallel text, as described in Section 13, we can build a paraphrasing model with no annotated parallel text. This overall paradigm has proven quite effective, and is now the basis for the widely used paraphrase database PPDB [20].[46]

One other difficult aspect of paraphrase generation is how to evaluate the generated paraphrases. We would like the paraphrase to be accurate and fluent, but we also need to ensure that they need to be significantly different from the original text. One example of an evaluation measure this is PINC [12], which is like BLEU but considers not only the BLEU score, but also the dissimilarity from the original input.

Interested readers can find an extensive survey in [2] or on `http://paraphrasing.org`.

15.3 Recognition/Generation of Continuous Inputs

While most of the previous sections have covered applications that take sequences of discrete inputs and generate sequences of discrete outputs, there are also a large number of works on modeling continuous inputs or ouputs, such as speech or images.

15.3.1 Sequence Generation from Continuous Inputs

One classical task that is a sequence-to-sequence modeling problem where the input sequence is continuous is **speech recognition** (often abbreviated ASR for "automatic speech recognition"). Classical approaches to speech recognition take a form similar to the WFST-based

[46] `http://paraphrase.org/`

symbolic translation models in Section 12 [43]. The main difference between recognition and translation is that now instead of creating a translation model $P(f \mid e)$ that gives us the likelihood of a source word given the target word, we create an acoustic model $P(x \mid y)$ that gives the probability of acoustic features x given a phoneme y. It is also common to flip this probabiltiy $P(y \mid x)$ These phonemes are then combined into words which are scored by the language model.

Acoustic models are now almost exclusively modeled using deep neural networks that either take in the acoustic features for a single frame x and predict its phoneme label y [42], or take in a whole sequence X, encode it with a recurrent network (such as bi-directional LSTMs [23]), and predict the probabilities based on this whole sequence worth of information. One interesting method that can be applied to these problems **connectionist temporal classification** (CTC), which automatically induces an alignment between phonemes and corresponding frames using dynamic programming, and uses the alignments to train the neural network [22]. There have also been some promising preliminary results on end-to-end speech recognition with neural networks [11], which take in a sequence of speech features, and directly try to predict the output as characters or words.

Speech recognition is generally evaluated using word error rate, which directly measures the number of insertions, deletions, or substitutions necessary to turn the output into the reference text.

Another example of continuous inputs is images, and **caption generation** models have been applied to transform images into captions [41]. One major difference between images and sequences is that images are consistent in two-dimensional space, and because of this, most image captioning methods use two-dimensional convolutional neural networks [34] to encode the input. Once the input is encoded, it is common to use attention-based models similar to those described in Section 8, often with additional improvements [29, 56]. It is also possible to extend these models to describing videos, which requires additional management of information across multiple time-steps, which can be done by using a recurrent neural network over these time steps [56]. Another hybrid task that considers both text and visual input is **visual question answering**, in which the input is both an image (visual) and a question (textual), and the model is required to output an answer [3].

15.3.2 Generating Continuous Outputs

All of the tasks above can also be reversed into a task that takes a discrete input sequence and outputs a continuous output such as speech or images.

Text-to-speech conversion, or **speech synthesis**, is the generation of speech from text, and models to do so generally stich together existing wave forms in a coherent way [26], or generate speech using models such as hidden Markov models [65] and deep neural networks [64]. One method that has recently proven effective in the speech synthesis area uses dialated convolutional neural networks, which use convolutions with gradually increasing spans in the decoder portion of the network [53].[47] There are also methods for **voice conversion**, which map a sequence of speech frames to another sequence of speech frames in the voice of another speaker [52].

Speech synthesis models are often evaluated using **mel-frequency cepstral distortion** [32], which is a measure of difference between reference speech and the generated speech. This

[47] These dialated convolutional networks have also proven useful in modeling text.[27]

is an incomplete measure, however, and manual listening tests are often employed as well.

Another example of generating continuous outputs is image generation from captions. This can be done with both recurrent [24] neural networks, or **deconvolutional networks** [63], which reverse the order of the convolution to generate from compressed representations to individual image pixels. One method that has evolved from this image generation task to be used in a number of other areas is **generative adversarial networks** (GANs) [21]. The idea of GANs is that in addition to our *generator* neural network, we also have a *discriminator* network that tries to distinguish between generated and proposed outputs. The training objective of the generator is then modified to both assign high probability to true images, and allow the generator to learn to generate images that "fool" the discriminator. This makes images that are more natural, as any particularities of generated images that may be picked up on by the discriminator will be explicitly penalized.

15.4 Models of Structured Data

Finally, there are sequence-to-sequence models that attempt to generate sequential data with an explicit structure, such as tree structure or graph structure.

One widely researched example of this is syntactic parsing, which tries to find the syntactic structure of sentences as shown in Section 14. In general, this syntactic structure is created using specialized algorithms such as the CKY algorithm, with refinements in how we learn the grammar, etc. [40]. [54] have also shown that generation of parse trees can be performed using standard encoder-decoder models, by linearizing the parse tree into a bracketed sequence representing the original tree structure. More sophisticated models combine structure with neural networks, leveraging syntactic constraints to help the models learn more effectively for the task at hand [50, 15]. Syntactic parsing is generally evaluated based on the accuracy of trees, specifically bracketing F-measure.

Another structured generation task is **semantic parsing**, which attempts to generate an analysis of a natural language utterance in a form that is easy to use for downstream applications. These semantic representations can take various forms, with one recently popular form being the **abstract meaning representation** (AMR) [5], a graph-based representation of who does what to whom. It is also common to generate task-specific representations, the most common of which being for performing question answering [61], giving commands [4], or even generating general-purpose programs [38]. In addition to methods tailored specifically to generating trees, there are also methods for performing semantic parsing using general-purpose sequence-to-sequence models, both symbolic [1] and neural [33]. The accuracy of semantic parses can be measured either using direct evaluation of the semantic structures themselves [8], or through extrinsic evaluation on how well the representations such as how well a question answerer can answer questions [13].

Natural language generation performs transformation in the opposite direction, generating natural language sentences from semantic representations. In many cases this can be done with rule-based models [46]. However, there is also significant work in generation using data-driven approaches such as tree-based symbolic models [17], or neural models [59]. Evaluation of generated language can be done by automatic metric such as BLEU, but it is common to rely on manual effort to provide a final evaluation.

15.5 Exercise

A potential exercise for this section would be to find and download a data set for one of these tasks, and run your sequence-to-sequence model on it and observe the results.

References

[1] Jacob Andreas, Andreas Vlachos, and Stephen Clark. Semantic parsing as machine translation. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 47–52, 2013.

[2] Ion Androutsopoulos and Prodromos Malakasiotis. A survey of paraphrasing and textual entailment methods. *Journal of Artificial Intelligence Research*, 38:135–187, 2010.

[3] Stanislaw Antol, Aishwarya Agrawal, Jiasen Lu, Margaret Mitchell, Dhruv Batra, C Lawrence Zitnick, and Devi Parikh. Vqa: Visual question answering. pages 2425–2433, 2015.

[4] Yoav Artzi and Luke Zettlemoyer. Weakly supervised learning of semantic parsers for mapping instructions to actions. *Transactions of the Association for Computational Linguistics*, 1:49–62, 2013.

[5] Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. Abstract meaning representation (amr) 1.0 specification. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 1533–1544, 2012.

[6] Colin Bannard and Chris Callison-Burch. Paraphrasing with bilingual parallel corpora. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 597–604, 2005.

[7] Regina Barzilay, Kathleen R. McKeown, and Michael Elhadad. Information fusion in the context of multi-document summarization. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 550–557, 1999.

[8] Shu Cai and Kevin Knight. Smatch: an evaluation metric for semantic feature structures. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 748–752, 2013.

[9] Chris Callison-Burch, Philipp Koehn, and Miles Osborne. Improved statistical machine translation using paraphrases. In *Proceedings of the 2006 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL)*, pages 17–24, 2006.

[10] Jaime Carbonell and Jade Goldstein. The use of mmr, diversity-based reranking for reordering documents and producing summaries. pages 335–336. ACM, 1998.

[11] William Chan, Navdeep Jaitly, Quoc Le, and Oriol Vinyals. Listen, attend and spell: A neural network for large vocabulary conversational speech recognition. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 4960–4964. IEEE, 2016.

[12] David Chen and William Dolan. Collecting highly parallel data for paraphrase evaluation. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 190–200, 2011.

[13] James Clarke, Dan Goldwasser, Ming-Wei Chang, and Dan Roth. Driving semantic parsing from the world's response. pages 18–27, 2010.

[14] Trevor Cohn and Mirella Lapata. Sentence compression beyond word deletion. In *Proceedings of the 22th International Conference on Computational Linguistics (COLING)*, pages 137–144, 2008.

[15] Chris Dyer, Adhiguna Kuncoro, Miguel Ballesteros, and Noah A. Smith. Recurrent neural network grammars. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, pages 199–209, 2016.

[16] Katja Filippova, Enrique Alfonseca, Carlos A. Colmenares, Lukasz Kaiser, and Oriol Vinyals. Sentence compression by deletion with lstms. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 360–368, 2015.

[17] Jeffrey Flanigan, Chris Dyer, Noah A. Smith, and Jaime Carbonell. Generation from abstract meaning representation using tree transducers. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, pages 731–739, 2016.

[18] Michel Galley, Chris Brockett, Alessandro Sordoni, Yangfeng Ji, Michael Auli, Chris Quirk, Margaret Mitchell, Jianfeng Gao, and Bill Dolan. deltableu: A discriminative metric for generation tasks with intrinsically diverse targets. *arXiv preprint arXiv:1506.06863*, 2015.

[19] Mahak Gambhir and Vishal Gupta. Recent automatic text summarization techniques: a survey. *Artificial Intelligence Review*, 47(1):1–66, 2017.

[20] Juri Ganitkevitch, Benjamin Van Durme, and Chris Callison-Burch. Ppdb: The paraphrase database. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, pages 758–764, 2013.

[21] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Proceedings of the 28th Annual Conference on Neural Information Processing Systems (NIPS)*, pages 2672–2680, 2014.

[22] Alex Graves, Santiago Fernández, Faustino Gomez, and Jürgen Schmidhuber. Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks. pages 369–376. ACM, 2006.

[23] Alex Graves and Jürgen Schmidhuber. Framewise phoneme classification with bidirectional lstm and other neural network architectures. *Neural Networks*, 18(5):602–610, 2005.

[24] Karol Gregor, Ivo Danihelka, Alex Graves, Danilo Jimenez Rezende, and Daan Wierstra. Draw: A recurrent neural network for image generation. *arXiv preprint arXiv:1502.04623*, 2015.

[25] Jiatao Gu, Zhengdong Lu, Hang Li, and Victor O.K. Li. Incorporating copying mechanism in sequence-to-sequence learning. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 1631–1640, 2016.

[26] Andrew J Hunt and Alan W Black. Unit selection in a concatenative speech synthesis system using a large speech database. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 1, pages 373–376. IEEE, 1996.

[27] Nal Kalchbrenner, Lasse Espeholt, Karen Simonyan, Aaron van den Oord, Alex Graves, and Koray Kavukcuoglu. Neural machine translation in linear time. *arXiv preprint arXiv:1610.10099*, 2016.

[28] Anjuli Kannan, Karol Kurach, Sujith Ravi, Tobias Kaufmann, Andrew Tomkins, Balint Miklos, Greg Corrado, László Lukács, Marina Ganea, Peter Young, et al. Smart reply: Automated response suggestion for email. *arXiv preprint arXiv:1606.04870*, 2016.

[29] Andrej Karpathy and Li Fei-Fei. Deep visual-semantic alignments for generating image descriptions. pages 3128–3137, 2015.

[30] Yuta Kikuchi, Graham Neubig, Ryohei Sasano, Hiroya Takamura, and Manabu Okumura. Controlling output length in neural encoder-decoders. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2016.

[31] Kevin Knight and Daniel Marcu. Summarization beyond sentence extraction: A probabilistic approach to sentence compression. *Artificial Intelligence*, 139(1):91–107, 2002.

[32] John Kominek, Tanja Schultz, and Alan W Black. Synthesizer voice quality of new languages calibrated with mean mel cepstral distortion. In *SLTU*, pages 63–68, 2008.

[33] Tomáš Kočiský, Gábor Melis, Edward Grefenstette, Chris Dyer, Wang Ling, Phil Blunsom, and Karl Moritz Hermann. Semantic parsing with semi-supervised sequential autoencoders. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1078–1087, 2016.

[34] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. pages 1097–1105, 2012.

[35] Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao, and Bill Dolan. A diversity-promoting objective function for neural conversation models. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, pages 110–119, 2016.

[36] Jiwei Li, Michel Galley, Chris Brockett, Georgios Spithourakis, Jianfeng Gao, and Bill Dolan. A persona-based neural conversation model. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 994–1003, 2016.

[37] Chin-Yew Lin. Rouge: A package for automatic evaluation of summaries. In *Text Summarization Branches Out: Proceedings of the ACL-04 Workshop*, pages 74–81, 2004.

[38] Wang Ling, Phil Blunsom, Edward Grefenstette, Karl Moritz Hermann, Tomáš Kočiský, Fumin Wang, and Andrew Senior. Latent predictor networks for code generation. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2016.

[39] Chia-Wei Liu, Ryan Lowe, Iulian Serban, Mike Noseworthy, Laurent Charlin, and Joelle Pineau. How NOT to evaluate your dialogue system: An empirical study of unsupervised evaluation metrics for dialogue response generation. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2122–2132, 2016.

[40] Takuya Matsuzaki, Yusuke Miyao, and Jun'ichi Tsujii. Probabilistic cfg with latent annotations. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 75–82. Association for Computational Linguistics, 2005.

[41] Margaret Mitchell, Jesse Dodge, Amit Goyal, Kota Yamaguchi, Karl Stratos, Xufeng Han, Alyssa Mensch, Alex Berg, Tamara Berg, and Hal Daume III. Midge: Generating image descriptions from computer vision detections. In *Proceedings of the 13th European Chapter of the Association for Computational Linguistics (EACL)*, pages 747–756, 2012.

[42] Abdel-rahman Mohamed, George E Dahl, and Geoffrey Hinton. Acoustic modeling using deep belief networks. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(1):14–22, 2012.

[43] Mehryar Mohri, Fernando Pereira, and Michael Riley. Speech recognition with weighted finite-state transducers. *Handbook on speech processing and speech communication, Part E: Speech recognition*, 2008.

[44] Hajime Morita, Ryohei Sasano, Hiroya Takamura, and Manabu Okumura. Subtree extractive summarization via submodular maximization. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 1023–1032, 2013.

[45] Graham Neubig, Shinsuke Mori, and Tatsuya Kawahara. A WFST-based log-linear framework for speaking-style transformation. In *Proceedings of the 10th Annual Conference of the International Speech Communication Association (InterSpeech)*, pages 1495–1498, 2009.

[46] Ehud Reiter and Robert Dale. Building applied natural language generation systems. *Natural Language Engineering*, 3(01):57–87, 1997.

[47] Alan Ritter, Colin Cherry, and William B. Dolan. Data-driven response generation in social media. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 583–593, 2011.

[48] Alexander M. Rush, Sumit Chopra, and Jason Weston. A neural attention model for abstractive sentence summarization. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 379–389, 2015.

[49] Lifeng Shang, Zhengdong Lu, and Hang Li. Neural responding machine for short-text conversation. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 1577–1586, 2015.

[50] Richard Socher, John Bauer, Christopher D. Manning, and Ng Andrew Y. Parsing with compositional vector grammars. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 455–465, 2013.

[51] Karen Sparck Jones and John I Tait. Automatic search term variant generation. *Journal of Documentation*, 40(1):50–66, 1984.

[52] Tomoki Toda, Alan W Black, and Keiichi Tokuda. Voice conversion based on maximum-likelihood estimation of spectral parameter trajectory. *IEEE Transactions on Audio, Speech, and Language Processing*, 15(8):2222–2235, 2007.

[53] Aäron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. Wavenet: A generative model for raw audio. *CoRR abs/1609.03499*, 2016.

[54] Oriol Vinyals, Łukasz Kaiser, Terry Koo, Slav Petrov, Ilya Sutskever, and Geoffrey Hinton. Grammar as a foreign language. In *Proceedings of the 29th Annual Conference on Neural Information Processing Systems (NIPS)*, pages 2773–2781, 2015.

[55] Oriol Vinyals and Quoc Le. A neural conversational model. *arXiv preprint arXiv:1506.05869*, 2015.

[56] Oriol Vinyals, Alexander Toshev, Samy Bengio, and Dumitru Erhan. Show and tell: A neural image caption generator. pages 3156–3164, 2015.

[57] Tong Wang, Ping Chen, John Rochford, and Jipeng Qiang. Text simplification using neural machine translation. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, pages 4270–7271. AAAI Press, 2016.

[58] Tsung-Hsien Wen, Milica Gasic, Nikola Mrkšić, Pei-Hao Su, David Vandyke, and Steve Young. Semantically conditioned lstm-based natural language generation for spoken dialogue systems. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1711–1721, 2015.

[59] Tsung-Hsien Wen, Milica Gasic, Nikola Mrkšić, Pei-Hao Su, David Vandyke, and Steve Young. Semantically conditioned LSTM-based natural language generation for spoken dialogue systems. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1711–1721, 2015.

[60] Tsung-Hsien Wen, David Vandyke, Nikola Mrksic, Milica Gasic, Lina M Rojas-Barahona, Pei-Hao Su, Stefan Ultes, and Steve Young. A network-based end-to-end trainable task-oriented dialogue system. *arXiv preprint arXiv:1604.04562*, 2016.

[61] Yuk Wah Wong and Raymond Mooney. Learning for semantic parsing with statistical machine translation. In *Proceedings of the 2006 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL)*, pages 439–446, 2006.

[62] Wei Xu, Alan Ritter, Bill Dolan, Ralph Grishman, and Colin Cherry. Paraphrasing for style. In *Proceedings of the 24th International Conference on Computational Linguistics (COLING)*, pages 2899–2914, 2012.

[63] Matthew D Zeiler, Dilip Krishnan, Graham W Taylor, and Rob Fergus. Deconvolutional networks. pages 2528–2535. IEEE, 2010.

[64] Heiga Zen, Andrew Senior, and Mike Schuster. Statistical parametric speech synthesis using deep neural networks. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 7962–7966. IEEE, 2013.

[65] Heiga Zen, Keiichi Tokuda, and Alan W Black. Statistical parametric speech synthesis. *Speech Communication*, 51(11):1039–1064, 2009.

[66] Zhemin Zhu, Delphine Bernhard, and Iryna Gurevych. A monolingual tree-based translation model for sentence simplification. In *Proceedings of the 23rd International Conference on Computational Linguistics (COLING)*, pages 1353–1361, 2010.