#QE questions - talking points

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. **Chapter 1 – Introduction**
   1. **Motivation**
      * What is non-autoregressive machine translation?
      * Why the need for NAT?
      * What are the problems or obstacles NAT face?
   2. **Report Outline**
2. **Chapter 2 Literature Review**
   1. **Background and History**
      1. *Autoregressive methods for sequence to sequence generation*
         * What are the defacto models for sequence to sequence generation?
         * What is an autoregressive factorization?
         * Why is this autoregressive factorization useful and popular?
         * Why was non-autoregressive methods not as vital then?
      2. *Transformer*
         * What is a transformer, and what makes it so popular?
         * What is the legacy of the transformer?
         * What has the transformer been applied to?
         * What is transfer learning?
         * What is an example of Transfer Learning (my paper)
      3. *What are the drawback of the transformer?*
         * What is the trade off of the transformer?
         * What is the trend of model sizes over the years?
   2. **Non-autoregressive transformer sequence to sequence models.**
      1. *Problem Definition*
         * How is Non-autoregressive sequence generation formulated?
      2. *Obstacles to realizing a deployable non-autoregressive sequence to sequence model*
         * What is the problem of length prediction, and how does it arise?
         * How is length prediction modelled?
         * What is the multimodality problem?
         * How does the multimodality problem manifest?
         * What is an example of the multimodality problem?
   3. **Solutions and Approaches**
      1. *Sequence level knowledge distillation (SLKD)*
         * What is SLKD?
         * How does SLKD tackle the multimodality problem?
      2. *Fertilities*
         * What are fertilities?
         * What are the critiques of fertilities?
      3. *Iterative Generation*
         * What is Iterative generation?
         * What is left-to-right non-autoregressive sequence generation?
         * What is arbitrary position non-autoregressive sequence generation?
      4. *Supplementing generation with external knowledge*
         * What kind of external knowledge is provided to the model?
      5. *Post-Editing*
         * What is post-editing?
         * How do the models perform post-editing?
3. **Chapter 3 Exploring the Use of Transfer Learning**
   1. **Motivation**
      * + What are the limitations of the proposed models?
        + Why do I think using transfer learning is a good idea?
   2. **Transfer Learning**
      * + What is the benefit of using transfer learning?
      1. *Adaptor Layers*
         * What are Adaptor Layers
   3. ***N-gram language modelling***
      * + What is N-gram language modelling
        + How is it useful?
   4. ***Research questions***
      * + What are our research questions?
4. **Methodology**
   1. **Past and Future N-gram Language Modeling Task**
      * + What is the past and future n-gram language modelling task?
        + Why do we design it this way?
   2. **Model Architecture**
      * + What is the model architecture, and what is the difference from the standard transformer?
   3. **Decoding Methods**
      * + What decoding methods are used for inference?
        + What is the Argmax Decoding Method
        + What is the Hysteresis Decoding Method? Why do we need it?
        + What is Beam search?
   4. **What are our implementation details?**
      * + What is our dataset?
        + How is training done and hyperparameters?
        + How is inference done?
   5. **Experiments**
      1. What is the difference when using the future ngram objective versus the past and future ngram objective?
         * What is the effectiveness of argmax sequence refinement?
         * What is the effectiveness of past-future n-gram prediction task?
      2. What is the influence of using different types of data?
         * How does transfer learning affect training and results??
         * What is the effectiveness of sequence level knowledge distillation?
      3. Effect of decoding methods
         * How does argmax decoding work and why is it insufficient?
         * Why is Hysteresis Decoding needed?
         * How does hysteresis decoding work?
         * What are the effects of range and manual threshold hysteresis decoding?
      4. **Contribution of different layers**
         * How does freezing the encoder affect the results?
         * Which layers are the most important for turning an autoregressive model to a non-autoregressive model?
5. **Conclusion**
   1. **Conclusions**
      * + What can we conclude from this?
   2. **Future Work**

* How can we incorporate graph information in the encoder?
* Can Non-autoregressive sequence generation be applied to other tasks? For instance automated audio captioning.