**Ideias principais dos papers de NLP:**

1. Tokenization as the initial phase in NLP:
   1. Abordar um pouco como a tokenização é essencial para NLP
   2. E como em línguas como inglês e português que apresentam espaçamento isso é algo “trivial”
   3. Mas fazer um contraponto da importância disso e a dificuldade em outras línguas como chinês que não tem delimitadores claros
   4. Dificuldade em separar morfemas ou itens com significado útil
   5. Abordagem de sinônimos, ele cita *hand* como um sinônimo para casamento ou jogo de cartas e como é raro os dois significados aparecerem juntos em um mesmo contexto
   6. Outro problema são palavras polimórficas como *furniture shop* em que claramente se estabelece uma relação entre as duas partes, é um tipo específico de loja e não de mobília
   7. Sobre a questão da língua chinesa as abordagens usadas são: um dicionário que relaciona as palavras aos seus símbolos e estratégias para remover ambiguidades
   8. Noção que um token é semelhante a um átomo de Dalton, sendo assim a menor parte, indivisível e com um significado
2. GloVe - Global Vectors for Word Representation:
   1. a
3. BERT - Pre-training of Deep Bidirectional Transformers for Language Understanding:
   1. a
4. Attention Is All You Need:
   1. a
5. A Primer in BERTology - What We Know About How BERT Works:
   1. a
6. Efficient Estimation of Word Representations in Vector Space:
   1. a