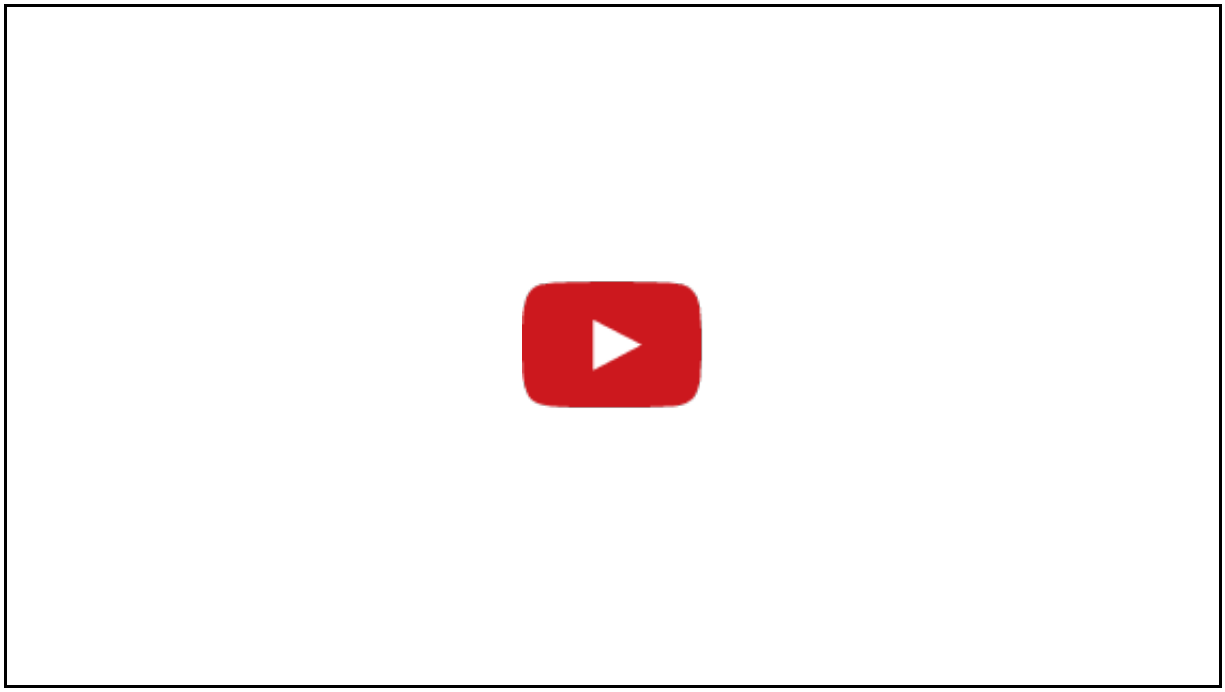
Week 2: Video Lectures

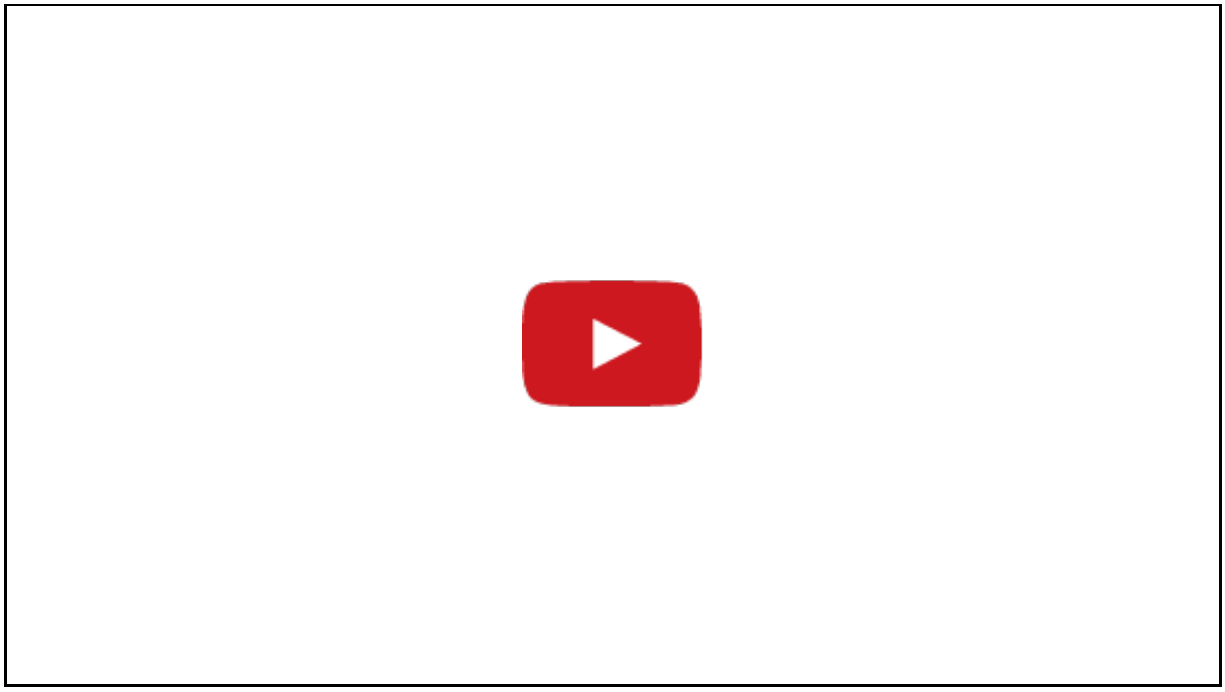
1. Alignment (I) - Introduction



Content: SMT first studies were focused on developing algorithms able to aligned words given parallel sentences. Word alignment is one of the most relevant tasks when building a SMT system. This lecture covers main concepts such as the noisy channel model and lexical translation. From these concepts, the lecture points out the first ideas of word alignment.

Lecturer: Lluís Formiga

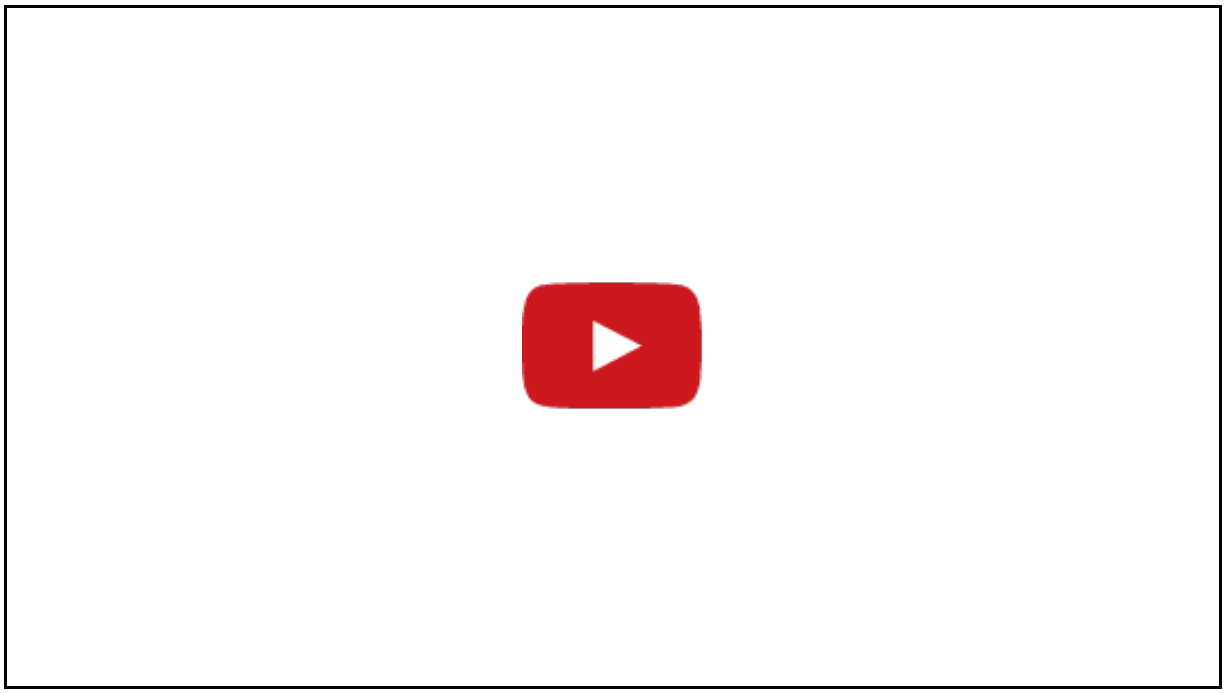
2. Alignment (II) - IBM Model 1 and EM



Content: A standard way to compute word alignment is by means of the IBM Models. This lecture covers the simplest IBM Model, which is Model 1 and also the expectation maximization algorithm

Lecturer: Lluís Formiga

3. Alignment (III) - IBM Model 2 and HMM



DISCLAIMER: The algorithm given for IBM Model 2 is an adaptation of the complete observed data version of the algorithm (i.e. we want to learn from a manually aligned corpus). Please, refer to the new version of the slides (**2-3-model2-hmm.pdf (https://mooc.upc.edu/courses/2/files/347/download?verifier=CXnvwRk8McsifiE1fLOm8291WxypGmH1hzq6DtxV&wrap=1)** **(https://mooc.upc.edu/courses/2/files/347/download?**

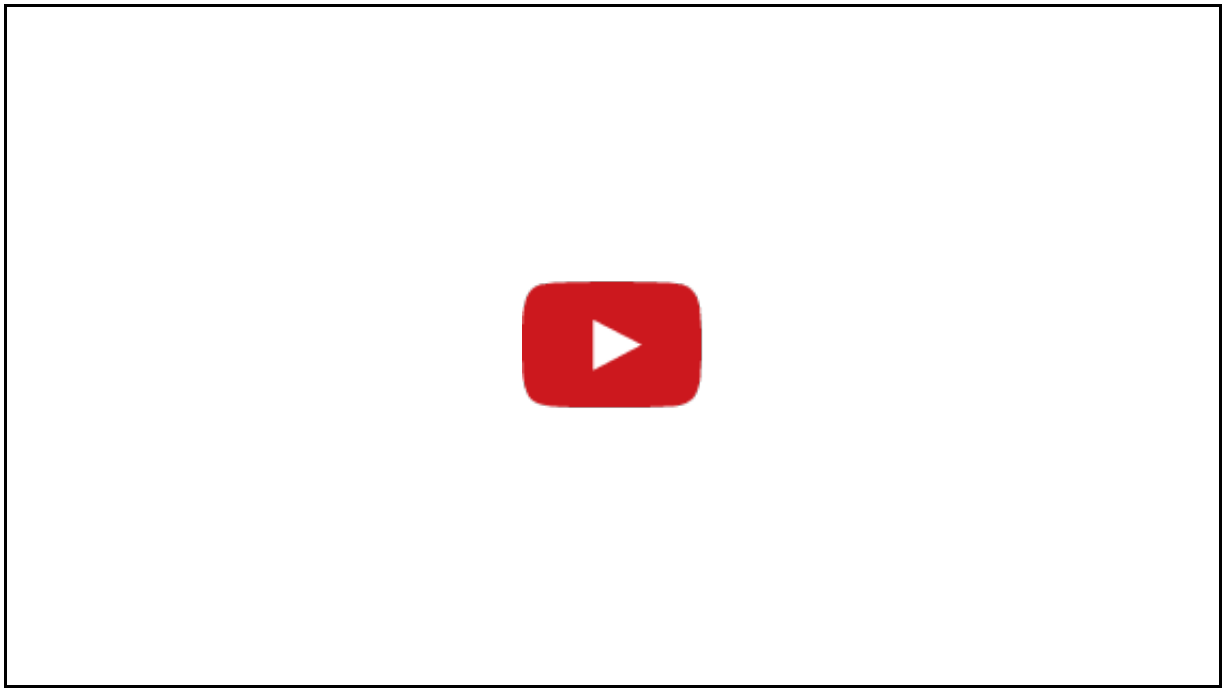
verifier=CXnvwRk8McsifiE1fLOm8291WxypGmH1hzq6DtxV&wrap=1) (https://mooc.upc.edu/courses/2/files/347/download?verifier=CXnvwRk8McsifiE1fLOm8291WxypGmH1hzq6DtxV&wrap=1)) to have the partially observed data version of the algorithm (i.e. we learn from a parallel corpus only aligned at the sentence level).

Content: This lecture continues with the theory behind word alignment by explaining IBM Model 2 and Hidden Markov Models.

Lecturer: Lluís Formiga

OPTIONAL -- NOT NEEDED FOR NEITHER ASSIGNMENT NOR QUIZ

4. Alignment (IV) - IBM Higher Models

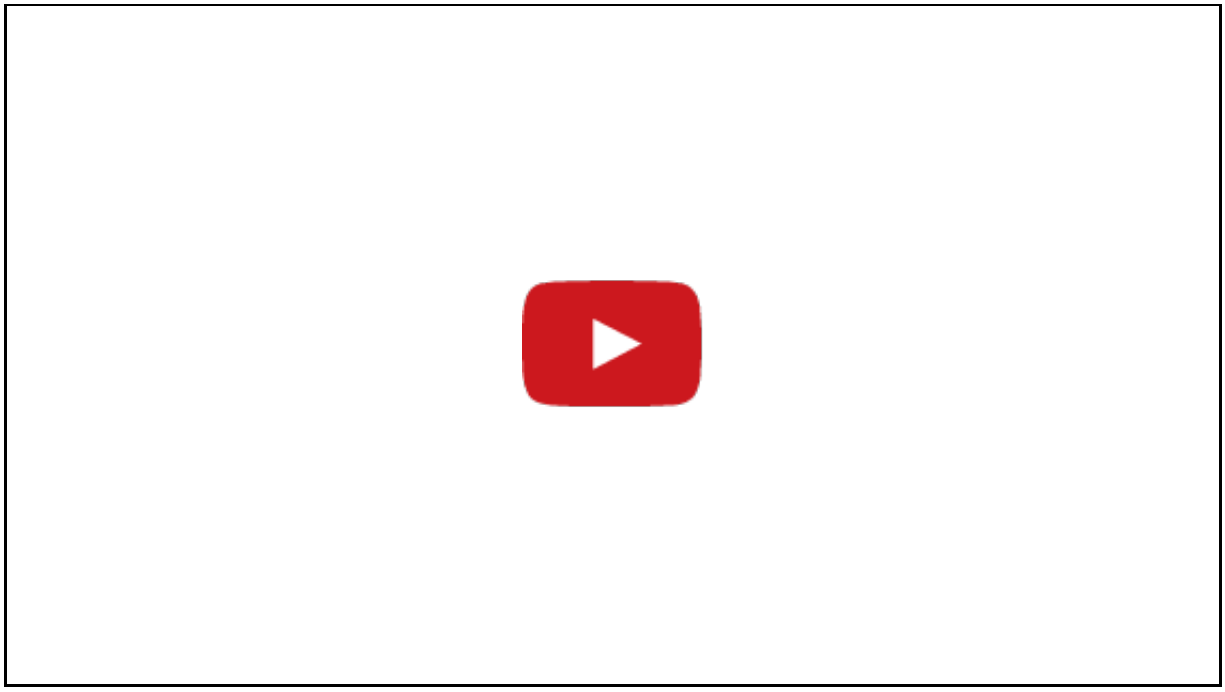


Content: *This lecture covers the rest of the IBM Models: Model 3, 4 and 5.*

Lecturer: *Lluís Formiga*

OPTIONAL -- NOT NEEDED FOR NEITHER ASSIGNMENT NOR QUIZ

5. Alignment (V) - Applications, Quality and Problems of Word Alignment

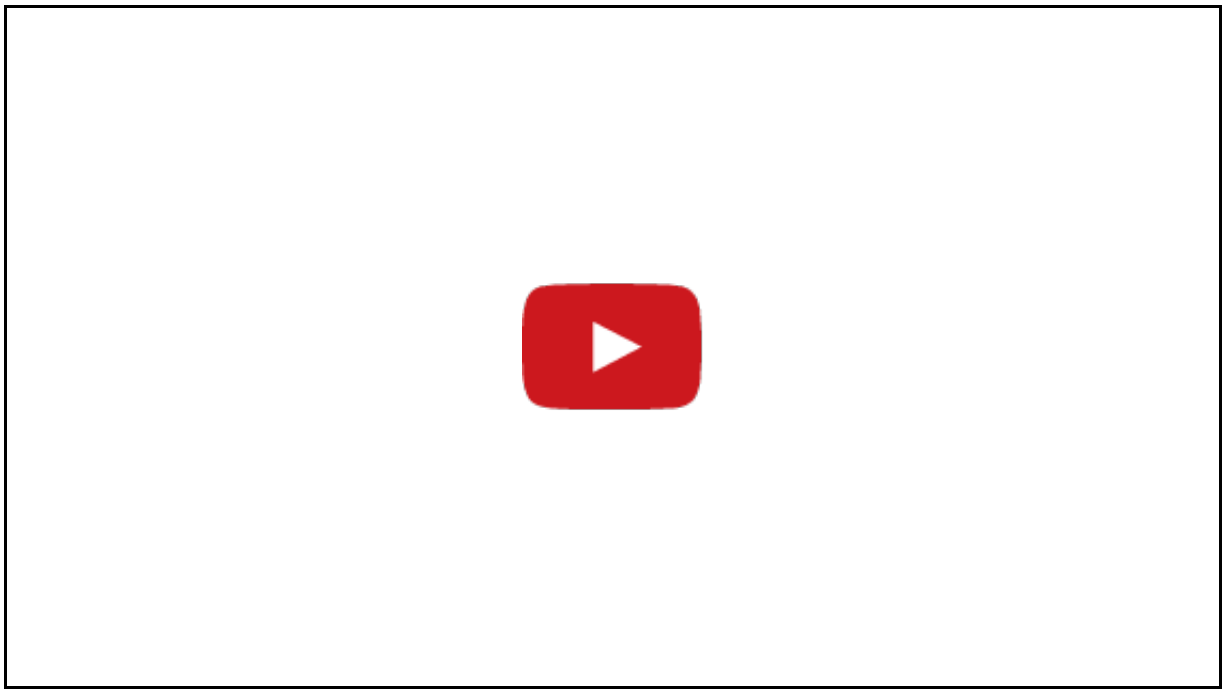


Content: *This lecture covers issues such as making word alignment incremental or multi-threading. Also the lecture explains how the measure of quality for word alignment which is the Alignment Error Rate.*

Lecturer: *Lluís Formiga*

---

6. Growing Heuristics, Symmetrization and Phrase-Models



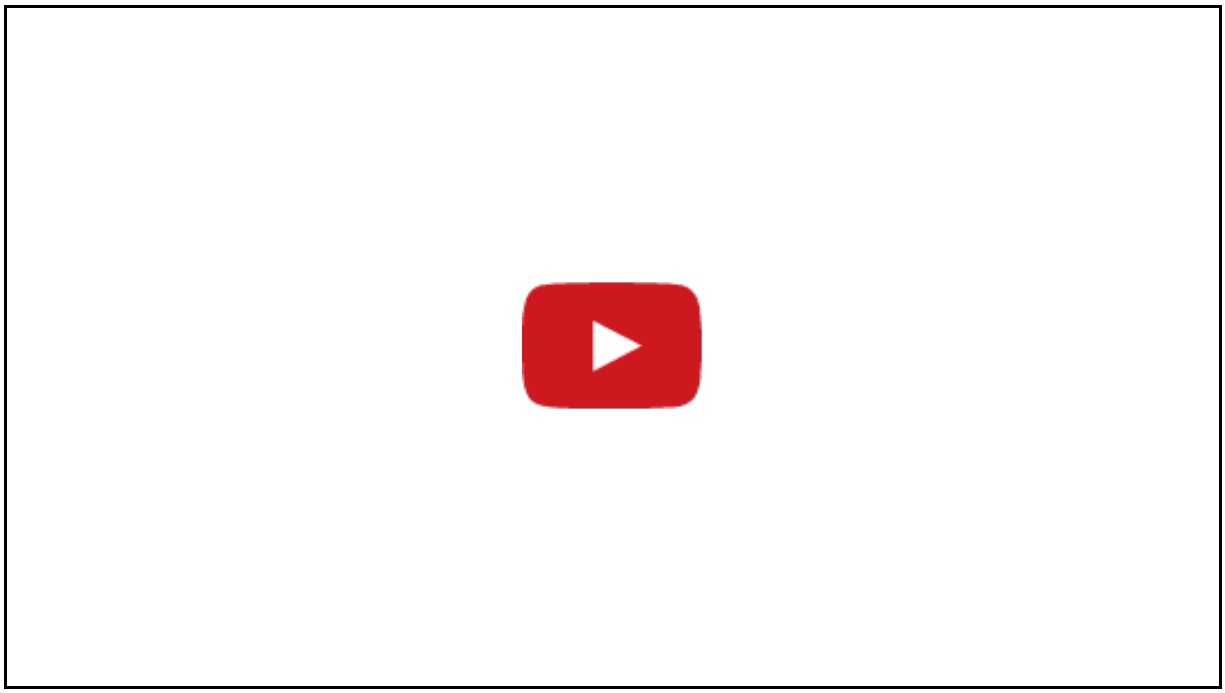
DISCLAIMER: In this video, only the grow-diag-final version of the algorithm is shown. For the advanced programming assignment, we recommend to use the grow-diag-final-AND version of the algorithm, which is detailed in the new version of the slides: **2-6-symmetrization.pdf (https://mooc.upc.edu/courses/2/files/359/download?**

verifier=GdvPKxqLc9ZMYPaZtJJRbpRmmwGC7HtE4OZrlCCj&wrap=1)
(https://mooc.upc.edu/courses/2/files/359/download?
verifier=GdvPKxqLc9ZMYPaZtJJRbpRmmwGC7HtE4OZrlCCj&wrap=1)
(https://mooc.upc.edu/courses/2/files/359/download?
verifier=GdvPKxqLc9ZMYPaZtJJRbpRmmwGC7HtE4OZrlCCj&wrap=1)

Content: The IBM Models explained in the earlier lectures are usually performed in both directions of translation (from source to target and target to source). This lecture shows standard procedures of symmetrization of word alignment. Then, the lecture shows the evolution from word-based translation to phrase-based translation. The main advantage is that phrases allow to translating sequences of words.

Lecturer: Lluís Formiga

7. Phrase-Extraction and Scoring



Content: Ok, now we have the word alignment, but how do we build a translation model based on phrases? This lecture focuses on how phrases are extracted from word-aligned sentences and how they are scored. A relevant criterion for extraction is the coherence with word alignment, which means that a phrase cannot contain a word that is aligned with a word outside the phrase unit.

Lecturer: Lluís Formiga