

# Error Coding of Second-Language Learner Texts Based on Mostly Automatic Alignment of Parallel Corpora

Dan Rosén<sup>1</sup>, Mats Wirén<sup>2</sup> & Elena Volodina<sup>1</sup>



## Svala: Svenska som Andraspråk – Länkning och Annotering (Swedish as a Second Language – Linking and Annotation)

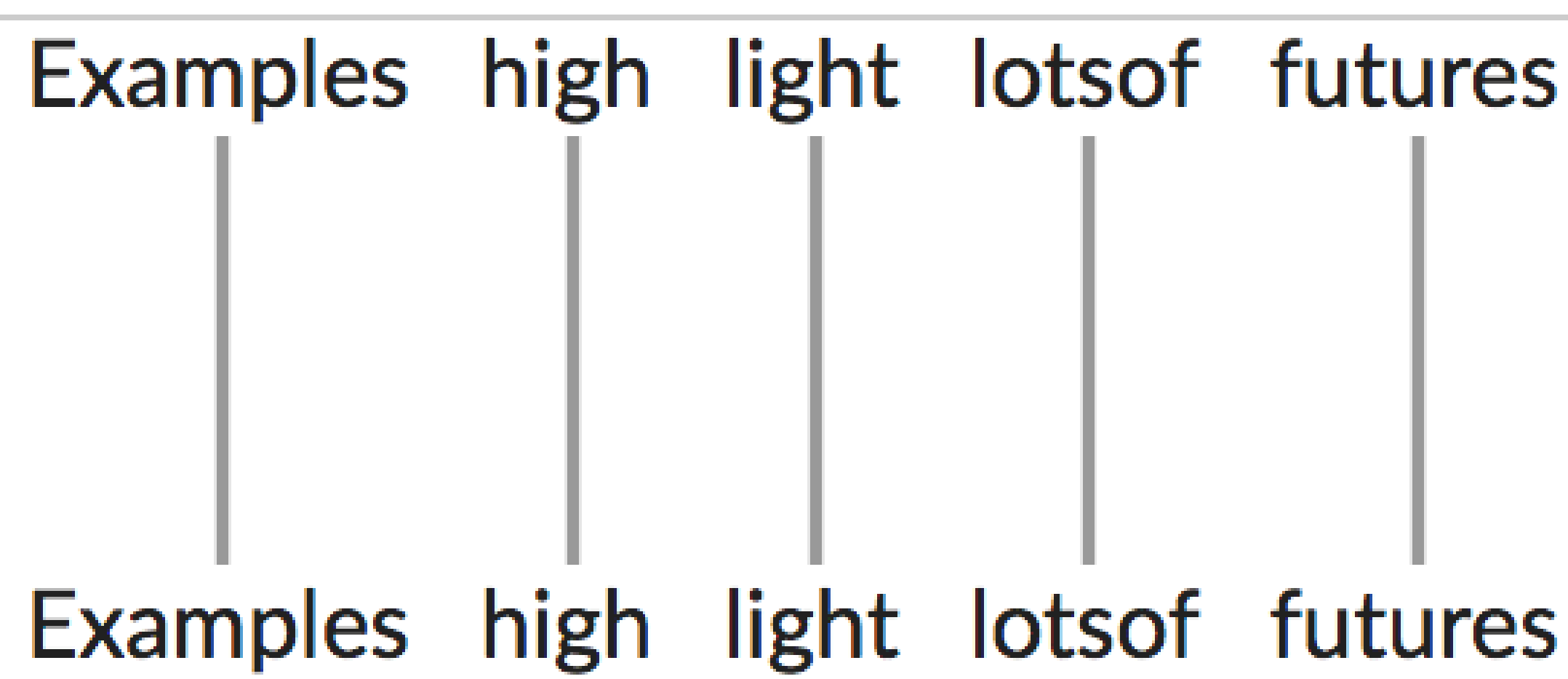
1. Correction (normalisation) of learner text and annotation of corrections are separated
2. A parallel corpus of the learner and corrected texts is automatically generated

## Error coding in Svala is made in three steps

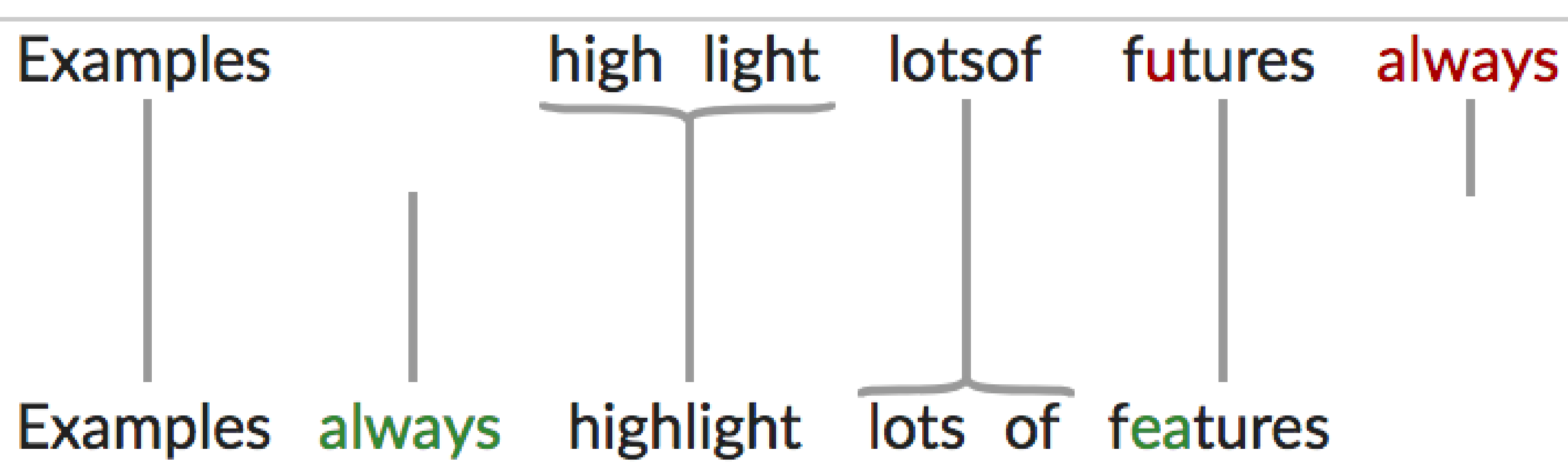
1. Error correction (normalisation) of the learner text according to the target hypotheses.  
Automatic alignment
2. Manual alignment of unattached or misaligned links
3. Annotation of corrections

The screenshot shows the Svala interface with a 'Source text' field containing 'Examples high light lotsof futures always' and a 'Target text' field containing 'Examples always highlight lots of features'. A 'copy to target' button is visible. On the left, a lexical list includes categories like Lexical, Orthographic, Morphological, Syntactical, Punctuation, and Unidentified, with 'S-WO' selected. Below the text fields, a diagram shows the alignment of words between source and target with syntactic labels like S-WO, O-COMP, and L.

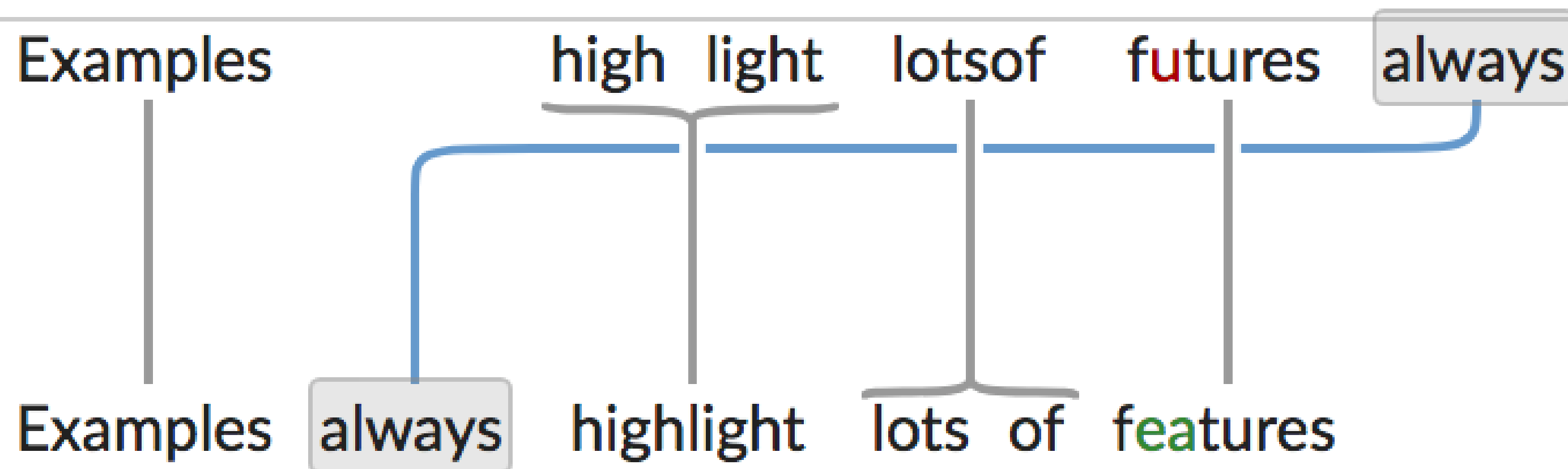
## 0. Initial alignment



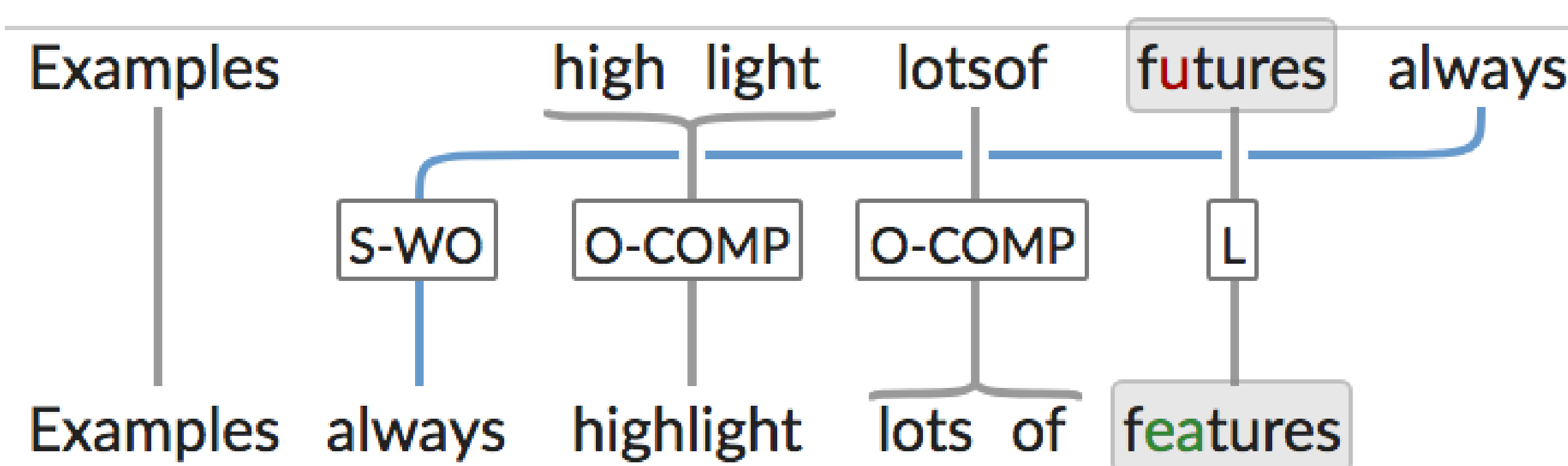
## 1. Automatic alignment after error correction



## 2. Manual alignment



## 3. Annotation of corrections



## Anonymisation with Svala

