

**Original text** “\_” and “/” denote a space and a subword boundary, respectively.

**Scriptio continua languages (Japanese):**

私は形態素解析器の研究をしています。

*I morphological analyzers on research am doing*

**English:**

I\_am\_doing\_research\_on\_morphological\_analyzers.

**Step 1:** Morphological analysis (Splitting into “word-level” semantic units)

私\_は\_形態素\_解析\_器\_の\_研究\_を\_し\_て\_い\_ます\_。

**Step 2:** Subword tokenization

私 / は / 形態 / ##素 / 解析 / 器 / の / 研究 / を / し / て / い / ます / 。

I / am / doing / research / on / morphological / analyze / ##rs / .