Samuel MORSE minimized the number of on-off clicks for non-verbal communication.

This MORSE minimizes the vocabulary size for Natural Language Processing systems.
1. Morpheme Segmentation
Morpheme Segmentation

Hopefully
Not a trivial task

Players
Playing
Beijing
Butterflies
Applications

Machine Translation

Model:
- Quickly
  - Rapidement
- Sad
  - Triste

Test:
- Sadly

Model:
- Quick
  - Rapide
- ly
  - ment
- Sad
  - Triste

Test:
- Sadly
  - Tristement

Model:
- Quickly
  - Rapidement
- Sad
  - Triste

Test:
- Sadly

Model:
- Quick
  - Rapide
- ly
  - ment
- Sad
  - Triste

Test:
- Sadly

Model:
- Quickly
  - Rapidement
- Sad
  - Triste

Test:
- Sadly

Model:
- Quick
  - Rapide
- ly
  - ment
- Sad
  - Triste

Test:
- Sadly

Model:
- Quickly
  - Rapidement
- Sad
  - Triste

Test:
- Sadly

Model:
- Quick
  - Rapide
- ly
  - ment
- Sad
  - Triste

Test:
- Sadly

Model:
- Quickly
  - Rapidement
- Sad
  - Triste

Test:
- Sadly

Model:
- Quick
  - Rapide
- ly
  - ment
- Sad
  - Triste

Test:
- Sadly

Model:
- Quickly
  - Rapidement
- Sad
  - Triste

Test:
- Sadly

Model:
- Quick
  - Rapide
- ly
  - ment
- Sad
  - Triste

Test:
- Sadly

Model:
- Quickly
  - Rapidement
- Sad
  - Triste

Test:
- Sadly

Model:
- Quick
  - Rapide
- ly
  - ment
- Sad
  - Triste

Test:
- Sadly

Model:
- Quickly
  - Rapidement
- Sad
  - Triste

Test:
- Sadly

Model:
- Quick
  - Rapide
- ly
  - ment
- Sad
  - Triste

Test:
- Sadly

Model:
- Quickly
  - Rapidement
- Sad
  - Triste

Test:
- Sadly

Model:
- Quick
  - Rapide
- ly
  - ment
- Sad
  - Triste

Test:
- Sadly

Model:
- Quickly
  - Rapidement
- Sad
  - Triste

Test:
- Sadly

Model:
- Quick
  - Rapide
- ly
  - ment
- Sad
  - Triste

Test:
- Sadly

Model:
- Quickly
  - Rapidement
- Sad
  - Triste

Test:
- Sadly

Model:
- Quick
  - Rapide
- ly
  - ment
- Sad
  - Triste

Test:
- Sadly

Model:
- Quickly
  - Rapidement
- Sad
  - Triste

Test:
- Sadly

Model:
- Quick
  - Rapide
- ly
  - ment
- Sad
  - Triste

Test:
- Sadly

Model:
- Quickly
  - Rapidement
- Sad
  - Triste

Test:
- Sadly

Model:
- Quick
  - Rapide
- ly
  - ment
- Sad
  - Triste

Test:
- Sadly

Model:
- Quickly
  - Rapidement
- Sad
  - Triste

Test:
- Sadly

Model:
- Quick
  - Rapide
- ly
  - ment
- Sad
  - Triste

Test:
- Sadly

Model:
- Quickly
  - Rapidement
- Sad
  - Triste

Test:
- Sadly

Model:
- Quick
  - Rapide
- ly
  - ment
- Sad
  - Triste

Test:
- Sadly

Model:
- Quickly
  - Rapidement
- Sad
  - Triste

Test:
- Sadly

Model:
- Quick
  - Rapide
- ly
  - ment
- Sad
  - Triste

Test:
- Sadly

Model:
- Quickly
  - Rapidement
- Sad
  - Triste

Test:
-Sadly

Model:
- Quick
  - Rapide
- ly
  - ment
- Sad
  - Triste

Test:
- Sadly

Model:
- Quickly
  - Rapidement
- Sad
  - Triste

Test:
- Sadly

Model:
- Quick
  - Rapide
- ly
  - ment
- Sad
  - Triste

Test:
- Sadly

Model:
- Quickly
  - Rapidement
- Sad
  - Triste

Test:
- Sadly

Model:
- Quick
  - Rapide
- ly
  - ment
- Sad
  - Triste

Test:
- Sadly

Model:
- Quickly
  - Rapidement
- Sad
  - Triste

Test:
- Sadly

Model:
- Quick
  - Rapide
- ly
  - ment
- Sad
  - Triste

Test:
- Sadly

Model:
- Quickly
  - Rapidement
- Sad
  - Triste

Test:
- Sadly

Model:
- Quick
  - Rapide
- ly
  - ment
- Sad
  - Triste

Test:
- Sadly

Model:
- Quickly
  - Rapidement
- Sad
  - Triste

Test:
- Sadly

Model:
- Quick
  - Rapide
- ly
  - ment
- Sad
  - Triste

Test:
- Sadly

Model:
- Quickly
  - Rapidement
- Sad
  - Triste

Test:
- Sadly

Model:
- Quick
  - Rapide
- ly
  - ment
- Sad
  - Triste

Test:
- Sadly

Model:
- Quickly
  - Rapidement
- Sad
  - Triste

Test:
- Sadly

Model:
- Quick
  - Rapide
- ly
  - ment
- Sad
  - Triste

Test:
- Sadly

Model:
- Quickly
  - Rapidement
- Sad
  - Triste

Test:
- Sadly

Model:
- Quick
  - Rapide
- ly
  - ment
- Sad
  - Triste

Test:
- Sadly

Model:
- Quickly
  - Rapidement
- Sad
  - Triste

Test:
- Sadly

Model:
- Quick
  - Rapide
- ly
  - ment
- Sad
  - Triste

Test:
- Sadly

Model:
- Quickly
  - Rapidement
- Sad
  - Triste

Test:
- Sadly

Model:
- Quick
  - Rapide
- ly
  - ment
- Sad
  - Triste

Test:
- Sadly

Model:
- Quickly
  - Rapidement
- Sad
  - Triste

Test:
- Sadly

Model:
- Quick
  - Rapide
- ly
  - ment
- Sad
  - Triste

Test:
- Sadly

Model:
- Quickly
  - Rapidement
- Sad
  - Triste

Test:
- Sadly

Model:
- Quick
  - Rapide
- ly
  - ment
- Sad
  - Triste

Test:
- Sadly

Model:
- Quickly
  - Rapidement
- Sad
  - Triste

Test:
- Sadly

Model:
- Quick
  - Rapide
- ly
  - ment
- Sad
  - Triste

Test:
- Sadly

Model:
- Quickly
  - Rapidement
- Sad
  - Triste

Test:
- Sadly

Model:
- Quick
  - Rapide
- ly
  - ment
- Sad
  - Triste

Test:
- Sadly

Model:
- Quickly
  - Rapidement
- Sad
  - Triste

Test:
- Sadly

Model:
- Quick
  - Rapide
- ly
  - ment
- Sad
  - Triste

Test:
- Sadly

Model:
- Quickly
  - Rapidement
- Sad
  - Triste

Test:
- Sadly

Model:
- Quick
  - Rapide
- ly
  - ment
- Sad
  - Triste

Test:
- Sadly

Model:
- Quickly
  - Rapidement
- Sad
  - Triste

Test:
- Sadly

Model:
- Quick
  - Rapide
- ly
  - ment
- Sad
  - Triste

Test:
- Sadly

Model:
- Quickly
  - Rapidement
- Sad
  - Triste

Test:
- Sadly

Model:
- Quick
  - Rapide
- ly
  - ment
- Sad
  - Triste

Test:
- Sadly

Model:
- Quickly
  - Rapidement
- Sad
  - Triste

Test:
- Sadly

Model:
- Quick
  - Rapide
- ly
  - ment
- Sad
  - Triste

Test:
- Sadly

Model:
- Quickly
  - Rapidement
- Sad
  - Triste

Test:
- Sadly

Model:
- Quick
  - Rapide
- ly
  - ment
- Sad
  - Triste

Test:
- Sadly

Model:
- Quickly
  - Rapidement
- Sad
  - Triste

Test:
- Sadly

Model:
- Quick
  - Rapide
- ly
  - ment
- Sad
  - Triste

Test:
- Sadly

Model:
- Quickly
  - Rapidement
- Sad
  - Triste

Test:
- Sadly

Model:
- Quick
  - Rapide
- ly
  - ment
- Sad
  - Triste

Test:
- Sadly

Model:
- Quickly
  - Rapidement
- Sad
  - Triste

Test:
- Sadly

Model:
- Quick
  - Rapide
- ly
  - ment
- Sad
  - Triste

Test:
- Sadly

Model:
- Quickly
  - Rapidement
- Sad
  - Triste

Test:
- Sadly
Here at Toyota World, we have the cheapest cars in town. We are proudly called the first and last stop.
Previous Work
Letter Successor Variety (Harris, 1970)

Helplessly
Morfessor (Creutz and Lagos, 2005)

Help: 2387
Helping: 1586
Helper: 498
Helps: 2437

Jump: 1847
Jumping: 1664
Jumper: 1290
Jumps: 2987
Downsides

Freshman

Butterflies

Butterflies
Locally Semantic

Cosine similarity

- car
- caring

- car
- cars

(Schone and Jurafsky, 2000)
(Narasimhan et al., 2015)
(Luo et al., 2017)
Distinguishing criteria

car → cars
player → players
runner → runners
goal → goals
play → plays
fine → fines
wheel → wheels
hand → hands
laptop → laptops
lab → labs
MORSE

Input: Word Embeddings

Segmentation: Optimization Problem

Unsupervised Morphology Learning

4 hyperparameters: Small tuning dataset
Step 1

Learning Morphology
Collecting candidate morphological rules

Vocabulary: jump  play  buy  jumping  playing  buying  jumper  player  buyer  .....  and  stand

(suf, ∅, ing):  (jump, jumping)  (play, playing)  (buy, buying)
(suf, ∅, er):  (jump, jumper)  (play, player)  (buy, buyer)
(pre, ∅, st):  (and, stand)  (ore, store)  (one, stone)

(Soricut and Och, 2015)
Signals

Orthographic

| Word   | Orthographic Signal |
|--------|---------------------|
| quick  | quickly             |
| beautiful | beautifully        |
| confident | confidently       |
| wrong  | wrongly             |

Semantic

Word Embeddings

| Word   | Semantic Signal |
|--------|-----------------|
| quick  | quickly         |
| beautiful | beautifully   |
| confident | confidently |
| wrong  | wrongly         |
| confidently |            |
What makes a good rule?

**Signal 1: Orthography**

Rule = (suf, ∅, ly)  Size = 8723

(quick, quickly)  (beautiful, beautifully)
(confident, confidently)
........................................ (wrong, wrongly)

Rule = (pre, ∅, st)  Size = 16

(ore, store) ......
(amp, stamp)
What makes a good rule?

**Signal 2: Semantics**

- quick
- beautiful
- confidently
- wrong
- wrongly
- confident
- quickly
- beautifully
- one
- amp
- stamp
- stone
- store
- ore
- stand
- and
What makes a good member of a rule?

**Scope: Vocabulary-Wide**

- quick
- quickly
- confidently
- confident
- confident ly
- wrong
- wrongly
- beautiful
- beautifully
- on
- only
What makes a good member of a rule?

**Scope: Local**

- **Correct**: confident
- **Correct**: confidently

- **Incorrect**: only
- **Incorrect**: on
Step 2

Segmenting
Linear Optimization Problem

uncaring

t_1

(t_1, uncar)

(t_2, uncar)

(t_3, uncar)

(t_4, uncar)

(ring, uncar)

(caring, uncar)

(uncare, uncar)
Iterate

caring

un + caring

(car, caring)  (care, caring)  (carol, caring)
Iterate

un + care + ing

(car, care) (ca, care) (re, care)
Experiments
Experimental Setup

**Training**

**Languages**

**Gold Datasets**

Morpho Challenge

- jumping jump ing
- playing play ing
- jumps jump s
- calls call s
- rooms room s
Experiments

| Language | Morfessor | MORSE |
|----------|-----------|-------|
| English  | 64.35     | 70.32 |
| Turkish  | 31.01     | 38.07 |
| Finnish  | 14.98     | 34.06 |
Morpho Challenge downsides

Non-compositional

Business

Trivial instances

Turning-point

Human error

Turning

Player’s
Experiments

**New Dataset: SD17**

- 2000 words
- Compositional
- 91% inter-annotator agreement
- In canonical (butterfly + ies) and non-canonical version (butterfl + ies)
Results on SD17

- Morfessor: 57.31
- MORSE (tuned on MC): 81.01
- MORSE (tuned on SD17): 83.96

F-score
Against state-of-the-art
Negative Dataset

- 100 words like: honeymoon, passport, outdoors
- Checks for robustness

![Bar chart showing #Segmentations for Morfessor (43) and MORSE (7)]
Looking forward

- Robustness to highly agglutinative languages
- Extending to other languages (non-concatenative)

\[ \text{\textbf{k \dot{a} t \text{	extsc{A}} b \ a}} \]
Looking forward

- Morphological mappings across languages

| English  | French  |
|----------|---------|
| (suf, φ, ly) | (suf, φ, ment) |
| (suf, φ, s)  | (suf, φ, s) |
|           | (suf, φ, es) |
Links

https://morse.mybluemix.net

https://github.com/yoonlee95/morse_segmentation
Thank you

Questions?
Effect of Hyperparameters

Recall

Precision
Prerequisite

Morpho-syntactic regularities in word vectors

**Valid rule** with an **invalid instance**
(suf, ø, ing) (s, sing)

play → playing
jump → jumping
cry → screaming

s → sing

**Invalid rule**
(pre, ø, s)

scream → scream
store → tore

lay → slay

mile → smile
Demo

corse.mybluemix.net