Jujeop : Korean Puns for K-pop Stars on Social Media

Soyoung Oh*, Jisu Kim*, Seungpeel Lee*, Eunil Park†, Sungkyunkwan University

{sori424, rlawltn908, leepeel}@g.skku.edu, eunilpark@skku.edu

* Equally contributed † Corresponding author

What is Jujeop?

Comment has a **sense of humor** by praising K-pop stars with exaggerations and flashy modifiers.

Jujeop types

Type1. Fragmenting words to create a twist

언니. 왜 맨날 똑같은 티만 입어? 프리티!

Sis, Why do you always wear the same **Tee**? **pretTee**!

Type2. Homophones and Homographs

너 영어이름을 조지 부시로 해줘...
내 마음을 조지고 부시니까.

Change your English name to **George Bush**...
because you **beat and break** my heart.

Type3. Repetition

그만 보고 싶다. 그만 보고 싶다.

I want to stop seeing. I want to stop seeing.

Type4. Nonsense

언니 사실 혼혈이라면서요, 한국과 천국

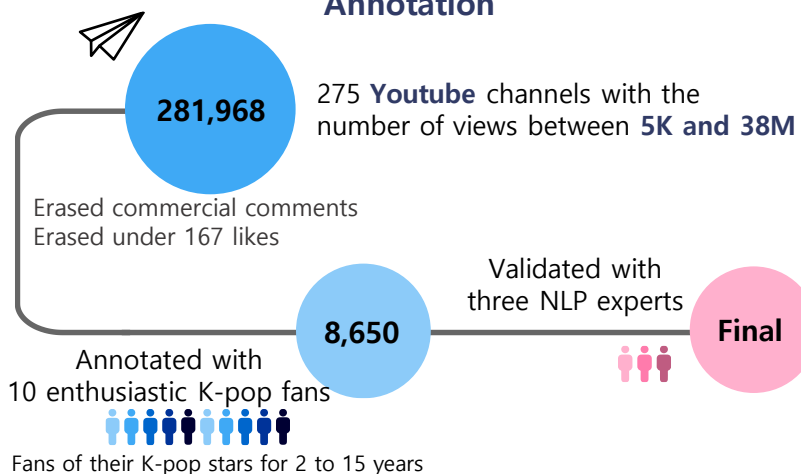Sis, I heard you're **Mixture**. **Korean** and **Heaven**.

✎ Visit our Github Page

Dataset and more examples of each type are publicly available. Baseline codes for experiments are released.
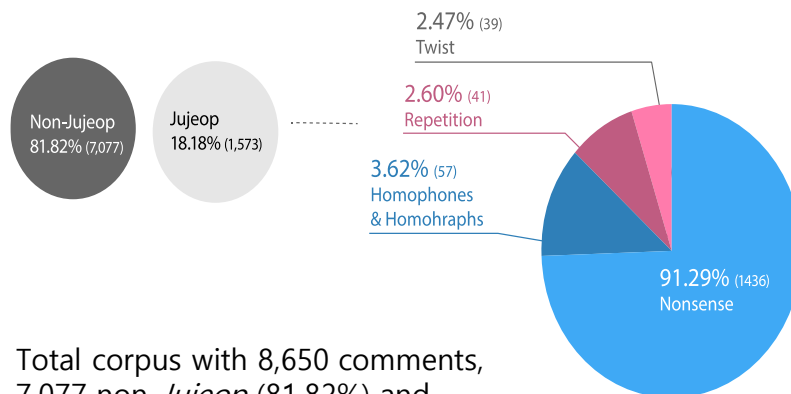
https://github.com/merry555/Jujeop

Annotation

281,968

275 **Youtube** channels with the number of views between **5K and 38M**

Erased commercial comments
Erased under 167 likes

8,650

Annotated with
10 enthusiastic K-pop fans

Fans of their K-pop stars for 2 to 15 years

Validated with three NLP experts

Final

Krippendorff's α on four types of Jujeop comments **0.532**

Corpus Statistics

Non-Jujeop 81.82% (7,077)

Jujeop 18.18% (1,573)

2.47% (39) Twist

2.60% (41) Repetition

3.62% (57) Homophones & Homohraphs

91.29% (1436) Nonsense

Total corpus with 8,650 comments, 7,077 non-Jujeop (81.82%) and 1,573 Jujeop (18.18%).
Within Jujeop comments, each type of the Jujeop comment is as described in the pie chart.

Experiments

Classification

Jujeop classification

- Binary classification for non-Jujeop VS Jujeop

Jujeop type classification

- Multi-class classification for Jujeop types

F1-score	Jujeop (2-ary)	Jujeop type (4-ary)
CNN	67.94%	62.63%
BiLSTM	69.91%	56.96%
KoBERT	**72.91%**	**77.18%**

Clustering

Confusion matrix of k-means clustering with part-of-speech tagged Jujeop comments

	Type 1	Type 2	Type3	Type 4
Type 1	1	4	9	9
Type 2	13	34	10	18
Type 3	23	17	20	18
Type 4	2	2	2	5

Future Work

Building class balanced dataset with given definition by
- a crowdsourcing experiment
- a sentence generation by deep learning approaches