Break It Down For Me
A Study in Automated Lyric Annotation

Lucas Sterckx[1], Jason Naradowsky[2], Bill Byrne[3], Thomas Demeester[1], Chris Develder[1]

lucas.sterckx@ugent.be - @lusterck

1: INTEC, Ghent University, Belgium
2: Language Technology Lab, University of Cambridge, UK
3: Department of Engineering, University of Cambridge, UK

emnlp2017

Automated Lyric Annotation

How does it feel?
To be without a home
Like a complete unknown,
Like a rolling stone

The proverb "A rolling stone gathers no moss" refers to people who are always on the move, never putting down roots or accumulating responsibilities and cares.

Figure 2: A lyric annotation for "Like A Rolling Stone" by Bob Dylan.

- We introduce the task of **automated lyric annotation** . Compared to many traditional NLP systems, which are trained on newswire or similar text, an automated system capable of explaining abstract language, or finding alternative text expressions for slang (and other unknown terms) would exhibit a deeper understanding of the nuances of language.
- Applications:
  - In addition to providing lyric annotations, such systems can lead to improved **NLP analysis of informal text** (blogs, social media, novels and other literary works of fiction),
  - better handling of **genres with heavy use of jargon** (scientific texts, product manuals),
  - **increased robustness to textual variety** in more traditional NLP tasks and genres.

Dataset

# Lyric Annotation pairs	803,720
# Lyric tokens	13,677,332
# Annotation tokens	38,255,671
⊘ Tokens per Lyric	15
⊘ Tokens per Annotation	43
$\lvert V_{\text{lyrics}}\rvert$	124,022
$\lvert V_{\text{annot}}\rvert$	260,427

Figure 3: Generated by users of the **Genius** online lyric database. Properties of gathered dataset (V_{lyrics} and V_{annot} denote the vocabulary for lyrics and annotations, ⊘ denotes the average amount).

Alignment and Context Sensitivity

- **No single, predefined and shared global goal.**
- We distinguish between two types of annotations:
  - **Context independent** annotations are independent of their surrounding context and can be interpreted without it, e.g., explain specific metaphors or imagery or provide narrative while normalizing slang language.
  - **Context sensitive** annotations provide broader context beyond the song lyric excerpt, e.g., background information on the artist.

Type	% of annotations	Examples
Context independent	34.8%	[L] `Gotta patch a lil kid tryna get at this cabbage` [A] He's trying to ignore the people trying to get at his money. [L] `You know it's beef when a smart brother gets stupid` [A] You know an argument is serious when a rational man loses rational.
Context sensitive	65.2%	[L] `Cause we ain't break up, more like broke down` [A] The song details Joe's break up with former girlfriend Esther. [L] `If I quit this season, I still be the greatest, funk` [A] Kendrick has dropped two classic albums.

Figure 4: Examples of context independent and dependent pairs of lyrics [L] and annotations [A].

Baseline Annotators

- **Statistical Machine Translation:** We train a standard phrase-based SMT model to translate lyrics to annotations, using GIZA++ for word alignment and Moses for phrasal alignment, training, and decoding.
- **Seq2Seq:** A recurrent neural network (RNN) encodes the source sequence to a single vector representation. A separate decoder RNN generates the translation conditioned on this representation of the source sequence's semantics. We utilize **Seq2Seq with attention**, which allows the model to additionally condition on tokens from the input sequence during decoding.

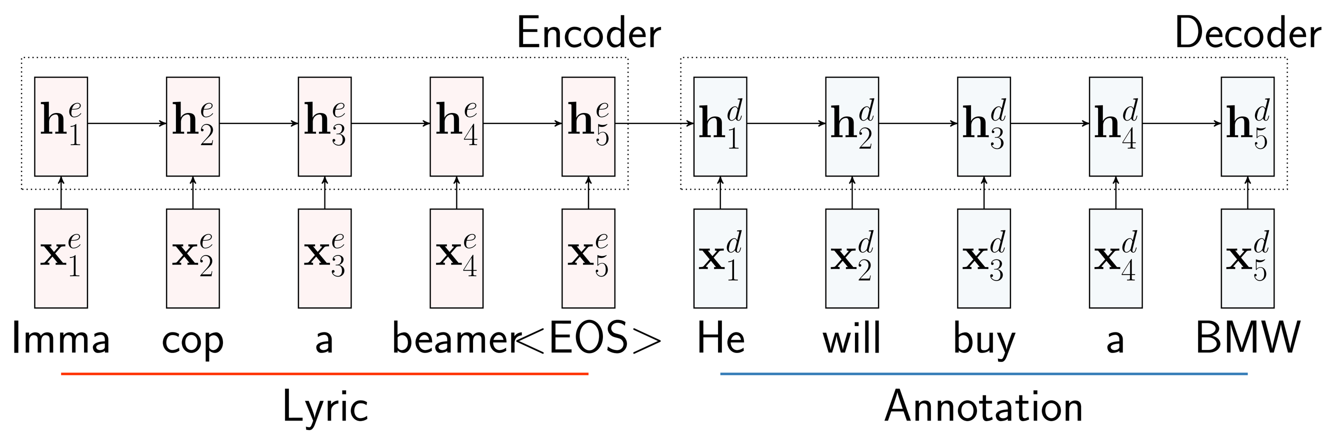

Figure 5: Seq2Seq model for automated lyric annotation

- **Retrieval:** We treat the training corpus as a database of lyrics' excerpts with corresponding annotations, and at test time select the annotation assigned to the most similar lyric. We use standard TF-IDF weighted cosine distance as similarity measure between lyrics' excerpts.

Output Examples

Lyric:	`and when I slept, I dream g's, son I need some`
Human:	he dreams about large amounts of money because he is in need of it .
SMT:	and when he slept, and he's dream, g's son, he needs some
Seq2Seq:	when he sleeps , he dreams of getting money.
Retrieval:	g's refer to the ghetto gangsters in london that the kid has become involved with
Lyric:	`Mossberg pump I'm ridin' shotgun literally`
Human:	riding shotgun means to ride up front in the passenger seat
SMT:	Mossberg pump - action shotgun, he's saying he's ridin' shotgun literally.
Seq2Seq:	Mossberg is a firearms manufacturer. He is riding shotgun with his shotgun.
Retrieval:	he fully respects the people mentioned earlier and is riding for them.
Lyric:	`and my eyes red cause of all that haze`
Human:	his eyes turn red due to lack of hydration from the weed one smokes (in this case , smoking purple haze)
SMT:	and his eyes are red because of all that haze.
Seq2Seq:	his eyes are red because of all the weed he smokes.
Retrieval:	his eyes are red as a sign of the fatigue of travel and , possibly , from drugs and alcohol .

Figure 6: Lyrics excerpts with annotations from Genius ('Human') and automated annotators.

Evaluation

	Properties		Automated Evaluation				Human Evaluation	
	Length Ratio	Profanity/Tok.	BLEU	iBLEU	METEOR	SARI	Fluency	Information
Human	1.19	0.0027	-	-	-	-	3.93	3.53
SMT (Sent.)	1.23	0.0068	<u>6.22</u>	1.44	<u>12.20</u>	<u>38.42</u>	3.82	3.31
Seq2Seq (Sent.)	1.05	0.0023	5.33	<u>3.64</u>	9.28	36.52	3.76	3.25
Seq2Seq	1.32	0.0022	5.15	3.46	10.56	36.86	3.83	<u>3.34</u>
Retrieval	1.18	0.0038	2.82	2.27	5.10	32.76	<u>3.93</u>	2.98

Figure 7: Quantitative evaluation of different automated annotators.

Conclusion and Future Work

- We presented and released the **Genius dataset** to study the task of **Automated Lyric Annotation**. As a first investigation, we studied automatic generation of context independent annotations as **machine translation and information retrieval**. Our baseline system tests indicate that our corpus is suitable to train machine translation systems.
- Standard SMT models are capable of automated lyric annotation but tend to keep close to the structure of the song lyric. Seq2Seq models demonstrated potential to generate **more fluent and informative text**, dissimilar to the lyric.
- A large fraction of the annotations is heavily based on context and background knowledge, one of their most appealing aspects. As future work we suggest **injection of structured and unstructured external knowledge and explicit modeling of references**.