

LARD: Large-scale Artificial Disfluency Generation

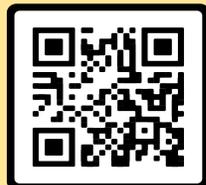
Introduction

- Virtual assistants and spoken dialogue systems are increasingly used in many applications
- Disfluencies:** Interruptions, self-corrections, false-starts, repetitions etc.
- Disfluency detection:** Detection of disfluent regions in spoken language transcripts

Motivation

- Existing datasets do not contain sufficiently all the different types of disfluencies
- Example:** Switchboard [1] contains only 40K/160K utterances with more than 50% of repetitions (most trivial class)
- Existing augmentation techniques use simplistic rules and are not capable of generating all different kinds of disfluencies

Generating artificial disfluencies from fluent text



CHECK IT OUT

Repetition Algorithm

Given a fluent sequence:

Select randomly 1-3 consecutive words to repeat

I want to find a flight.



I want **to find to find** a flight.

Replacement Algorithm

Given a fluent sequence:

- Randomly extract a repair candidate (noun, verb, adjective)
- Generate synonyms and antonyms for the selected candidate
- Replace with a synonym or antonym with/without a repair cue

I am looking for a salon in San Mateo.



I am looking for **a beauty shop no for a salon** in San Mateo.

Restart Algorithm

Given **two or more** fluent sequences:

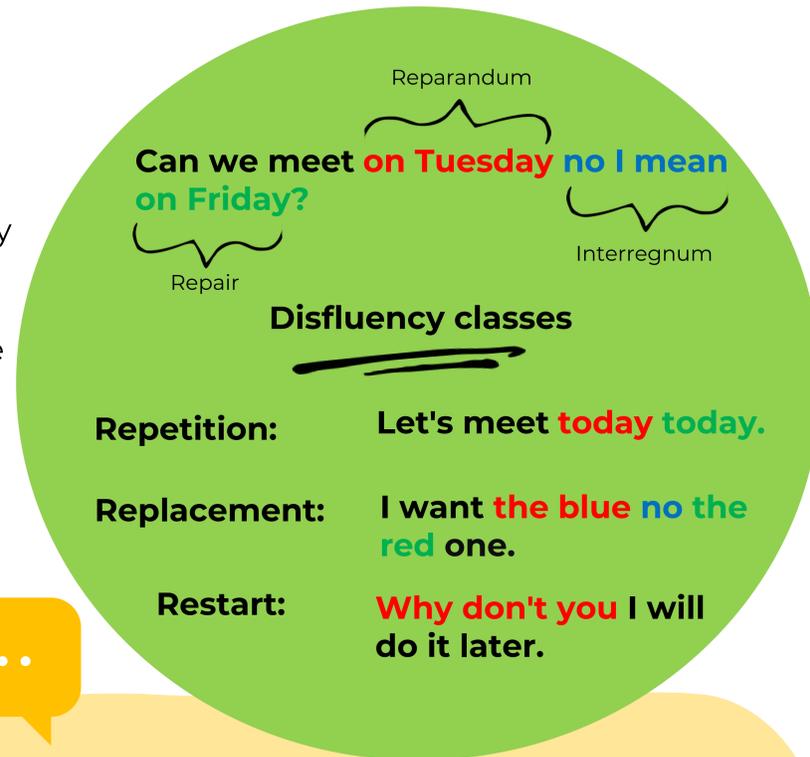
- Randomly pick two sequences
- Split the first sequence in a random position
- Combine the broken sequence with the second unbroken one

Do you want to check-out on March 11th?

When is the check-out date?



do you want to when is the check-out date?



LARD Dataset

Table 1: LARD statistics

Dataset Statistics	
# repetitions	23398
# replacements	23398
# restarts	23398
# fluencies	23398
# total	95992

A new **large-scale artificial and balanced dataset for disfluency detection** based on an existing fluent dataset: Schema-Guided Dataset [2]

Results

Table 2: Experimental results on LARD dataset

	Prec	Rec	F1	BLEU
Detection	97.63	97.61	97.62	-
Classification	97.31	97.30	97.29	-
Extraction	98.12	96.60	97.30	-
Correction	-	-	-	86.48

Table 3: Accuracy (%) for different disfluency classes (repetitions, replacements and restarts) and models trained on different datasets.

	Switchboard (detection)	LARD (detection)	LARD (classification)
Repetitions	85.42	99.57	99.5
Replacements	54.52	99.67	98.39
Restarts	19.6	95.08	93.89

References

- [1] Godfrey, J. J., Holliman, E. C., & McDaniel, J. (1992, March). SWITCHBOARD: Telephone speech corpus for research and development. In Proc. ICASSP 1992 (Vol. 1, pp. 517-520).
- [2] Rastogi, A., Zang, X., Sunkara, S., Gupta, R., & Khaitan, P. (2020, April). Towards scalable multi-domain conversational agents: The schema-guided dialogue dataset. In *Proceedings of AAAI 2020* (Vol. 34, No. 05, pp. 8689-8696).

Acknowledgements

This work is partially funded by the European Commission as part of its H2020 Programme, under the contract number 870930-IA WELCOME Project.

welcome-h2020.eu

