

# Investigating non lexical markers of the language of schizophrenia in spontaneous conversations

Chuyuan Li, Maxime Amblard, Chloé Braud, Caroline Demily, Nicolas Franck, Michel Musiol

{chuyuan.li, maxime.amblard, michel.musiol}@univ-lorraine.fr

chloe.braud@irit.fr

{caroline.demily, nicolas.franck}@ch-le-vinatier.fr



## Introduction

- Schizophrenia is a severe mental illness, affects about 1% of the world's adult population [2]
  - *Positive thought disorder*: disorganized language output such as *derailment* and *tangetiality*
  - *Negative thought disorder*: poverty of speech and language, known as *alogia* [3]
- Contribution of the study: the first SCZ detection in **French dialogues**; proposition of a **delexicalized model**; revelation of special **language features** of SCZ which are confirmed by psychologists

## Corpus

- SLAM project [4], free exchanges between 1 psychologist (PSY) and
  - 18 patients of Schizophrenia (SCZ)
  - 23 controls: students (STU)
 → **lexical bias**
- 2 balanced groups: gender, age, IQ, nb. years of studies, 3 cognitive tests (WAIS-III, TMT, CVLT)

## Translated Examples

### PSY-SCZ

psy - So now you are going to a workshop hum, what is it?

scz - Yes, so I went to a **therapeutic** workshop... what do they call it...

psy - Therapeutic education... right

### PSY-STU

psy - What do you want to do after?

stu - Uh I would like to do the **master** of psychopathy of the cognition and the interactions.

psy - Mmh mmh.

## Related Work and Our Approach

### 1. Automatic Classification of SCZ:

	data type	language	feature	result
Strous et al., 2009	written	en	lexical	Acc. = 83.3%
Mitchell et al., 2015	<i>tweets</i>	en	lexical	Acc. = 82.3%
Kayi et al., 2018	written and <i>tweets</i>	en	syntactic	F1 = 81.65%
Allende-Cid et al., 2019	narrative texts	en	morpho-synt.	F1 = 82.8%
Amblard et al., 2020 [1]	<b>clinic conversations</b>	<b>fr</b>	lexical	Acc. = 93.7%

⇒ Corpus of different nature: difficult to compare

### 2. Our Approaches:

**Varying dialogue size:** Tackle with data sparsity, introduce *more or less* context

- Indiv. setting: classify individual speech turn → **no context**
- Full setting: concatenation of all speech turns → **full context**
- W-n setting ( $n \in \{128, 256, 512\}$ ): window of at least  $n$  words → **partial context**

**Comparing representations:** Minimize lexical bias, test with *less lexicalized* features

- Dialogical features: Open Class Repair (OCR): "pardon?", "huh?"; Backchannel (BC): "yeah", "hum mmh"; Connectives (Conn): "because", "but"
- Morpho-syntactic features:  $n$ -gram Part-of-speech (POS) and treelet

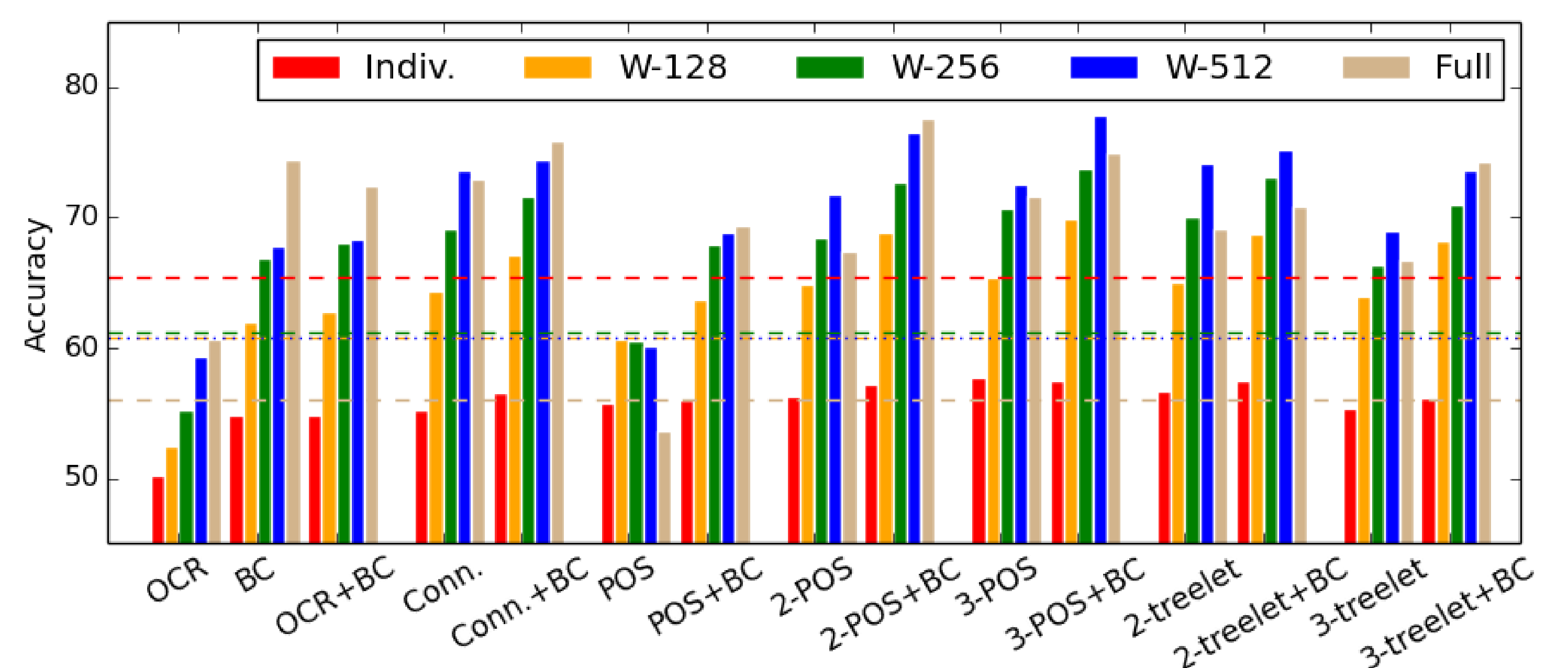
**Classifiers:** Naive Bayes, Logistic regression, SVM, Random Forest, Perceptron

## Results and Analysis

Traits	Full	Indiv.	W-128	W-256	W-512
bow	<b>93.66</b>	<b>72.43</b>	-	-	-
ngram	85.61	69.59	-	-	-
OCR	60.62	50.17	52.43	55.19	59.28
BC	74.48	54.79	62.01	66.89	67.86
Connectives	72.44	55.28	64.05	69.68	73.57
POS	53.66	55.80	60.63	60.48	60.09
2-POS	67.36	56.33	64.85	68.53	71.74
3-POS	71.65	56.53	65.39	70.66	72.55
2-treelet	69.19	56.73	65.02	70.11	74.19
3-treelet	66.78	55.34	63.95	66.39	69.03
1-2-3-POS	69.01	58.36	66.19	72.03	72.67
POS+2-3-treelet	66.59	57.77	65.52	69.11	72.39
3-POS+BC	<b>74.93</b>	<b>57.46</b>	69.92	73.75	<b>77.86</b>

⇒ Performance drops without **lexical information**

⇒ **Morpho-syntactic features**: very good indicators



⇒ OCR: poor results due to few occurrences

⇒ BC: improve results syst., especially when combined with POS

⇒ Conn: good indicators

## Conclusion and Future Work

### Final conclusion

- Test different representations for context and linguistic features; Study of "high level" features

### Future Work

- Take into account the full interaction with a neural hierarchical architecture
- Extension to other tasks: impact of speaker's features on the dialogical structure, e.g. emotion, other mental disorders

## References

- [1] Maxime Amblard, Chloé Braud, Chuyuan Li, Caroline Demily, Nicolas Franck, and Michel Musiol. Investigating learning methods applied to language specificity of persons with schizophrenia. In *Traitement Automatique des Langues Naturelles (TALN, 27e édition)*, 2020.
- [2] APA. *DSM-5-Manuel diagnostique et statistique des troubles mentaux*. 2015.
- [3] Gina R Kuperberg. Language in schizophrenia part 1: an introduction. *Language and linguistics compass*, 4(8):576–589, 2010.
- [4] Manuel Rebuschi, Maxime Amblard, and Michel Musiol. Using SDRT to analyze pathological conversations. Logicality, rationality and pragmatic deviances. In *Interdisciplinary Works in Logic, Epistemology, Psychology and Linguistics: Dialogue, Rationality, and Formalism*. 2014.