The Treatment of Some Deviation Cases in Non-Standard Data Using Artificial Intelligence and Natural Language Processing Applied to Arabic Language

Dr. Bilal Shafei

An-Najah National University

Dr. Bilal Shafei

❑ Director of Language Resource Centre (LRC)

❑ Doctor of Philosophy Franche-Comté University

❑ Visiting professor: Columbia University

❑ Visiting professor: Paris 6 –LIP6

The Treatment of Some Deviation Cases in Non-Standard Data Using Artificial Intelligence and Natural Language Processing Applied to Arabic Language

The Treatment of Some Deviation Cases in Non-Standard Arabic Data

I - Introduction

- **Deviance definition**

- **Deviance or Ill-formed input**

- **Types of Deviance**

- **Examples of Deviance**

II - Could Artificial Intelligence produce systems without deviance?

Example:

- هذا البيوت ـــ هذا البيت ـ هذه البيوت,

- وترجل الفرس بجانب النهر

II - Could Artificial Intelligence produce systems without deviance?

ترجل الفرس بجانب النهر وبدأ يصطاد السمك ليأكله

لا توجد أخطاء في النص

III - Does AI have the tools to analyze deviance?

- Expert Systems and NLP

  1) Different forms (vowels or not)

  2) Right to left script

  3) Agglutination

  4) Different levels of Arabic

  5) Regional languages

  6) Different periods and

  7) Continuous text level

III - Does AI have the tools to analyze deviance?

- **<u>Dictionaries of forms :</u>**
  - without vowels could reach 500, 000 entries
  - with vowels will be about 1,000, 000.
  - agglutinated forms 100 million
  - grammatical classes :

III - Does AI have the tools to analyze deviance?

- grammatical classes :

  - $\sum$verb at the 3rd person masculine singular/active,

  - $\sum$Verb at the 3rd person masculine singular/passive,

  - $\sum$substantive masculine plural,

  - $\sum$substantive masculine singular

  - $\sum$imperative verb 2nd person masculine singular.

IV - The indicators that could help to understand the deviance

❖ *Indicators given by the subject of deviance case*

❖ *Indicators given by actor of the case*

❖ *Indicators that will generate the forms*

❖ *Classification of indicators*

❖ *Treatment of indicators*

V - How can we explain indicators?

➢ **Describing a collection of rules and conditions**

➢ **Taking into consideration all the other factors**

➢ **Explaining indicators by a human expert**

➢ **Constructing a model for the treatment of the indicators**

VI - Conclusion

- Not easy to deal with deviance cases in the non-standard Arabic

- Highlight issues related to deviance cases and how the AI can be useful

- Other studies in the different fields of NLP applied to Arabic

- Using different formal analyses to handle the problem

Thanks

Questions and/or

Comments