International conference NooJ 2017
Kenitra, Morocco, May 18-20

Arabic NooJ Parser: Nominal Sentence Case

Nadia GHEZAIEL HAMMOUDA & Kais HADDAR

Miracl Laboratory, University of Sfax, Tunisia

Outline

- Introduction
- Previous work
- Typology of nominal sentence
- Proposed method
- Experimentation and evaluation
- Conclusion and perspectives

Introduction

Parsing Arabic corpora is an important task aiming to:

- Understand Arabic language
- Enrich and enhance the electronic resources
- Increase the efficiency of application for Arabic corpora like translation or recognition of named entities

... Introduction

Arabic is considered one of the difficult languages to analyze due to its morphological and syntactical characteristics

There are two types of sentences in Arabic:

- verbal sentence
- nominal sentence

... Introduction

Different forms of nominal sentence exist

Interaction between nominal and verbal sentences

Formalization of rules requires much effort to guarantee several qualities:

- efficiency
- robustness
- extensibility

… Introduction

Transducers have proved their usefulness in a wide variety of applications in NLP

Transducer cascades made possible to carry out a robust and a highly precise syntactic analyzes on different corpora

... Introduction

Transformation of recursive graph of transducers into transducer cascade is very interesting

The transformation is a difficult task due to :

- difference between application levels
- Interaction of linguistic phenomena

… Introduction

For the cascade:

- The order of transducers should respect different constraints
- Constraints can be deduced from studies done on Arabic corpus

... Introduction

Our objective is to construct an Arabic parser implemented in NooJ

To do that, we:

- study, essentially, the Arabic nominal sentences but also other sentence forms
- establish a set of rules recognizing nominal sentences that can be generalized to treat any sentence type
- implement the transducer cascade in NooJ

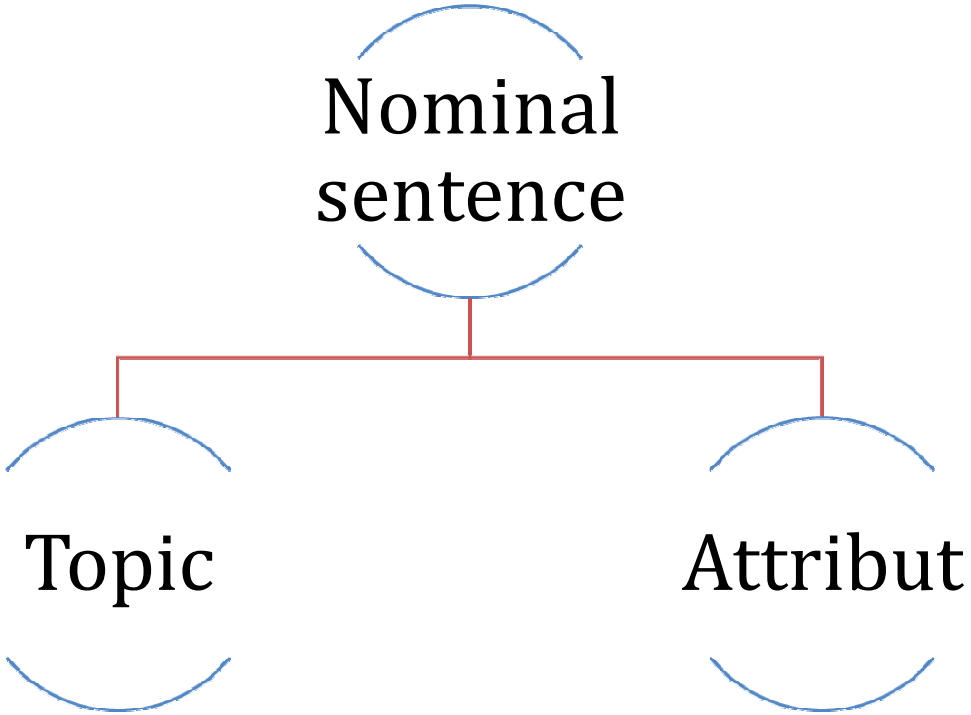
Previous work

	Approach	Formalism	Tool
Habash et al., 2009	Hybrid	Production rules and numeric algorithms	Arabic parser
Abdelkarim et al., 2006	Linguistic	HPSG	HPSG parser for nominal sentences
Boukédi et al., 2014	Linguistic	HPSG	HPSG Parser for Coordination
BenSalem et al., 2017	Hybrid	Property grammar	Arabic Parser for ATB sentences

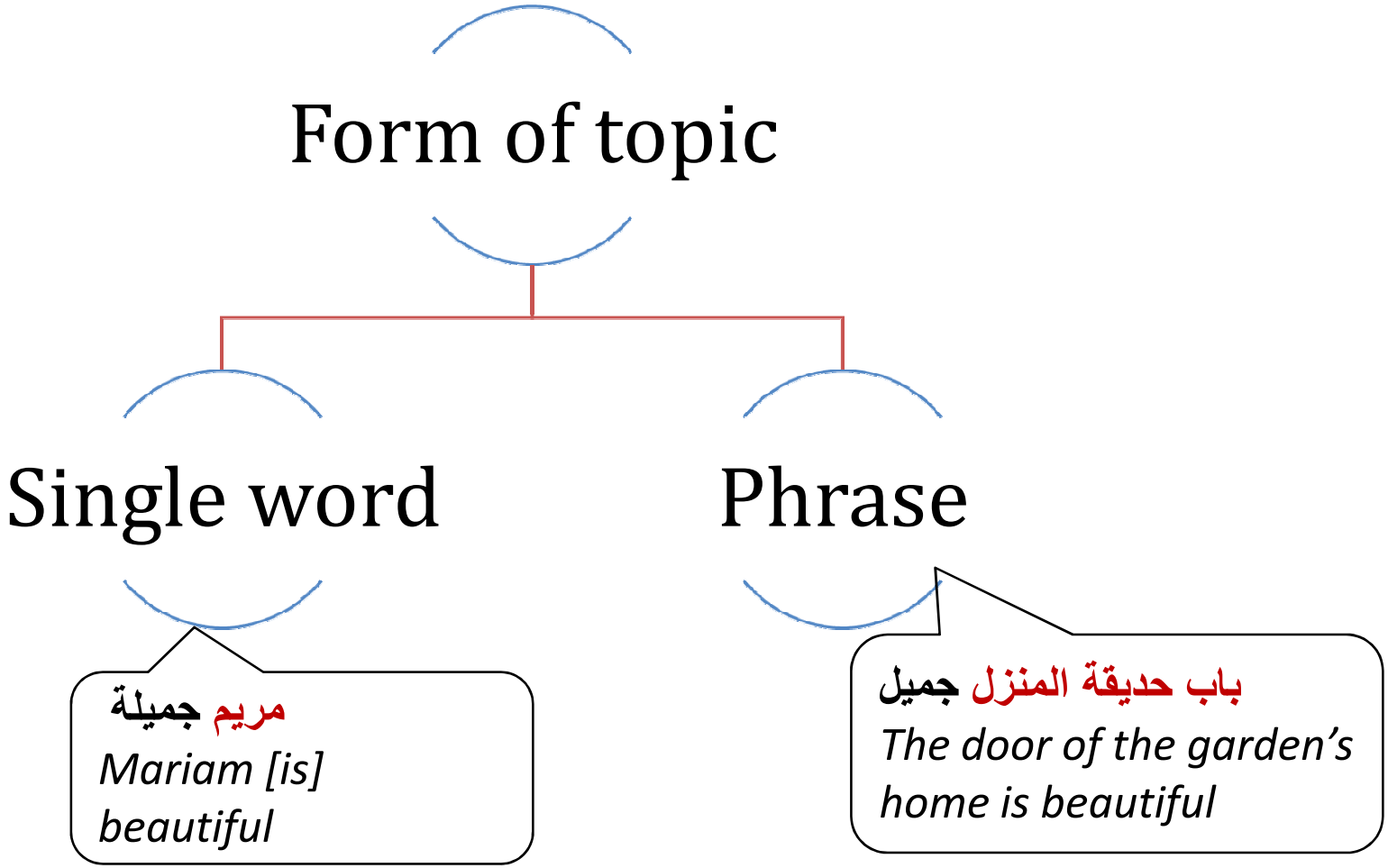
… Previous work

- ☹ Lack of complete linguistic approach for all sentence type
- ☹ Absence of NooJ Arabic parser for sentence complex structure
- ☹ Parser execution time is very high

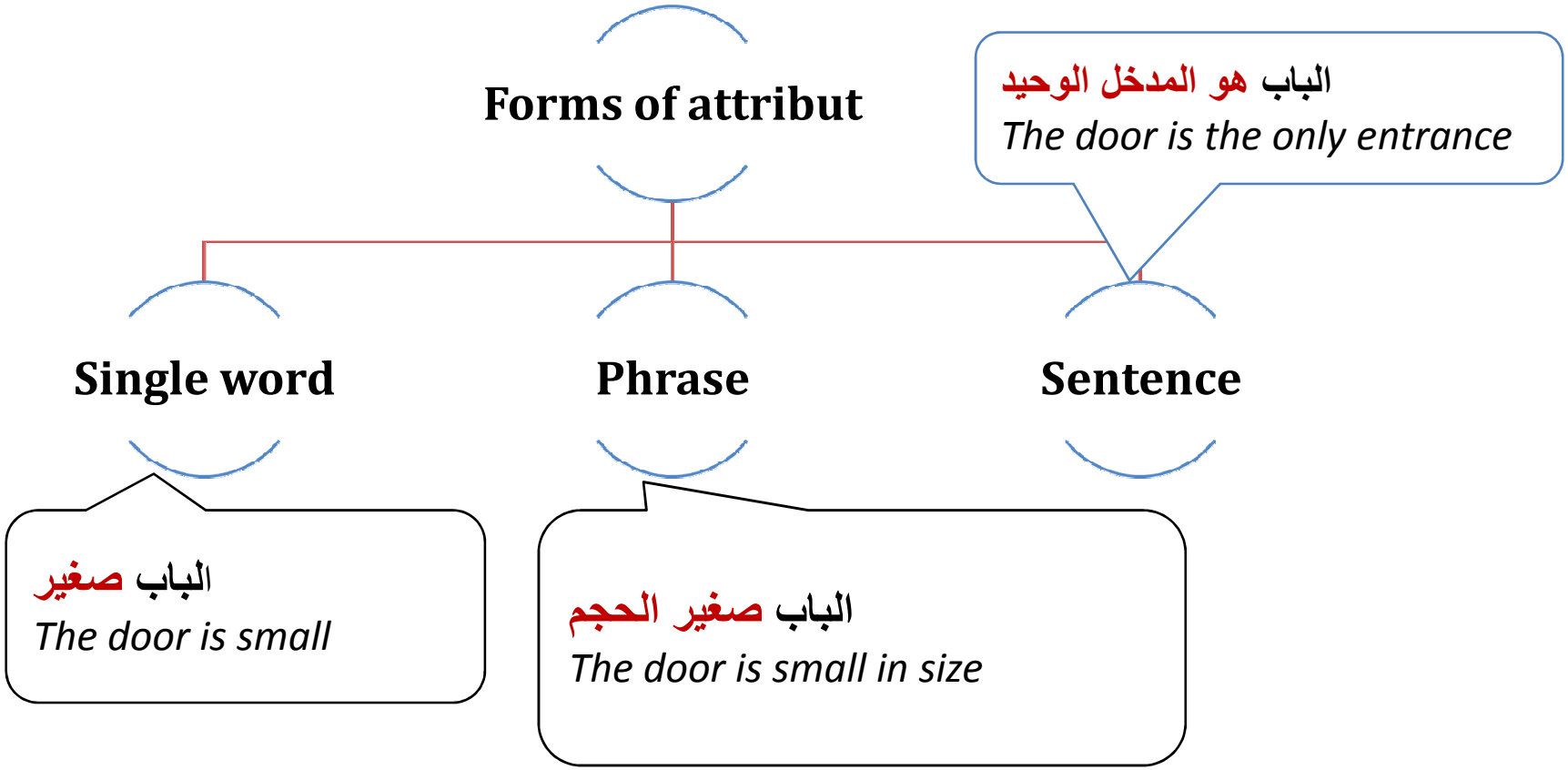
Typology of nominal sentence



… Typology of nominal sentence



… Typology of nominal sentence



… Typology of nominal sentence

A nominal sentence can contain a verbal sentence with different forms

Verb

الفتاة تجري

Verb + adjective

The girl runs

الفتاة تجري سريعا

Verb + noun

The girl running quickly

الرجل يساعد المحتاجين

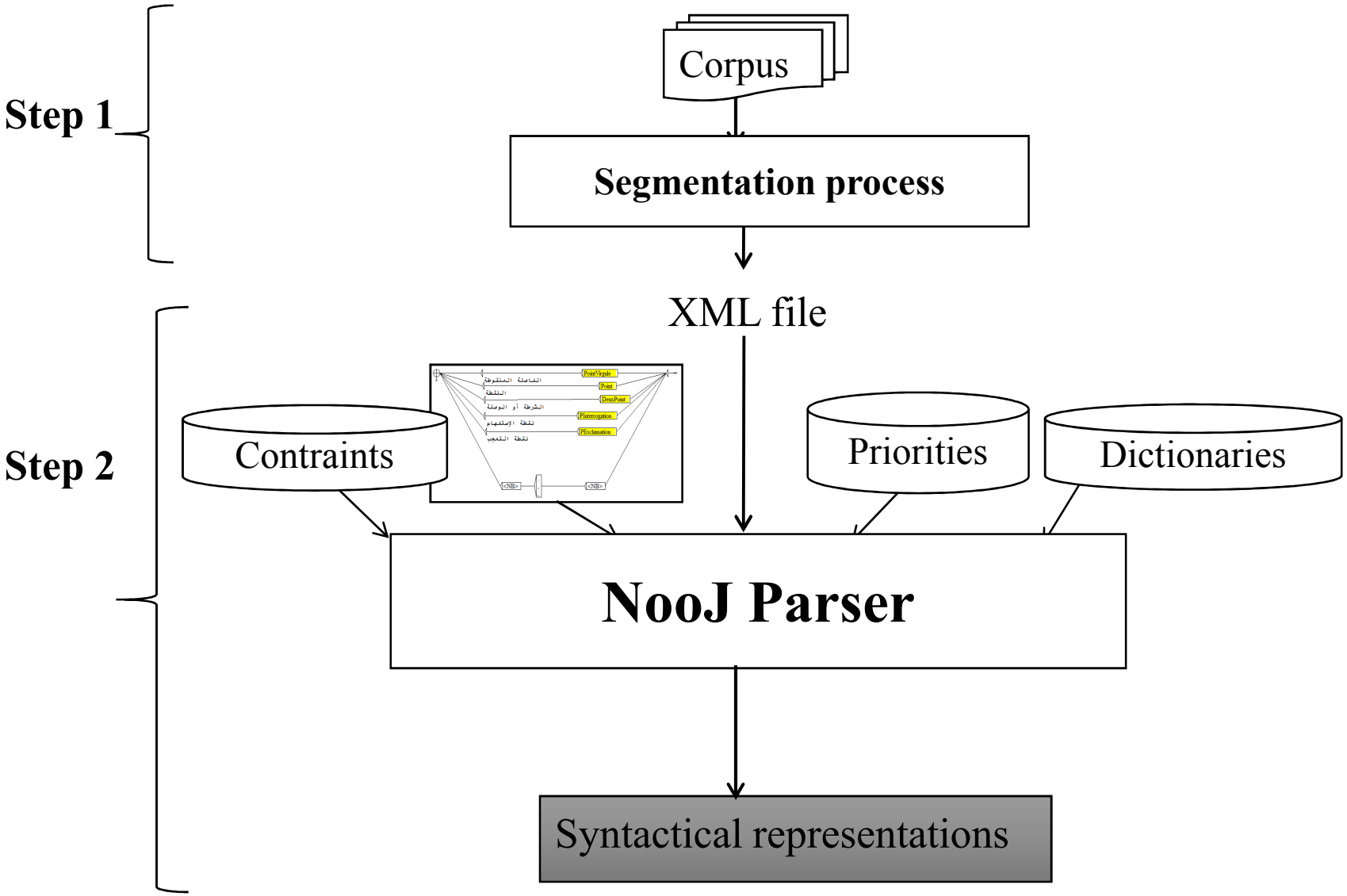
Verb + noun + adjective + noun

The man helps poor people

المديرة تمنح التلاميذ المميزين الجوائز

The director gives prices to distinguished students

Proposed method



… Proposed method

High level of granularity is used to solve some lexical ambiguities and reduce the complexity of parser

Illustrative example

Noun, مدرسة	Niveau 0
Indefinite Noun, مدرسة	Niveau 1
Indefinite Nominative Noun, مدرسة	Niveau 2

Tag set for annotation

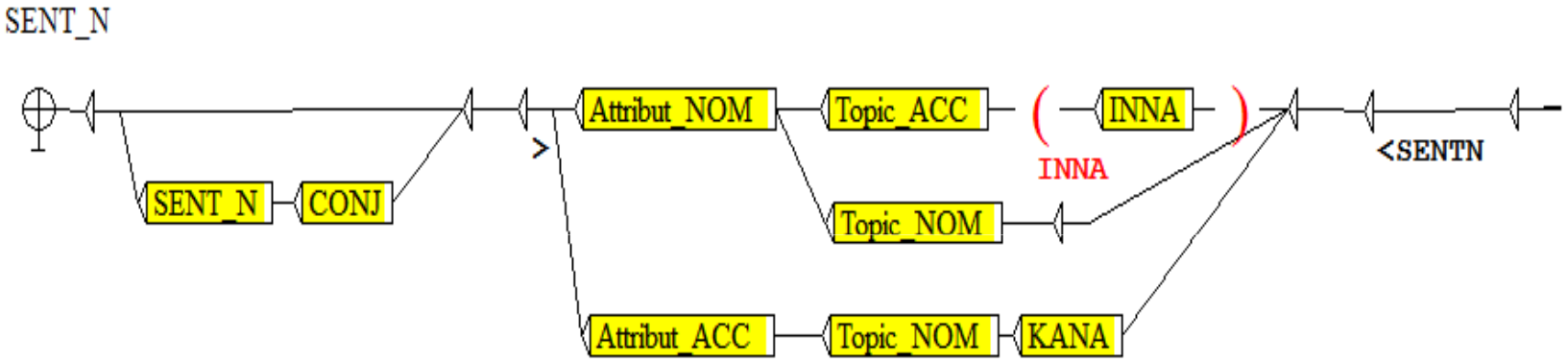
NN	**Indefinite Nominative Noun *u***
NTN	**Indefinite Nominative Noun *un***
NND	**Definite Nominative Noun *u***
NA	**Indefinite Accusative Noun *a***
NTA	**Indefinite Accusative Noun *an***
NAD	**Definite Accusative Noun *a***
NG	**Indefinite Genitive Noun *i***
NTG	**Indefinite Genitive Noun *in***
NGD	**Definite Genitive Noun *i***

Experimentation and evaluation

Morphological grammars	Numbers
Verb inflected form patterns	113
Inflected relative pronoun patterns	8
Broken plural patterns	10
Agglutination's grammars	3

… Experimentation and evaluation

Recursive graph for parsing



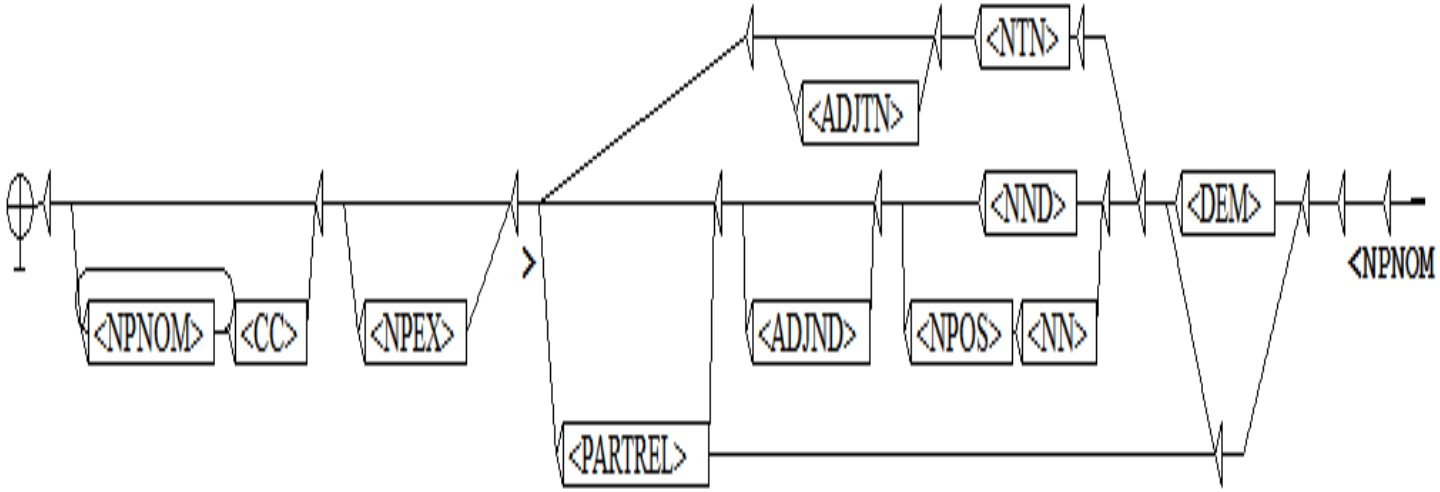
Transducer for a nominal sentence

Each state of the transducer is considered as a sub-graph

… Experimentation and evaluation

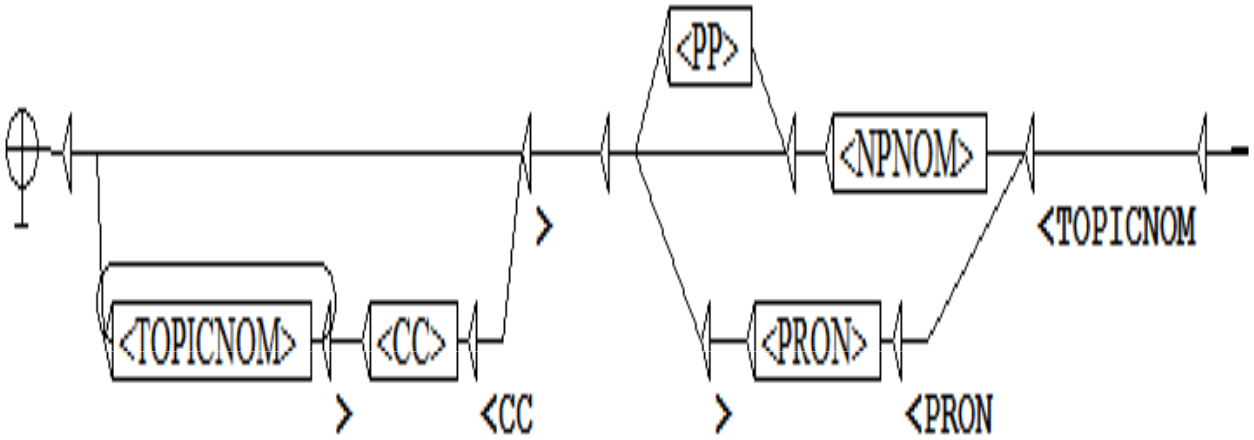
Cascade for parsing

Each nominal sentence component is implemented by a separated transducer



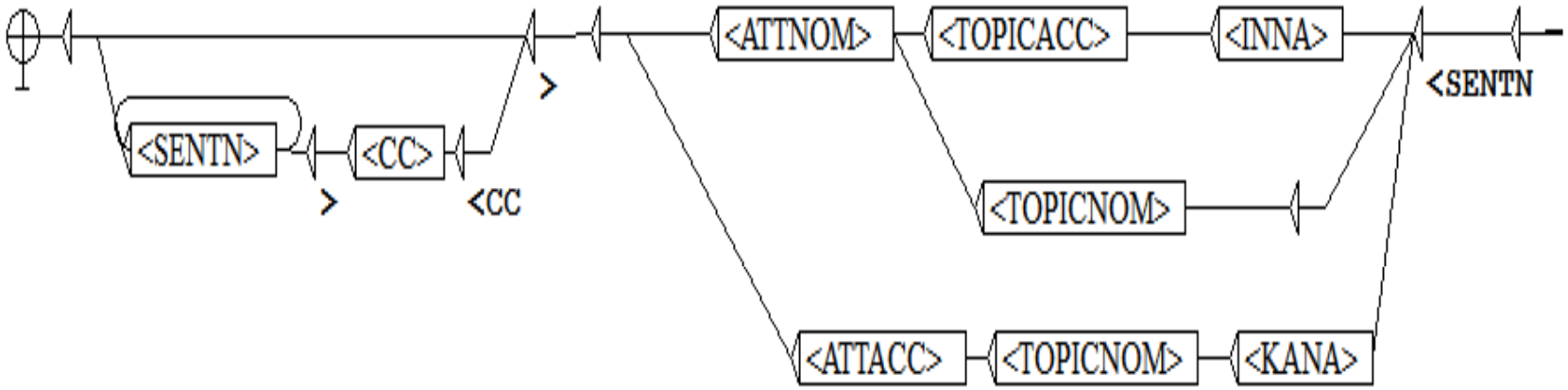
Transducer for nominative NP

… Experimentation and evaluation



Transducer for a topic

… Experimentation and evaluation



Transducer for a nominal sentence

… Experimentation and evaluation

- The call order of transducers is fixed inspired from our carried out study
- Starting by phrases until gathering all the sentence
- Particles → Phrases → Sentences

… Experimentation and evaluation

Syntactic transducer cascade:

Total number of graphs is 50 called in a fixed order

Syntactic Resources:

Order	Grammar
1	Cas_CONJ.nog
2	Cas_Daref.nog
3	Cas_PrepZamen.nog
4	Cas_PrepPART.nog
5	Cas_DEM.nog
6	Cas_KANA.nog
7	Cas_INNA.nog
8	Cas_ProREL.nog
9	Cas_TOOL.nog
10	Cas_NG.nog

… Experimentation and evaluation

Corpus	Number	Percentage
Sentences	5900	100%
Totally disambiguated	4840	83%
Partial disambiguated	1060	18%
Failed disambiguation	0	0%

… Experimentation and evaluation

Recursive Graph

Corpus	Precision	Recall	F-mesure
5900 sentences	0,6	0,7	0,62

Cascade

Corpus	Precision	Recall	F-mesure
5900 sentences	0,74	0,82	0,77

Conclusion and perspectives

Study different types of Arabic nominal sentence

Establish a transducer's cascade to analyze Arabic corpora

Compare our new method with a recursive one

As perspectives:

Increase the granularity's level

Generalizing our method by covering other linguistic phenomena

Thank you for your attention

Questions ?