



Craiova, 11-12 September 2017



International Conference on  
Human-Computer Interaction

# IMPROVING AN NP-CHUNKER BY EXPLOITING THE INNER NP DEPENDENCY STRUCTURE

Mihaela Colhon  
Department of Computer  
Science  
University of Craiova  
mcolhon@inf.ucv.ro

Dan Cristea  
Faculty of Computer Science  
"Alexandru Ioan Cuza" University of Iași

Institute for Computer Science  
Romanian Academy - the Iași branch  
dcristea@info.uaic.ro

# Motivation

A way to incorporate into user interfaces knowledge expressed in natural language or modalities for communicating in a human to human style can be achieved by implementing NLP tasks. In the context of HCI the NLP issues range from various *speech recognition* systems, to *natural language interfaces* for various applications, or a multitude of *machine translation* or *information retrieval* systems.



## Motivation (cont.)

As an exemplification of the necessity of reliable NLP tools, let us consider the following *question-answer* example:

- the input: "[..] Prices of sunflower oil rose by 27% due to low production [..]"
- and the question: "For which food the price has increased by 27%?"

The answer "sunflower oil" is within the nominal group "Prices of sunflower oil "

→

the correct identification of the nominal group is absolutely necessary in order to provide the correct answer.

# Google vs. Bing


## Who understands NP better?

Google

All Images News Videos Maps More Settings Tools


About 838,000 results (0.54 seconds)

Romania / President

Klaus Iohannis 

More about Klaus Iohannis [Feedback](#)

**Președintele României**  
[www.presidency.ro/](http://www.presidency.ro/) [Translate this page](#)  
Președintele României, domnul Klaus Iohannis, a susținut miercuri, 30 august a.c., la Palatul



Web Images Videos News My saves

1,270,000 RESULTS

**Președintele României - presidency.ro**  
[www.presidency.ro](http://www.presidency.ro/)  
O reuniune plină de emoție astăzi, la Sibiu, cu foștii mei elevi de la Colegiul Național „Samuel von Brukenthal”, la 20 de ani de la absolvire.

<b>Decrete Si ACTE Oficiale</b> Documente în baza Legii 544/2001 . Pentru a vizualiza versiunea ...	<b>Comunicate De Presă</b> Documente în baza Legii 544/2001 . Pentru a vizualiza versiunea ...
<b>Consilierii Președintelui</b> Departamentul Relații cu Autoritățile Publice și Societatea Civilă - ...	<b>Cancelaria Ordinelor</b> NOUI Republicarea Legii nr. 29/2000 . privind sistemul național de decorații ...

[See results only from presidency.ro](#)

**Președintele României - Wikipedia**  
[https://ro.wikipedia.org/wiki/Președintele\\_României](https://ro.wikipedia.org/wiki/Președintele_României)  
Președintele României este șeful de stat al României. El are rolul de a reprezenta statul român în relațiile internaționale, conform Constituției, este ...  
Istoric · Atribuțiile ... · Alegerea Președintelui · Jurământul ...

**Contact - presidency.ro**

**Related searches**  
klaus iohannis presedintele romaniei  
adresa presedintelui romaniei  
presedintele romaniei 2016  
ioannis presedintele romaniei  
presedintia romaniei cotroceni  
primul presedinte al romaniei  
administratia prezidentiala romania  
iohannis presedinte

# Google vs. Bing

## Who understands NP better?

Google

sediul guvernului romaniei

All Images Maps News Videos More Settings Tools

About 530,000 results (0.67 seconds)

- Guvernul României - Wikipedia**  
[https://ro.wikipedia.org/wiki/Guvernul\\_României](https://ro.wikipedia.org/wiki/Guvernul_României) Translate this page  
Guvernul se compune din un lider numit prim-ministru și un număr variabil de ... **Guvernul României** ocupă în București, ca **sediul** al activităților sale Palatul ...  
Prezentare generală · Structură · Aparatul de lucru al ... · Minister
- Palatul Victoria - Wikipedia**  
[https://ro.wikipedia.org/wiki/Palatul\\_Victoria](https://ro.wikipedia.org/wiki/Palatul_Victoria) Translate this page  
Palatul Victoria este un palat din București, situat în Piața Victoriei și **sediul** al **Guvernului României**. A fost construit în 1937 după planurile arhitectului Duiiu ...
- Palatul Victoria - Guvernul României**  
[gov.ro/guvernul/palatul-victoria/istoric-ul-cladirii](http://gov.ro/guvernul/palatul-victoria/istoric-ul-cladirii) Translate this page  
**Guvern** - Palatul Victoria ... **sediul** Ministerului de Externe și al Consiliului de Miniștri și a devenit, în 1990, **sediul** al primului **guvern** al **României** post-comuniste.

Bing

sediul guvernului romaniei

Web Images Videos News My saves

594,000 RESULTS

- gov.ro - Învățarea în Parlament a Guvernului Tudose**  
[www.gov.ro](http://www.gov.ro)  
Ministrii **Guvernului României**. Prim-ministrul conduce Guvernul și coordonează activitatea membrilor acestuia, cu respectarea atribuțiilor legale care le revin.  
Guvernul României · Procesul LEGISLATIV · Cabinetul De Miniștri · Guvern · Știri
- Guvernul României - Wikipedia**  
[https://ro.wikipedia.org/wiki/Guvernul\\_României](https://ro.wikipedia.org/wiki/Guvernul_României)  
Numirea **guvernului** se face de Președintele **României** pe baza votului de încredere acordat de către parlament. ... este **sediul** **Guvernului**.  
Prezentare generală · Structură · Procesul legislativ · Cronologia prim ...
- Palatul Victoria - Enciclopedia României - prima ...**  
[enciclopediaromaniei.ro/wiki/Palatul\\_Victoria](http://enciclopediaromaniei.ro/wiki/Palatul_Victoria)  
**sediul** **Guvernului României**: Localitate București: Arhitect Duiiu Marcu: Adresa Piața Victoriei nr. 1 ... a fost construită pentru a fi **sediul** Ministerului de ...
- Perchezitii DNA la sediul Guvernului! Ar fi vizat un ...**  
[www.ebihoreanul.ro/stiri/ultima-or-31-35/perchezitii-dna-la-sediul...](http://www.ebihoreanul.ro/stiri/ultima-or-31-35/perchezitii-dna-la-sediul...)  
Descinderi, luni, la **sediul** **Guvernului României**, din Piața Victoriei, din Capitală. etariatul General al ...

# Google vs. Bing

## Who understands NP better?

Google

US government headquarters

All Images Maps News Videos More Settings Tools

About 190,000,000 results (0.95 seconds)

Federal government of the United States / Headquarters



Washington, D.C., United States

Plan a trip and points of interest

b

US government headquarters

Web Images Videos News My saves

71,500,000 RESULTS

- USA.gov - Official Site  
<https://www.usa.gov>  
USA.gov can help you start your search for government information by topic and agency.  
Federal Agencies · Unclaimed Money · Travel and Immigration · Benefits and Grants
- A-Z Index of U.S. Government Departments and Agencies  
<https://www.usa.gov/federal-agencies>  
The .gov means it's official. Federal government websites often end in .gov or .mil. Before sharing sensitive information, make sure you're on a federal government ...  
Government Agencies · U.S. Air Force · U.S. Army · Explore Topics
- Federal government of the United States - Wikipedia  
[https://en.wikipedia.org/wiki/Federal\\_Government\\_of\\_the\\_United\\_States](https://en.wikipedia.org/wiki/Federal_Government_of_the_United_States)  
Headquarters: The White House: Main ... The terms "Government of the United States of America" or "United States Government" are often used in official documents to ...  
Naming · History · Legislative branch · Executive branch · Judicial branch
- USAJOBS - The Federal Government's Official Jobs Site  
<https://www.usajobs.gov>  
USAJOBS is the Federal Government's official one-stop source for Federal jobs and employment information.

# Resume

**The General Problem.** Reliable dependency parsing is a notorious difficult problem in Natural Language Processing.

what we do: pattern-based parsing of Noun Phrase (NP) chunks

why NP chunks?

- because NPs cover a significant part of a sentence
- their structure can be determined with enough trust
- usually NPs have contiguous structures in the overall sentence tree

why pattern-based?

- because corpus evidence shows that similar sequence of tags in similar contexts are linked by the same dependency relations

# Abstract

This paper presents an on-going research started in 2014 on a treebank dependency parsing mechanism that is restricted to NP chunks which addresses Romanian language.

The present study refines the originally proposed syntactic patterns by considering not only the two terms of the dependency relation but also the terms surrounding these two that will be treated as contexts: *left*, *middle*, *right* depending on their position from the dependency terms.

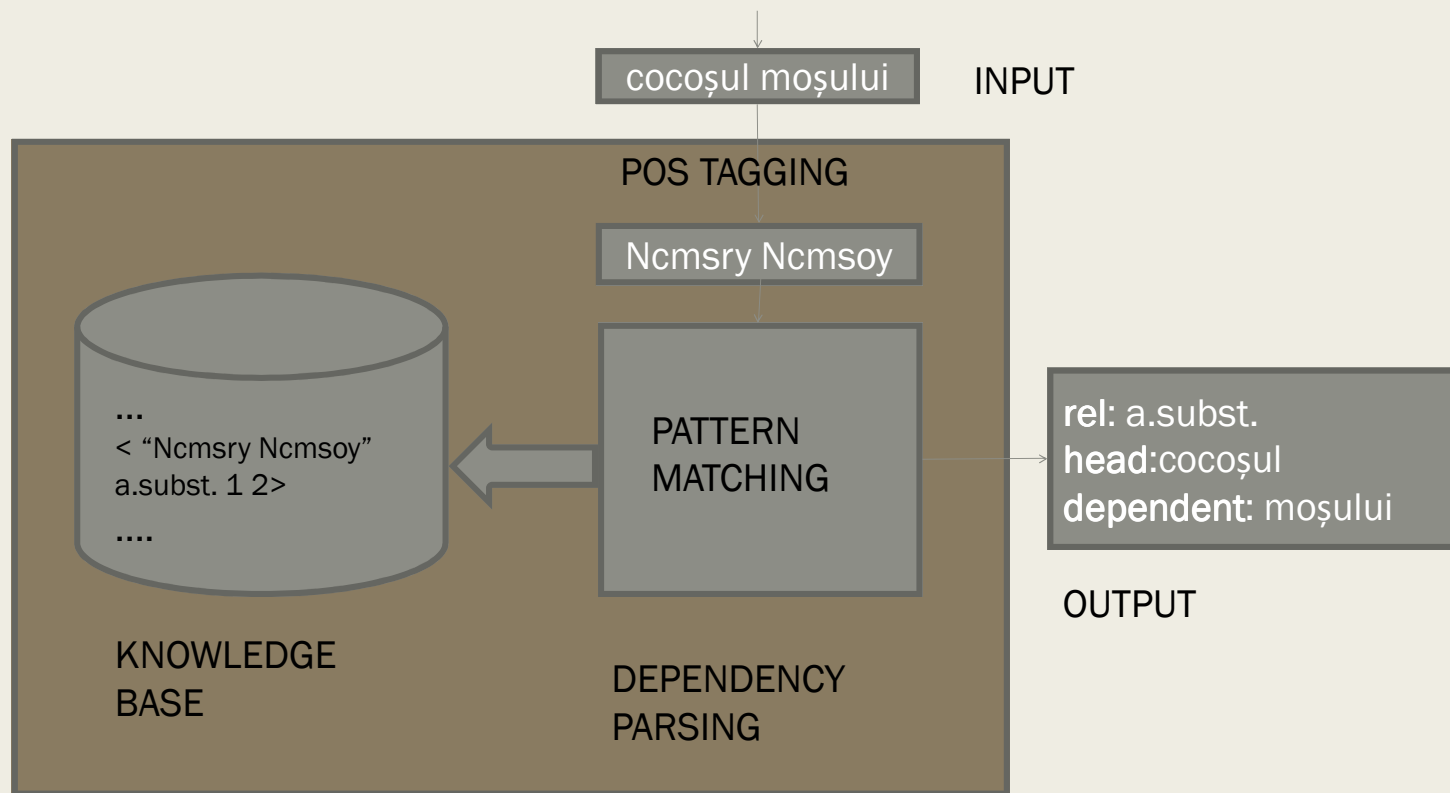
We work on Romanian, but the method is general enough to be applicable to other languages as well.



# Nouns and Noun Phrases

- A NOUN is a word that names a person, place, or thing.
- Phrases occurring in Object and Subject positions are NOUN PHRASES.
- In most phrases, there is a core word, called the HEAD of the phrase.
- We name a phrase by the category of the head.
  - *Example: “That big fish” is a NOUN PHRASE because its head is a noun – “fish”.*

# Pattern-based Parsing



# Pattern-based parsing. Automatic Construction of Dependency Grammar Rules

The information taken from the corpus for each NP sequence:

<<NP chunk> <dependency link> <idx1> <idx2>>

- <NP chunk> the sequence of MSD tags,
- <dependency link> the dependency relation,
- <idx1> and <idx2> the positions of the head word and respectively, the dependent word

## Example:

<NP>

<word deprel="sbj." form="Holul" head="3" id="1" postag="Ncmsry"/>

<word deprel="a.subst." form="blocului" head="1" id="2" postag="Ncmsoy"/>

</NP>



< "Ncmsry Ncmsoy" a.subst. 1 2>  
a.subst.(Holul-1,blocului-2)

The corpus contains 374 sentences, 6778 tokens and 650 NP structures.

# Pattern-based parsing

The mechanism: generates and apply corpus-based patterns for a dependency parsing application.

<sentence id="Oro.1.2.3.">

<NP>

<word deprel="sbj." form="Holul" head="3" postag="Ncmsry"/>

<word deprel="a.subst." form="blocului" head="1" postag="Ncmsoy"/>

</NP>

<word form="mirosea" head="0" postag="Vmii3s"/>

<word deprel="c.i." form="a" head="3" postag="Spsa"/>

<NP>

<word deprel="prep." form="varză" head="4" postag="Ncfsrn"/>

<word deprel="a.adj." form="călită" head="5" postag="Afpfsrn"/>

</NP>

<word deprel="coord." form="și" head="4" postag="Crssp"/>

<word deprel="coord." form="a" head="7" postag="Spsa"/>

<NP>

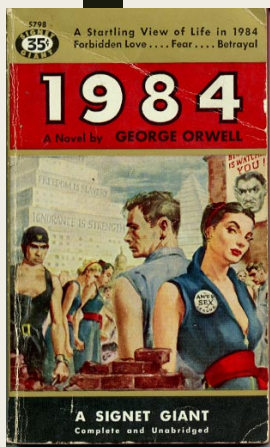
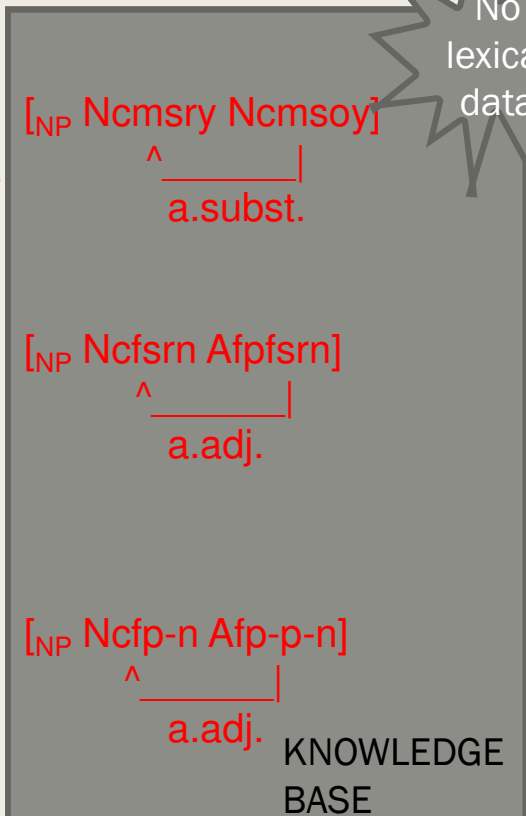
<word deprel="prep." form="preșuri" head="8" postag="Ncfp-n"/>

<word deprel="a.adj." form="vechi" head="9" postag="Afp-p-n"/>

</NP>

<word deprel="punct." form="." head="3" postag="punct"/>

</sentence>



# Extracting the Patterns. Examples


A Romanian noun phrase

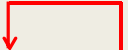
„un fond de rezervă” (English ”a reserve fund”):


- 4 MSD tag sequence:

Timsr Ncms-n Spsa Ncfsrn

- represented in 3 patterns, each identifying a particular dependency relation

  
Timsr Ncms-n Spsa Ncfsrn

  
Timsr Ncms-n Spsa Ncfsrn

  
Timsr Ncms-n Spsa Ncfsrn

# Patterns Generalization

Objectives:

- reduce the total numbers of patterns found in the corpus
- cover similar constructions for which no occurrence has been found in the corpus

Generalization mechanism:

- for the same dependency construction
  - collapse different contextual representations into a single expression.
  - collapse similar MSD tags into a single condensed tag

# Patterns Generalization. Example

A syntactic pattern of the form :

Nc[fm]sry Ncfsoy (Pp3fso.|Apfson)?

codifies six NP morphological structures:

- Ncfsry Ncfsoy Pp3fso-
- Ncmsry Ncfsoy Pp3fso-
- Ncfsry Ncfsoy Apfson
- Ncmsry Ncfsoy Apfson
- Ncfsry Ncfsoy
- Ncmsry Ncfsoy.

# Examples of patterns consisting of 3 MSD tags

acoperișul Blocului Victoria En: the Victoria block's roof



Ncmsry Ncmsoy Ncfsry



**Ncmsry** Ncmsoy Ncfsry



a.subst.

Ncmsry **Ncmsoy** Ncfsry



a.subst.

< "Ncmsry Ncmsoy Ncfsry" a.subst. 1 2 >

< "Ncmsry Ncmsoy Ncfsry" a.subst. 2 3 >



# Examples of patterns consisting of 3 MSD tags

En: the Two Minutes

**cele Două Minute**



Tdfpr Mcfp-l Ncfp-n



Tdfpr **Mcfp-l** Ncfp-n



det.

< "Tdfpr Mcfp-l Ncfp-n" det. 2 3 >

Tdfpr Mcfp-l **Ncfp-n**



a.adj.

< "Tdfpr Mcfp-l Ncfp-n" a.adj. 3 2 >

# Examples of patterns consisting of 3 MSD tags

En: *of the Inner Party*

En: *the solid breast*

(Tsms)? Ncms[or]y Afpms-n

**al Partidului Interior**

pieptul solid



Tsms Ncmsoy Afpms-n

Ncmsoy Afpms-n



a.adj.

a.adj.

< "(Tsms)? Ncms[or]y Afpms-n" a.adj. 1 2 >



# Evaluation

Accuracy			
	Precision	Recall	F-measure
Dependency relations			
ALL	0.63	0.74	0.68
a.adj.	0.54	0.86	0.66
a.subst.	0.67	0.72	0.69
det.	0.90	0.82	0.86
prep.	0.74	0.66	0.70
Unlabelled relations	0.72	0.78	0.75

# Thank you!

[mcolhon@inf.ucv.ro](mailto:mcolhon@inf.ucv.ro), [dcristea@info.uaic.ro](mailto:dcristea@info.uaic.ro)

