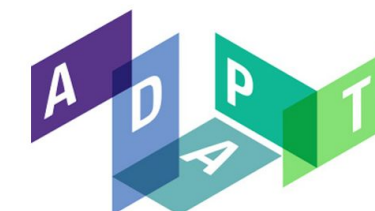




SCAN ME



Introduction

1. We present the first system for generating Wikipedia texts in Irish on demand (English available too).
2. Our system is based on rules, and models Irish grammar and Irish words.
3. Below we show step by step what happens when you request a text using the demo.
4. Use the demo yourself: scan the code on the top left corner!



Why not using ChatGPT or machine translation to produce Irish texts?

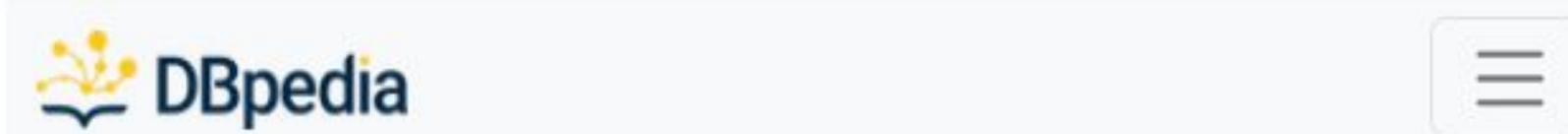
1. Large Language models used by ChatGPT or machine translation can produce very fluent text but the contents are **not always accurate**. Accuracy is crucial when generating factual knowledge.
2. Large Language models can **consume an incredible amount of resources** and **can be expensive** to use. Our system has a total disk space of ~10MB and runs with <1GB of RAM.
3. Large Language models are **black boxes**. We like languages, so we like understanding what is happening when a text is created. Rule-based systems allow for full control of the process.

1 - Get information about chosen entity on DBpedia

DBpedia is a structured repository that underlies the infoboxes in Wikipedia!

It allows us to access info needed for generation:

- **Properties**: birth date, birth place, etc.
- **Class** information: Person, Band, Location, etc.
- **Gender** information.
- **Irish** Named entity labels



About: Douglas Hyde

An Entity of Type: [animal](#), from Named Graph: <http://dbpedia.org>, within Data Space: [dbpedia.org](#)

Douglas Hyde (17 janvier 1860 - 12 juillet 1949) est un homme d'État, poète, écrivain et professeur irlandais, premier président d'Irlande du 25 juin 1938 au 24 juin 1945.



Property	Value
dbo:almaMater	<ul style="list-style-type: none">dbr:Trinity_College_Dublin
dbo:birthDate	<ul style="list-style-type: none">1860-01-17 (xsd:date)
dbo:birthPlace	<ul style="list-style-type: none">dbr:Castlereaudbr:County_Roscommon
dbo:deathCause	<ul style="list-style-type: none">dbr:Alzheimer's_diseasedbr:Pneumonia
dbo:deathDate	<ul style="list-style-type: none">1949-07-12 (xsd:date)
dbo:deathPlace	<ul style="list-style-type: none">dbr:Dublindbr:Phoenix_Parkdbr:Little_Ratna
dbo:nationality	<ul style="list-style-type: none">dbr:Irish_people
dbo:party	<ul style="list-style-type: none">dbr:Independent_politicians_in_Ireland
dbo:profession	<ul style="list-style-type: none">dbr:Politiciandbr:Linguisticsdbr:Academic

On the right side, we use the 4 properties in the red frames for illustration: almaMater, BirthDate, BirthPlace, profession.

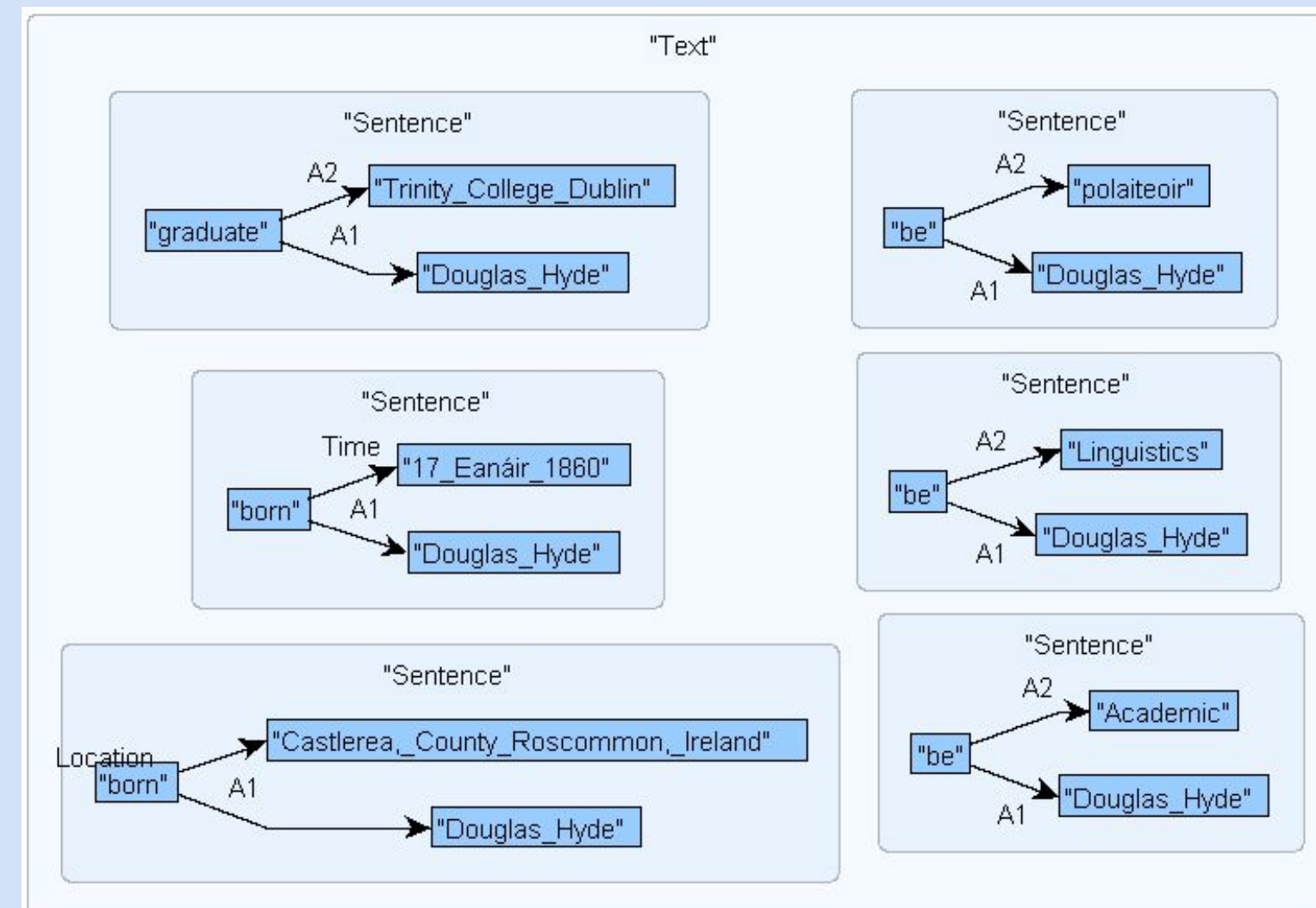
2 - Generate text for selected properties

We use the FORGe generator for the semantic and syntactic processing and modelling the lexicon.

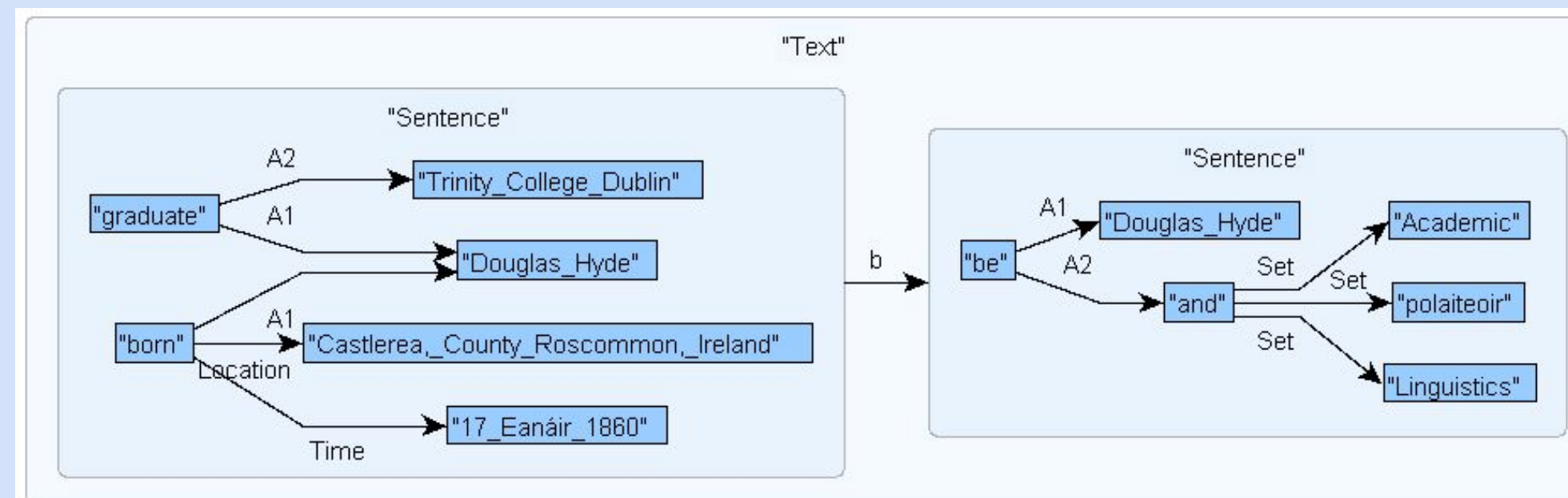
We use Irish NLP tools for the morphological processing.

2.1: Semantic processing

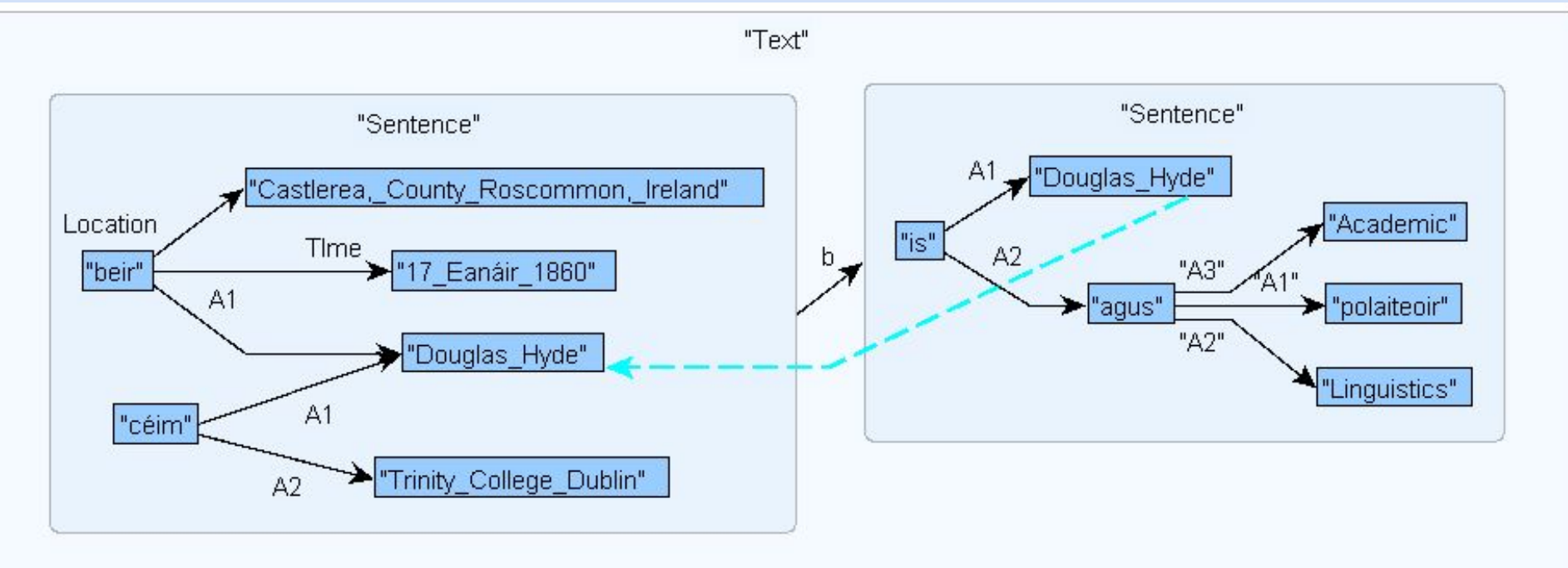
Map each property/value to an abstract linguistic structure.



Then package the properties into sentences.



Then map linguistic units to Irish content words.



Lexicon

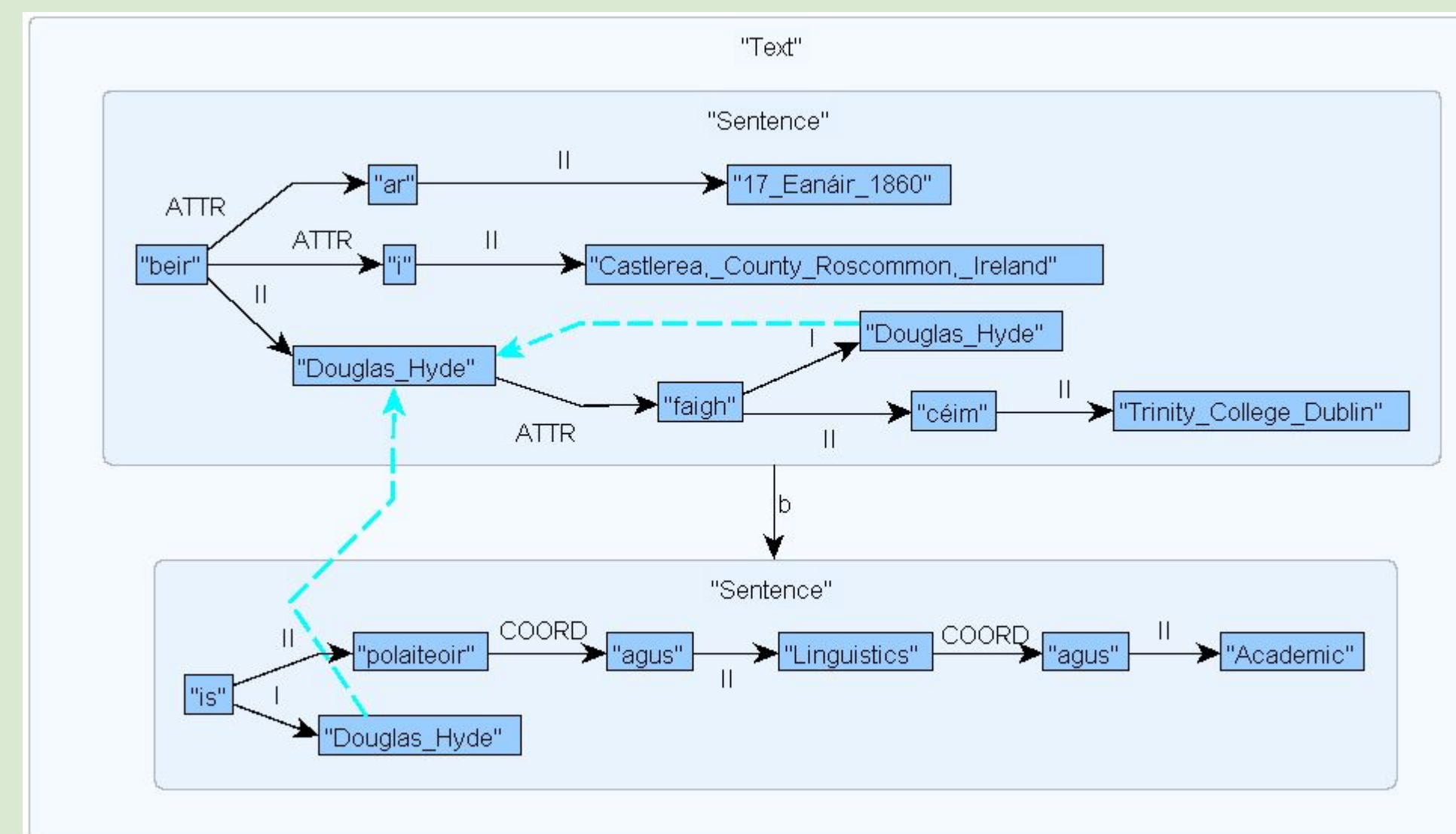
To go from semantic to syntax, we use lexical resources that describe the Irish words, in which we store:

- features such as part of speech and gender;
- if they have semantic participants, and how many;
- which other words they typically combine with.
- etc.

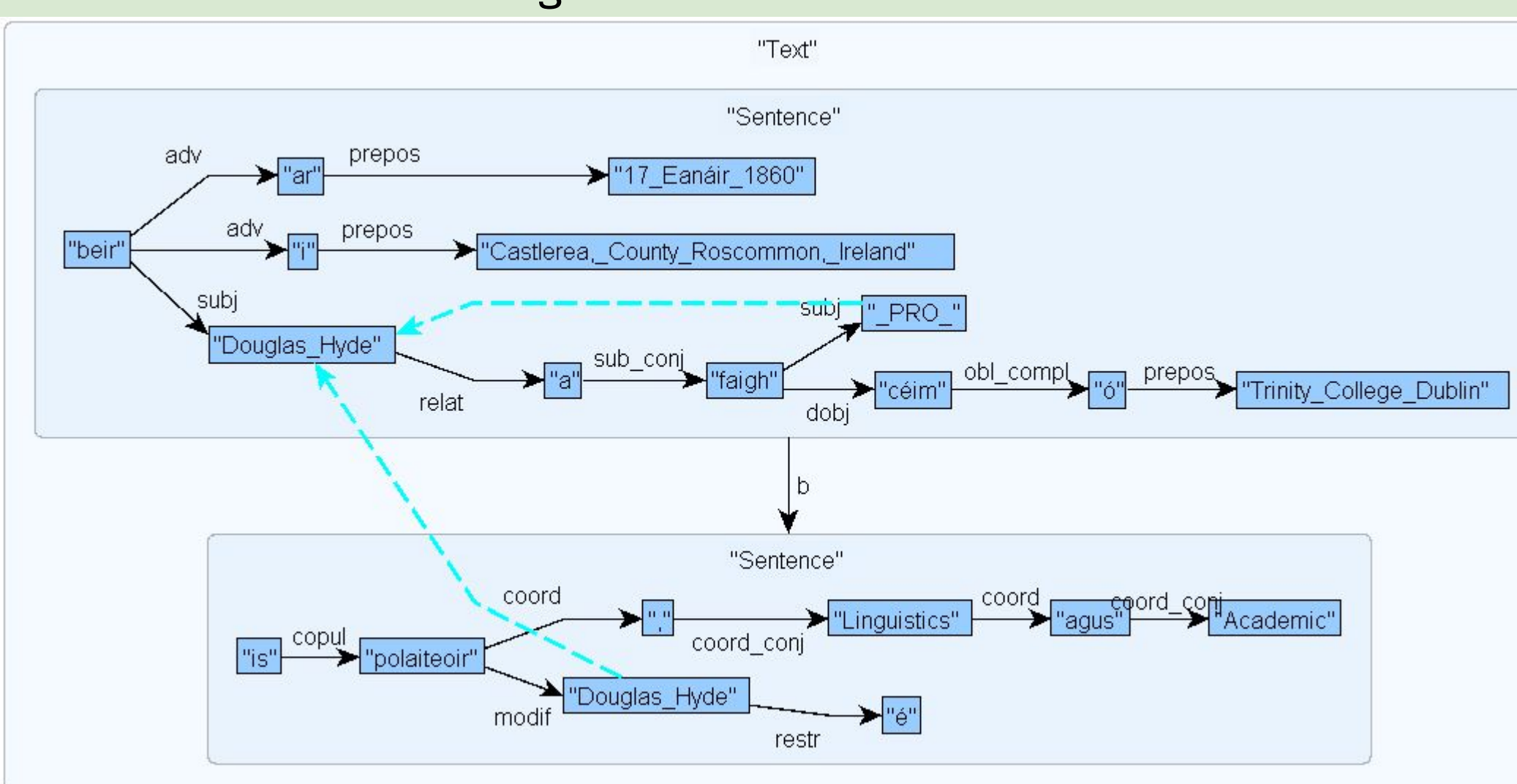
```
"céim_MN_01": noun, {  
  // degree (graduate)  
  project = "WebNLG"  
  lemma = "céim"  
  gender = "FEM"  
  Gper1 = "faigh_VB_01"  
  gp = {  
    A1 = I  
    A2 = II  
    I = {}  
    II = { prep = "ó" }  
  }  
}
```

2.2 Syntactic processing

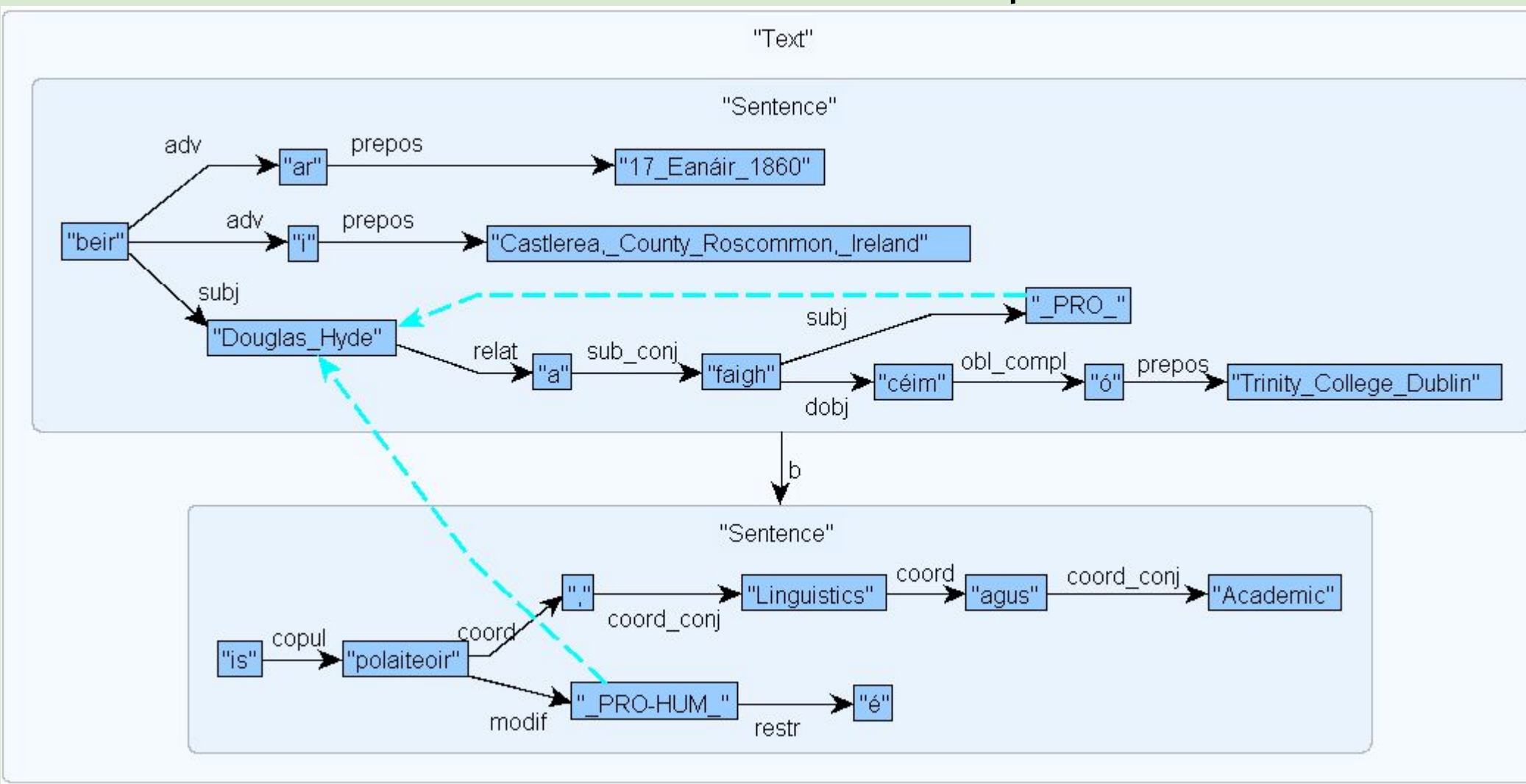
Define the structure of each sentence.



Then introduce grammatical words and relations.



Then introduce cross-sentence pronouns.



2.3 Morphological processing

Before starting, we define the order of the words. Then, for each word, we get the inflected form.

FORGe	Morphology	Post-processing
bci+Verb+PastInd+Auto Douglas_Hyde+Noun+Masc+Com+Sg	rugadh Douglas_Hyde	Rugadh Douglas Hyde
faigh+Verb+PastInd céim+Noun+Fem+Com+Sg	fuair céim	fuair céim
ó Trinity_College_Dublin	ó Trinity_College_Dublin	ó Trinity College Dublin
Castlereau_County_Roscommon_Ireland ar 17_Eanáir_1860	Castlereau_County_Roscommon_Ireland ar 17_Eanáir_1860	gCastlereau_County_Roscommon_Ireland ar 17_Eanáir_1860
is+Cop+Past polaitoir+Noun+Masc+Com+Sg	ba polaitoir	Ba polaitoir
é Linguistics+Noun+Masc+Com+Sg agus Academic+Noun+Masc+Com+Sg	é Linguistics agus Academic	é Linguistics agus Academic

The final text looks like the following:

Rugadh Douglas Hyde, a fuair céim ó Thrinity College Dublin, i gCastlereau, County Roscommon, Ireland ar 17 Eanáir 1860. Ba polaiteoir é, Linguistics agus Academic.

Or in English:

Douglas Hyde, who graduated from Trinity College Dublin, was born in Castlereau, County Roscommon, Ireland on January 17, 1860. He was a Politician, a Linguistics and a Academic.

Limitations - Future - Reference

1. We fully depend on the DBpedia quality:
 - Data errors! (e.g. *Linguistics* above).
2. It is an early version of our system:
 - We will improve the quality of texts.
 - We will increase the output variety.
 - We will cover more properties.
3. For more information about the system:
 - Simon Mille, Elaine Ui Dhoonchadha, Stamatia Dasiopoulou, Lauren Cassidy, Brian Davis, and Anya Belz. 2023. DCU/TCDFORGe at WebNLG'23: Irish rules! Technical report, ADAPT, Dublin City University. WebNLG 2023 System Description.