

# NounAtlas

Omar Bayoumi

Sapienza University of Rome

Elective in Artificial Intelligence - Narrative Understanding And Storytelling (NUANS)

Minihomework 1b

bayoumi.1747042@studenti.uniroma1.it

## 1 Introduction

Some verbs, linked to **VerbAtlas** frames, in **WordNet** have a derivationally related form to some nominal synsets. The goal is to create a **mapping from nominal synsets to VerbAtlas frames**.

## 2 First phase - Unambiguous nominal synsets

The **first phase** was: for each verb connected to a **VerbAtlas** frame, visit the derivation tree of the **hypernyms** of the nominal synsets of the derivationally related forms up to *Entity*. From this visit, if one of the *Event* synsets is present within the path then **the nominal synset is a nominal event synset** and thus is linked to the source **VerbAtlas** frame.

From this initial approach, I was able to isolate all the verbal synsets that have and do not have a derivationally related nominal form. Then, the nominal synsets were divided into two sets: those that reach *Event* and those that do not. For both sets, there was a division between **ambiguous** and **unambiguous** synsets [Figure 1].

A nominal synset is defined as ambiguous if it's derived from two different verbal synsets associated with **two** different **VerbAtlas** frames.

## 3 Second phase - Ambiguous nominal synsets

In this phase, after collecting the unambiguous synsets in a safe way, **some ambiguous synsets** were taken and tried to disambiguate them using a **statistical approach**.

### 3.1 Statistic manipulation

For each **nominal unambiguous synset**, all synsets in their paths to *Event* were collected and saved. So, for each **VerbAtlas** frame associated with the nominal unambiguous synset, the set of synsets present

in the paths was associated with the **frequency of how often they appeared**, e.g. the synset *Entity*, since it is reached by all, for each **VerbAtlas** frame will have the **maximum value**, while for the frame *SMELL*, the synset *sensory\_activity.n.01* appears three times only.

### 3.2 Disambiguate

The technique used to disambiguate is to use the statistics mentioned above. Specifically, if a synset in the path from ambiguous nominal synset to *Entity* is **associated with a particular frame within the statistics**, then most likely, **the reference frame is the same**. For this reason, synsets that appear in at least *one-third* of the frames are ignored since they would result in "*hint*" synsets that are too ambiguous. Since several synsets in the statistics might have **associated the same frame**, another step was necessary to resolve the ambiguities: if a synset in the statistics appears in **several frames**, only the frame that has the **highest frequency** on this synset will keep it. Instead, if a synset appears less than a **threshold value**, then it is most likely a synset that is **too unique to be reliable** and therefore is not considered [Figure 2].

## 4 Conclusion

In conclusion, I found **2515** unambiguous nominal synsets in the first phase. All the manipulations in the second phase, in the end, resulted, **after manually choosing the threshold at 20**, in the **disambiguation of 148 nominal synsets out of 965**. Since this is a statistical method, and by reading the definitions of the nominal synsets they do indeed seem to be rightly associated with the different frames, the results are not *100%* reliable and perhaps a **neural network approach**, trained on the definitions, would lead to better results.

VERBAL SYNSETS	13767
Without related forms:	4833
With related forms:	8934
NOMINAL SYNSETS	11751
Event synsets:	3480
Ambiguous:	965
Non ambiguous:	2515
Not event synsets:	8271
Ambiguous:	1860
Non ambiguous:	6411

Figure 1: Number of synsets analyzed and categorized.

## SYNSET DISAMBIGUATION

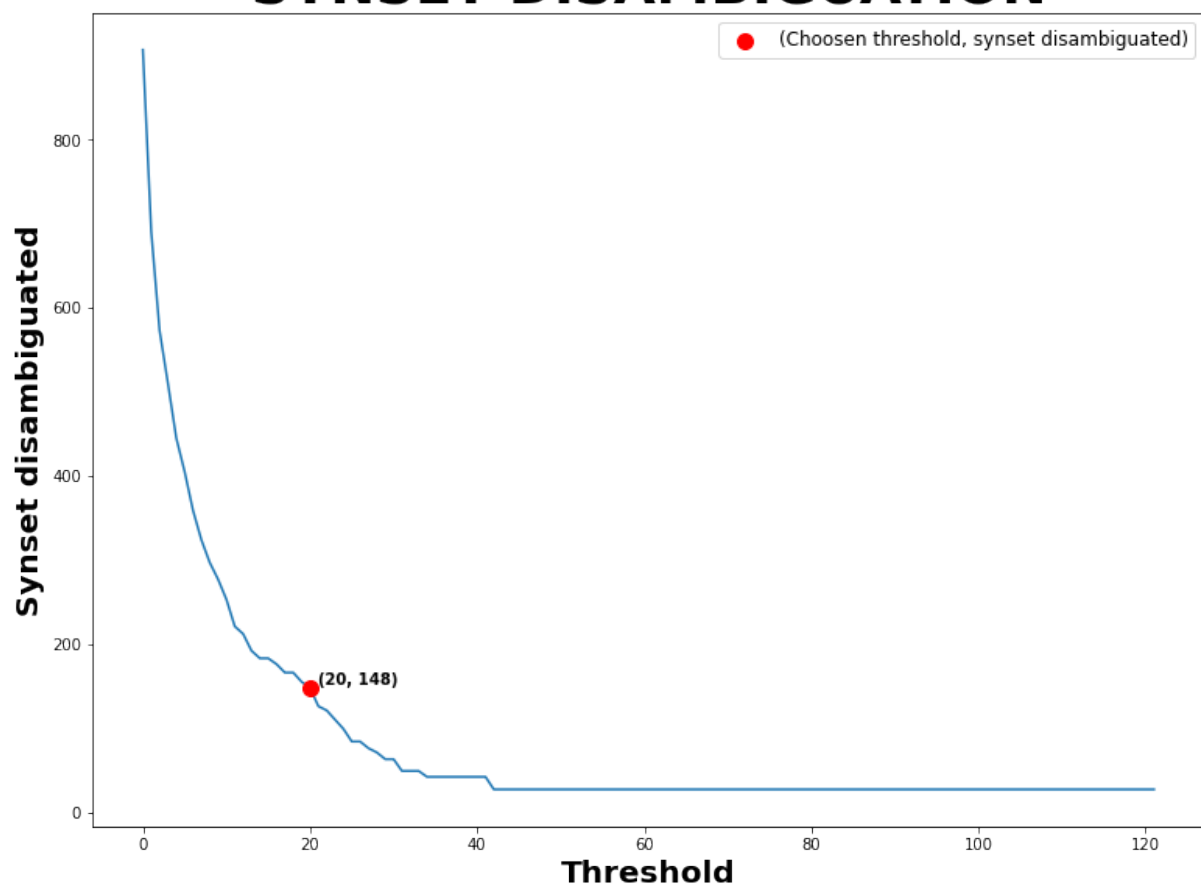


Figure 2: Number of disambiguated synsets for each threshold level. The selected threshold is shown in red.