Programmable Semantic Shift Analysis Using ChainNet and CoreLex

Student Name: Barbora Matlachová, Dominik Fránek
Username: bmahdalova, dfranek161
Email: mahdba00@upol.cz, frando02@upol.cz

January 2026

1 Introduction

The original objective of this assignment was to investigate lexical–conceptual mappings and morphological patterns, with particular emphasis on phenomena such as zero-derivation and semantic links between categories like PERSON and ANIMAL. While the assignment suggested approaches such as cross-lingual comparison or manual inspection of Wiktionary definitions, this project adopts a computational perspective focused on semantic shifts in English as a foundational step.

Specifically, we aim to automate the detection of semantic category shifts (e.g., ANIMAL → HUMAN) by integrating two structured lexical resources: ChainNet and CoreLex. The motivation behind this approach is to replace manual, example-driven analysis with a programmable and scalable pipeline capable of systematically identifying metaphorical mappings across the lexicon.

By linking ChainNet—which encodes metaphorical and metonymic relations between WordNet senses—with CoreLex—which assigns WordNet noun senses to a small set of coarse-grained semantic types—we seek to automatically characterize both *that* a semantic shift occurs and *what kind* of shift it represents.

2 Background

This project builds upon the methodology introduced in *Metonymy is more multilingual than metaphor: Analysing tropes using ChainNet* (Bond & Maudslay, GWC 2025), which

demonstrates how tropes such as metaphor and metonymy can be studied computationally using structured lexical data.

Three primary resources were used:

- **OMW English 1.4 (ChainNet):** A version of WordNet extended with approximately 12,000 links representing tropes, including metaphor and metonymy. These links encode directional sense-to-sense mappings, allowing us to detect when a semantic shift has occurred.
- **CoreLex:** A semantic ontology that reorganizes WordNet noun synsets into 39 coarse-grained semantic types (e.g., ANIMAL, HUMAN, ARTIFACT). This abstraction enables systematic comparison of source and target domains in metaphorical mappings.
- **Python `wn` Library:** A library for parsing and querying WordNet-formatted XML files, used to ingest and manipulate the ChainNet-enhanced OMW dataset.

3 Approach

The project evolved from an initial exploratory phase based on raw text and JSON files into a more robust pipeline using the `wn` library and structured XML resources.

3.1 Problem Definition

The assignment required examining senses with different hypernyms—specifically PERSON and ANIMAL—while making the process programmable. To achieve this, we needed to map the sense identifiers used in ChainNet to the semantic categories defined in CoreLex.

A major challenge was interoperability between resources. ChainNet uses custom sense identifiers (e.g., `omw-en-chestnut-07772274-n`), while CoreLex relies on standard WordNet synset names (e.g., `chestnut.n.03`). Early attempts to rely on raw ChainNet JSON files failed due to the absence of synset offsets, making reliable alignment impossible.

3.2 Solution Pipeline

After locating and ingesting the correct `omw-en_1.4_cn.xml` file, we implemented a Python script (`analyze_shifts_final.py`) that bridges the datasets through the following steps:

1. **Resource Ingestion:** The ChainNet XML file was installed into the local `wn` database using `wn.add()`.

2. **Relation Extraction:** For a target list of polysemous words, the script queried all senses containing metaphorical relations.

3. **ID Normalization and Resource Alignment:** The 8-digit synset offset was extracted from the ChainNet sense ID using regular expressions and passed to NLTK to retrieve the corresponding standard WordNet synset name.

4. **Semantic Type Lookup:** The standardized synset name was queried in the CoreLex mapping to retrieve the source and target semantic types.

4 Results

The implemented system successfully identified and categorized lexical–conceptual mappings involving semantic shifts. A representative sample of the results is shown in Table 1.

Table 1: Sample Lexical–Conceptual Mappings

Word	**Source Type**	**Target Type**	**Shift Description**
pig	ANIMAL	HUMAN	Zoomorphic metaphor for behavior
head	BODY_PART	LEADER	Physical top → social authority
star	CELESTIAL_BODY	HUMAN	Brightness → fame
mouth	BODY_PART	ARTIFACT	Shape-based aperture mapping
chestnut	PLANT	BODY_PART	Visual similarity (horse anatomy)

These results demonstrate that the pipeline can automatically identify PERSON/ANIMAL and related semantic mappings, directly fulfilling the assignment's requirement for a programmable solution.

5 Discussion

5.1 Analysis of Semantic Shifts

The observed mappings support the *directionality principle of metaphor*, whereby meaning tends to shift from concrete domains (e.g., ANIMAL, BODY_PART) toward more abstract or social domains (e.g., HUMAN, LEADERSHIP).

No reverse mappings (e.g., HUMAN → ANIMAL) were observed among the prototype senses examined, suggesting that these shifts are systematic rather than random.

5.2 Technical Challenges

The primary technical difficulty involved handling incompatible resource formats and missing identifiers. The key breakthrough was recognizing that the `wn` library could ingest the ChainNet XML directly, combined with a custom normalization step to align internal identifiers with CoreLex-compatible synset names.

6 Future Work

With a working pipeline that identifies shifts such as ARTIFACT → ACT, future work could integrate a morphological database to examine whether these semantic shifts correlate with zero-derivation in English. This would also enable cross-lingual comparison, contrasting English zero-derivation with overt morphological marking in other languages.

7 Conclusion

This project demonstrates that lexical–conceptual mappings traditionally identified through manual inspection can be detected automatically using structured lexical resources. By integrating ChainNet with the CoreLex ontology, we developed a programmable pipeline that systematically identifies and categorizes semantic shifts in English metaphors.

Regarding contributions, the code represents collaborative work, resulting from extensive brainstorming and problem-solving.The collaborative work included understanding the assignment objectives, analyzing and reconciling the ChainNet and CoreLex datasets, designing and implementing the semantic-shift detection pipeline, and verifying the results to ensure accuracy and completeness. The original report was written by Dominik Fránek, and it was subsequently refined to its final form by Barbora Mahdalová.

While the computational analysis and report content were developed independently by the authors, minor assistance from AI tools was used to improve language clarity and formatting.

Code Availability

The Python script `analyze_shifts_final.py` and all associated data files have been submitted to the project repository.