TOWARDS A DUTCH FRAMENET-STYLE SEMANTIC ROLE LABELER

Master's thesis, Chantal van Son (ReMa Linguistics)

Disambiguating entities and their roles in texts based on background knowledge

Friday 12 December 2014

INTRODUCTION

1. John broke the window.
2. The window was broken by John.
3. The window broke.

INTRODUCTION

1. John broke the window. [SUBJECT]
2. The window was broken by John. [PREPOSITIONAL PHRASE]
3. The window broke.

INTRODUCTION

1. John broke the window. [DIRECT OBJECT]
2. The window was broken by John. [SUBJECT]
3. The window broke. [SUBJECT]

INTRODUCTION

1. John broke the window. [DIRECT OBJECT]
2. The window was broken by John. [SUBJECT]
3. The window broke. [SUBJECT]

Syntactic parsing does not represent full meaning of sentence:

John = AGENT, "the breaker"

the window = PATIENT, "the thing being broken"

INTRODUCTION

- To summarize: a semantic role is the underlying conceptual relation that a syntactic constituent has with its predicate
- Semantic Role Labeling (SRL): automatic detection of semantic roles
  - Key task for answering "Who did What to Whom", "Where", "When", etc. in IE, QA, MT, Summarization, etc.

INTRODUCTION

- Statistical approaches for SRL driven by resources such as PropBank and FrameNet
- Different sets of roles:
  - PropBank: [John $_{ARG0}$] stole [the watch $_{ARG1}$].
  - FrameNet: [John $_{THEFT\#PERPETRATOR}$] stole [the watch $_{THEFT\#GOODS}$].

OVERVIEW

- Introduction to SRL
- Research question
- Methodology
  - PropBank vs. FrameNet
  - NewsReader pipeline
  - Predicate Matrix
- Problems/challenges to solve
- ~~Results~~

RESEARCH QUESTION

- Goal: FrameNet-style SRL for Dutch
  - [John $_{\text{THEFT\#PERPETRATOR}}$] stole [the watch $_{\text{THEFT\#GOODS}}$].
- Problem: Currently no FrameNet-like resource available for Dutch

RESEARCH QUESTION

- Goal: FrameNet-style SRL for Dutch
  - [John $_{\text{THEFT\#PERPETRATOR}}$] stole [the watch $_{\text{THEFT\#GOODS}}$].
- Problem: Currently no FrameNet-like resource available for Dutch

Research question:
How can we perform FrameNet-style SRL in Dutch without a resource that is annotated with this specific information?

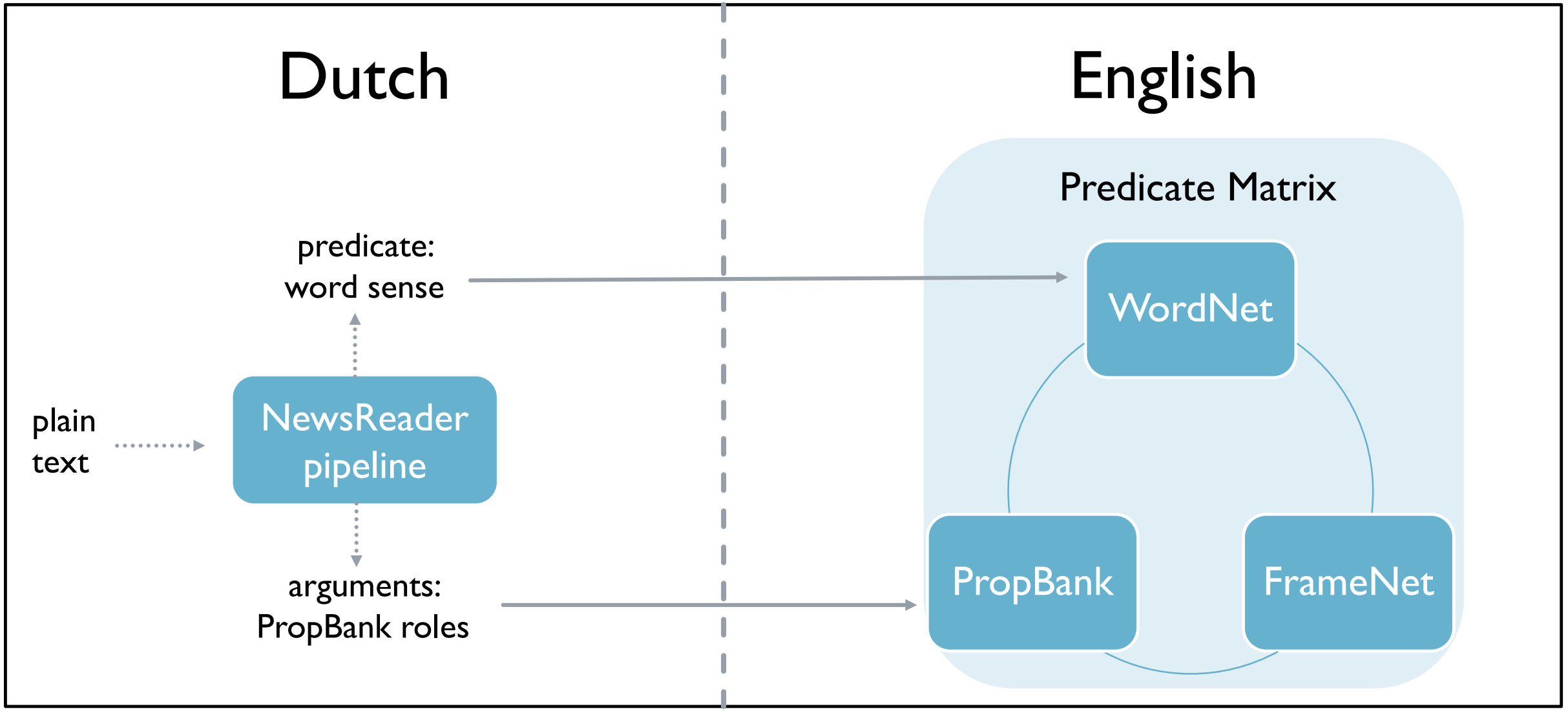
METHODOLOGY

- Use available tools and resources:
  - Sonar Semantic Role Labeler (SSRL): PropBank-style
  - Predicate Matrix
    - Integration of PropBank, FrameNet and WordNet
  - Mappings between Dutch/English WordNets

METHODOLOGY

- Use available tools and resources:
  - Sonar Semantic Role Labeler (SSRL): PropBank-style
  - Predicate Matrix
    - Integration of PropBank, FrameNet and WordNet
  - Mappings between Dutch/English WordNets

→ Map FrameNet roles on top of PropBank roles generated by SSRL by using the mappings between resources and between languages


Dutch
English
Predicate Matrix
predicate:
word sense
WordNet
plain
text
NewsReader
pipeline
PropBank
FrameNet
arguments:
PropBank roles

FRAMENET

- Based on Frame Semantics (Fillmore)
  - Word meanings are organized around schematic conceptual scenarios (frames), which represent prototypical situations or states including specific kinds of entities that can participate
- Primarily a lexicographical project:
  - Group words into frames and back up frame-semantic descriptions with annotated examples

Commerce_buy

Definition

These are words describing a basic commercial transaction involving a **BUYER** and a **SELLER** exchanging **MONEY** and **GOODS**, taking the perspective of the **BUYER** .The words vary individually in the patterns of frame element realization they allow. For example, the typical pattern for the verb BUY: **BUYER** buys **GOODS** from **SELLER** for **MONEY** .

Commerce_buy

Definition

These are words describing a basic commercial transaction involving a **BUYER** and a **SELLER** exchanging **MONEY** and **GOODS**, taking the perspective of the **BUYER** .The words vary individually in the patterns of frame element realization they allow. For example, the typical pattern for the verb BUY: **BUYER** buys **GOODS** from **SELLER** for **MONEY** .

Frame Elements

BUYER	The **BUYER** wants the **GOODS** and offers **MONEY** to a **SELLER** in exchange for them.
GOODS	The FE **GOODS** is anything (including labor or time, for example) which is exchanged for **MONEY** in a transaction.
SELLER	The **SELLER** has possession of the **GOODS** and exchanges them for **MONEY** from a **BUYER**.
MONEY (non-core)	**MONEY** is the thing given in exchange for **GOODS** in a transaction.
TIME (non-core)	When the event occurs.

Commerce_buy

Definition

These are words describing a basic commercial transaction involving a **BUYER** and a **SELLER** exchanging **MONEY** and **GOODS**, taking the perspective of the **BUYER** .The words vary individually in the patterns of frame element realization they allow. For example, the typical pattern for the verb BUY: **BUYER** buys **GOODS** from **SELLER** for **MONEY** .

Frame Elements

BUYER	The **BUYER** wants the **GOODS** and offers **MONEY** to a **SELLER** in exchange for them.
GOODS	The FE **GOODS** is anything (including labor or time, for example) which is exchanged for **MONEY** in a transaction.
SELLER	The **SELLER** has possession of the **GOODS** and exchanges them for **MONEY** from a **BUYER**.
MONEY (non-core)	**MONEY** is the thing given in exchange for **GOODS** in a transaction.
TIME (non-core)	When the event occurs.

Lexical Units

buy.v, buyer.n, purchase.v, purchase_((act)).n, purchaser.n

Commerce_buy

Definition

These are words describing a basic commercial transaction involving a **BUYER** and a **SELLER** exchanging **MONEY** and **GOODS**, taking the perspective of the **BUYER** .The words vary individually in the patterns of frame element realization they allow. For example, the typical pattern for the verb BUY: **BUYER** buys **GOODS** from **SELLER** for **MONEY** .

Frame Elements

BUYER	The **BUYER** wants the **GOODS** and offers **MONEY** to a **SELLER** in exchange for them.
GOODS	The FE **GOODS** is anything (including labor or time, for example) which is exchanged for **MONEY** in a transaction.
SELLER	The **SELLER** has possession of the **GOODS** and exchanges them for **MONEY** from a **BUYER**.
MONEY (non-core)	**MONEY** is the thing given in exchange for **GOODS** in a transaction.
TIME (non-core)	When the event occurs.

Lexical Units

buy.v, buyer.n, purchase.v, purchase_((act)).n, purchaser.n

Annotated example

Yesterday, Abby bought a car from Robin for $5,000.

PROPBANK

- Semantic role annotation of the Penn Treebank corpus
- More practical aim: provide training data for supervised SRL
- Verb-specific and theory-neutral approach
  - Arg0 t/m Arg5
  - Modifiers, e.g. ArgM-TIME, ArgM-LOC
- Lexicon to facilitate annotation; later evolved into a resource on its own

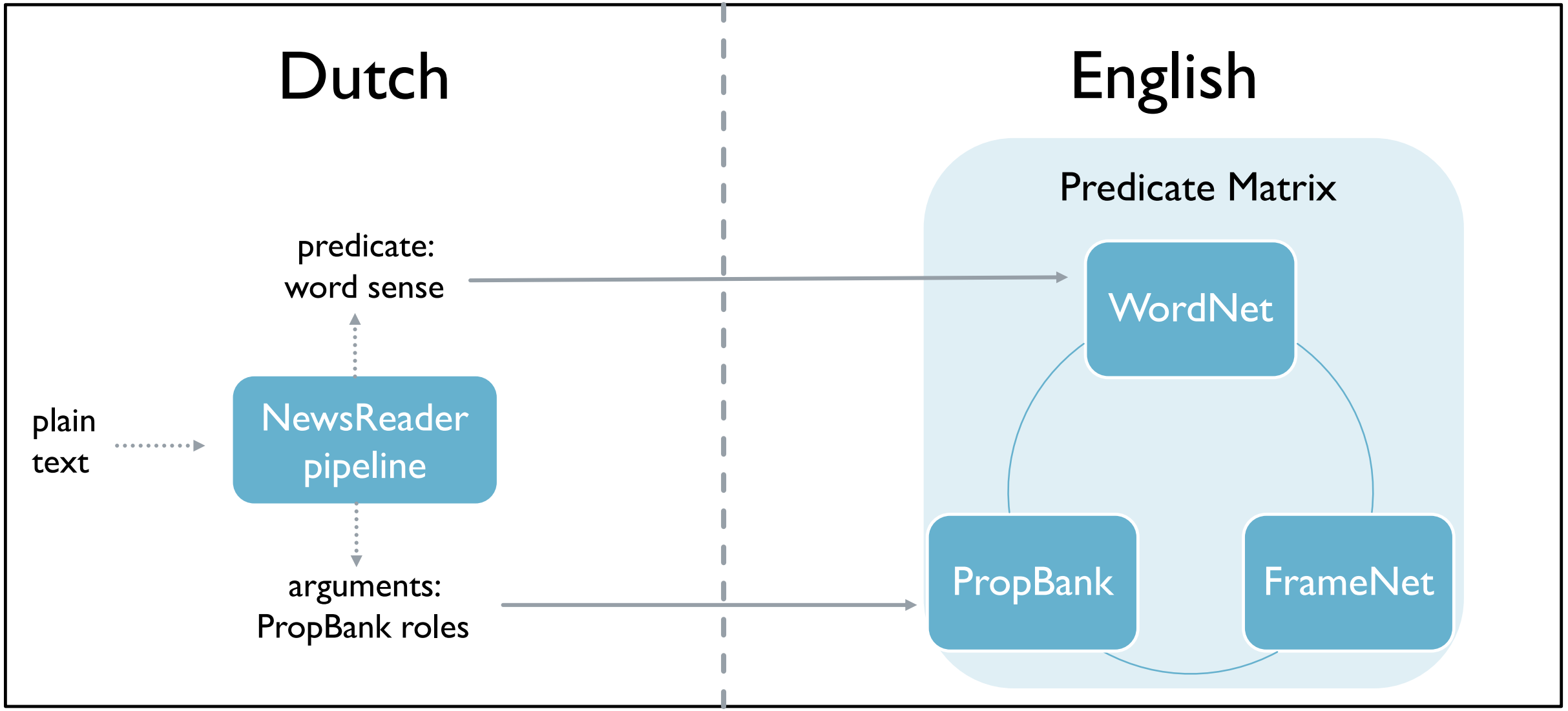
Roleset ID: purchase.01

Roles

Arg0-PAG: *purchaser*
Arg1-PPT: *thing purchased*
Arg2-DIR: *seller*
Arg3-VSP: *price paid*
Arg4-GOL: *benefactive*

Example: With benefactive:
John purchased a dozen roses for his mother.

Arg0: John
Rel: purchased
Arg1: a dozen roses
Arg4: for his mother


Dutch
English
Predicate Matrix
predicate:
word sense
WordNet
plain
text
NewsReader
pipeline
PropBank
FrameNet
arguments:
PropBank roles

NEWSREADER PIPELINE

1) Tokenizer
2) Morphosyntactic parser (Alpino)
3) Word Sense Disambiguation
4) Ontotagger (Predicate Matrix)
5) Semantic Role Labeler (SSRL)

NEWSREADER PIPELINE

1) Tokenizer
2) Morphosyntactic parser (Alpino)
3) Word Sense Disambiguation
4) Ontotagger (Predicate Matrix)
5) Semantic Role Labeler (SSRL)

NEWSREADER PIPELINE

1) Tokenizer
2) Morphosyntactic parser (Alpino)
3) Word Sense Disambiguation
4) Ontotagger (Predicate Matrix)
5) Semantic Role Labeler (SSRL)

Volgens Apple is dat **vissen** naar wachtwoorden een gebruikelijke praktijk.

```
<!--vissen-->
<term id="t_91" type="open" lemma="vissen" pos="verb" morphofeat="WW(inf,nom,zonder,zonder-n)">
  <span>
    <target id="w92"/>
  </span>
  <externalReferences>
    <externalRef resource="Cornetto" reference="d_v-317255" confidence="0.17348981"/>
    <externalRef resource="Cornetto" reference="d_v-317258" confidence="0.09935643"/>
    <externalRef resource="Cornetto" reference="r_v-9597" confidence="0.04414937"/>
    <externalRef resource="Cornetto" reference="r_v-9598" confidence="0.17084068"/>
  </externalReferences>
</term>
```

NEWSREADER PIPELINE

1) Tokenizer
2) Morphosyntactic parser (Alpino)
3) Word Sense Disambiguation
4) Ontotagger (Predicate Matrix)
5) Semantic Role Labeler (SSRL)

PREDICATE MATRIX

- Predicate information of multiple resources integrated
- Each row of the Predicate Matrix represents the mapping of a role over the different resources and includes all the aligned knowledge about its corresponding verb

odwn-eq_synonym:d_v-3757-v odwn-synset:kopen;aanschaffen; vn:get-13.5.1 vn:13.5.1 vn:buy wn:buy%2:40:00 mcr:ili-30-02207206-v fn:Commerce_buy fn:buy.v pb:buy.01 mcr:0 mcr:economy mcr:Buying mcr:Dynamic;Possession; mcr:possession mcr:get%2:40:00 wn:102 wn:024 SEMLINK fn-pb-role:Buyer#0 SEMLINK;FN_FE fn-pb-role:Means#3 fn-pb-role:Duration#4 fn-pb-role:Means#2 fn-pb-role:Goods#1 vn:obtain-13.5.2 vn:13.5.2 vn:obtain-13.5.2-1 vn:13.5.2-1 vn:purchase wn:purchase%2:40:00 fn:purchase.v pb:purchase.01 wn:34 SEMLINK;PREDICATE_MAPPING SEMLINK;PREDICATE_MAPPING;FN_FE fn-pb-role:Buyer#3 fn-pb-role:Money#3

```
<term id="t1.30" lemma="voegen" pos="V.verb" type="open">
  <span>
    <!--voegen-->
    <target id="w1.30"/>
  </span>
  <externalReferences>
    <externalRef confidence="0.388571" reference="nld-21-d_v-7822-v" resource="cdb2.0-nld-all.infv.0.0.no-allwords">
      <externalRef resource="predicate-matrix1.1">
        <externalRef reference="fn:Rope_manipulation" resource="fn"/>
        <externalRef reference="pb:tie.01" resource="pb"/>
        <externalRef reference="fn-pb-role:Agent#0" resource="fn-pb-role"/>
        <externalRef reference="fn-pb-role:Means#2" resource="fn-pb-role"/>
        <externalRef reference="fn-pb-role:Instrument#3" resource="fn-pb-role"/>
        <externalRef reference="fn-pb-role:Agent#1" resource="fn-pb-role"/>
        <externalRef reference="fn-role:Location" resource="fn-role"/>
      </externalRef>
      <externalRef resource="predicate-matrix1.1">
        <externalRef reference="fn:Attaching" resource="fn"/>
        <externalRef reference="pb:connect.01" resource="pb"/>
        <externalRef reference="fn-pb-role:Agent#0" resource="fn-pb-role"/>
        <externalRef reference="fn-pb-role:Goal#2" resource="fn-pb-role"/>
        <externalRef reference="fn-pb-role:Item#1" resource="fn-pb-role"/>
        <externalRef reference="pb:link.01" resource="pb"/>
        <externalRef reference="pb:tie.01" resource="pb"/>
        <externalRef reference="fn-pb-role:Instrument#3" resource="fn-pb-role"/>
        <externalRef reference="fn-role:Goal" resource="fn-role"/>
        <externalRef reference="fn:tie.v" resource="fn"/>
        <externalRef reference="fn-pb-role:Connector#1" resource="fn-pb-role"/>
      </externalRef>
```

NEWSREADER PIPELINE

1) Tokenizer
2) Morphosyntactic parser (Alpino)
3) Word Sense Disambiguation
4) Ontotagger (Predicate Matrix)
5) Semantic Role Labeler (SSRL)

```
<predicate id="pr24">
  <span>
    <target id="t_182"/>
  </span>
  <role id="r44" semRole="ArgM-NEG">
    <span>
      <target id="t_178" head="yes"/>
    </span>
  </role>
  <role id="r50" semRole="Arg0">
    <span>
      <target id="t_175" head="yes"/>
      <target id="t_176"/>
      <target id="t_177"/>
      <target id="t_178"/>
    </span>
  </role>
  <role id="r52" semRole="Arg1">
    <span>
      <target id="t_176"/>
      <target id="t_177" head="yes"/>
    </span>
  </role>
</predicate>
```

CURRENT BASELINE

- Looks for predicate in SRL layer
- Finds WSD and PM data for this predicate in term layer
- Assigns all frames of all word senses to the predicate
- Assigns all FrameNet roles to the arguments if they correspond to PropBank roles

Fine-tuning: filtering frames and roles (e.g. by WSD-score, frequency of frame)

FUTURE WORK: CHALLENGES

- Quality of Predicate Matrix: wrong or missing information
  - Association between PropBank and FrameNet roles not given
  - Frames and/or roles not given at all
- Coverage of FrameNet
- Differences PropBank/FrameNet
  - Metaphorical usages and support constructions
- From Dutch to English
  - Quality of WordNet mappings
  - English Frames

THANK YOU FOR YOUR ATTENTION.

QUESTIONS?