

# BabelNet: A Wide-coverage Multilingual Semantic Network

Presented by: Yizhong Wang  
Presentation date: 2015.6.3

# Article Info

- Title:** BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network
- Authors:** Roberto Navigli\*, Simone Paolo Ponze
- Journal:** Artificial Intelligence
- Date:** 1 July 2012



# Outline

## ○ Introduction

- Objectives
- WordNet
- Wikipedia

## ○ Construction details

- Mapping algorithm
- Disambiguation context
- Translating to multilingual

## ○ Evaluation & System

- Mapping evaluation
- Translation evaluation
- Online system

# Outline

## ○ **Introduction**

- [Objectives](#)
- [WordNet](#)
- [Wikipedia](#)

## ○ **Construction details**

- Mapping algorithm
- Disambiguation context
- Translating to multilingual

## ○ **Evaluation & System**

- Mapping evaluation
- Translation evaluation
- Online system

# Glossary

Term	Definition	Example
sense	The exact meaning of an expression	The word “play” has many senses.
synonym	words or phrases that mean exactly or nearly the same	“play” & “drama”
hypernym	A is a generalization of B	“Color” is a hypernym of “red”.
hyponym	A is a specification of B	“Soccer” is a hyponym of “sport”.
gloss	explanation	Play: “A dramatic work intended for performance by actors on a stage”

# Objectives

# Natural language is highly ambiguous

Yesterday, I saw an underground rock.



# We need knowledge of an expression's semantics (Lexical Knowledge)

1. unstructured terminologies and glossaries
2. machine-readable dictionaries
3. computational lexicons and ontologies,  
such as WordNet

**Monolingual**

**English**

**Poor coverage**

# The Multilingual, Big-Picture Goal

“Underground  
rock concert”

“언더그라운드 락  
콘서트”

“Underground rock  
formation”

“지하 암석”

[semantic representation]



[semantic representation]



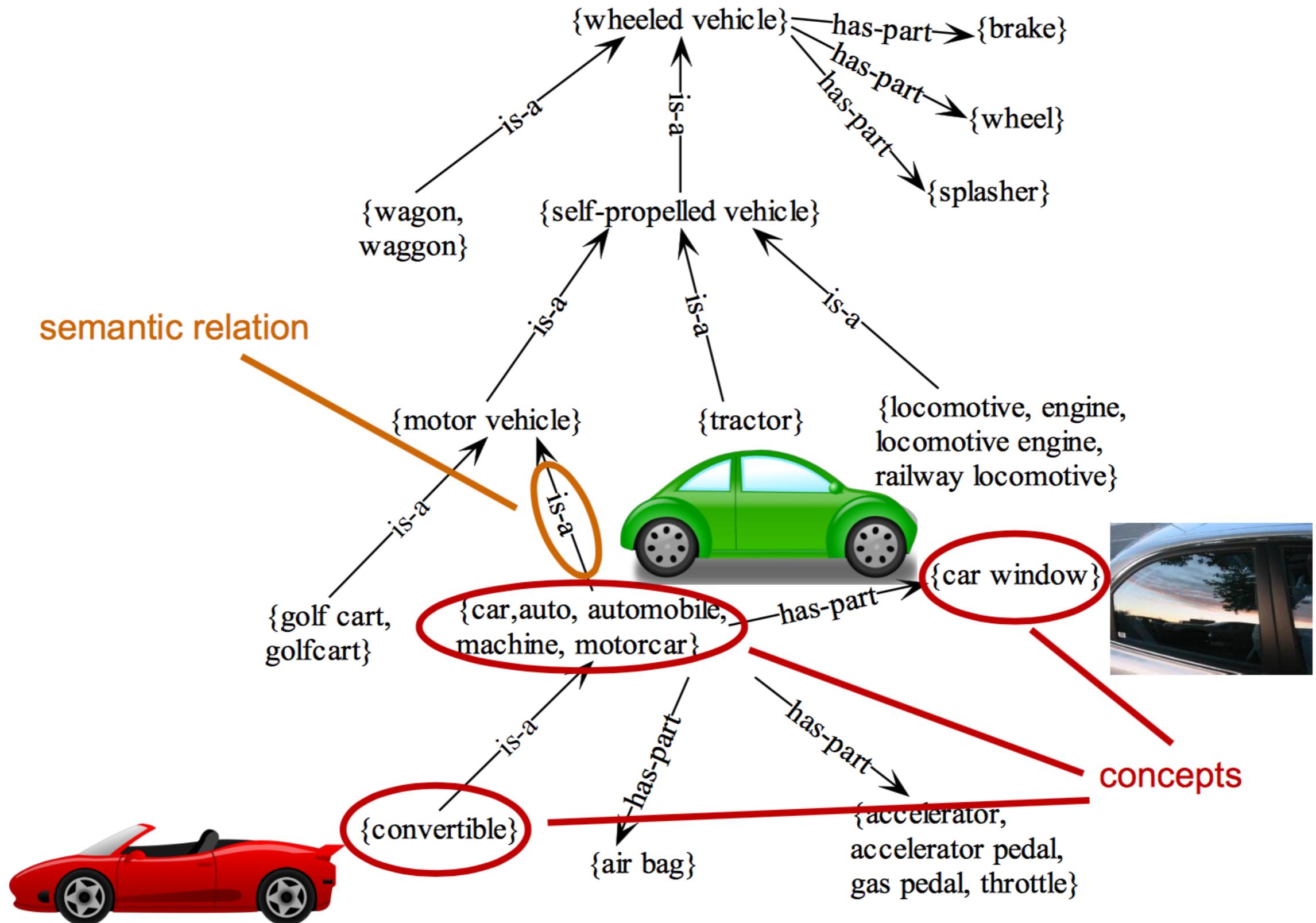
NLP  
Application

# WordNet

# WordNet Elements

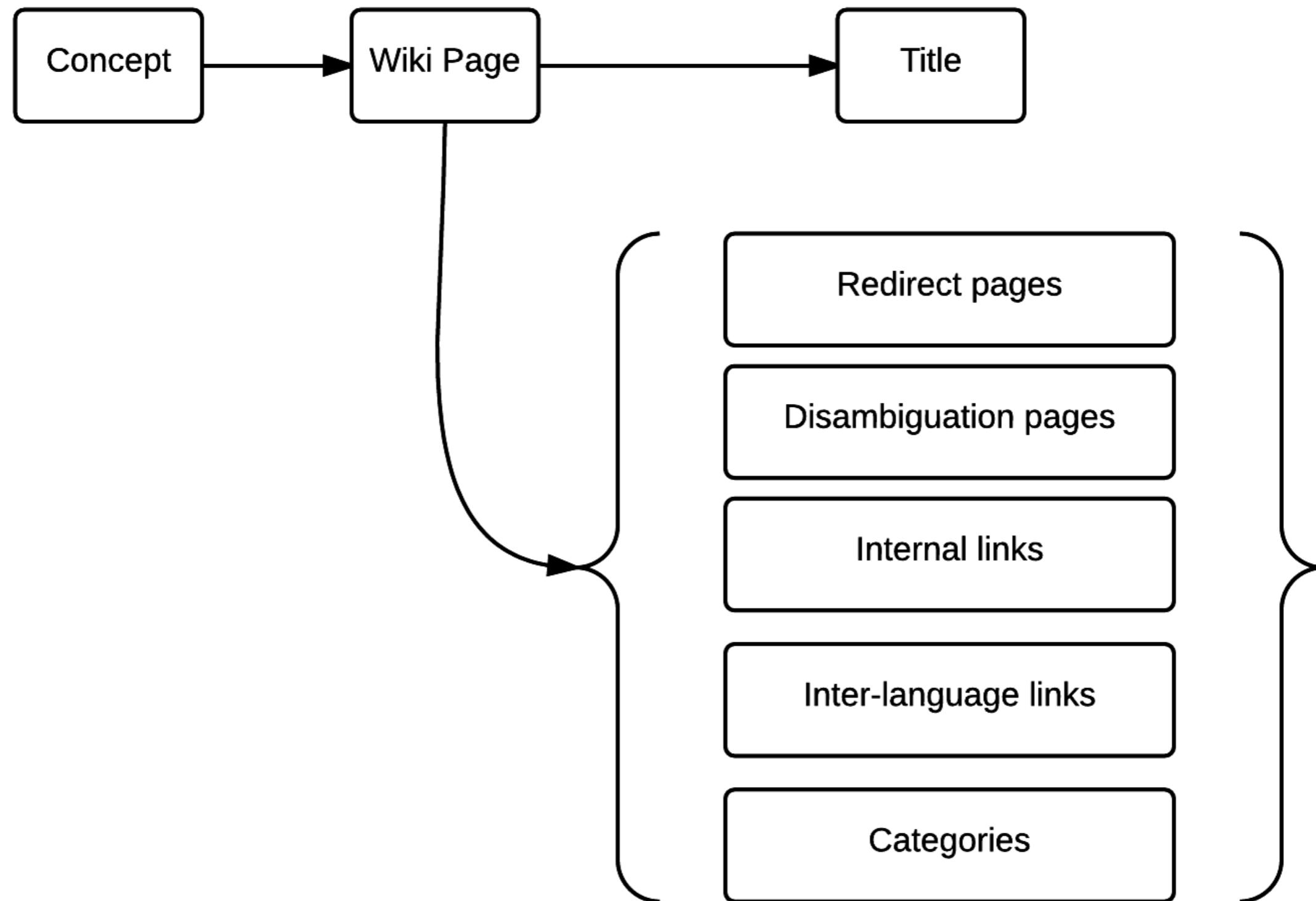
- Each concept is encoded as a **synset** (synonym set), which is a collection of synonymous senses
  - $\text{Senses}(\text{play}) = \{\{\text{play}_n^1, \text{drama}_n^1, \text{dramatic play}_n^1\}, \{\text{play}_n^8, \text{child's play}_n^2\}\}$
- **Semantic relations** between synsets
  - Hypernymy ( $\text{car}_n^1$  **is-a**  $\text{motor vehicle}_n^1$ )
  - Meronymy ( $\text{car}_n^1$  **has-a**  $\text{car door}_n^1$ )
  - Entailment, similarity, attribute, etc.
- **Lexical relations** between word senses
  - Antonymy ( $\text{good}_a^1$  antonym of  $\text{bad}_a^1$ )
  - Pertainymy ( $\text{dental}_a^1$  pertains to  $\text{tooth}_n^1$ )
  - Nominalization ( $\text{service}_n^2$  nominalizes  $\text{serve}_v^4$ )

# WordNet Example



# Wikipedia

# Wikipedia Structure



# Title

- The lemma of the concept defined
- An optimal label in parentheses which specifies its meaning if the lemma is ambiguous

Article [Talk](#) Read Edit View history Search

## Play (theatre)

From Wikipedia, the free encyclopedia

A **play** is a form of [literature](#) written by a [playwright](#), usually consisting of [dialogue](#) between [characters](#), intended for theatrical performance rather than just [reading](#). Plays are performed at a variety of levels, from [Broadway](#), [Off-Broadway](#), regional theater, to [Community theatre](#), as well as University or school productions. There are rare dramatists, notably [George Bernard Shaw](#), who have had little preference whether their plays were performed or read. The term "play" can refer to both the written works of playwrights and to their complete theatrical performance.<sup>[1]</sup>

**Contents** [hide]

1 Genres  
  1.1 Comedy



# Redirect pages

- These pages are used to forward to the Wikipage containing the actual information about a concept of interest.
- This models synonymy.

[View](#) (previous 50 | next 50) ([20](#) | [50](#) | [100](#) | [250](#) | [500](#))

- [Theater/play \(redirect page\)](#) ([links](#))
- **[Stage play \(redirect page\)](#) ([links](#))**
- [Stage plays \(redirect page\)](#) ([links](#))
- [Script \(performing arts\) \(redirect page\)](#) ([links](#))
- [Theatre play \(redirect page\)](#) ([links](#))
- [Straight play \(redirect page\)](#) ([links](#))
- [Playlet \(redirect page\)](#) ([links](#))
- [Stageplay \(redirect page\)](#) ([links](#))
- [Play \(theater\) \(redirect page\)](#) ([links](#))

## Stage play

From Wikipedia, the free encyclopedia

Redirect page

↳ [Play \(theatre\)](#)

# Disambiguation pages

- This page collects links for a number of possible concepts that an arbitrary expression could be referred to.
- This models homonymy and polysemy.

## Play

---

From Wikipedia, the free encyclopedia

(Redirected from [PLAY](#))

**Play** may refer to:

- [Play \(activity\)](#), enjoyed by animals, including humans
- [Play \(theatre\)](#), structured literary form of theatre

### In films [edit]

---

- [Play \(2005 film\)](#), a Chilean film directed by Alicia Scherson
- [Play](#), a 2009 short film directed by [David Kaplan](#)
- [Play \(2011 film\)](#), a Swedish film directed by Ruben Östlund
- [The Play \(film\)](#), a 2013 Bengali film

# Internal links

- The hypertexts that links to other Wikipages, which refer to related concepts.

## Play (theatre)

---

From Wikipedia, the free encyclopedia

A play is a form of literature written by a playwright, usually consisting of dialogue between characters, intended for theatrical performance rather than just reading. Plays are performed at a variety of levels, from Broadway Off-Broadway regional theater to Community theatre, as well as University or school productions. There are rare dramatists, notably George Bernard Shaw, who have had little preference whether their plays were performed or read. The term "play" can refer to both the written works of playwrights and to their complete theatrical performance.<sup>[1]</sup>

# Inter-language links

- The counterparts of the concept contained within wikipedias in other languages.

The screenshot shows a Wikipedia page with a sidebar on the left containing language links. A red box highlights the 'Languages' section and the list of language names. The main content area contains sections for 'Comedy', 'Farce', and 'Satirical', each with a 'Main article' link and a brief description. The page has a standard header with 'Download as PDF' and 'Printable version' options.

Download as PDF  
Printable version

Languages

- Afrikaans
- العربية
- Azərbaycanca
- Български
- Català
- Чӑвашла
- Deutsch
- Eesti
- Español
- Esperanto
- فارسی
- Français
- Frysk
- 한국어
- Հայերեն
- Bahasa Indonesia
- Italiano
- עברית

Comedy [edit]  
*Main article: Comedy*  
Comedies are plays which are d  
characters, and strange circums  
one of the two original play types  
William Shakespeare's play "A N  
Night Live".<sup>[2][3]</sup>

Farce [edit]  
*Main article: Farce*  
A generally nonsensical genre o  
includes William Shakespeare's

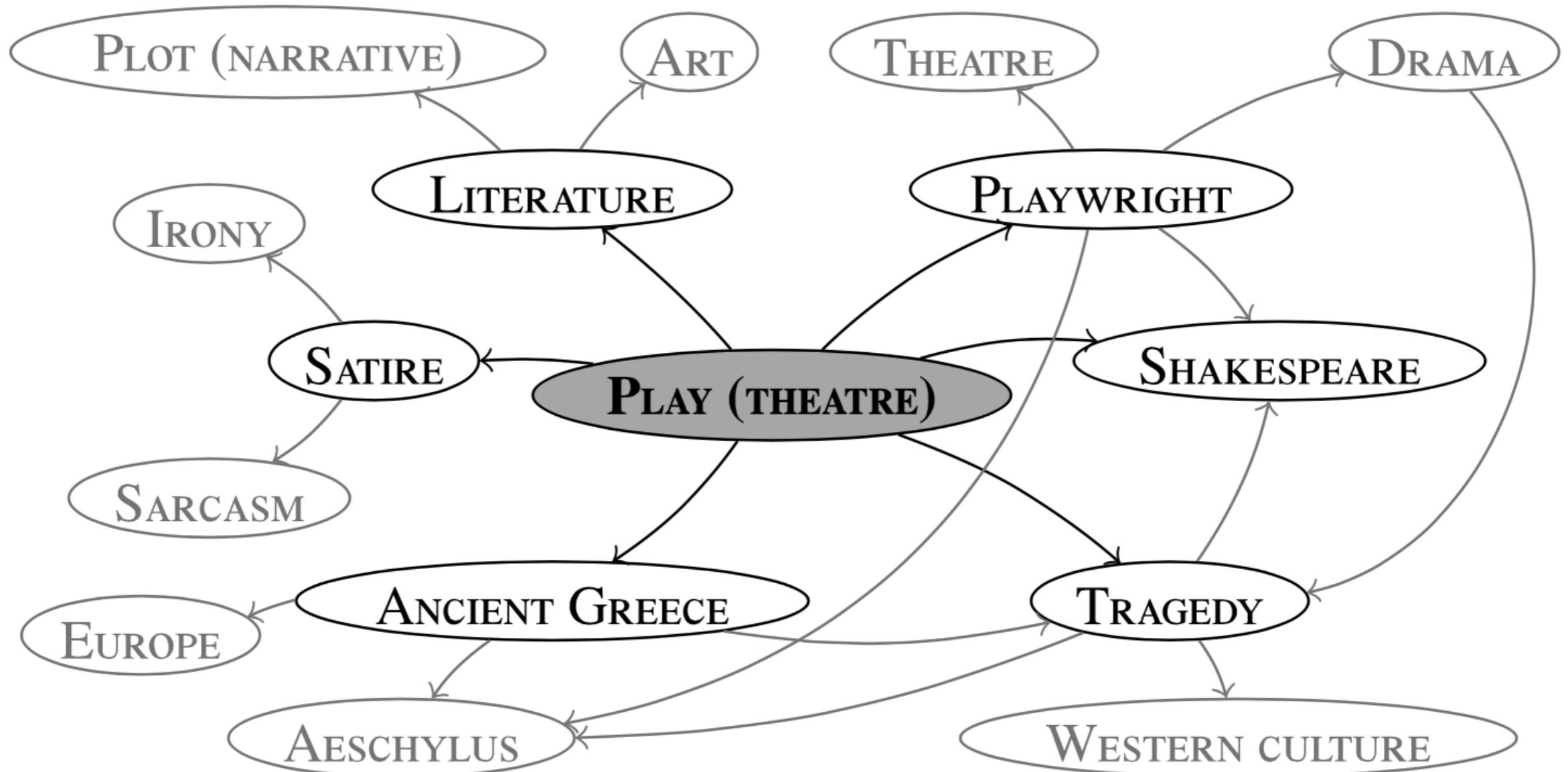
Satirical [edit]  
*Main article: Satire*  
A satire play takes a comic look

# Categories

- Wikipages can be assigned to one or more categories, which are also concepts and have corresponding Wikipages.

The screenshot shows a portion of a Wikipedia page. At the top, there is a light blue header bar with the text "Authority control" and links for "GND: 4461958-3" and "NDL: 00562448". Below this, a red rectangular box highlights a section containing the text "Categories: Drama | Theatre | Performing arts". Further down the page, a grey footer area contains the text "This page was last modified on 20 May 2015, at 02:31." and "Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. The page is part of the Wikimedia Foundation, Inc., a non-profit organization."

# Wikipedia graph



# Outline

## ○ Introduction

- Objectives
- WordNet
- Wikipedia

## ○ Construction details

- Mapping algorithm
- Disambiguation context
- Translating to multilingual

## ○ Evaluation & System

- Mapping evaluation
- Translation evaluation
- Online system

# Mapping algorithm

# Condition 1: One-to-one correspondence

- Establishing links between Wikipages and WordNet senses:

$$\mu : Senses_{Wiki} \rightarrow Senses_{WN} \cup \{\varepsilon\}$$

$Senses_{Wiki}(w)$  : All the Wikipages with title lemma as  $w$ .

$Senses_{WN}(w)$  : The set of senses of the lemma of  $w$  in WordNet.

- If  $|Senses_{Wiki}(w)| = |Senses_{WN}(w)| = 1$  :  
the correspondence between Wikipage and WordNet  
senses is one-to-one, so we can directly link them.

## Condition 2 : No direct correspondence

- There are cases when:

$$|Senses_{WN}(w_1)| = 0$$

- But  $w_1$  has redirection page of  $w_2$  such that:

$$|Senses_{Wiki}(w_2)| = |Senses_{WN}(w_2)| = 1$$

## Condition 3: Ambiguous Correspondences

- No redirection one-to-one correspondence
- $|Senses_{Wiki}(w)| > 1 \quad or \quad |Senses_{WN}(w)| > 1$
- We should find the most likely correspondence:

$$\begin{aligned}\mu(w) &= \operatorname{argmax}_{s \in Senses_{WN}(w)} p(s|w) = \operatorname{argmax}_s \frac{p(s, w)}{p(w)} \\ &= \operatorname{argmax}_s p(s, w).\end{aligned}$$

# Disambiguation context

# Bag-of-words Disambiguation Context

Context of a Wikipage		Context of a WordNet sense	
Type	Example	Type	Example
Title labels	Play ( <a href="#">Theater</a> )	Synonymy	drama <sub>n</sub> <sup>1</sup> , dramatic play <sub>n</sub> <sup>1</sup>
Outgoing links	<a href="#">literature</a> , <a href="#">comedy</a> , etc	Hypernymy (Generalization)	dramatic composition
Redirections	<a href="#">Playlet</a> redirects to Play (Theater)	Hyponymy (Specification)	
Categories	<a href="#">Plays</a> , <a href="#">Drama</a> , <a href="#">Theater</a>	Gloss words	“A dramatic <a href="#">work</a> intended for performance by actors on <a href="#">stage</a> ”

# Joint probability estimation

- Joint probability of Wikipage  $w$  and WordNet sense  $s$ :

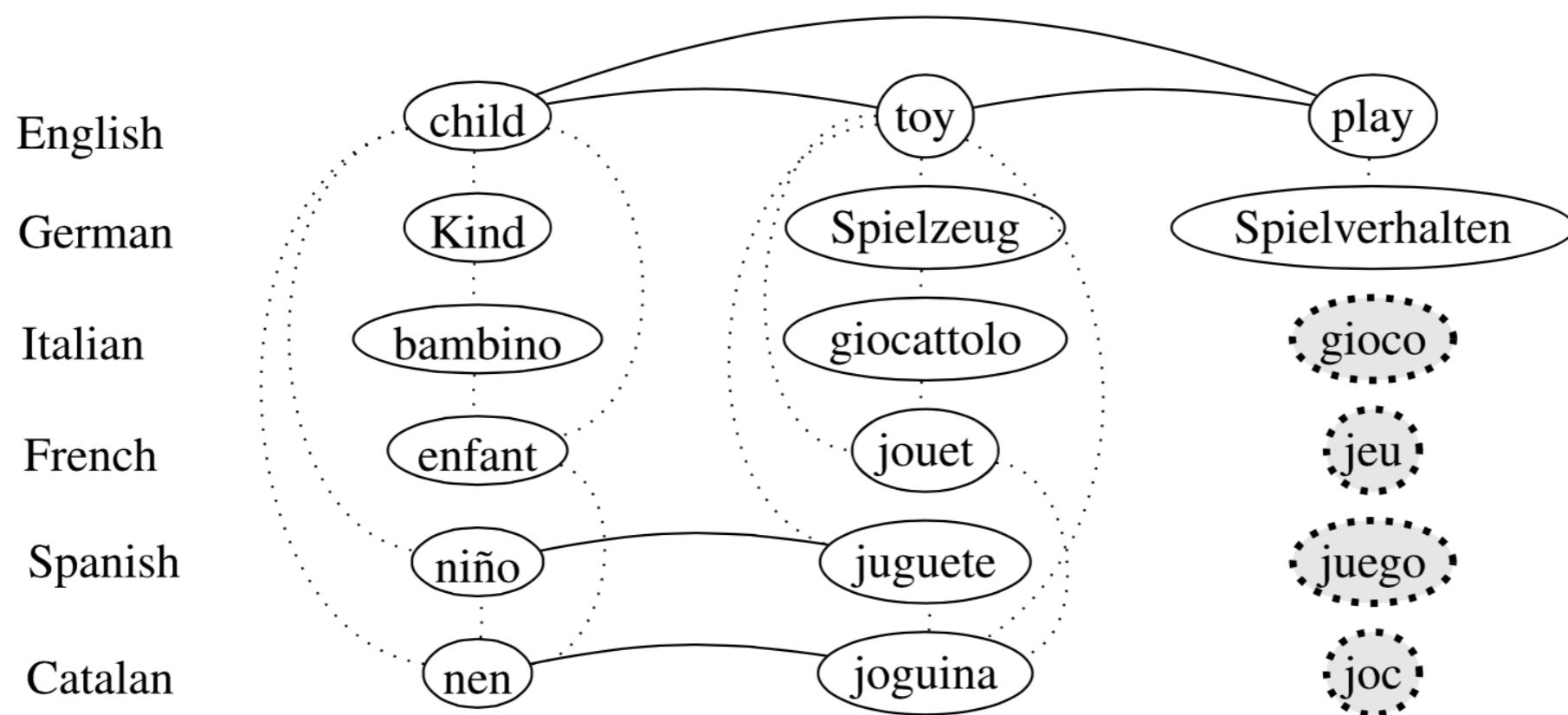
$$p(s, w) = \frac{\text{score}(s, w)}{\sum_{\substack{s' \in \text{Senses}_{\text{WN}}(w), \\ w' \in \text{Senses}_{\text{Wiki}}(w)}} \text{score}(s', w')}.$$

- where  $\text{score}(s, w) = |\text{Ctx}(s) \cap \text{Ctx}(w)| + 1$

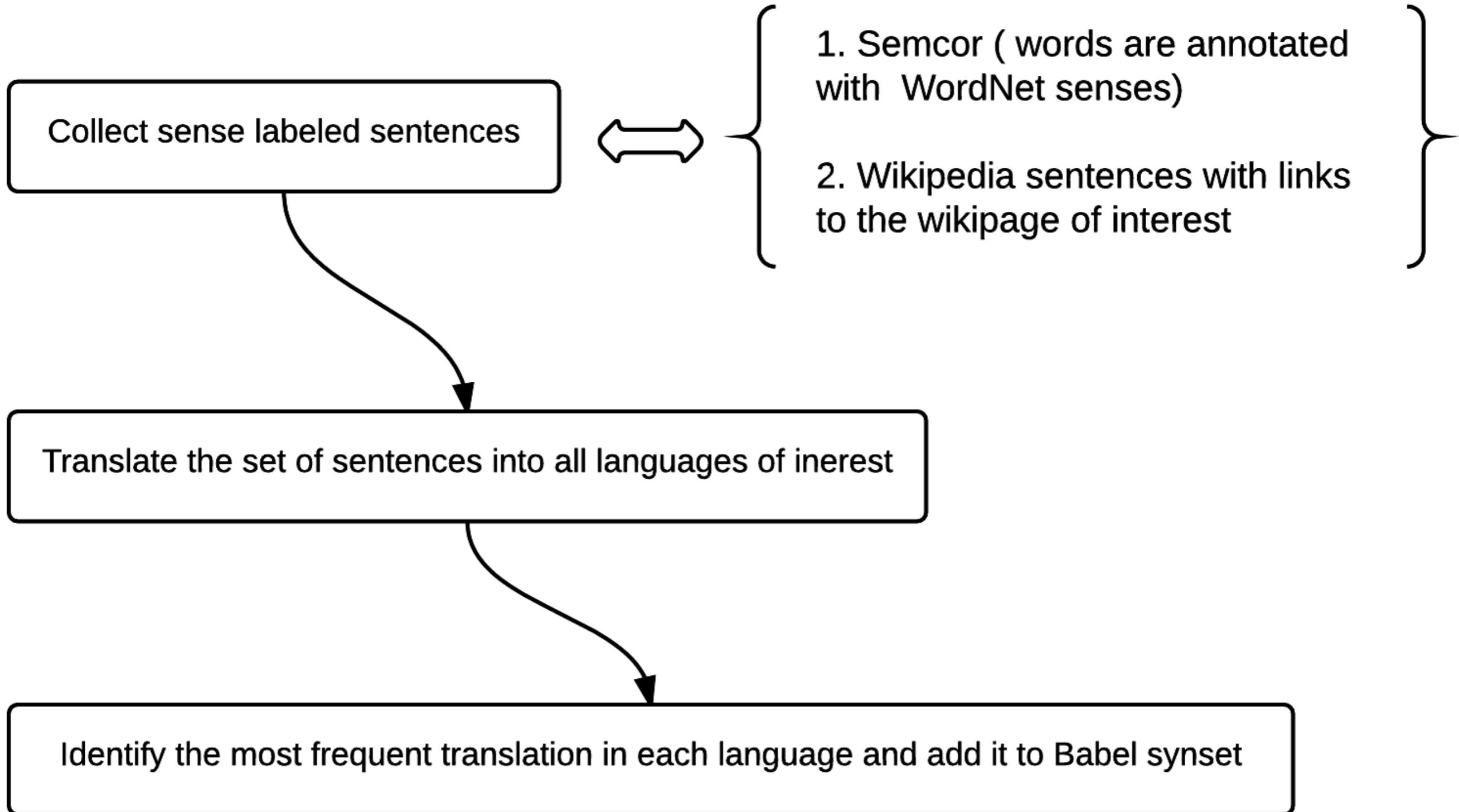
**Translating to  
multilingual**

# Babel synset

- Original WordNet synset.
- Wikipage w
- Redirections to w
- Pages linked via wikipage inter-language links (translation)
- Redirections to the inter-language linked pages (translation)



# Machine translation methodology



# Outline

## ○ Introduction

- Objectives
- WordNet
- Wikipedia

## ○ Construction details

- Mapping algorithm
- Disambiguation context
- Translating to multilingual

## ○ Evaluation & System

- Mapping evaluation
- Translation evaluation
- Online system

# Mapping evaluation

# Gold Standard dataset

- A dataset which includes all lemmas whose senses are contained both in WordNet and Wikipedia.
- Select a random sample of 1000 wikipages.
- Manually annotate correct WordNet senses for each page.
- An empty sense label is given, if no correct mapping.

# Mapping performance

WordNet Context	Precision	Recal	F <sub>1</sub>	All
taxonomic	89.7	47.8	62.3	72.6
gloss	87.6	51.8	65.1	74.0
taxonomic +gloss	87.5	65.6	75.0	80.9
Random baseline	24.2	46.9	31.9	24.2
MFS baseline	25.4	49.2	33.5	25.4

# Translation evaluation

# Datasets

	Catalan	French	German	Italian	Spanish
Word senss	64,171	44,265	15,762	57,255	83,114
Synsets	40,466	31,742	9,877	32,156	55,365

Note: all wordnets are linked to the English WordNet

# Evaluation measures

Synset coverage:

$$\text{SynsetCov}(\mathcal{B}, \mathcal{F}) = \frac{\sum_{S_F \in \mathcal{F}} \delta(S_B, S_F)}{|\{S_F \in \mathcal{F}\}|}$$

Word sense coverage:

$$\text{WordCov}(\mathcal{B}, \mathcal{F}) = \frac{\sum_{S_F \in \mathcal{F}} \sum_{s_F \in S_F} \delta'(S_B, s_F)}{|\{s_F \in S_F : S_F \in \mathcal{F}\}|}$$

Extra synset coverage:

$$\text{SynsetExtraCov}(\mathcal{B}, \mathcal{F}) = \frac{\sum_{S_E \in \mathcal{E} \setminus \mathcal{F}} \delta(S_B, S_E)}{|\{S_F \in \mathcal{F}\}|}$$

Extra word sense coverage:

$$\text{WordExtraCov}(\mathcal{B}, \mathcal{F}) = \frac{\sum_{S_E \in \mathcal{E} \setminus \mathcal{F}} \sum_{s_E \in S_E} \delta'(S_B, s_E)}{|\{s_F \in S_F : S_F \in \mathcal{F}\}|}$$

# Translating Performance

	Catalan	French	German	Italian	Spanish
SynsetCov	73.3	92.9	73.4	73.7	76.6
WordCov	64.0	86.0	57.6	52.9	66.4
SynsetExtraCov	142	159	902	213	102
WordExtraCov	340	514	2298	419	384

# System Demonstration

# Thanks,

## Q & A