

Introduction
oooo

Literature
oooooo

Simple Case
oo

Complex Case
oooooooooooooooooooooooooooo

Attempts at Visualisation of Etymological Information

Armin Hoenen

Goethe Universität Frankfurt, Germany

Introduction
oooo

Literature
oooooo

Simple Case
oo

Complex Case
oooooooooooooooooooooooooooo

- 1 Introduction
- 2 Literature
- 3 Simple Case
- 4 Complex Case

Introduction

Guiding Questions:

- Which forms of visualization exist for word histories?
 - Which forms are useful, which possible?
 - What added value can be gained?
 - In how far is automatization an option?

Formulation of an agenda of prospective next steps.

Text as Visualization

Text itself is a visualization of speech.

Then, extra-textual devices such as punctuation, linebreaks, table grids etc. are entirely visual.

Graphical formatting of text

- Separator sequences: ';' or '>', 'developed into' (OED online), sound changes
 - tables and other structural compositions

→ Primarily text-based

Here, we look for more sophisticated and less text-bound visualizations, where **non-textual** visual elements are perceptually dominant.

Ressources

de Melo (2014), [?]

Recently, some of these reference works, e.g. the ones in the Leiden Indo European Etymological Dictionary series, have also been made available as databases. Unfortunately, other than the resources listed below, we are not aware of any open, freely available machine-readable versions of such works. Additionally, most such reference works are restricted to a single language or a set of closely related languages.

One other publication explicitly focussed on automatically visualizing etymology.

Ressources

formal representation framework

GRAPH THEORY used to represent etymological relationships.
Semantic Web community most active, ontological graphs,
typed edges.

display and encoding of

- time as dimension
 - multiple languages (across time)
 - uncertainty/alternative hypotheses.

De Melo 2014

De Melo, G. (2014). Etymological wordnet: Tracing the history of words. In LREC, pages 1148–1154.

Etymological Semantic Web Triples

deu: Stapel **rel:etymological_origin_of** fra: étape

deu: Stapelspeicher **rel:has_derived_form** deu: Stapelspeichers

deu: Stapelspeichers **rel:is_derived_from** deu: Stapelspeicher

deu: Stapler **rel:etymological_origin_of** deu: Gabelstapler

deu: Star **rel:etymology** eng: star

De Melo 2014

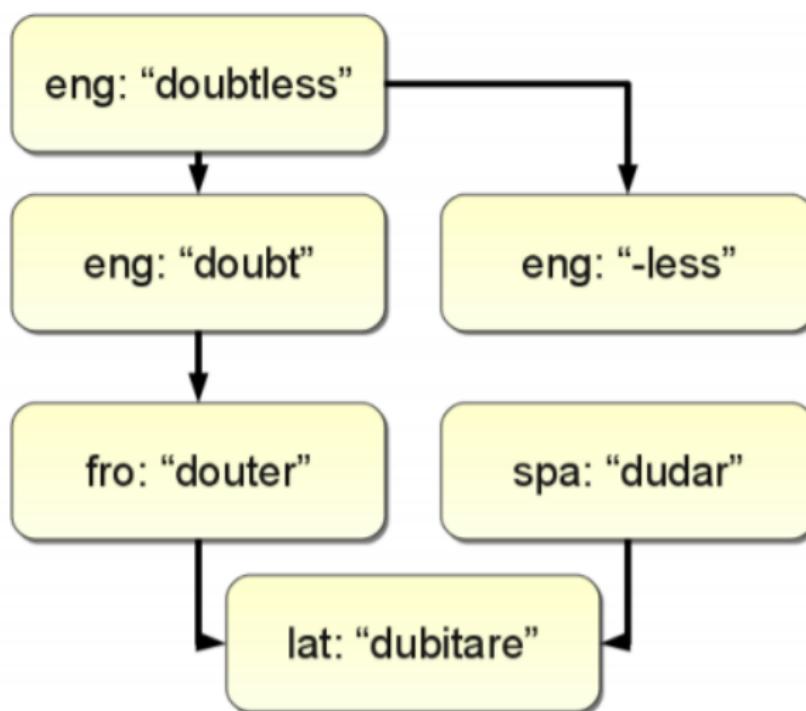


Figure 1: Excerpt from Etymological Wordnet

De Melo 2014

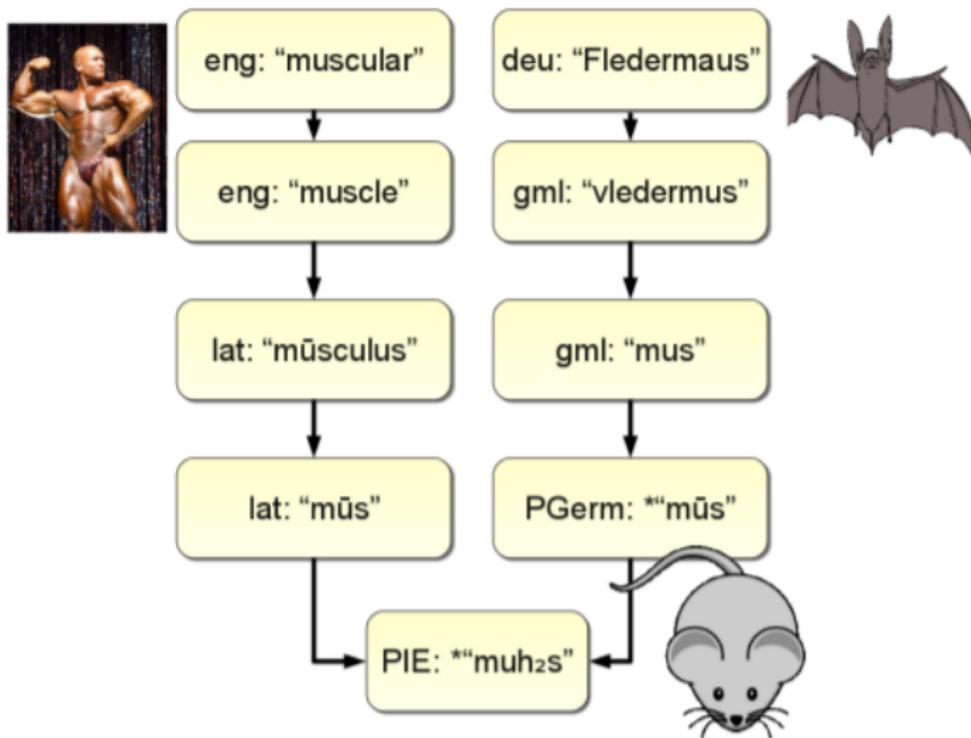


Figure 3: Connections in Etymological Wordnet

De Melo 2014

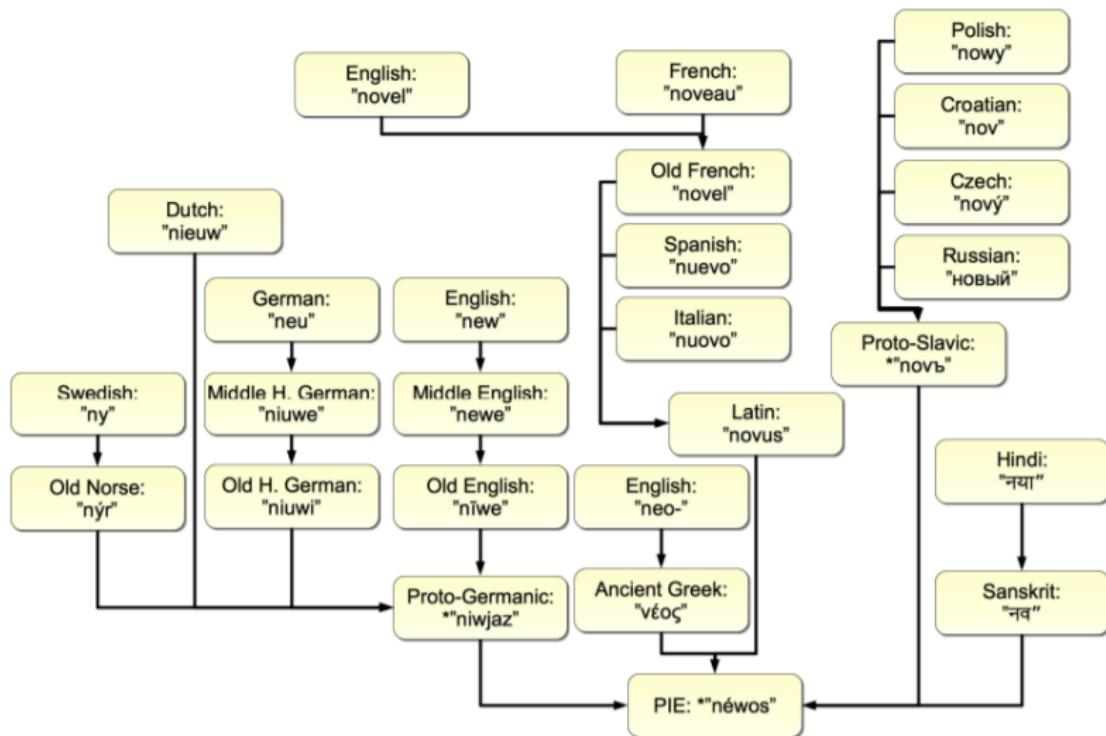


Figure 4: Descendants of a Word in Etymological Wordnet

Dixit & Karrfelt

Dixit, C. and Karrfelt, F. (2016). Visualizing etymology: A radial graph displaying derivations and origins.

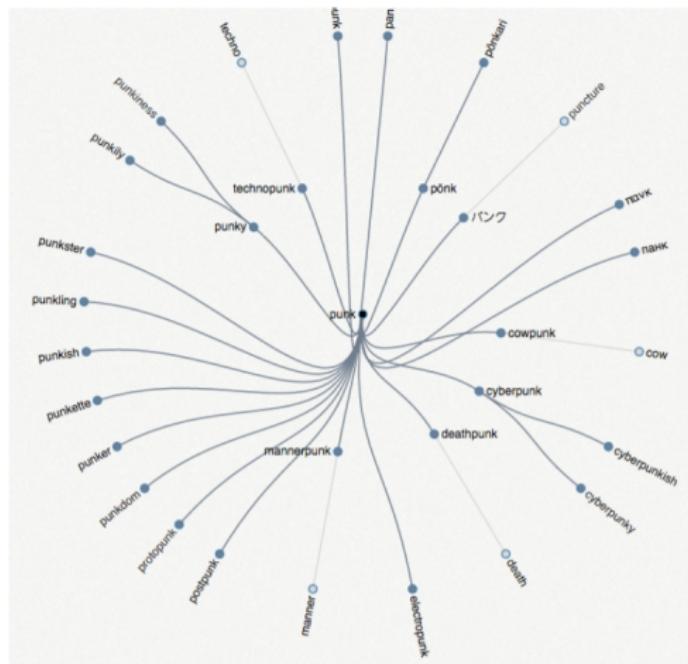


Figure 8 Graph for "punk" depth 2.

Resumee on Existing Approaches

- few ressources
- few models
- text and tables already some form of visualization

Suitable for simple, linear, monocausal relations, hypothesis comparsion.

AGENDA++

create new ressources

Comparing Simple Word Histories

"Schatulle" after DWDS



"Schatulle" after Duden



"Schatulle" after DWB



...information extraction (compare DBpedia): Comparability

Simple Word Histories

WB	edges	longest path	language shifts
DWDS	4	3	3
DUDEN	3	2	2
DWB	2	2	2

Implications for automatical retrieval

- the more edges are present, the more sophisticated the informational content
- Pathlength and loaning and reloaning may be subject to economy principles

Visualisation Complex Cases

The history of 'society', SHA-KAI: 社会
free after Yanabu, Akira 柳父 章 1991, abridged

Initial Problem

A word representing the meaning of a complex non local group does not exist in Meiji-era Japan (from 1868). Opening up a closed country and intellectual concern with Western culture leads to many new word formations.

Narrative Word History - Phase I

No translation equivalent & no equivalent reality (post-feudal society)

- early translations as in dictionaries:
 - circle of friends (nakama), meet (atsumaru), government/assembly (seifu)
 - probably significant: Ekaijiten dictionary 1847/48: KAI (meeting), kessha ('byakurensha' where SHA expresses *people with common goals*)
 - **2 senses of society**: locally delimited (S in NGS) - all (in a state etc.) OED Englisch; **translations only reflect sense 1**
 - Neologisms: ningen kousai, from **ningen** (mankind) and **kousai** (locally limited relationships - family, master and servant)

Narrative Word History - Phase IIa

Termformation & Formation of (consciousness about) the concurrent reality

- in the *broad public* groups form, presenting and discussing Western cultural artifacts, e.g. the bungakusha (literary circle) and term themselves X-and-Y-**SHA**, comp. Eikaiiten byakurenSHA
- very important: **MeirokuSHA**, 6 intellectuals, which issue journal **Meirokuzasshi** influencing/reflecting word formation

Narrative Word History - Phase IIb

Manifestation of the term SHAKAI (SHA = group; KAI = meet)

- SHAKAI and KAISHA possible generalizations of X-and-Y-SHAs
 - KAISHA - (any) -SHA: main characteristic of SHAs (head): they meet KAI; today KAISHA means company
 - SHAKAI - the phenomenon of meeting in these groups (SHAKAI enzetsu), head: KAI
- new antonyms; SHAKAI newly relates negatively to SEKEN (simple people) getting pejorative (known etymological phenomenon)
- translators start to consistently use SHAKAI for 'society' (Sense 2)

Alternatives

Parallelisms? Loans?

Geographical Society – Geographical = Society

ByakurenSHA – Byakuren = SHA

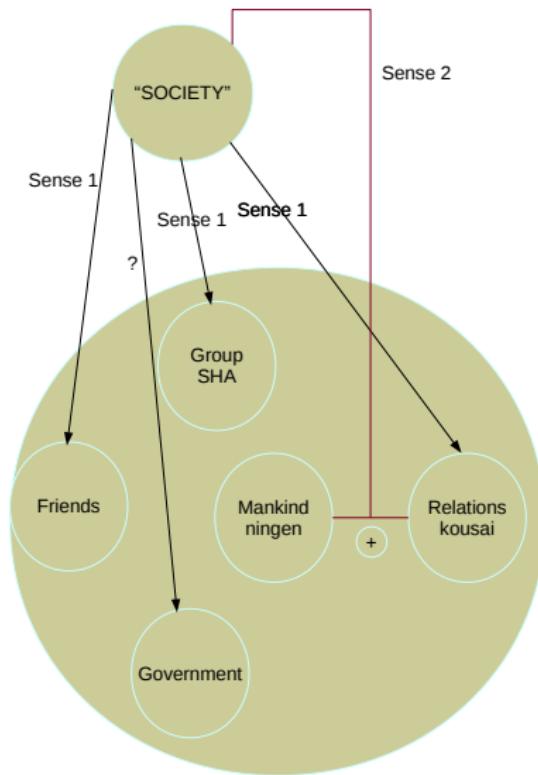
Bad option for Japanese, since SHA is cranberry-like.

Compounds are more common. SHA alone means (earth-god-)shrine. Similar reasons for SHASHA.

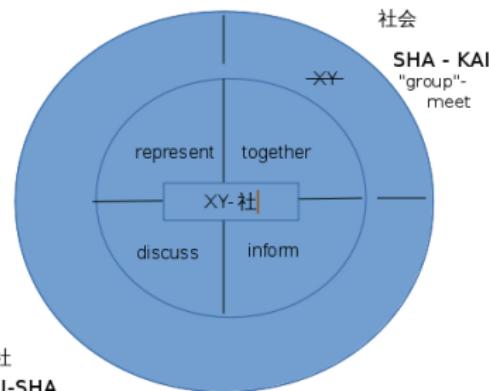
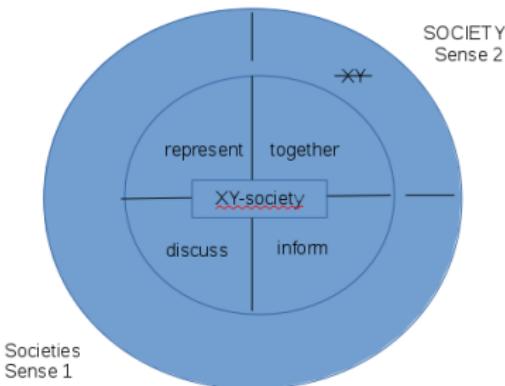
A loan 'sosaiti' is attested, but did not win the etymology race.

Co-incidence of new societal phenomena and introduction of new thoughts and abstracta are important for this word history.

Shakai – first attempt

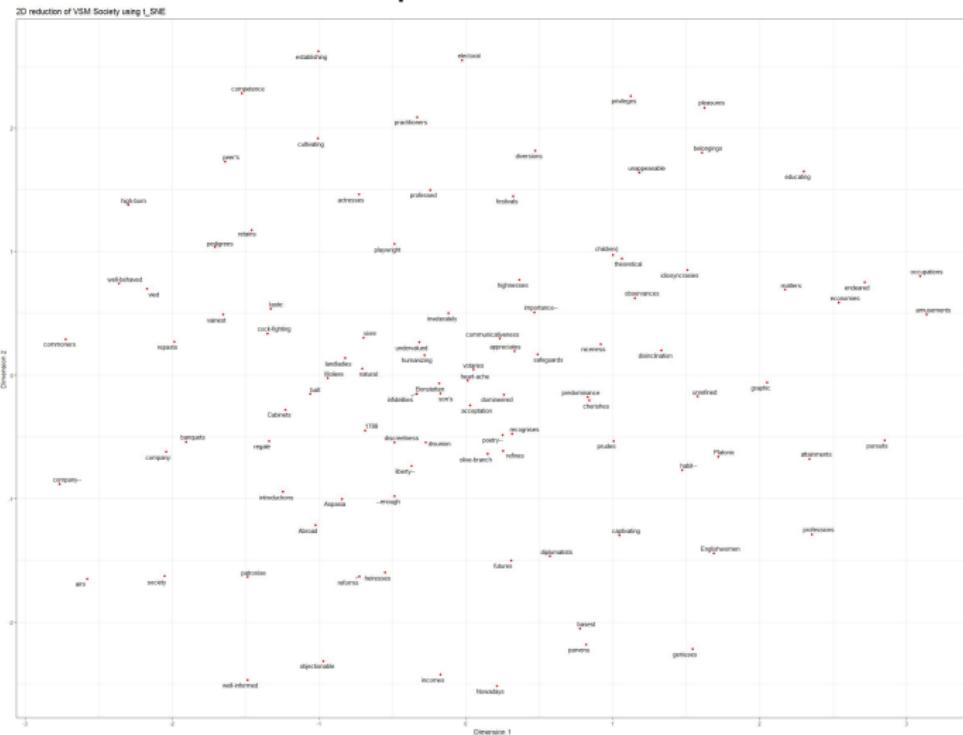


Shakai – second attempt



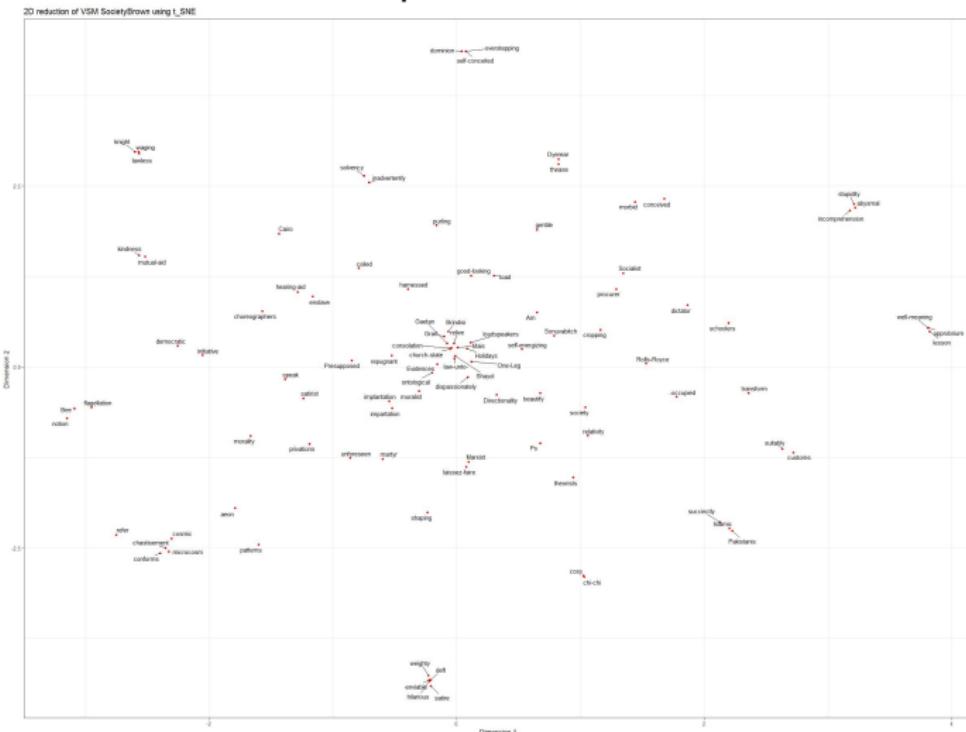
Word Embeddings

Word vector spaces entail semantics.



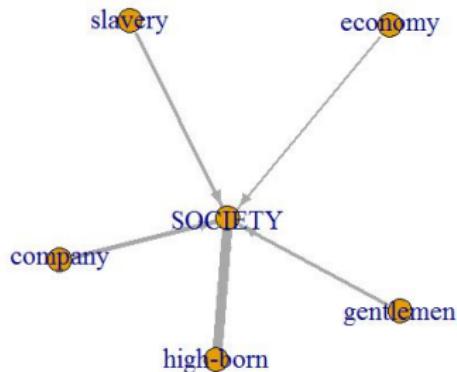
Word Embeddings

Word vector spaces entail semantics.



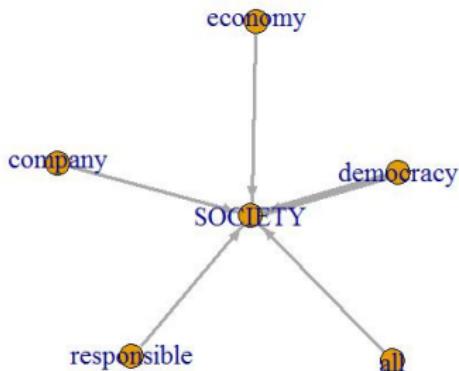
Secondary Visualisations

Closest Tokens as trees or weighted graphs.



Word Embeddings

Closest Tokens as trees or weighted graphs.



Word Embeddings

Problems

- noisiness in transformations (tsne reduction, tokenization, sample bias, etc.)
 - necessarily contrastive, not applicable for neologisms

Problems

Design and Abstraction

- many variables, parameters and sources: formalization?
- different histories very heterogeneous
- non-linguistic context
- ambiguity
- intersection of graph theory & set theory ?
- defining a boundary between simple and complex, how and where
- Getting to bigger etymological phenomena – pejorativization of older words, narrowing of scope etc.

Generalisation

Why a Japanese example?

"Similar" phenomena in the wake of social change cross linguistically attested. Well documented in Japan.

AGENDA++

- find empirical threshold and complexity measure for article
- think about formalization and framework of visualization

Conclusion

Simple Cases

Extractable (IE). Visualizable through graph theory eventually with typed edges. Semantic web based resources available.

Find a good criterion for distinguishing simple etymological articles/histories from complex ones.

Complex Cases

Big gains possible through breaking through forced sequentiality of text. Models and modelling frameworks to be formulated. Semantic web ontological presentation possible. Combination of visualizations for graph and set theory.

AGENDA

- create/extract new resources
- IE template, definitions
- Gold standard for extraction
- find empirical threshold and complexity measure for article
- think about formalization and framework of visualization of complex histories

Dynamic formats, maps

Introduction
oooo

Literature
oooooo

Simple Case
oo

Complex Case
oooooooooooooooooooo●oooooooooooo

New formats?

MANGA

漫画

Etymology Box 1

Translation: comic

Source: Japan

Compound: 'whimsical' + 'picture'

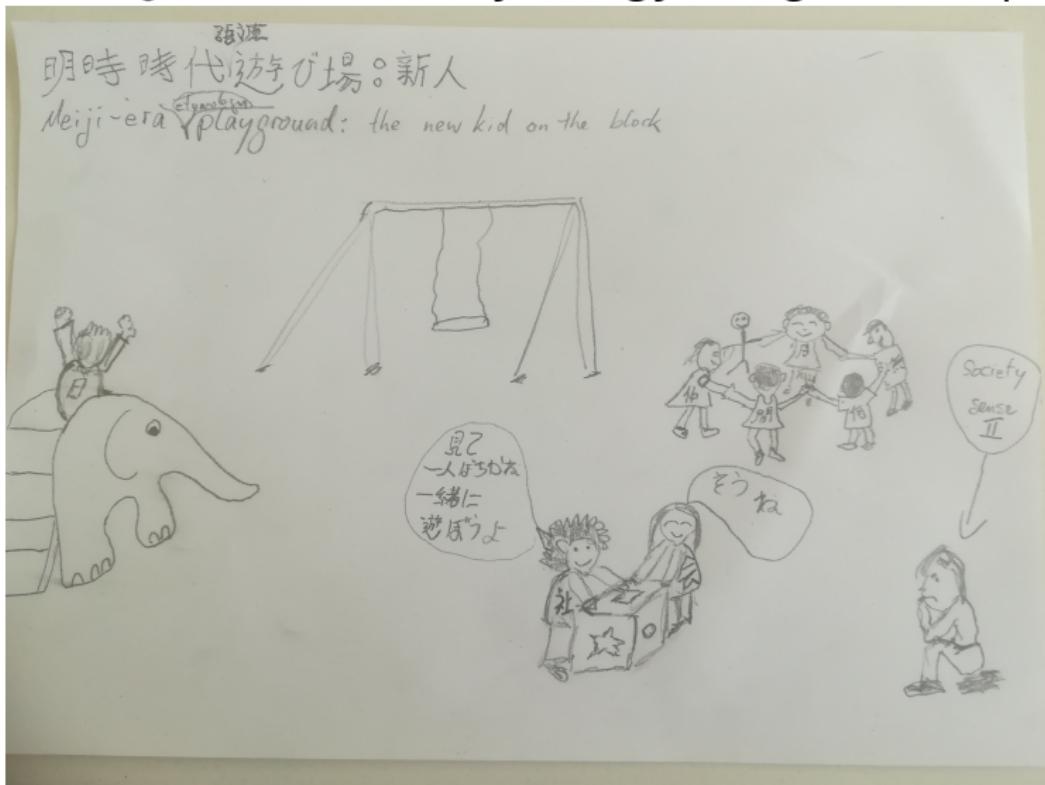
Birth: since late 18th century

Famous manga-artists: Hokusai, Tezuka, Hasegawa, ...

Popular mangas: One Piece, Dragon Ball, Naruto, Kaiji, ...

The arrival of a new sense and a new genre ?

Is a new genre in arrival? **Etymology mangas...** let's play



Reflection - Perception

Playground

added value?

- building on narrative content (comprehensible only if read)
- relatively demanding in production

The birth of a new word (birth metaphor is the most obvious)

