



Everyone is an NLP expert now

Natural Language Processing

Fall 2023

Programming Languages

C, C++, Java, Python, ...

- unambiguous
- fixed (relatively)
- designed
- learnable?
- known semantics

Natural Languages

French, English, Korean, Chinese, Tagalog, ...

- ambiguous
- evolving
- transmitted
- learnable
- unknown semantics

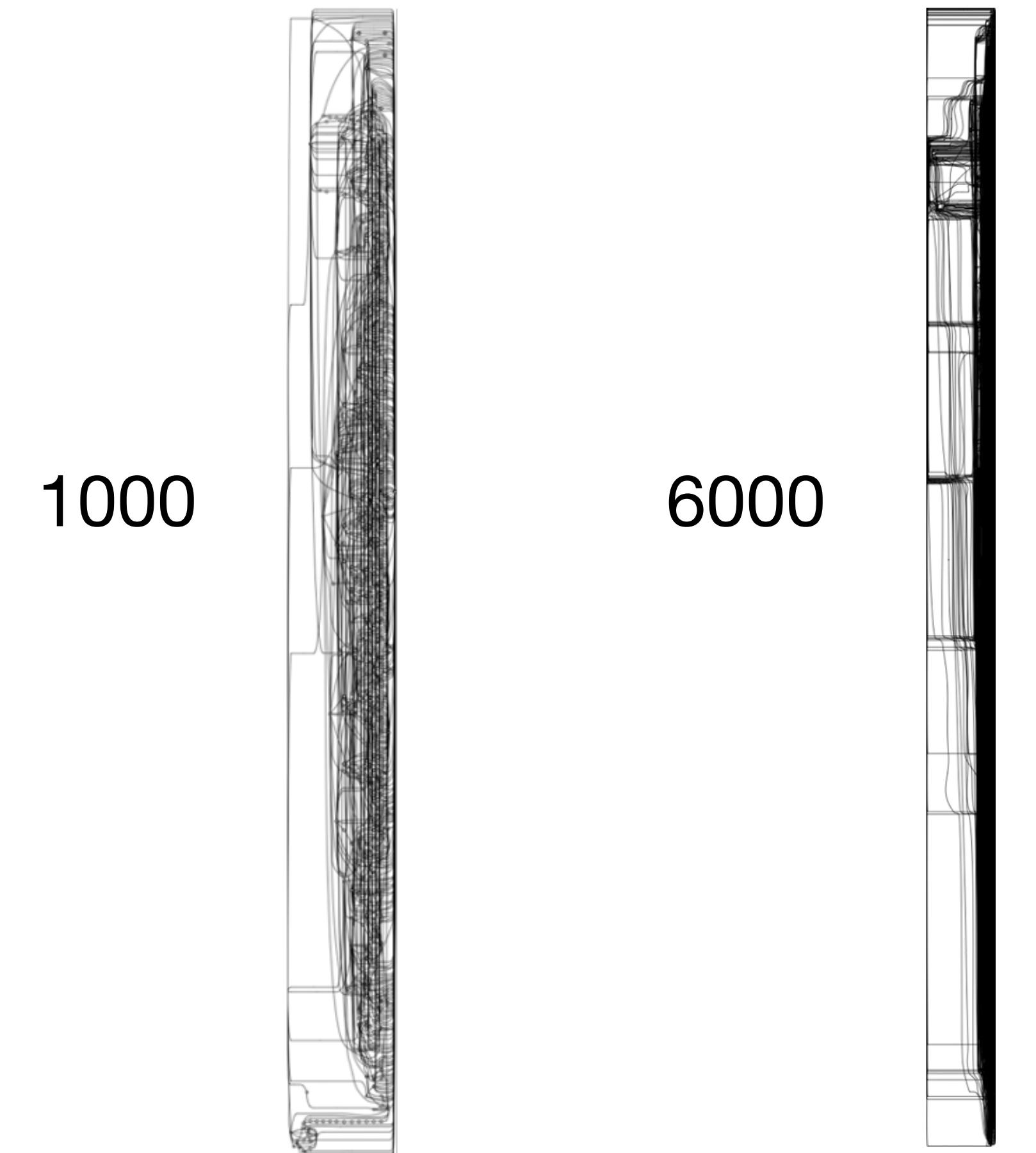
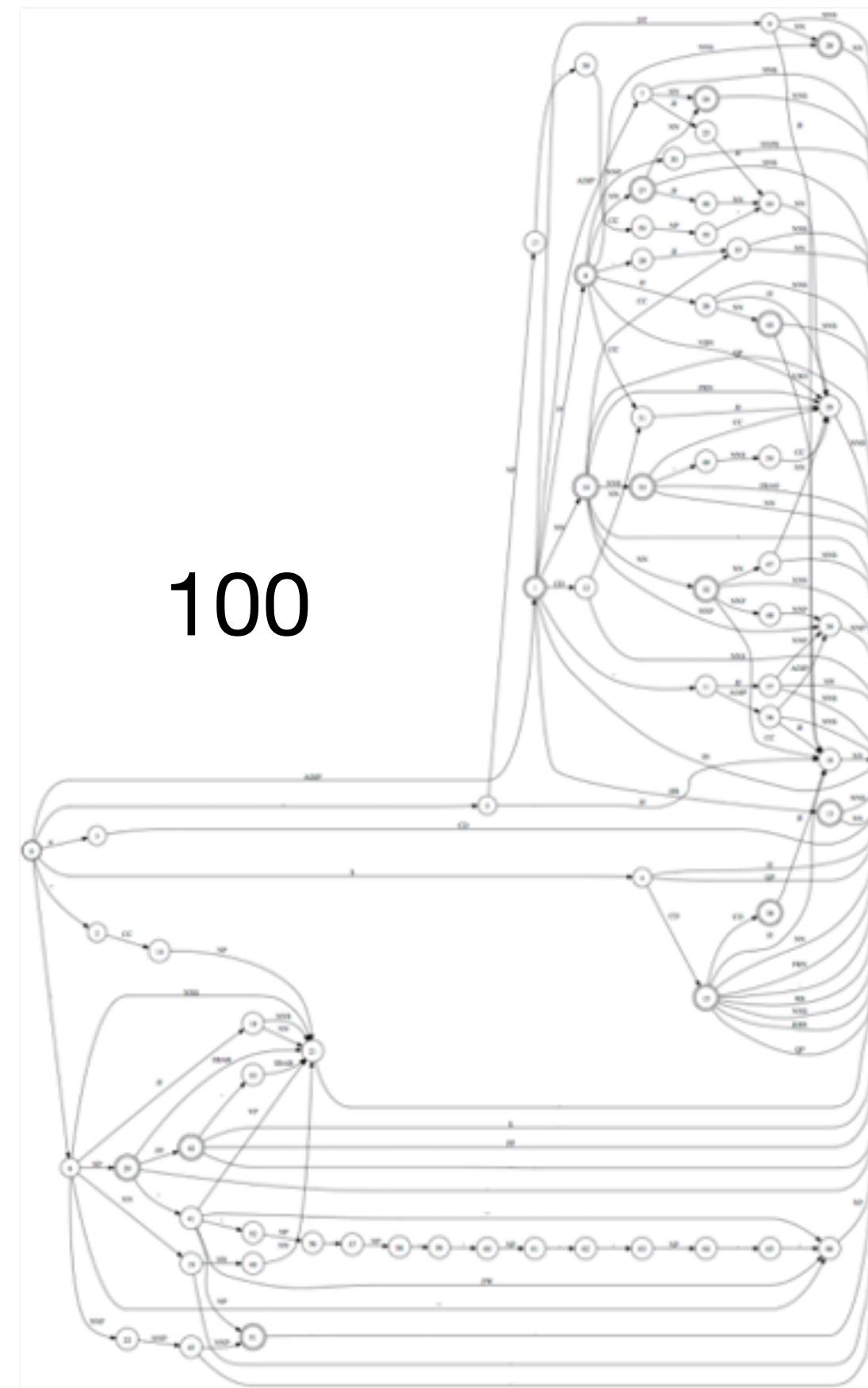
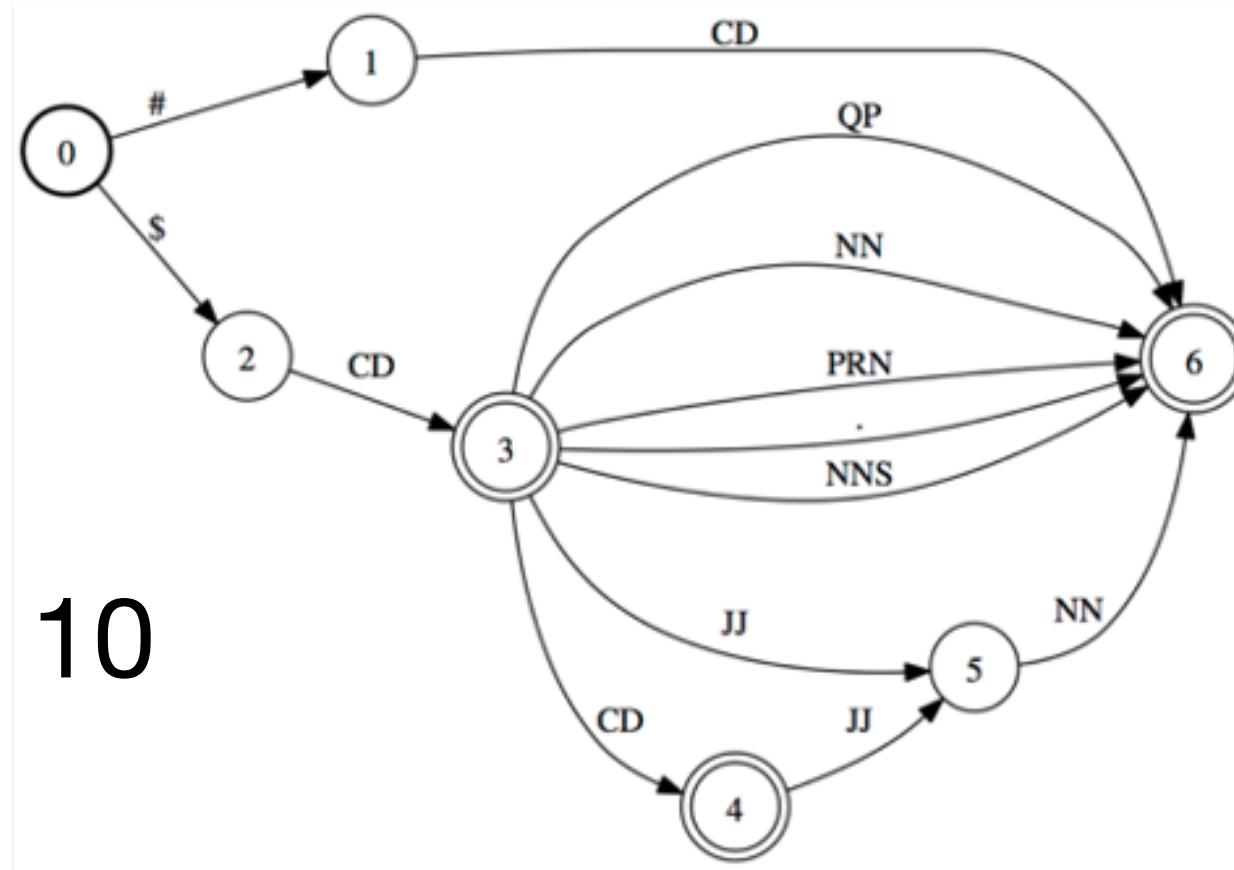
Why is NLP computationally hard?

Language is Complex

\$ 3 monthly fee

\$ 1 per-share pricing

- ✓ \$ number adjective noun
- ✗ \$ noun number adjective



Language is Complex

Edit the detailed description

oil painting of a man watching a woman in a park with a telescope



How many different meanings does this sentence have?

Language is Ambiguous

- Lung cancer in women mushrooms
 - Mushrooms is noun or a verb?
- Ban on nude dancing on governor's desk
 - Similar to “if-then-else” ambiguity
- Island Monks Fly in Satellite to Watch Pope Funeral
 - “get delivery” vs. “fly into space”
- British Left Waffles on Falkland Islands
 - Is it talking about the "British Left" political parties or "Left" as a verb?

Language is ambiguous

The screenshot illustrates the ambiguity of language through Google Translate's bidirectional translation feature.

Top Translation (English to Korean):

- Input: "The man saw the woman in the park with a telescope"
- Output: "남자는 공원에서 망원경으로 여자를 보았다."
- Details: The output is annotated with a blue arrow pointing from the English input to the Korean output.

Bottom Translation (Korean to English):

- Input: "남자는 공원에서 망원경으로 여자를 보았다."
- Output: "The man looked at the woman through a telescope in the park."
- Details: The output is annotated with a blue arrow pointing from the Korean input to the English output.

Interface Elements:

- Language selection dropdowns: English-Detected, English, Spanish, French (top left) and Korean-Detected, Korean, English, Spanish, French (bottom left).
- Speaker icons and "Look up details" links.
- Progress indicators: 50 / 5,000 (top right), 23 / 5,000 (bottom right).
- Feedback options: "Send feedback" (top right, bottom right).

Language is ambiguous

The screenshot illustrates the ambiguity of language through four Google Translate interface snippets arranged in a 2x2 grid.

Top Left: English - Detected → English, Spanish, French. Input: "The man saw the woman in the park with a telescope". Output: "Mies näki naisen puistossa kaukoputkella". A blue arrow points from this output to the bottom right output.

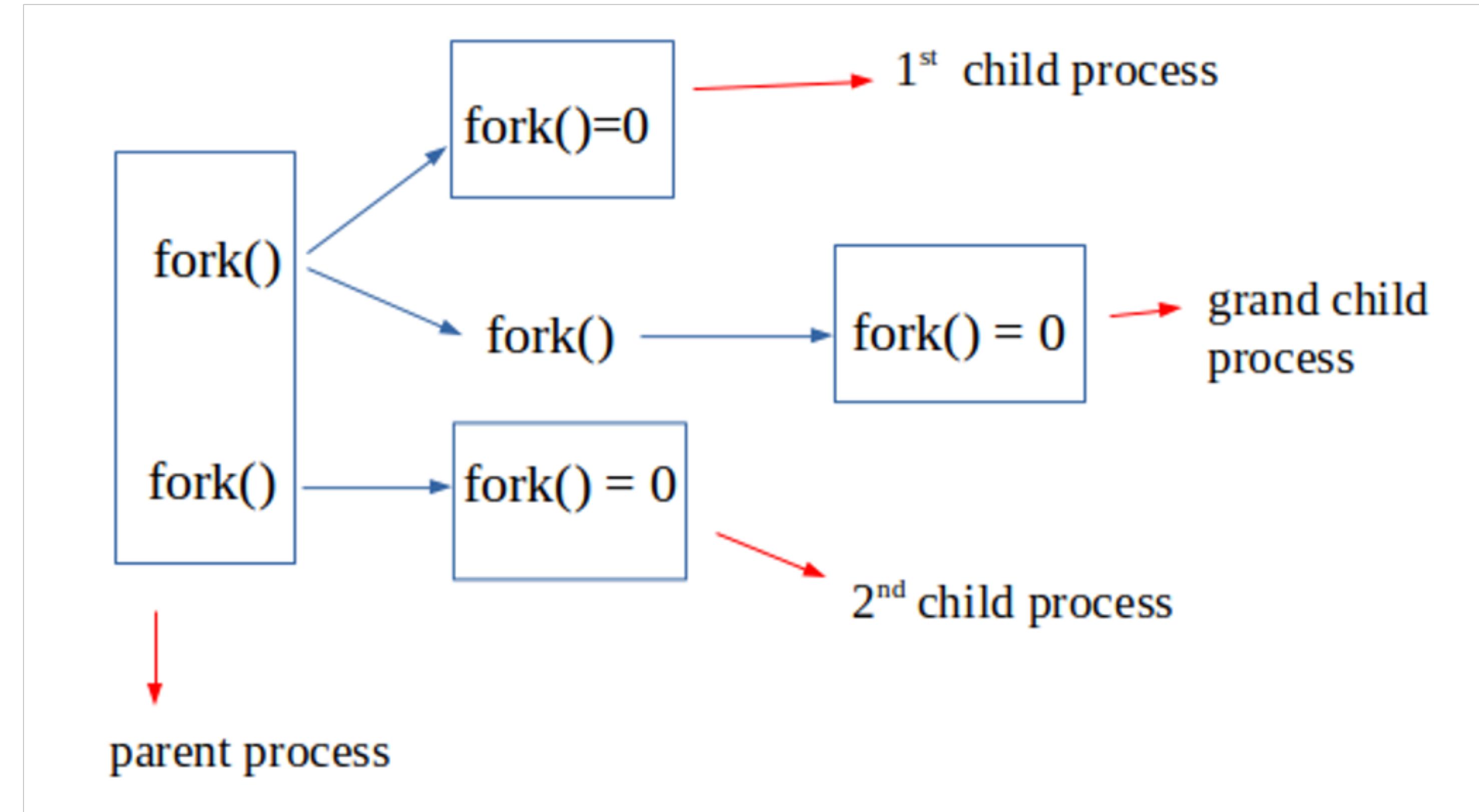
Top Right: English → Finnish, Korean, English. Input: "The man saw the woman in the park with a telescope". Output: "Mies näki naisen puistossa kaukoputkella". A blue arrow points from this input to the bottom left output.

Bottom Left: Finnish - Detected → English, Spanish, French. Input: "Mies näki naisen puistossa kaukoputkella". Output: "The man saw the woman in the park with binoculars". A blue arrow points from this output to the top right output.

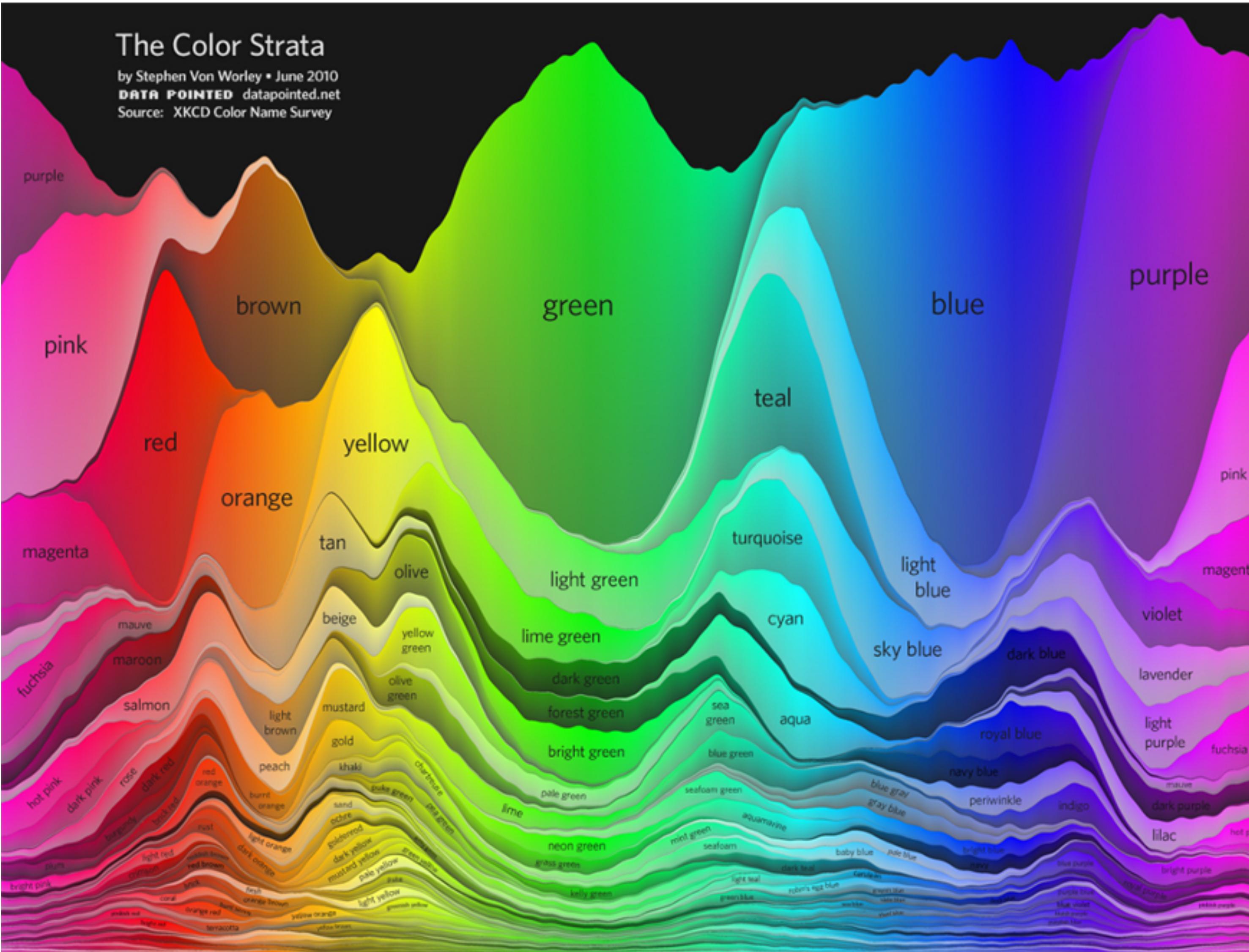
Bottom Right: English → Korean, English, Spanish. Input: "The man saw the woman in the park with a telescope". Output: "The man saw the woman in the park with binoculars".

Each snippet includes standard Google Translate features like "Look up details", audio icons, and a "Send feedback" button.

Language is ambiguous



Language is ambiguous



Language is ambiguous



Language is ambiguous

The screenshot shows the Google Translate interface comparing the Chinese character '锅' (guō) with its English translation 'pot'. Both terms are listed as nouns.

Left Panel (Chinese):

- Input: 锅
- Output: Guō
- Details: Noun
- Synonyms: 锅, 壶, 罐, 煲, 盆, 锅子
- Actions: Look up details, Microphone icon, Speaker icon, Edit icon

Right Panel (English):

- Input: pot
- Output: Noun
- Synonyms: 锅, 杠, 平锅
- Actions: Look up details, Speaker icon, Copy icon, Share icon, Expand all

More translations:

- pan: Noun, 锅, 杠, 平锅
- boiler: Noun, 锅炉, 锅, 镊
- cauldron: Noun, 釜, 锅, 镊, 甑
- caldron: Noun, 甑, 釜, 锅
- holloware: Noun, 锅

Language is ambiguous



Russian author Sasha Sokolov wrote *Between Dog and Wolf*, which had been deemed untranslatable from the original Russian since 1980. Alexander Boguslawski translated the work to English in December 2017.

<https://retrieverman.wordpress.com/2012/09/08/romeo-and-a-yellow-lab/>

<https://www.npr.org/2017/01/28/512199359/between-dog-and-wolf>

Language is ambiguous

- **Google's Computer Might Betters Translation Tool**
 - New York Times March 8, 2010
- **Number of Lothian patients made ill by drinking rockets**
 - Edinburgh Evening News, March 4, 2010
- **Violinist linked to JAL crash blossoms**
 - <https://languagelog.ldc.upenn.edu/nll/?p=1693>

Language is contextual

The town councillors refused to give the angry demonstrators a permit because they feared violence.

Who feared violence?

A) the town councillors

B) the angry demonstrators

Language is contextual

The town councillors refused to give the angry demonstrators a permit because they advocated violence.

Who advocated violence?

A) the town councillors

B) the angry demonstrators

Language is contextual

The screenshot shows the Google Translate interface. On the left, the English input field contains the sentence "she took the easy way out". On the right, the Korean output field displays the translation "그녀는 쉬운 길을 택했다". A blue speech bubble with the word "informal" points to the Korean word "택했다". The Korean input field at the top shows the detected language as "English - Detected". The Korean output field shows the detected language as "Korean". The interface includes standard translation controls like "Look up details", "Send feedback", and sharing icons.

informal

English - Detected English Spanish French

she took the easy way out

Look up details

25 / 5,000

Finnish English Korean

그녀는 쉬운 길을 택했다

geunyeoneun swiun gil-eul taeghaessda

Look up details

Send feedback

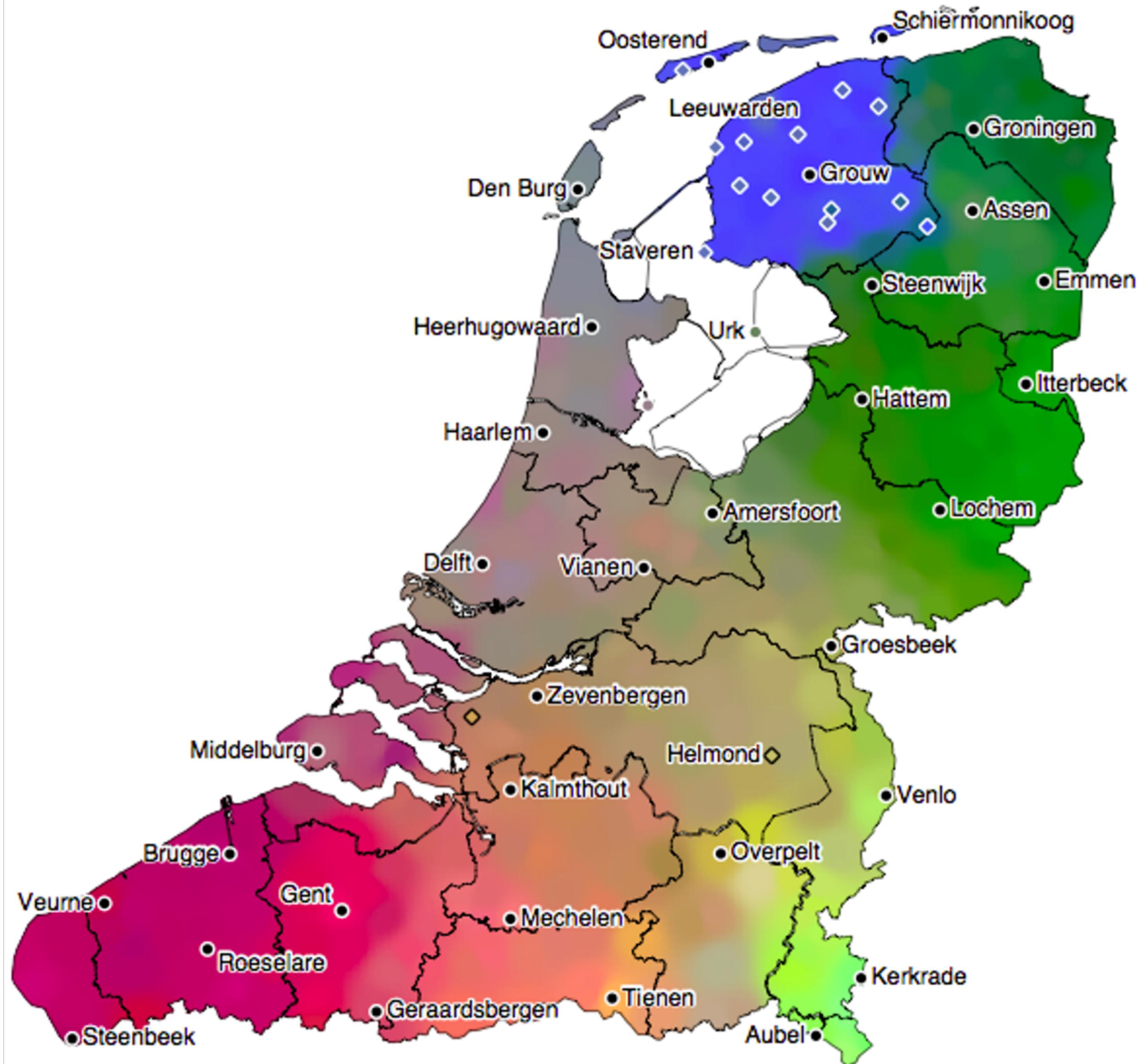
Language is contextual

The image shows a Google Translate interface. On the left, the English input field contains the sentence "her majesty, she took the easy way out". Below it are buttons for "Look up details", a microphone icon, a speaker icon, and a pen icon. The word count "38 / 5,000" is displayed. On the right, the Korean output field shows the translation "폐하, 그녀는 쉬운 길을 택했습니다". This phrase is circled in black. Below the Korean text is the phonetic transcription "pyeha, geunyeoneun swiun gil-eul taeghaessseubnida". There are also "Look up details", a speaker icon, and sharing icons. A blue speech bubble with the word "formal" points to the circled Korean phrase. The top navigation bar shows language pairs: English-Detected, English, Spanish, French, Finnish, English, Korean.

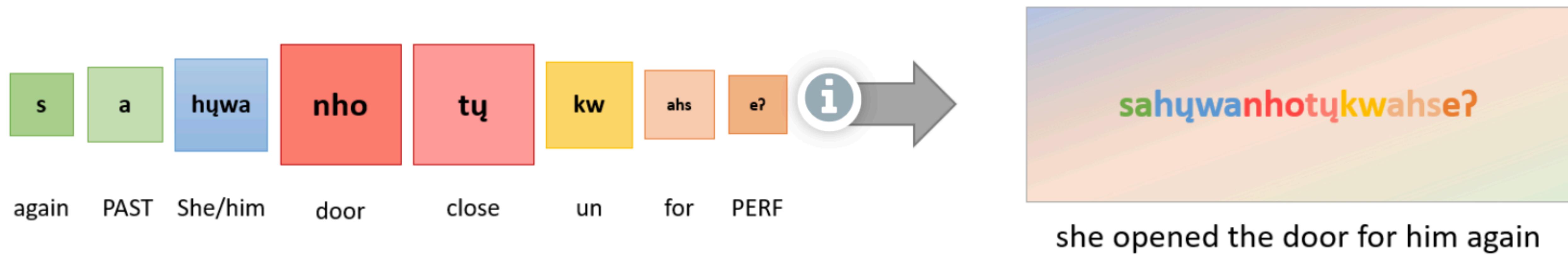
Linguistics

Sounds

*Pronunciation dialect map
of the Netherlands based
on phonetic edit-distance
(W. Heeringa Phd thesis,
2004)*



Words and morphemes



Mohawk or Kanien'kéha

Image credit: Dinesh Ramoo

Words and morphemes

- There is often ambiguity in how morphemes get combined
- Take the word in English: *delighted*
- Can you find two different ways to create this word using morphemes?

**BREAKTHROUGH:
A COMPUTER THAT
UNDERSTANDS YOU
LIKE YOUR MOTHER.**

Having to learn letter-perfect software languages can be frustrating to the average person trying to tap the power of a computer.

But practical thinkers at our McDonnell Douglas Computer Systems Company have created the first computer that accepts you as you are—human.

They emulated the two halves of the human brain with two-level software: One level with a dictionary of facts and a second level to interpret them. The resulting Natural Language processor understands everyday conversational English. So it knows what you mean, no matter how you express yourself. It also learns your idiosyncrasies, forgives your errors, and tells you how to find what you're looking for.

Now, virtually anyone who can read and write can use a computer.

We're creating breakthroughs not only in artificial intelligence but also in health care, space manufacturing and aircraft.

We're McDonnell Douglas.



**MCDONNELL
DOUGLAS**

© 1985 McDonnell Douglas Corporation

Syntax

(A computer that understands)
(you like your mother)

(A computer that understands you)
(like your mother)

1985 McDonnell Douglas Print Ad

<https://www.ebay.comitm/134242735420>

Syntax

Phenomenon	N	Acceptable Example	Unacceptable Example
ANAPHOR AGR.	2	<i>Many girls insulted <u>themselves</u>.</i>	<i>Many girls insulted <u>herself</u>.</i>
ARG. STRUCTURE	9	<i>Rose wasn't <u>disturbing</u> Mark.</i>	<i>Rose wasn't <u>boasting</u> Mark.</i>
BINDING	7	<i>Carlos said that Lori helped <u>him</u>.</i>	<i>Carlos said that Lori helped <u>himself</u>.</i>
CONTROL/RAISING	5	<i>There was <u>bound</u> to be a fish escaping.</i>	<i>There was <u>unable</u> to be a fish escaping.</i>
DET.-NOUN AGR.	8	<i>Rachelle had bought that <u>chair</u>.</i>	<i>Rachelle had bought that <u>chairs</u>.</i>
ELLIPSIS	2	<i>Anne's doctor cleans one <u>important</u> book and Stacey cleans a few.</i>	<i>Anne's doctor cleans one book and Stacey cleans a few <u>important</u>.</i>
FILLER-GAP	7	<i>Brett knew <u>what</u> many waiters find.</i>	<i>Brett knew <u>that</u> many waiters find.</i>
IRREGULAR FORMS	2	<i>Aaron <u>broke</u> the unicycle.</i>	<i>Aaron <u>broken</u> the unicycle.</i>
ISLAND EFFECTS	8	<i>Which <u>bikes</u> is John fixing?</i>	<i>Which is John fixing <u>bikes</u>?</i>
NPI LICENSING	7	<i>The truck has <u>clearly</u> tipped over.</i>	<i>The truck has <u>ever</u> tipped over.</i>
QUANTIFIERS	4	<i>No boy knew <u>fewer than</u> six guys.</i>	<i>No boy knew <u>at most</u> six guys.</i>
SUBJECT-VERB AGR.	6	<i>These casseroles <u>disgust</u> Kayla.</i>	<i>These casseroles <u>disgusts</u> Kayla.</i>

Table 1: Minimal pairs from each of the twelve linguistic phenomenon categories covered by BLiMP. Differences are underlined. N is the number of 1,000-example minimal pair paradigms within each broad category.

Semantics



(CONTINUED) You can't say that everyone's got a water buffalo
when everyone does not have a water buffalo!

Semantics



(CONTINUED) We're going to get nasty letters saying "Where's my water buffalo? Why don't I have a water buffalo?" And are you prepared to deal with that? I don't think so!

Not every viewer has a water buffalo (but some do)

Every single viewer does not have a water buffalo

Discourse

Plan

Premise

Premise: A new law grad returns home to start her career, but struggles with the broken justice system.

Setting

Setting: The story is set in a small town in the United States.

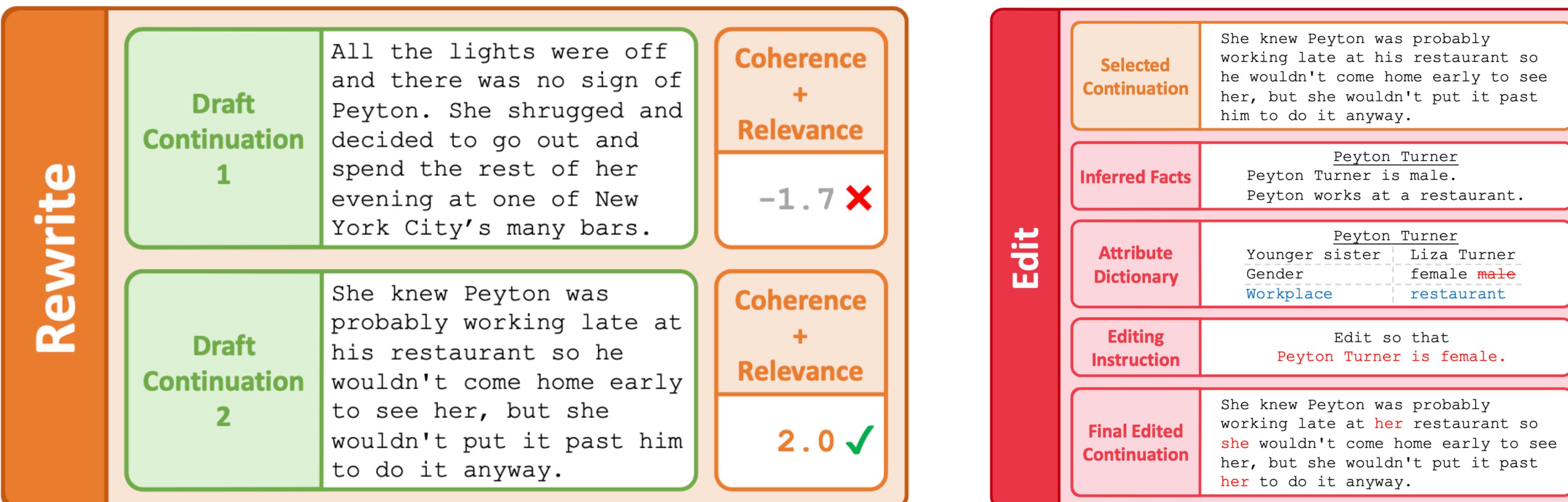
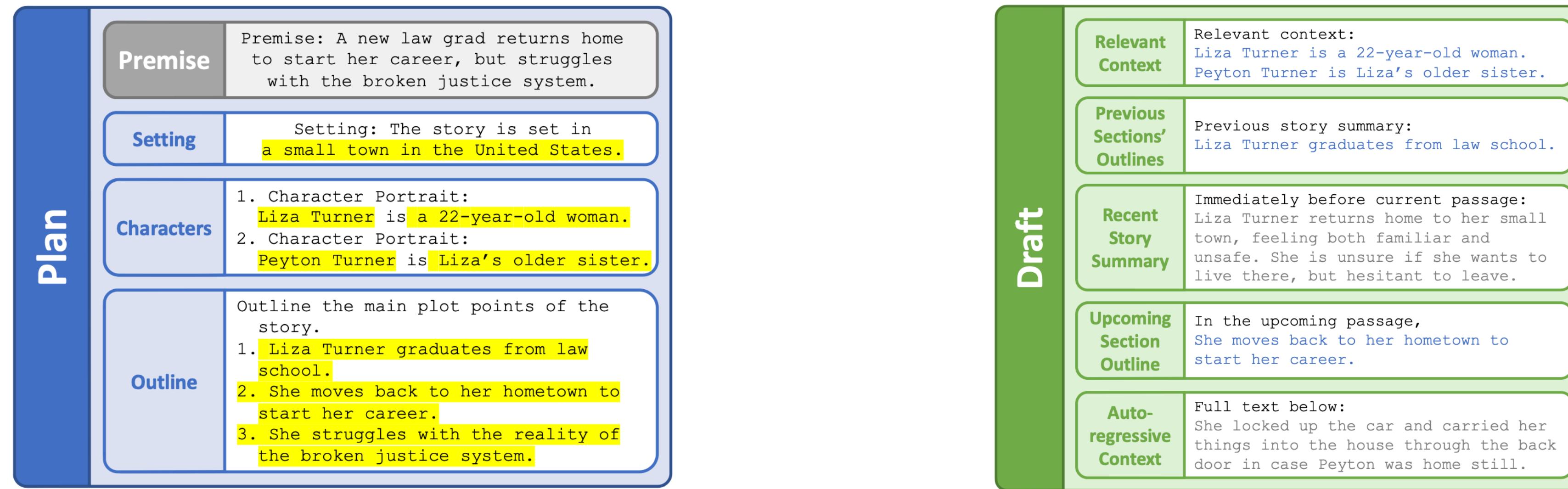
Characters

1. Character Portrait:
Liza Turner is a 22-year-old woman.
2. Character Portrait:
Peyton Turner is Liza's older sister.

Outline

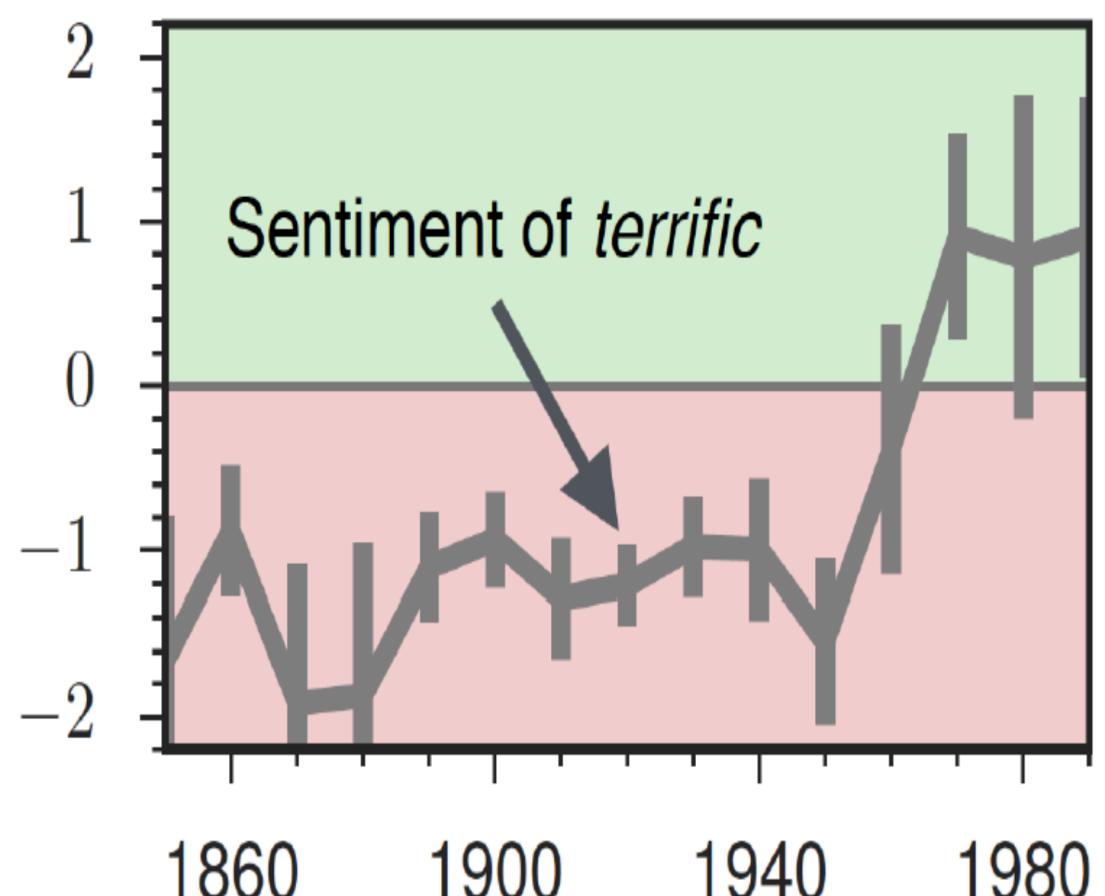
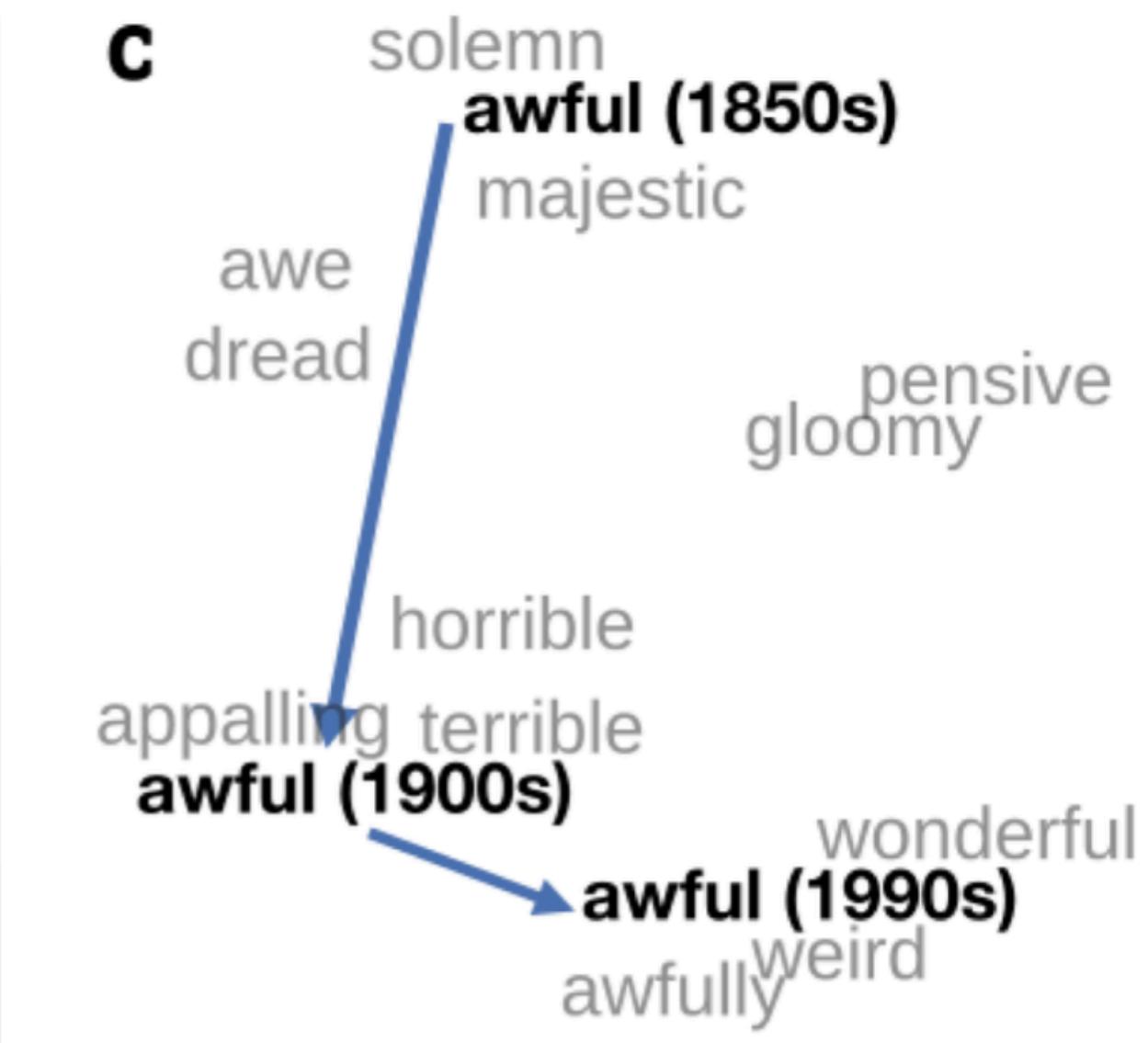
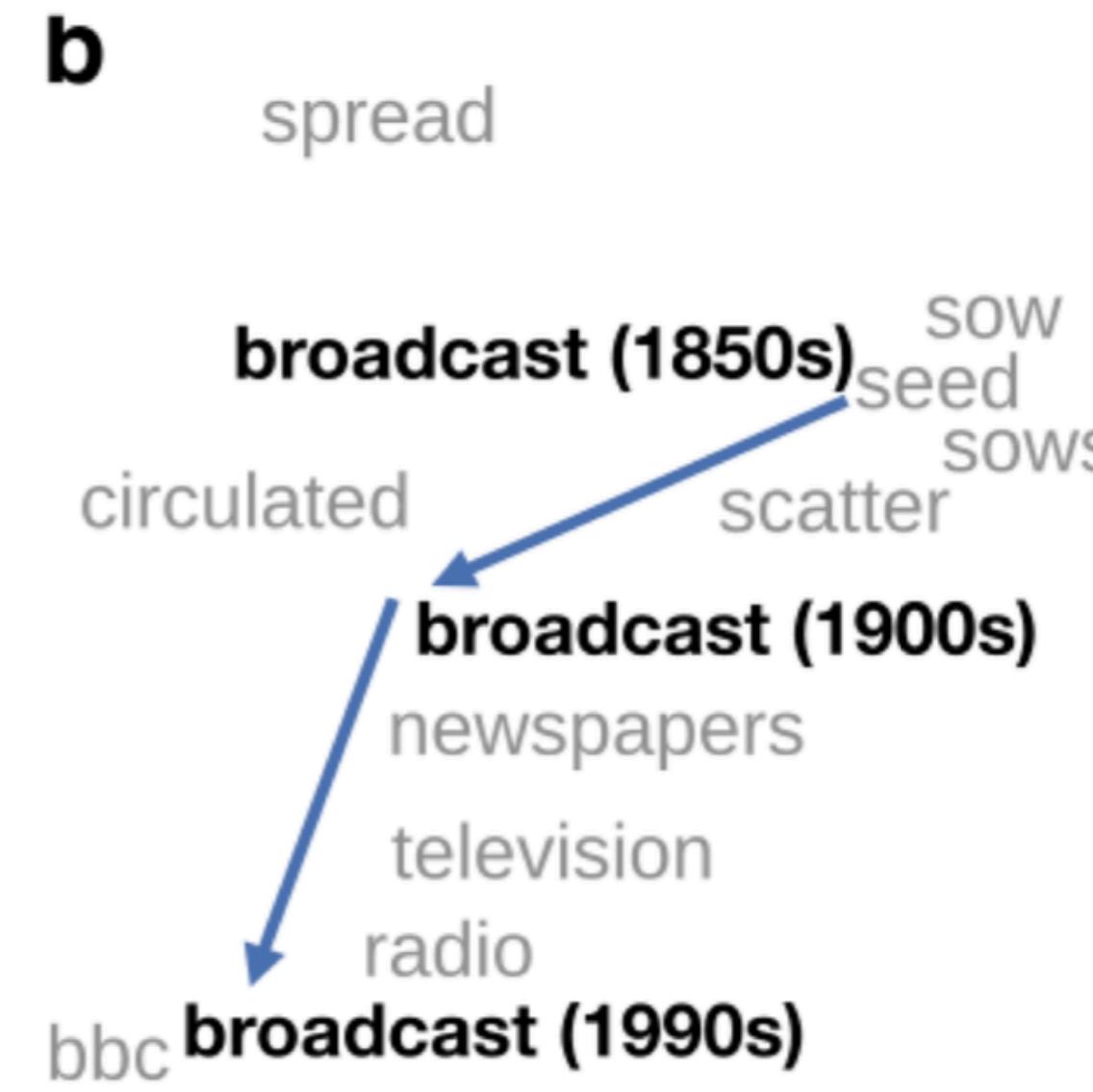
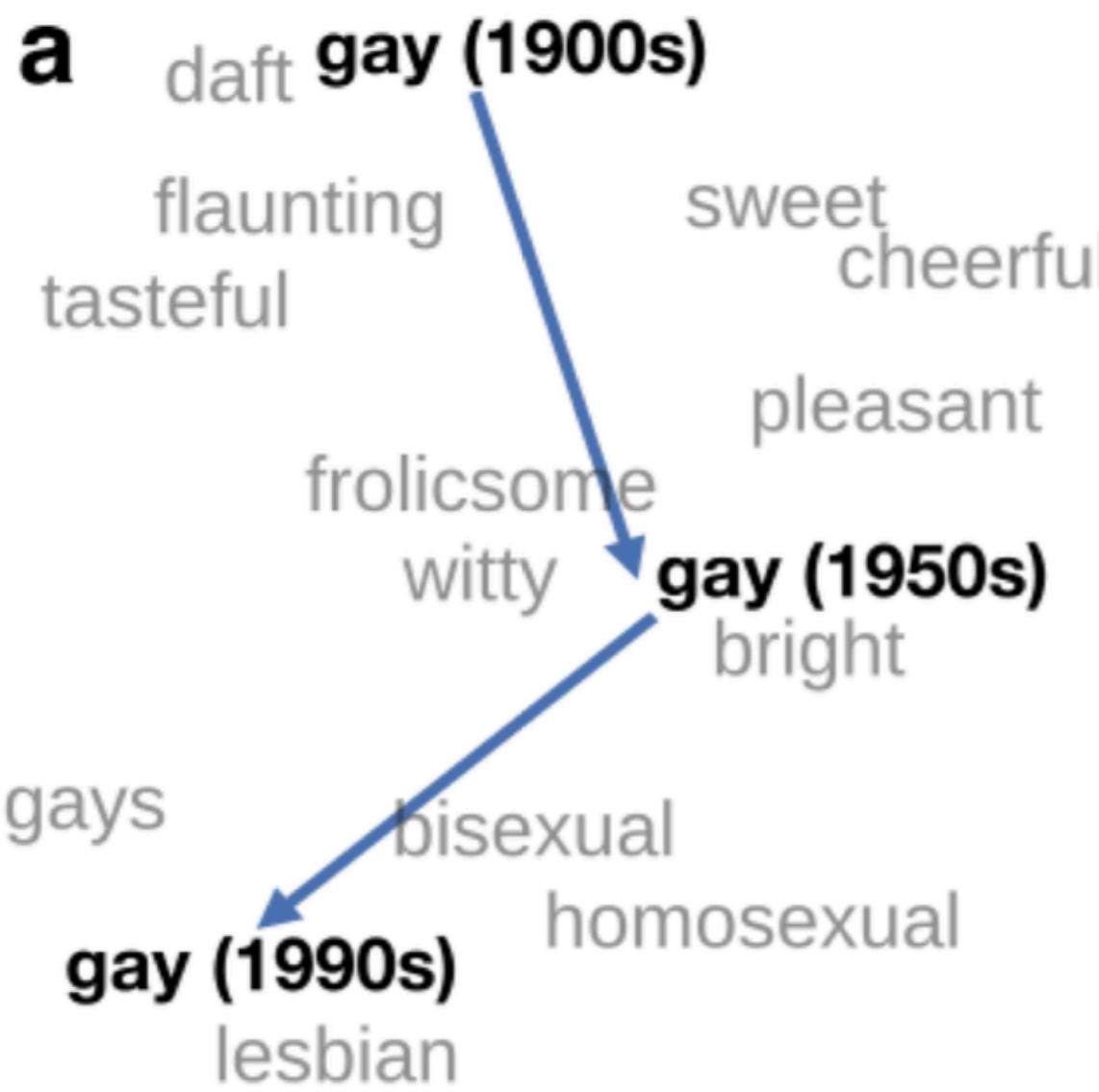
- Outline the main plot points of the story.
1. Liza Turner graduates from law school.
 2. She moves back to her hometown to start her career.
 3. She struggles with the reality of the broken justice system.

Discourse



Sociolinguistics

- Dialects
- Bias
- Toxicity
- Sentiment



Negative words change faster than positive words

Dan Jurafsky and James Martin.
Speech and Language Processing

Reasoning and Inference

That professor dresses even worse than Anoop

What can you infer from this?

Reasoning and Inference

Text	Judgments	Hypothesis
A man inspects the uniform of a figure in some East Asian country.	contradiction C C C C	The man is sleeping
An older and younger man smiling.	neutral N N E N N	Two men are smiling and laughing at the cats playing on the floor.
A black race car starts up in front of a crowd of people.	contradiction C C C C	A man is driving down a lonely road.
A soccer game with multiple males playing.	entailment E E E E E	Some men are playing a sport.
A smiling costumed woman is holding an umbrella.	neutral N N E C N	A happy woman in a fairy costume holds an umbrella.

- Material for some of these slides comes from ideas borrowed from Angel Chang, Danqi Chen, Karthik Narasimhan and others on the Slack group for teaching NLP.