How AI Cured Coronavirus and Delivered Universal Translation, and Other MT Myths and Magic

Martin Benjamin

18 November 2020 Translating and the Computer

ASLING TC42 online

Keynote Address

kamusi.org

kamusi is Swahili for *dictionary*

kamusi.org

Goal: A complete matrix of human expression across time and space

- As a knowledge resource
- As a data resource
- As a basis for any-to-any translation

kamusi.org

In service since 1994 - originally at **Yale Council on African Studies**
International NGO since 2009

- Registered non-profit in 🇺🇸 and 🇨🇭

Academic Home since 2013:
EPFL - Swiss Federal Institute of Technology in Lausanne
First at **LSIR** - Distributed Systems Information Laboratory
Now at the **Swiss EdTech Collider**

kamusi.org

White House Big Data Initiative (2013):

Launch Partner for Building the Data Innovation Ecosystem

Networking and Information Technology R&D Program

Office of Science and Technology Policy

ACALAN (Intergovernmental language agency for 55 member states of the African Union):

Platform for African Language Empowerment development partner

AFRICAN UNION

AFRICAN ACADEMY OF LANGUAGES

Artificial Intelligence (AI) & Machine Translation (MT) on the Road toward Universal Translation

1. AI facts and fantasies
2. Myths about AI and MT
3. Realistic fantasies about computation and translation

kamusi.org

Chapter 1: AI Facts and Fantasies

1. What is AI?
2. AI and COVID-19
3. AI and the weather
4. AI and online dating
5. AI and language

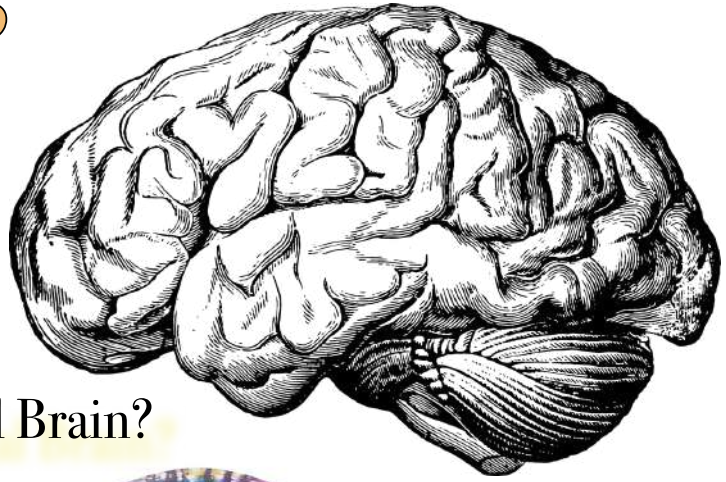
AI Facts and Fantasies: What is Artificial Intelligence?

1. What is AI?
2. AI and COVID-19
3. AI and the weather
4. AI and online dating
5. AI and language

kamusi.org

AI Facts and Fantasies: What is AI?

Artificial Muscle?



Artificial Brain?

kamusi.org

AI Facts and Fantasies: What is Artificial Intelligence?

Artificial Muscle

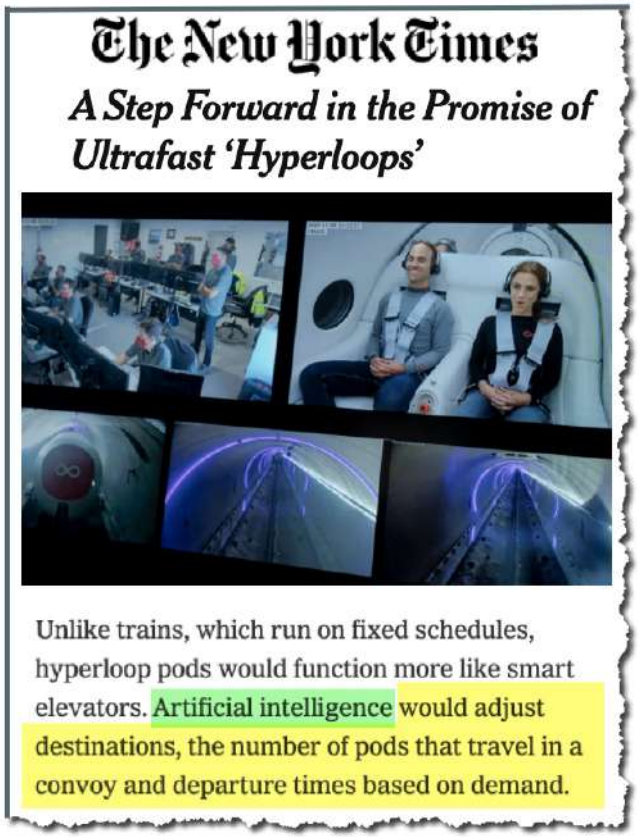
- Brute force calculations
- Faster than a human
- Instructed by a human

Example:

Calculate π to a million digits

3.14159265358979323846264338327950288419716939937510582097494459
23078164062862089986280348253421170679821480865132823066470938 4
46095505822317253594081284811174502841027019385211055596446229 4
89549303819644288109756659334461284756482337867831652712019091 4
56485669234603486104543266482133936072602491412737245870066063 1
55881748815209209628292540917153643678925903600113305305488204 6
65213841469519415116094330572703657595919530921861173819326117 9
31051185480744623799627495673518857527248912279381830119491298 3
36733624406566430860213949463952247371907021798609437027705392 1
71762931767523846748184676694051320005681271452635608277857713 4
27577896091736371787214684409012249534301465495853710507922796 8
92589235420199561121290219608640344181598136297747713099605187 0
72113499999983729780499510597317328160963185950244594553469083 0
26425223082533446850352619311881710100031378387528865875332083 8
14206171776691473035982534904287554687311595628638823537875937 5

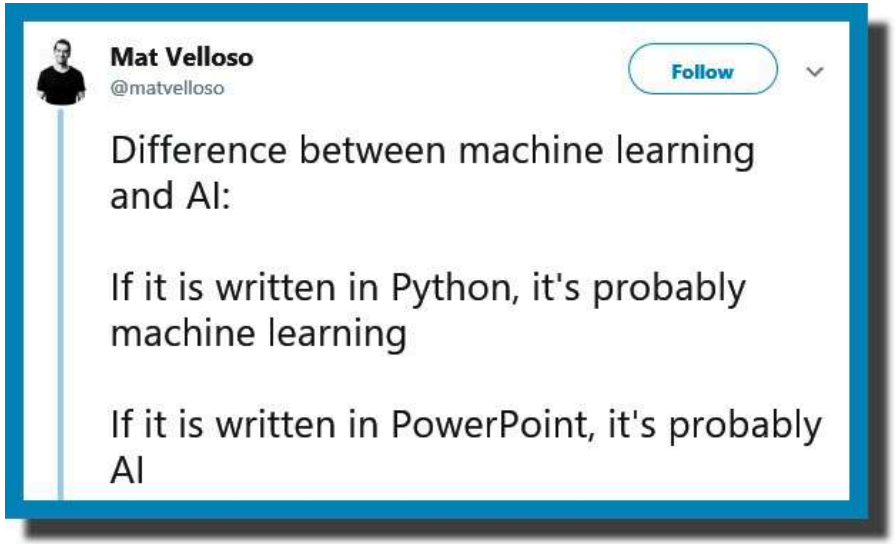
AI Facts and Fantasies: What is AI?



The New York Times

A Step Forward in the Promise of Ultrafast 'Hyperloops'

Unlike trains, which run on fixed schedules, hyperloop pods would function more like smart elevators. Artificial intelligence would adjust destinations, the number of pods that travel in a convoy and departure times based on demand.



Mat Velloso
@matvelloso
Follow

Difference between machine learning and AI:

If it is written in Python, it's probably machine learning

If it is written in PowerPoint, it's probably AI

kamusi.org

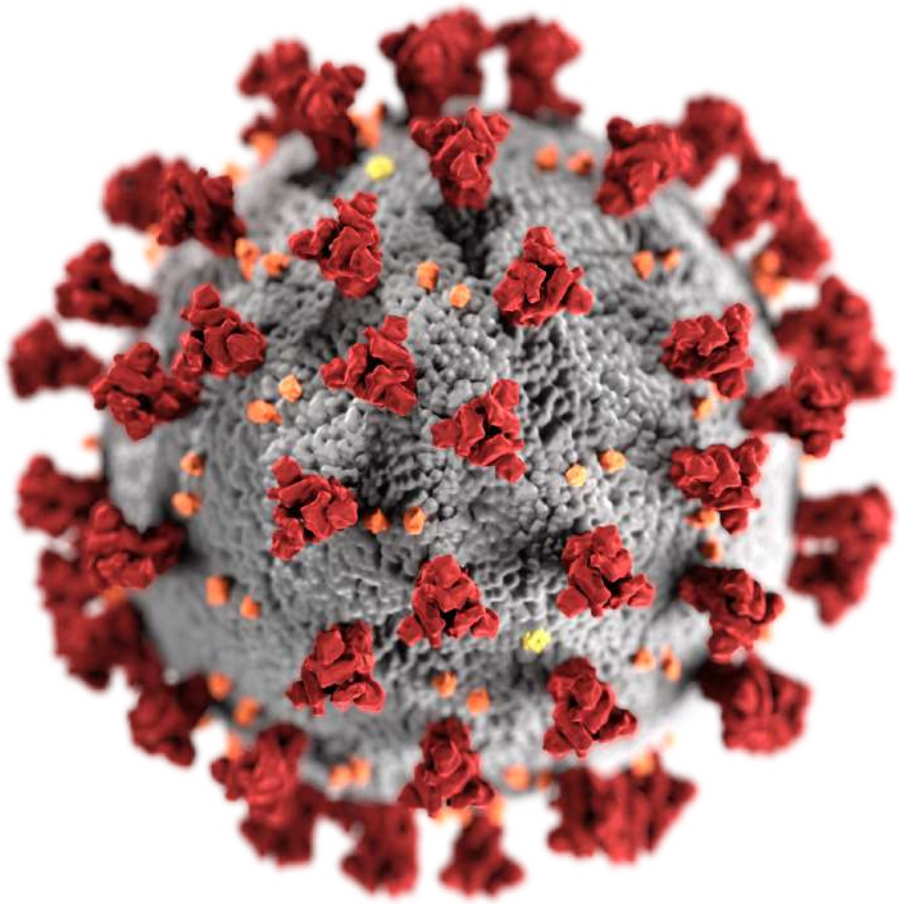
AI Facts and Fantasies: COVID-19

1. What is AI?
2. AI and COVID-19
3. AI and the weather
4. AI and online dating
5. AI and language

kamusi.org

AI Facts and Fantasies: COVID-19

- SPOILER: AI has not cured COVID-19
- What has AI done for COVID-19?



kamusi.org

AI Facts and Fantasies: COVID-19

- AI has not cured COVID-19
- What has AI done for COVID-19?

A Supercomputer Analyzed Covid-19 — and an Interesting New Theory Has Emerged

A closer look at the Bradykinin hypothesis

Artificial-intelligence tools aim to tame the coronavirus literature

Developers hope that tools for processing natural language will help biomedical researchers and clinicians to find the COVID-19 papers that they need.

Using AI to fast and effectively diagnose COVID-19 in hospitals

The European Commission will invest in the use of Artificial Intelligence to speed up the diagnosis of COVID-19 and improve future treatment of patients. A software developed to assist the work of medical staff by analysing images of pulmonary infections is introduced in 10 hospitals across Europe.

Artificial intelligence model detects asymptomatic Covid-19 infections through cellphone-recorded coughs

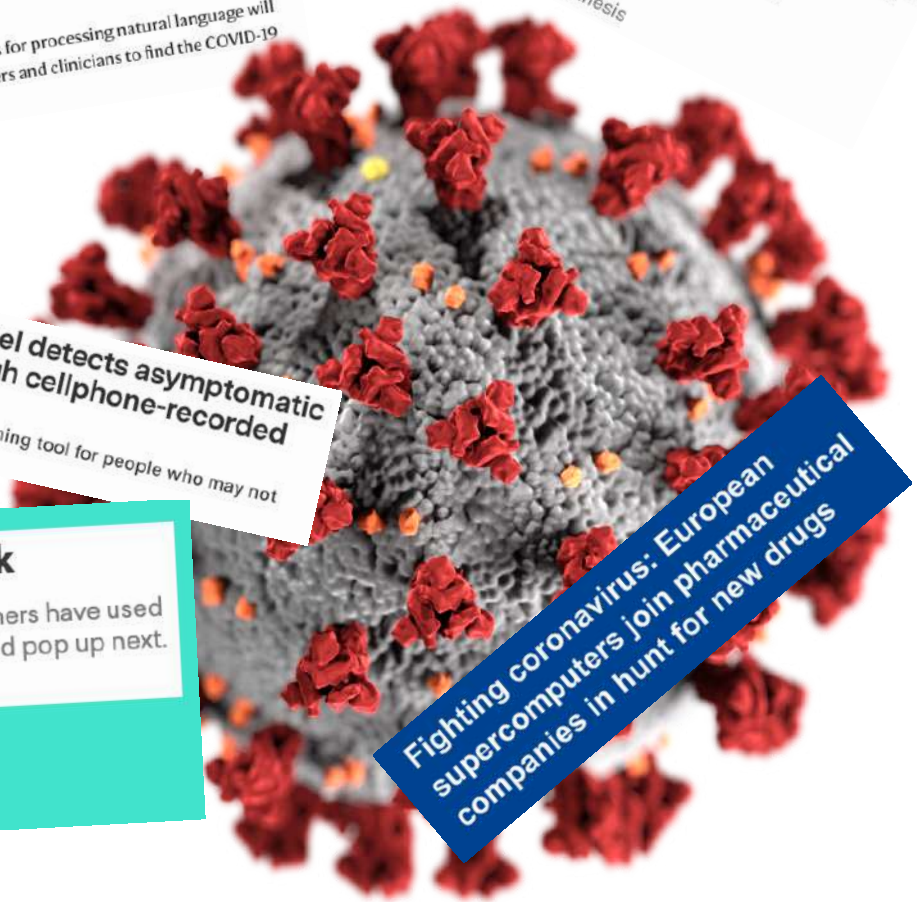
Results might provide a convenient screening tool for people who may not suspect they are infected.

How AI is battling the coronavirus outbreak

AI helped spot an early warning about the outbreak, and researchers have used flight traveler data to figure out where the novel coronavirus could pop up next.

By Rebecca Heilweil | Jan 28, 2020, 1:50pm EST

28 JANUARY, 2020

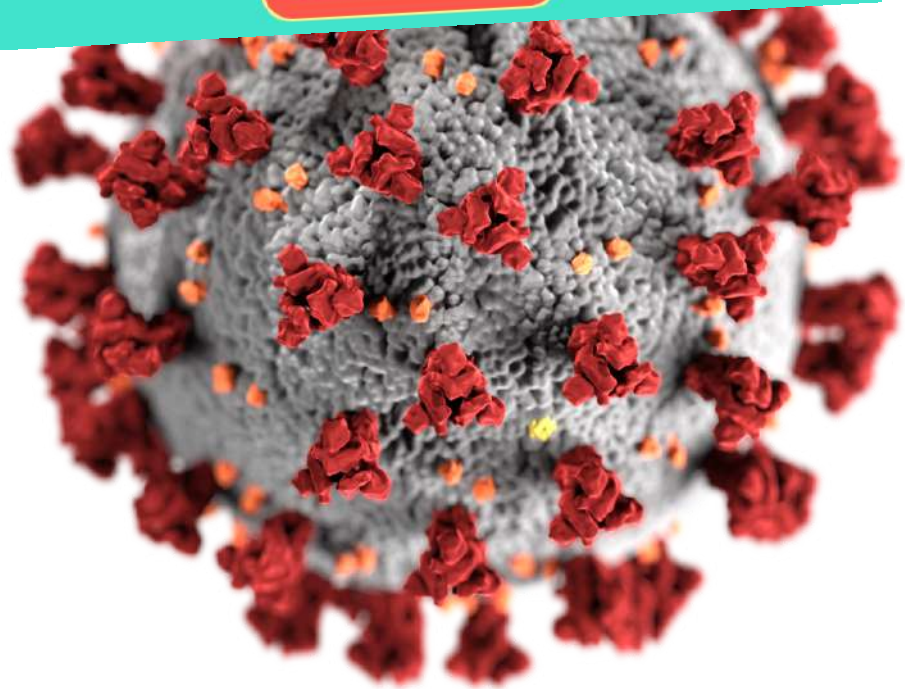


Fighting coronavirus: European supercomputers join pharmaceutical companies in hunt for new drugs

kamusi.org

AI Facts and Fantasies: COVID-19

- AI has not cured COVID-19
- What has AI done for COVID-19?



"In other words, our new AI overlords might actually help us survive the next plague."

- Using NLP to track 100,000 articles about 100 diseases in 65 languages
- Travel itineraries and flight paths
- Researching new drugs
- Detecting disease

Vox: https://www.vox.com/recode/2020/1/28/21110902/artificial-intelligence-ai-coronavirus-wuhan

kamusi.org

AI Facts and Fantasies: The Weather

1. What is AI?
2. AI and COVID-19
3. AI and the weather
4. AI and online dating
5. AI and language

kamusi.org

AI Facts and Fantasies: The Weather

Data to analyze for many locations:

• Temperature • Humidity • Pressure • Windspeed • Wind direction
• Cloud cover • etc

Patterns to discover:

- Does the flap of a butterfly's wings in Brazil set off a tornado in Texas? (Edward Lorenz)
- The relationship between certain parameters in Place A, and subsequent weather in Place B

New Paths:

- If certain parameters are changed in a model for Place A, what are the anticipated outcomes in Place B?

Machine Learning:

- Do future weather events occur as anticipated in the model?

- Analyzes data
- Discovers patterns
- Follows new paths based on those patterns

Machine Learning *if*

- a desired outcome is achieved (eg, win a game)

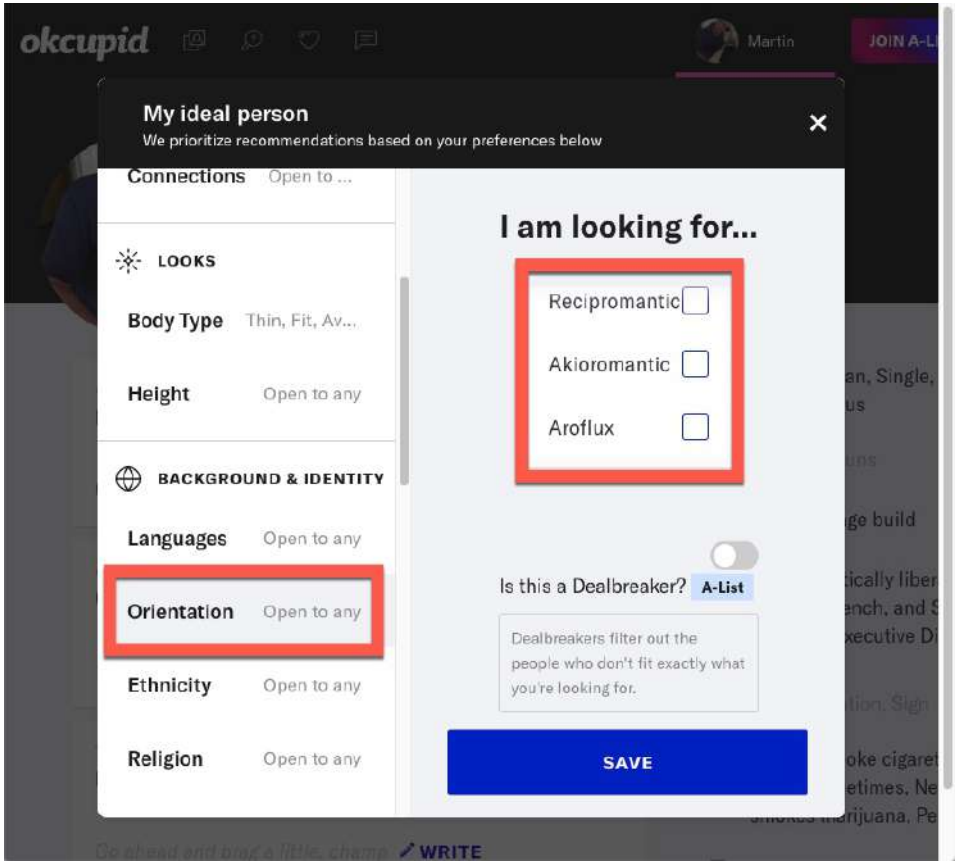
AI Facts and Fantasies: Online Dating

1. What is AI?
2. AI and COVID-19
3. AI and the weather
4. AI and online dating
5. AI and language

kamusi.org

AI Facts and Fantasies: Online Dating

Case Study: OkCupid



- ~~Analyzes data~~
- ~~Discovers patterns~~
- ~~Follows new paths based on those patterns~~

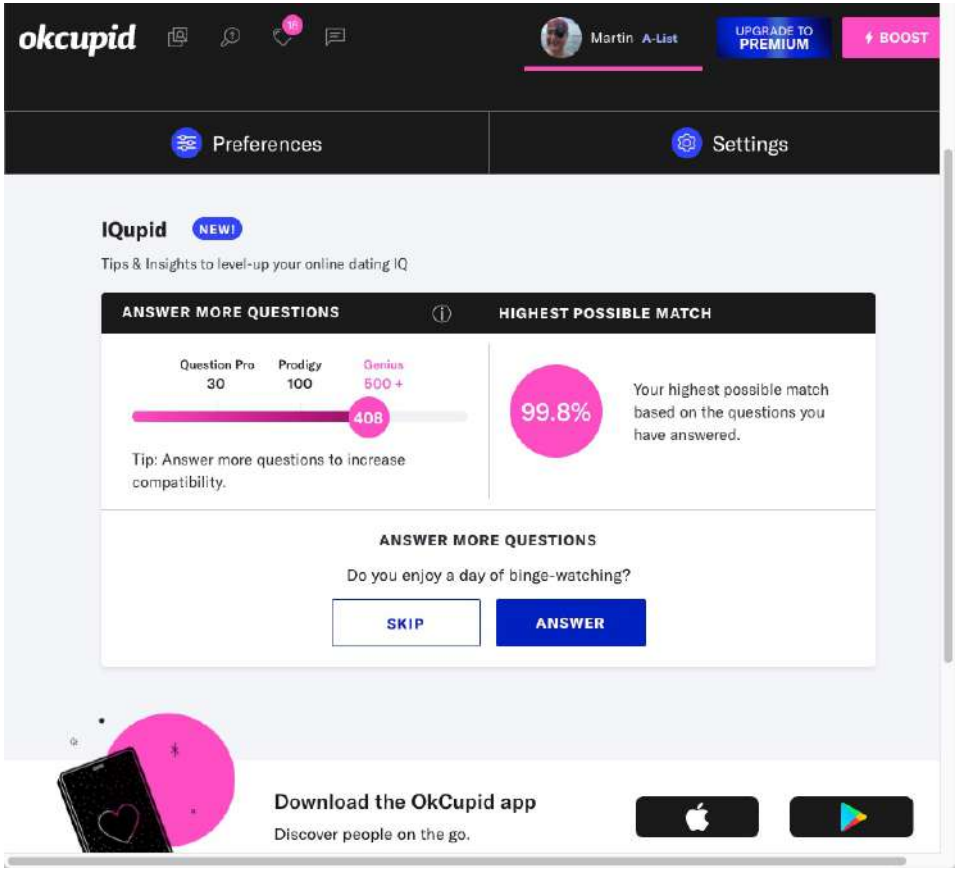
Machine Learning *if*

- ~~a desired outcome is achieved (eg, win a game)~~

kamusi.org

AI Facts and Fantasies: Online Dating

Case Study: OkCupid



- ~~Analyzes data~~
- ~~Discovers patterns~~
- ~~Follows new paths based on those patterns~~

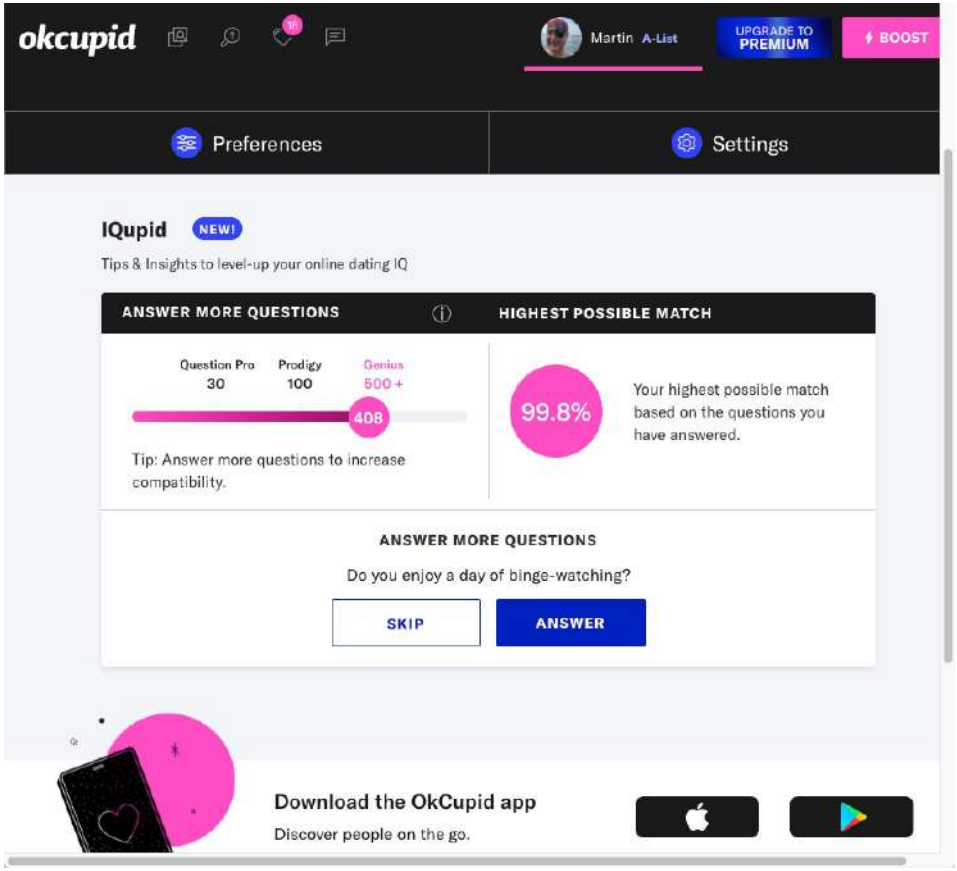
Machine Learning *if*

- ~~a desired outcome is achieved (eg, win a game)~~

kamusi.org

AI Facts and Fantasies: Online Dating

Case Study: OkCupid



- ~~Analyzes data~~
- ~~Discovers patterns~~
- ~~Follows new paths based on those patterns~~

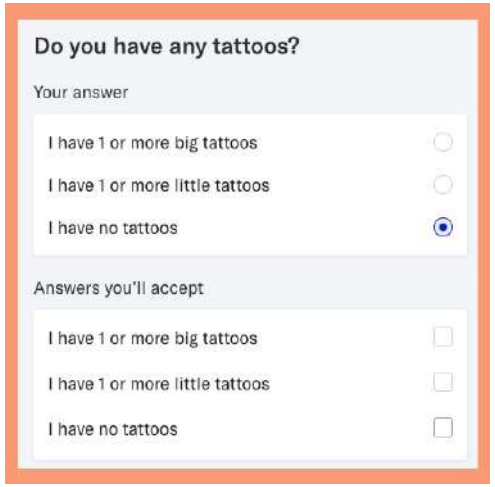
Machine Learning *if*

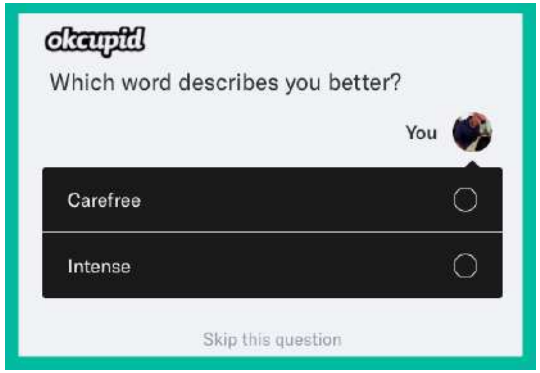
- ~~a desired outcome is achieved (eg, win a game)~~

kamusi.org

AI Facts and Fantasies: Online Dating

Case Study: OkCupid





- ~~Analyzes data~~
- ~~Discovers patterns~~
- ~~Follows new paths based on those patterns~~

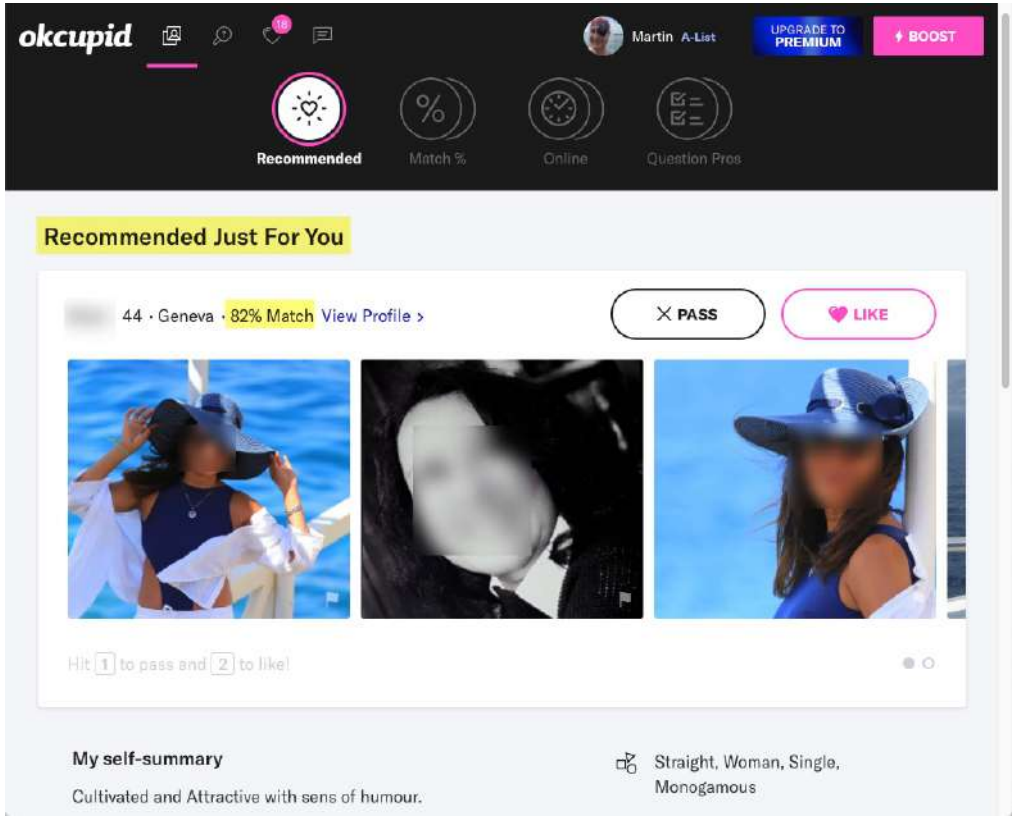
Machine Learning *if*

- ~~a desired outcome is achieved (eg, win a game)~~

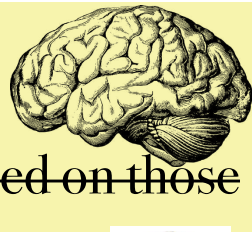
kamusi.org

AI Facts and Fantasies: Online Dating

Case Study: OkCupid



- ~~Analyzes data~~
- ~~Discovers patterns~~
- ~~Follows new paths based on those patterns~~



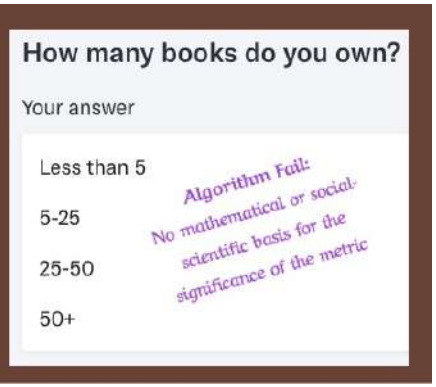
Machine Learning *if*

- ~~a desired outcome is achieved (eg, win a game)~~

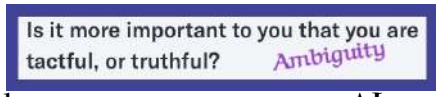
kamusi.org

AI Facts and Fantasies: Online Dating

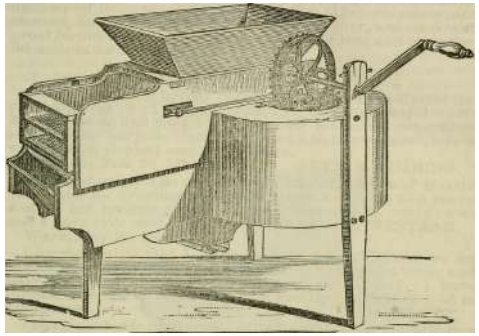
Case Study: OkCupid

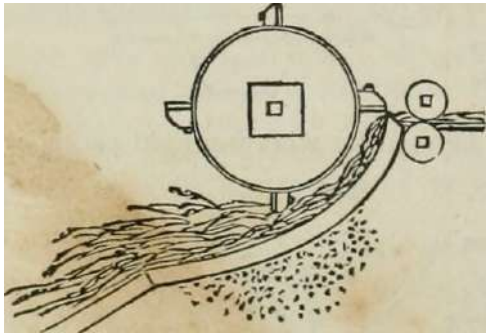


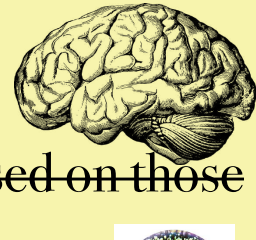
- Arbitrary metrics – questions have equal weight, but unequal (and unmeasurable) significance. (Eg, "Is pizza a top 5 food?" = "Would you date a single parent?")
- Ambiguity



- Machine is discovering equal answers, not patterns – AI would analyze the similarities among people one "likes" to find unseen preferences
- All data, no chemistry – and no way to predict or evaluate chemistry
- No outcome for a machine learning opportunity is either proposed or evaluated: • mutual "likes? • a chat? • a date? • a relationship? • marriage? • kids?







- ~~Analyzes data~~
- ~~Discovers patterns~~
- ~~Follows new paths based on those patterns~~

Machine Learning *if*

- ~~a desired outcome is achieved (eg, win a game)~~

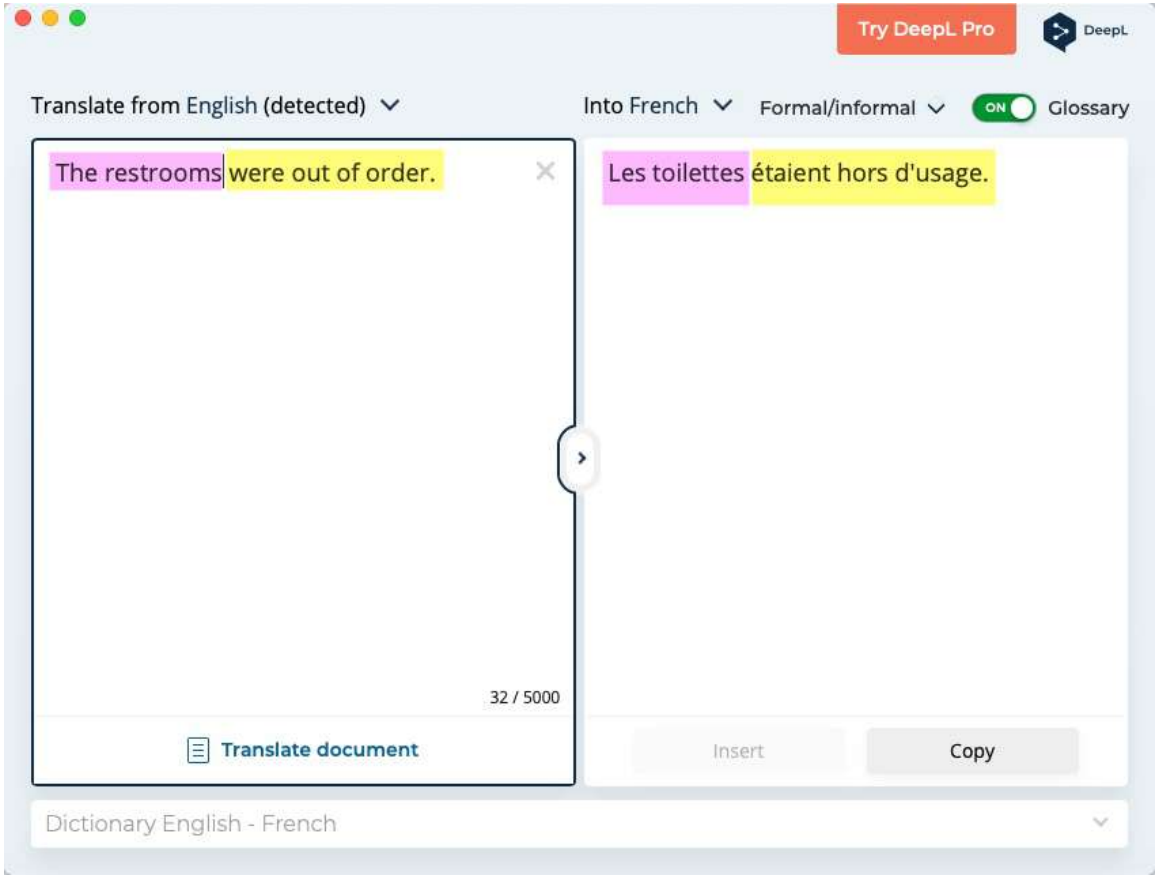
AI Facts and Fantasies: Language

1. What is AI?
2. AI and COVID-19
3. AI and the weather
4. AI and online dating
5. AI and language

kamusi.org

AI Facts and Fantasies: Language

Case Study: DeepL



- Analyzes data
- Discovers patterns
- Follows new paths based on those patterns

Machine Learning *if*

- a desired outcome is achieved (eg, win a game)

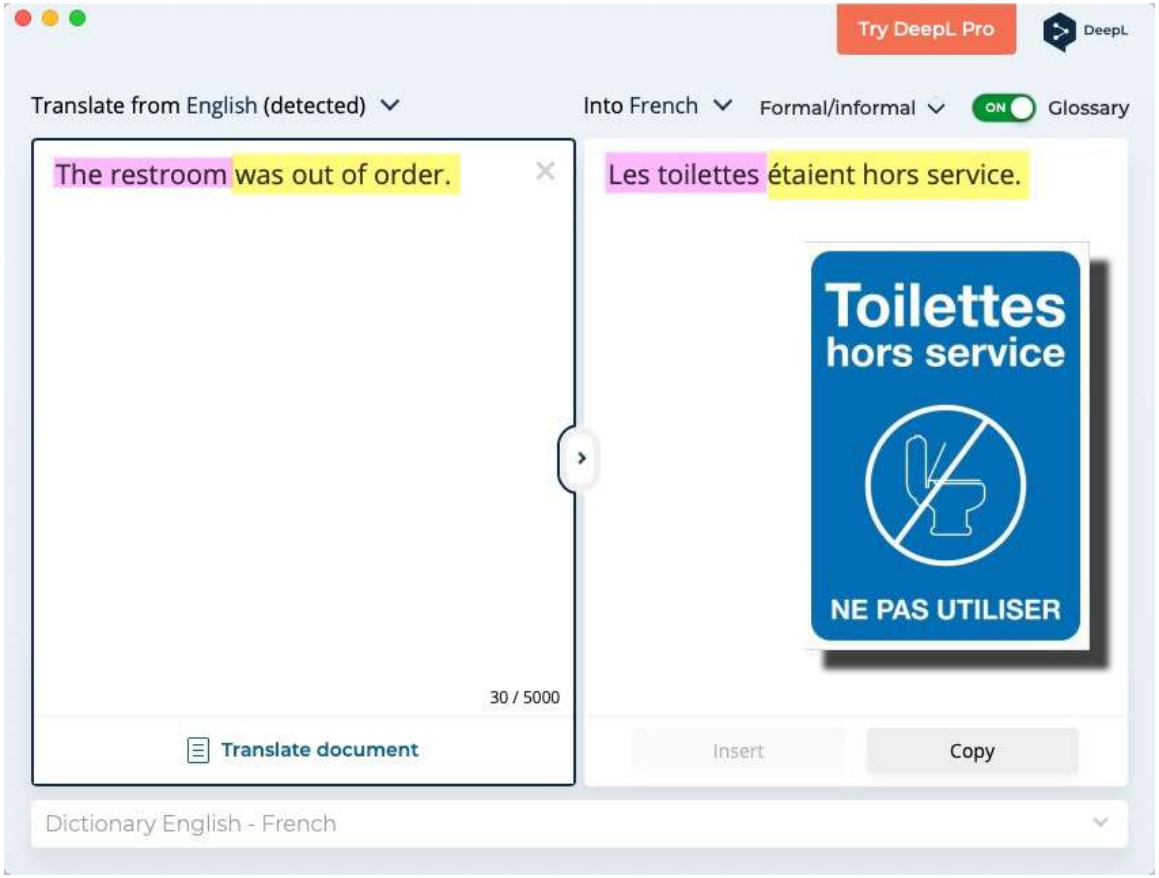
-- or --

- a human confirms the outcome (eg, that photo is indeed a cat)

kamusi.org

AI Facts and Fantasies: Language

Case Study: DeepL



- Analyzes data
- Discovers patterns
- Follows new paths based on those patterns

Machine Learning *if*

- a desired outcome is achieved (eg, win a game)

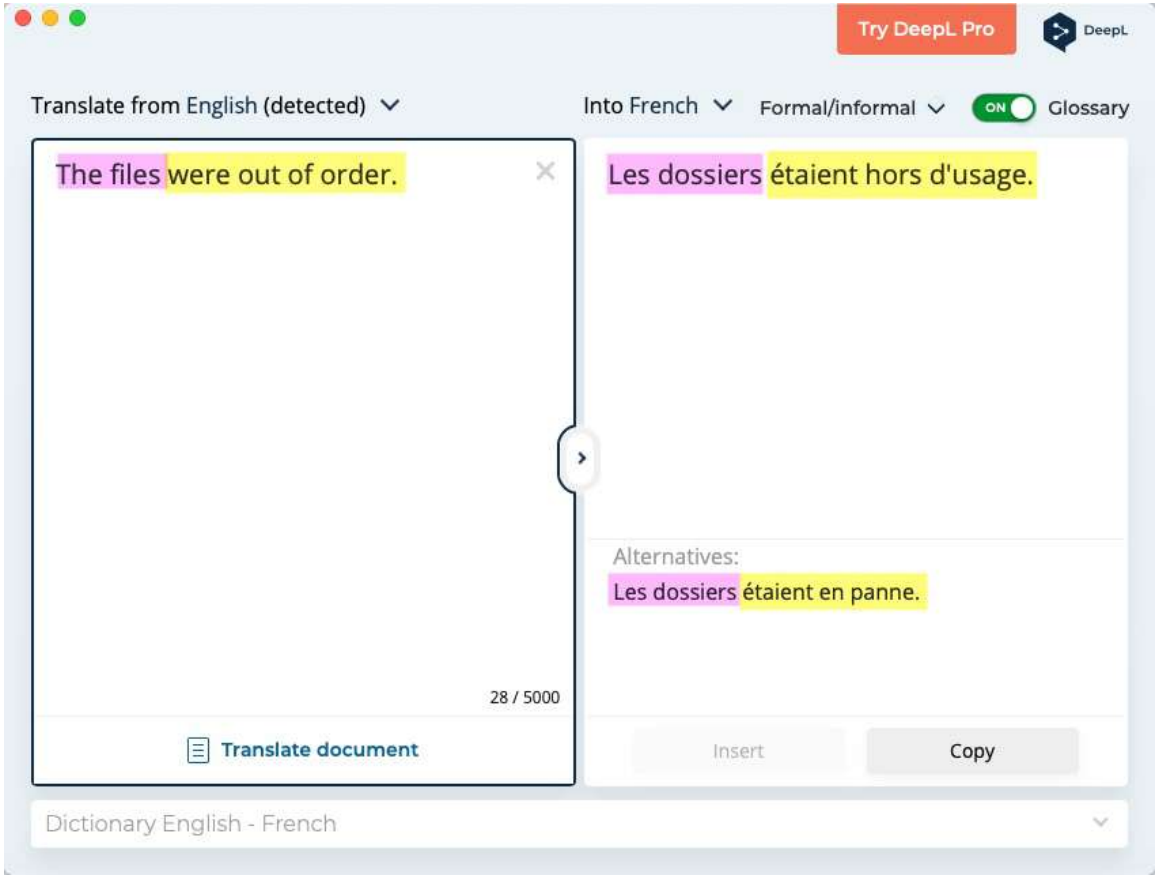
-- or --

- a human confirms the outcome (eg, that photo is indeed a cat)

kamusi.org

AI Facts and Fantasies: Language

Case Study: DeepL



- Analyzes data
- Discovers patterns
- Follows new paths based on those patterns

Machine Learning *if*

- a desired outcome is achieved (eg, win a game)

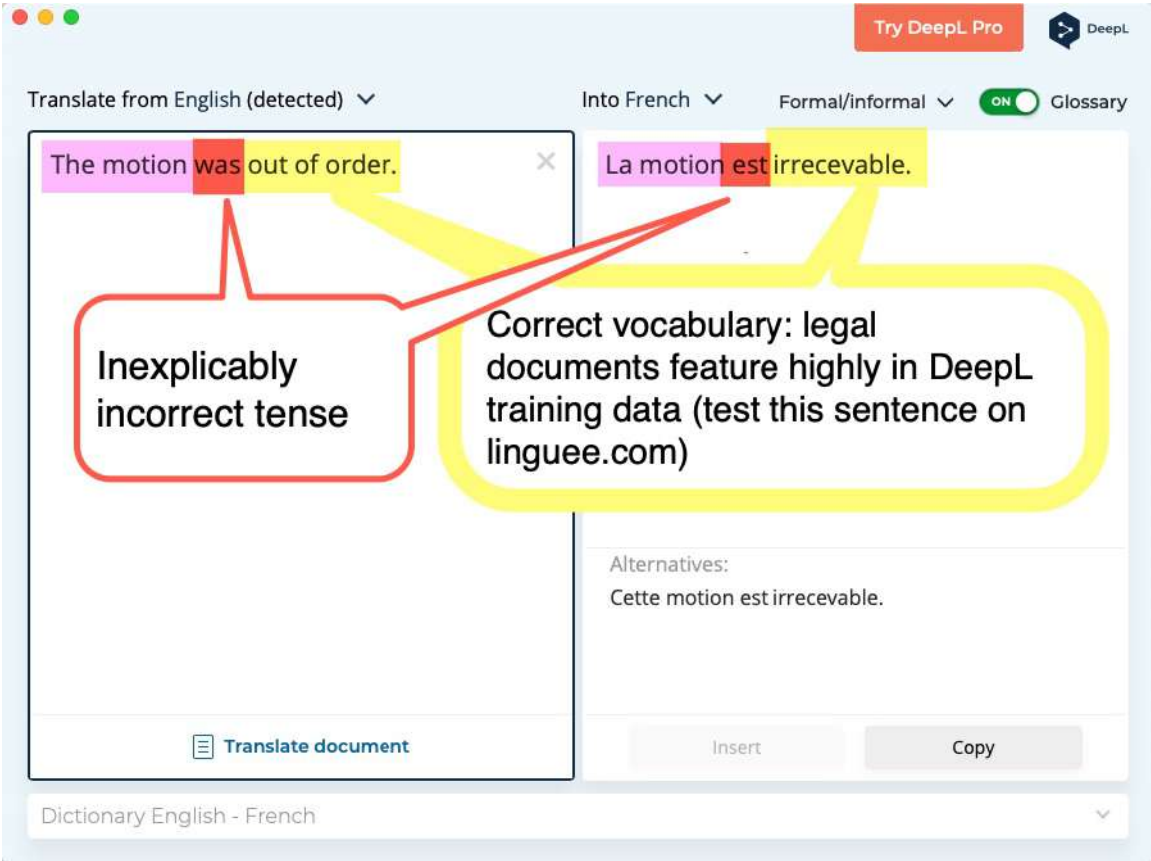
-- or --

- a human confirms the outcome (eg, that photo is indeed a cat)

kamusi.org

AI Facts and Fantasies: Language

Case Study: DeepL



- Analyzes data
- Discovers patterns
- Follows new paths based on those patterns

Machine Learning *if*

- a desired outcome is achieved (eg, win a game)

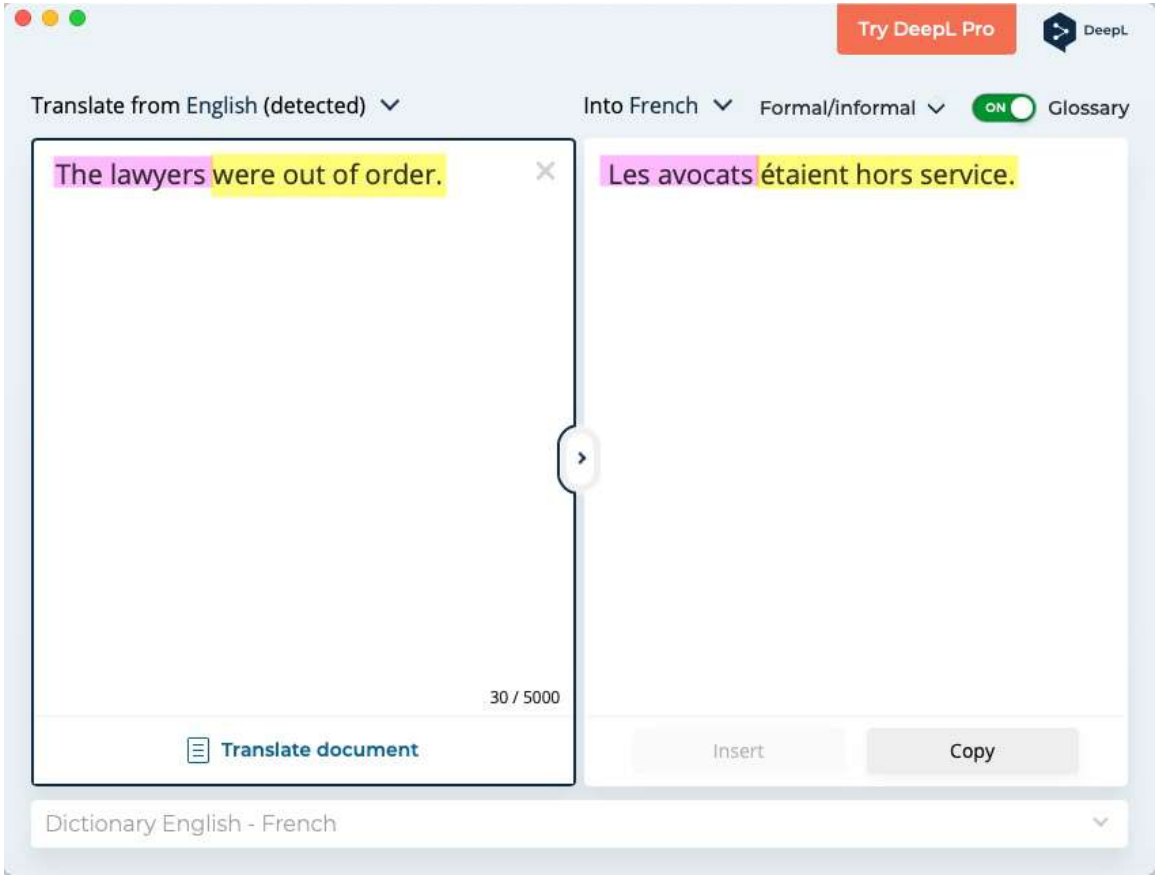
-- or --

- a human confirms the outcome (eg, that photo is indeed a cat)

kamusi.org

AI Facts and Fantasies: Language

Case Study: DeepL



- Analyzes data
- Discovers patterns
- Follows new paths based on those patterns

Machine Learning *if*

- a desired outcome is achieved (eg, win a game)

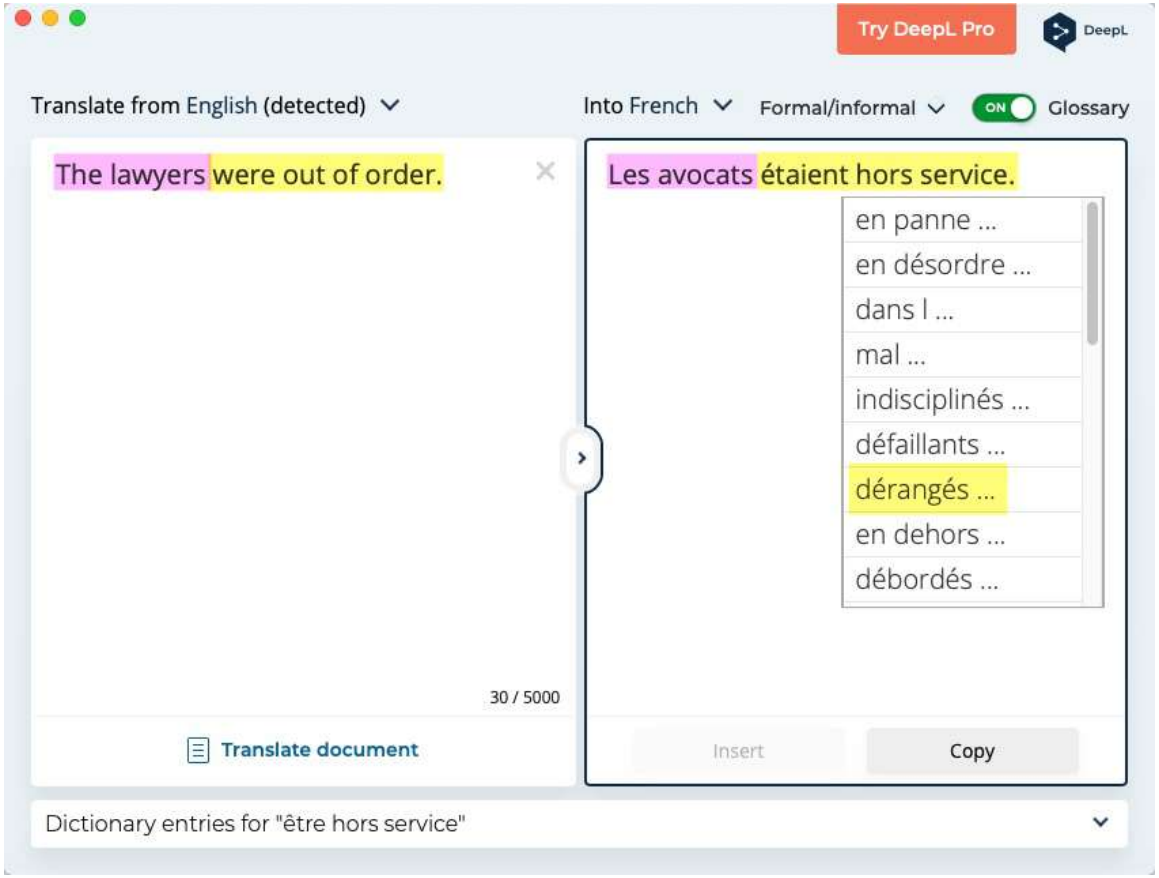
-- or --

- a human confirms the outcome (eg, that photo is indeed a cat)

kamusi.org

AI Facts and Fantasies: Language

Case Study: DeepL



- Analyzes data
- Discovers patterns
- Follows new paths based on those patterns

Machine Learning *if*

- a desired outcome is achieved (eg, win a game)

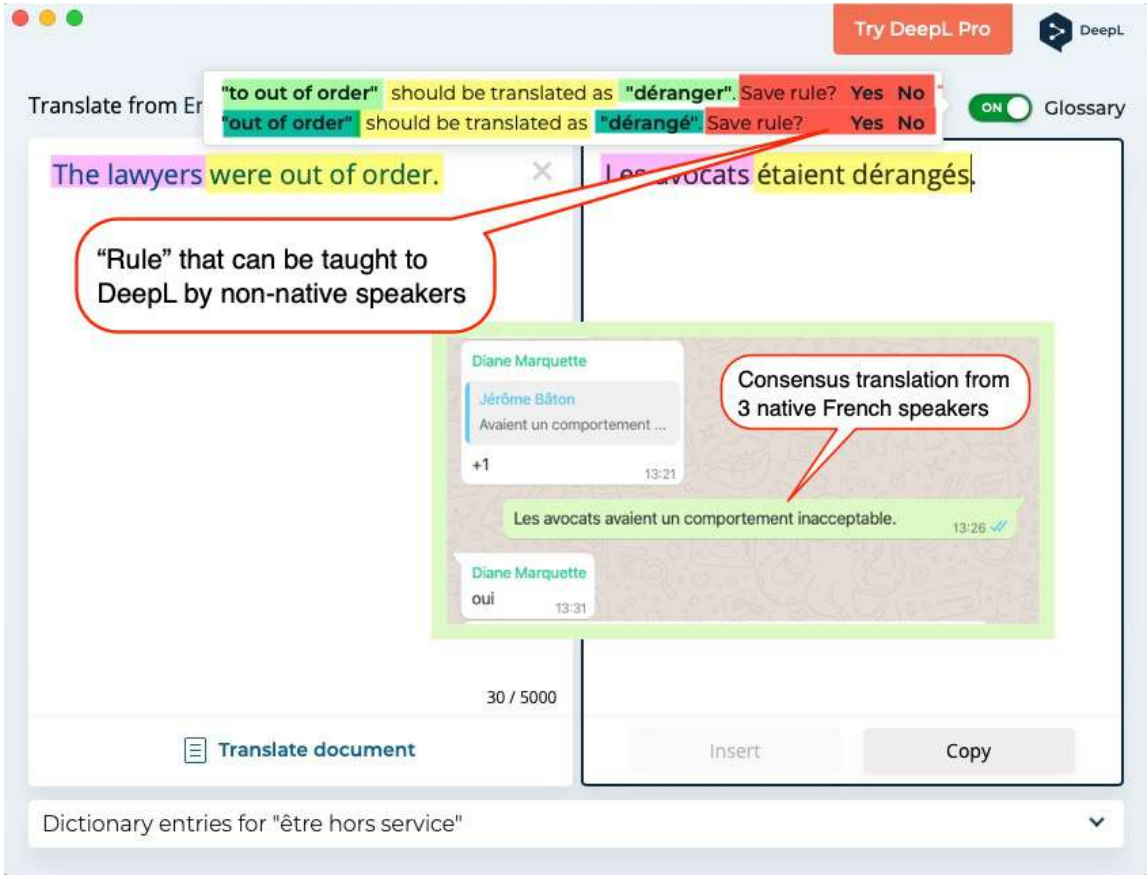
-- or --

- a human confirms the outcome (eg, that photo is indeed a cat)

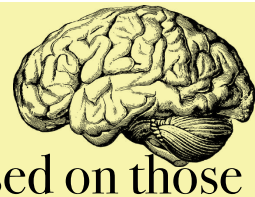
kamusi.org

AI Facts and Fantasies: Language

Case Study: DeepL



- Analyzes data
- Discovers patterns
- Follows new paths based on those patterns



Machine Learning *if*

- ~~a desired outcome is achieved~~ (eg, win a game)

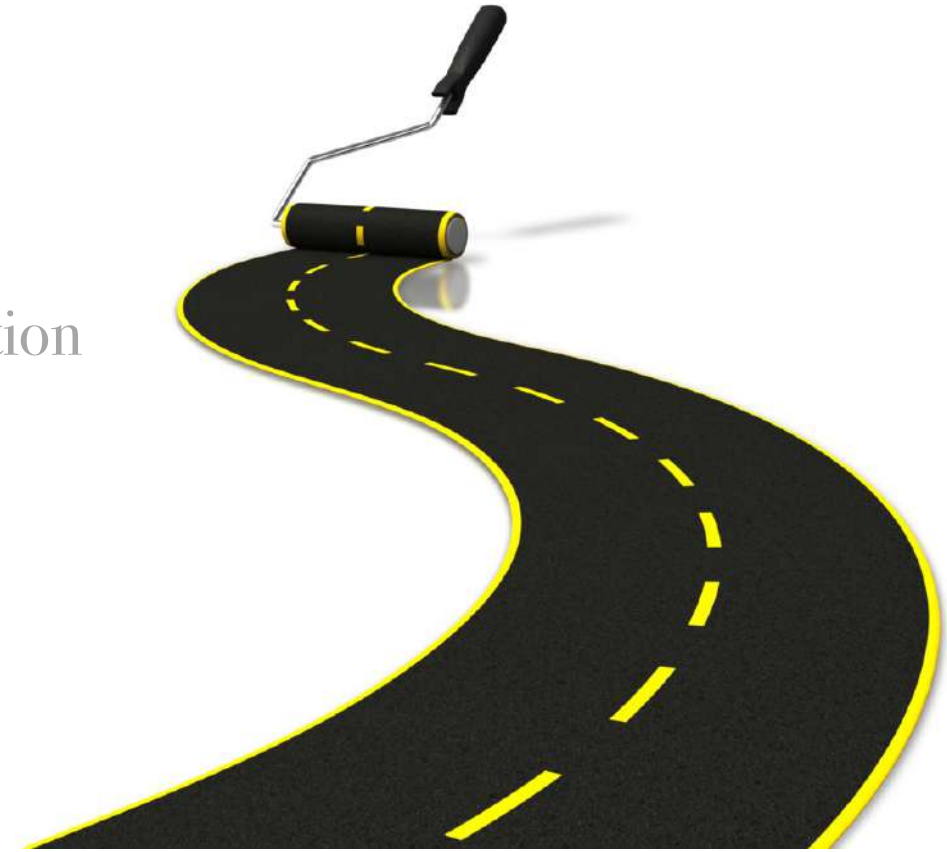
-- or --

- ~~a human confirms the outcome~~ (eg, that photo is indeed a cat)

kamusi.org

Artificial Intelligence (AI) & Machine Translation (MT) on the Road toward Universal Translation

1. AI facts and fantasies
2. Myths about AI and MT
3. Realistic fantasies about computation and translation



kamusi.org

Chapter 2: Myths about AI and MT

1. We have the data
2. We have the methods
3. Other biases and blind spots

kamusi.org

Myths about AI and MT

1. We have the data
2. We have the methods
3. Other biases and blind spots

kamusi.org

Myths about AI and MT: We have the data

Data

Words that have been **digitized** in a way that can be used within computer processes

- English
- Other high-investment languages
- Other high-population languages
- The other 98% of languages

kamusi.org

Myths about AI and MT: We have the data

Words that have been **digitized** in a way that can be used within computer processes

- 70 years of R & D
- All English words have some digital existence
- Many data sets include party terms (aka "multiword expressions")
- Natural Language Processing (NLP) can perform many analytical marvels (deduce parts of speech, verb conjugations, compose syntactically correct sentences…)
- Meaning and Shape are poorly associated (stay tuned!)

- We cannot interpret the data across systems (interoperability)
- We do not have reliable translations of English terms to most other languages (polysemy, party terms)

- Translation to and from English is only a small part of global translation needs

- English
- Other high-investment languages
- Other high-population languages
- The other 98% of languages

kamusi.org

Myths about AI and MT: We have the data

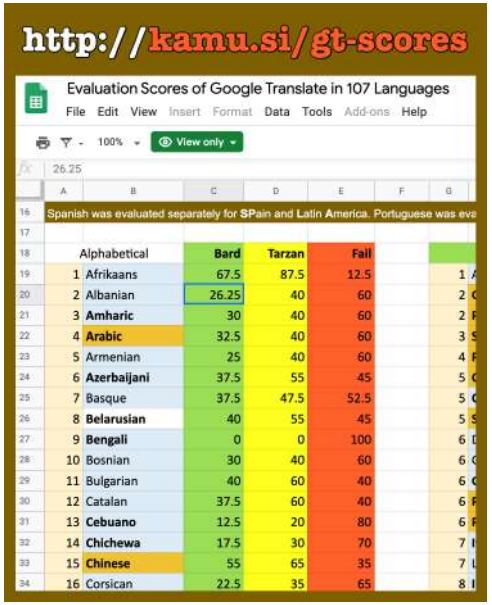
Words that have been **digitized** in a way that can be used within computer processes

- English
- Other high-investment languages
- Other high-population languages
- The other 98% of languages

- Major European languages (especially French, Italian, German, Spanish, Portuguese, Russian)
- Arabic, Chinese, Korean, and Japanese
- Tax money at work, e.g. Catalan, Estonian
- A few dozen somewhere on this part of the gradient

- Many words have some monolingual digital existence
- Some data sets with some party terms
- NLP inferior to English (claim based on analysis of which languages get research attention in computational linguistics)
- Parallel corpora with English
- Little data between non-English pairs

- We cannot interpret the data across systems (interoperability)
- MT is computed through spelling or word embeddings, disassociated from meaning



	Alphabetical	Bard	Tarzan	Fail
1	Afrikaans	67.5	87.5	12.5
2	Albanian	26.25	40	60
3	Amharic	30	40	60
4	Arabic	32.5	40	60
5	Armenian	25	40	60
6	Azerbaijani	37.5	55	45
7	Basque	37.5	47.5	52.5
8	Belarusian	40	55	45
9	Bengali	0	0	100
10	Bosnian	30	40	60
11	Bulgarian	40	60	40
12	Catalan	37.5	60	40
13	Cebuano	12.5	20	80
14	Chichewa	17.5	30	70
15	Chinese	55	65	35
16	Corsican	22.5	35	65

kamusi.org

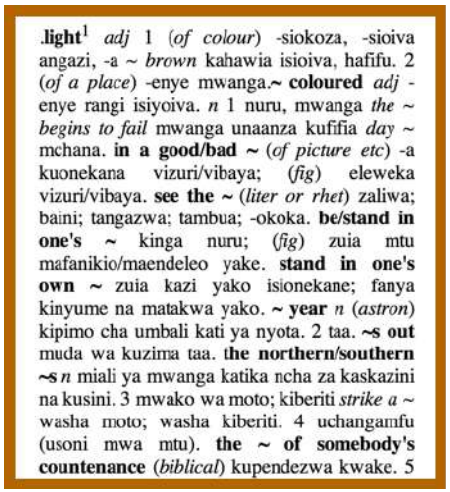
Myths about AI and MT: We have the data

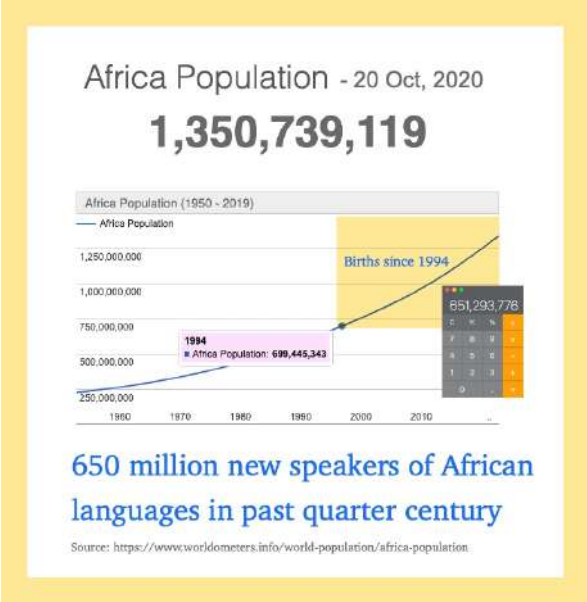
Words that have been **digitized** in a way that can be used within computer processes

- 9-figure languages, e.g. Hindi, Bengali, Indonesian, Swahili
- Fast growing languages throughout Africa and Asia
- 300+ embattled languages with more than 1,000,000 speakers

- Some words with some monolingual digital existence (e.g. brick-of-text dictionaries with some thousands of terms)
- Some basic bilingual dictionaries (lemmatic forms only), usually to English or French
- No NLP for most
- Some monolingual corpora higher on the gradient
- Zero interoperability
- MT to English exists for a few dozen, but is unusable

- English
- Other high-investment languages
- Other high-population languages
- The other 98% of languages


light[1] *adj* 1 (*of colour*) -siokoza, -sioiva angazi, -a ~ *brown* kahawia isioiva, hafifu. 2 (*of a place*) -enye mwanga.~ **coloured** *adj* -enye rangi isiyoiva. *n* 1 nuru, mwanga *the ~ begins to fail* mwanga unaanza kufifia *day ~* mchana. **in a good/bad ~** (*of picture etc*) -a kuonekana vizuri/vibaya; (*fig*) eleweka vizuri/vibaya. **see the ~** (*liter or rhet*) zaliwa; baini; tangazwa; tambua; -okoka. **be/stand in one's ~** kinga nuru; (*fig*) zuia mtu mafanikio/maendeleo yake. **stand in one's own ~** zuia kazi yako isionekane; fanya kinyume na matakwa yako. **~ year** *n* (*astron*) kipimo cha umbali kati ya nyota. 2 taa. **~s out** muda wa kuzima taa. **the northern/southern ~s** *n* miali ya mwanga katika ncha za kaskazini na kusini. 3 mwako wa moto; kiberiti *strike a ~* washa moto; washa kiberiti. 4 uchangamfu (usoni mwa mtu). **the ~ of somebody's countenance** (*biblical*) kupendezwa kwake. 5



kamusi.org

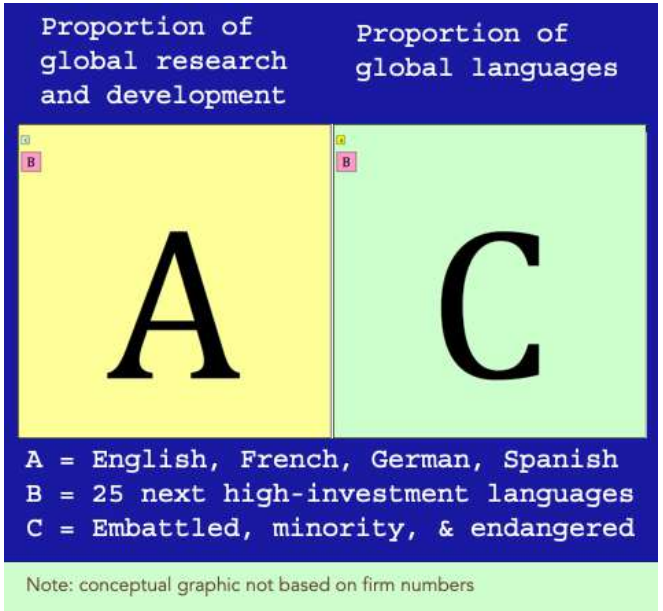
Myths about AI and MT: We have the data

Words that have been **digitized** in a way that can be used within computer processes

- Almost 7,000 languages
  (exact number discussed at http://kamu.si/7000-languages)

- Negligible digital existence of any sort of data for most languages
- No data = not possible in universe of "Universal Translation"
- Included in the universe of "Universal Translation" hype
- Important to document, preserve, and offer certain technological services – e.g., high quality dictionaries, language models
- Lesser call for text-to-text MT – low (if any) literacy, few (if any) texts
- Speech-to-speech translation could be valuable, is technologically possible (needs 7000 graduate students to collect and make sense of data)

- English
- Other high-investment languages
- Other high-population languages
- The other 98% of languages



kamusi.org

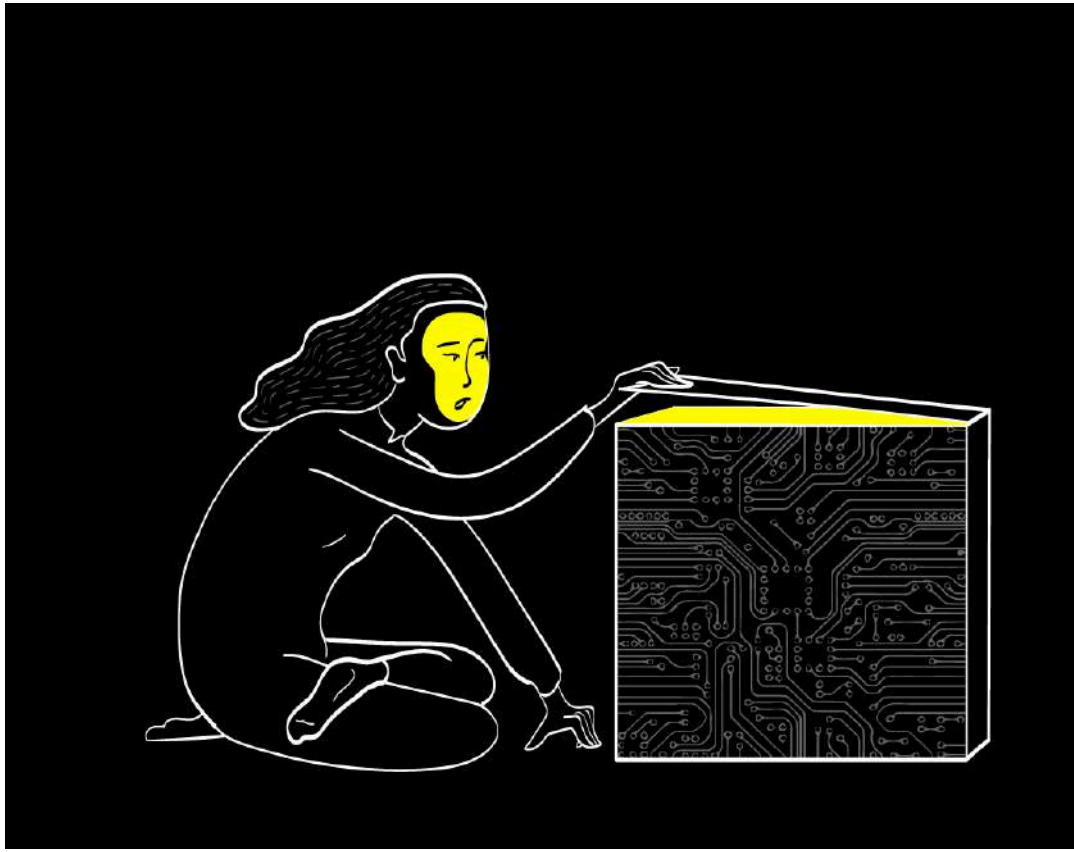
Myths about AI and MT

1. We have the data
2. We have the methods
3. Other biases and blind spots

kamusi.org

Myths about AI and MT: We have the methods

- Statistics (SMT)
- Corpora
- Neural Networks (NMT)
- Zero Shot
- Learning



kamusi.org

Myths about AI and MT: We have the methods

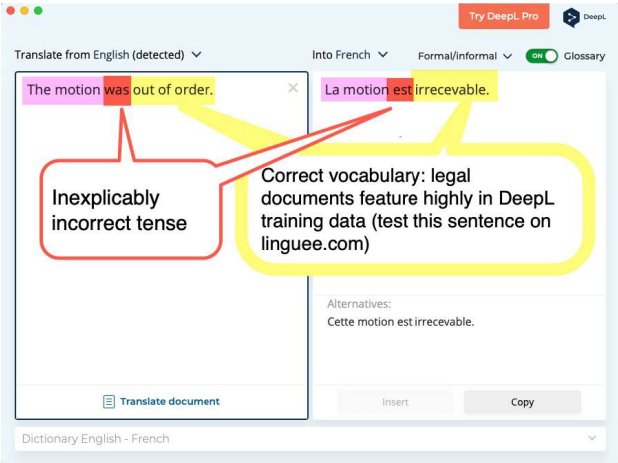
- Statistics (SMT)
- Corpora
- Neural Networks (NMT)
- Zero Shot
- Learning

- SMT is still a part of MT services (e.g., NMT can make no intelligent guesses about polysemy without context)
- Needs parallel data
- How SMT guesses (cartoon version)
  - "spring" = "primavera" in 40% of parallel sentences
  - other 60% "spring" = 10% bounciness, 10% a metal coil, 10% water flowing from the ground, 10% elastic force, 10% stretchiness, 10% a jump
  - Chance that SMT picks "springtime" sense = near 100%

kamusi.org

Myths about AI and MT: We have the methods

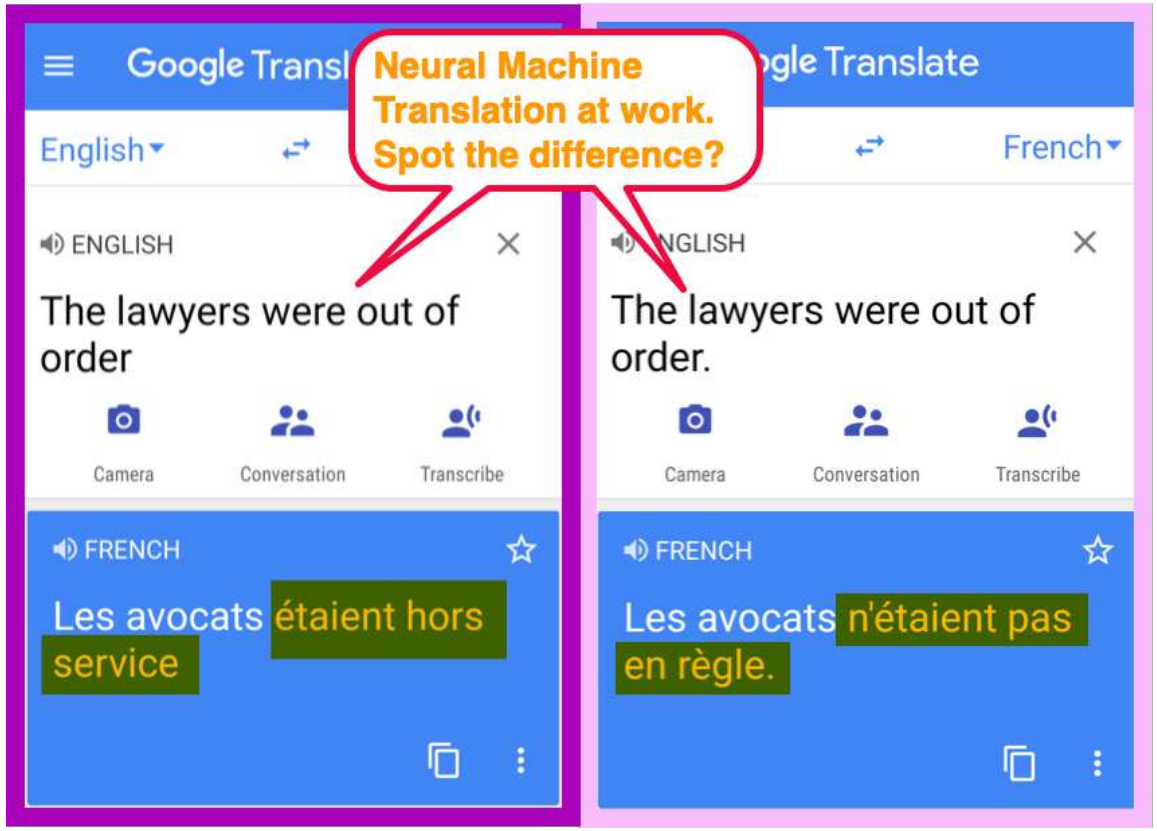
- Monolingual corpora can yield words, inflections, and expressions in one language
- Parallel corpora (translation gold) = extremely few pairs
- Limited topics – formal documents in the public domain
- Daily speech generally out of scope (including second-person constructions necessary for correspondence and text conversations)



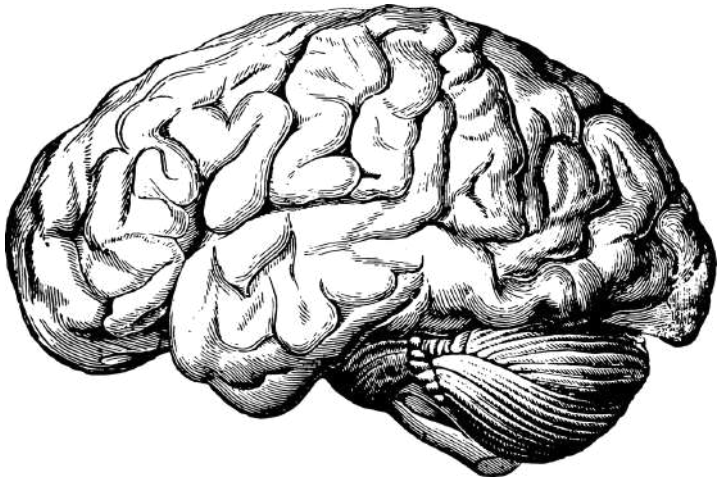
- Statistics (SMT)
- Corpora
- Neural Networks (NMT)
- Zero Shot
- Learning

kamusi.org

Myths about AI and MT: We have the methods



- Statistics (SMT)
- Corpora
- Neural Networks (NMT)
- Zero Shot
- Learning

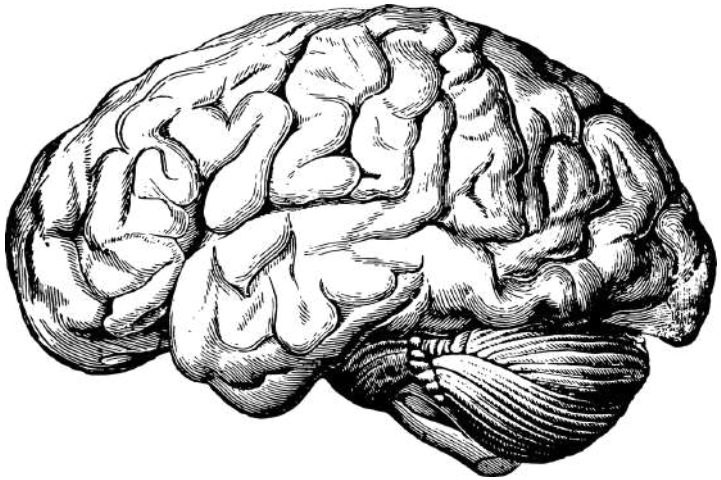


kamusi.org

Myths about AI and MT: We have the methods

- There is no brain. There are no neurons. "Neural network" is a marketing metaphor.
- NMT runs a lot of tests, can find hidden patterns
- With lots of parallel data, NMT performs nicely in certain circumstances:
  1. The language is at the upper tier of testing on teachyoubackwards.com
  2. The conversion is to or from English
  3. The text is well structured and written using formal language and short sentences
  4. The text relates to formal topics
  5. The translation is for casual purposes where misunderstanding cannot result in unpleasant consequences
- Even with lots of parallel data, lots of misses (e.g., our "out of order" examples, with context words, missed 4 out of 5 times)
- MUSA: The Make Up Stuff Algorithm
- Much more: http://kamu.si/myth2
- Most language pairs do not have parallel data, so…

- Statistics (SMT)
- Corpora
- Neural Networks (NMT)
- Zero Shot
- Learning



kamusi.org

Myths about AI and MT: We have the methods

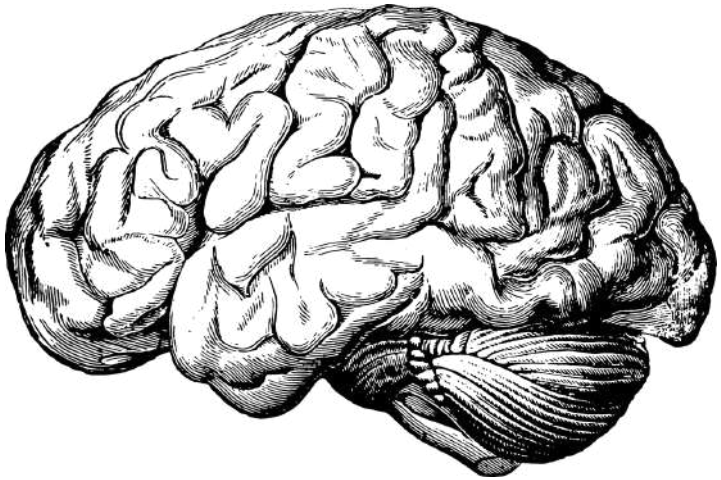
- MUSA: The Make Up Stuff Algorithm

(Facebook MT)

Real Somali ⇛

Fake English ⇛

- Statistics (SMT)
- Corpora
- Neural Networks (NMT)
- Zero Shot
- Learning



kamusi.org

Myths about AI and MT: We have the methods

- The idea:
  1. Align data from Language A with English.
  2. Align data from Language B with English.
  3. Shake.
  4. Remove English.
  5. A to B Translation!


Google's AI can translate language pairs it has never seen

If it can convert English to Korean and Japanese, it can automatically do Korean to Japanese

- Results (technically): phenomenally bad
- Results (media): phenomenally good
- Results (public perception): phenomenally good

- Much more: http://kamu.si/myth3

- Statistics (SMT)
- Corpora
- Neural Networks (NMT)
- Zero Shot
- Learning

kamusi.org

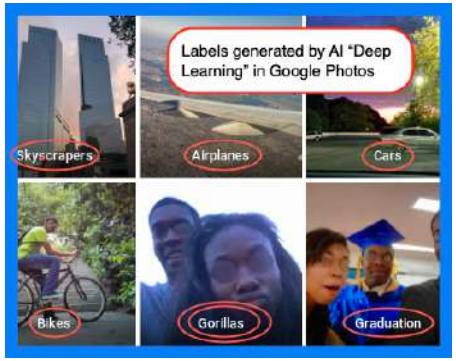
Myths about AI and MT: We have the methods

- Machine Learning
  - Needs a "gold standard" to compare results against a "ground truth"

  - else -
  - Needs human validation of results

  - else - Garbage In, Garbage Out
  - Only finds patterns within existing data
  - Risks locking info that is *sometimes* right as *always* right

- Learning from Users
  - 4 year "suggest an edit" experiment on Google Translate finds maximum 40% uptake
  - Crowdsourcing must be carefully fortified against bogus results – not the case for Google "Translate Community" DeepL "Rules"
  - Companies pay employees to teach AI for self-driving cars (so nobody dies), but not for most languages. That's a business choice, not science.
  - More: http://kamu.si/myth5



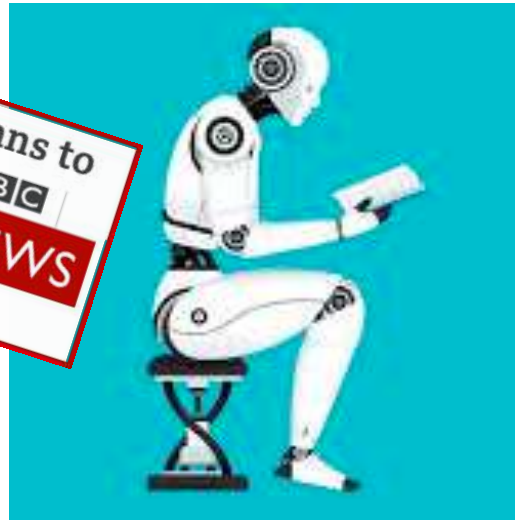
- Statistics (SMT)
- Corpora
- Neural Networks (NMT)
- Zero Shot
- Learning


Why Big Tech pays poor Kenyans to teach self-driving cars

When Artificial Intelligence works as intended, Silicon Valley types often say it's "like magic".

But it isn't magic. It's Brenda, a 26-year-old single mother who lives Kibera, Africa's largest slum, and perhaps the toughest neighbourhood on earth, where hundreds of thousands of people live in a space not too much bigger than London's Hyde Park.

BBC NEWS



kamusi.org

Myths about AI and MT

1. We have the data
2. We have the methods
3. Other biases and blind spots

kamusi.org

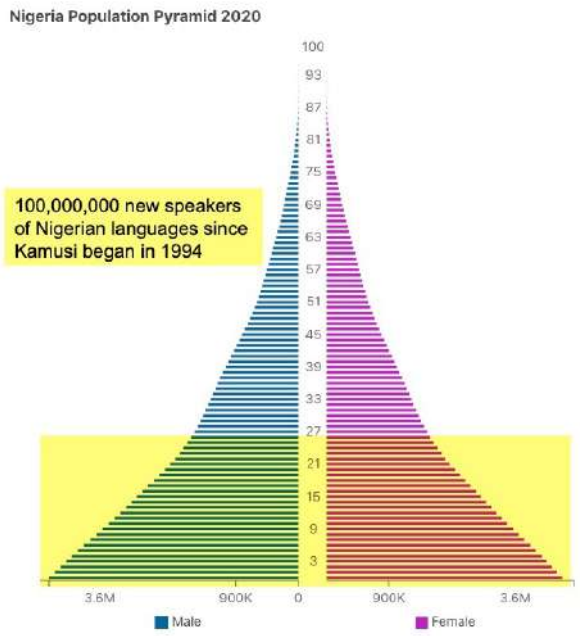
Myths about AI and MT: Biases and blind spots

- Follow the money
  - Corporate investment
  - Government support
  - Foundation grants
- Follow the research

- English ≠ Translation. Less important than people immersed in it think it is
- Elite language bias shared by leaders in Africa, India, etc
- Large tech-excluded languages are "Embattled" but not "Endangered" – growing, not going, with unserved language needs and untapped market potential

- Yes, language technology is inequitable. Yes, it matters.

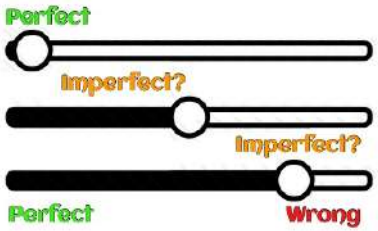
- White languages matter (+CKJ)
- Technology works
- Computer knows best



kamusi.org

Myths about AI and MT: Biases and blind spots

- We tend to trust technology to do what it claims
  - Maybe your GPS takes you down a cow path today, but you'll certainly follow it again tomorrow
  - Search for "gorilla", you'll get mostly gorillas
  - Would Google put Samoan in their service if it didn't work? How's your Samoan?
- It isn't perfect, but…



- Willing suspension of disbelief
- Rooting for "Team Translate" – we like tech victories, forgive (and forget) the defeats

- White languages matter (+CKJ)
- Technology works
- Computer knows best

"wild turkey" results in Google Images (top)
"turkish wilderness" result in Flickr (bottom)

kamusi.org

Myths about AI and MT: Biases and blind spots

- Don't mind the gaps – we fill in holes with our own understanding (fix tenses, odd words…)
- Confirmation bias – we can see that some of the translation is correct, so the rest must be okay
- Magic wand – results change as we type, so serious optimal calculations must be occurring. See "magic" in action: http://kamu.si/myth4
- Gaslighting – it looks suspicious, but who am I to challenge the engineers and linguists who are telling me it's right?

- White languages matter (+CKJ)
- Technology works
- Computer knows best

kamusi.org

Artificial Intelligence (AI) & Machine Translation (MT) on the Road toward Universal Translation

1. AI facts and fantasies
2. Myths about AI and MT
3. Realistic fantasies about computation and translation

kamusi.org

Chapter 3:
Realistic Fantasies about Computation and Translation

- Smurfs and Ducks
- Kam4D
- SlowBrew

kamusi.org

Realistic Fantasies about Computation and Translation

- Smurfs and Ducks
- Kam4D
- SlowBrew

kamusi.org

Realistic Fantasies about Computation and Translation

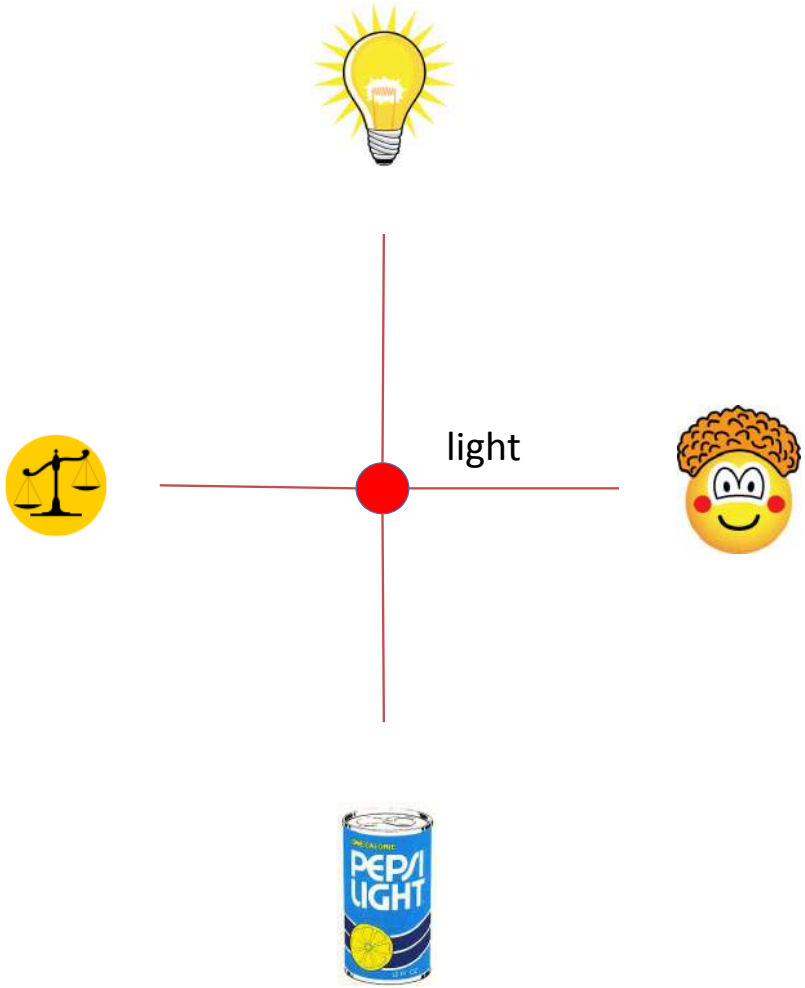
SMURF =
Spelling/ Meaning
Unit Reference

- Smurfs and Ducks
- Kam4D – kamu.si/kam4d
- SlowBrew

DUCKS = Data
Unified Concept
Knowledge Set

kamusi.org

light

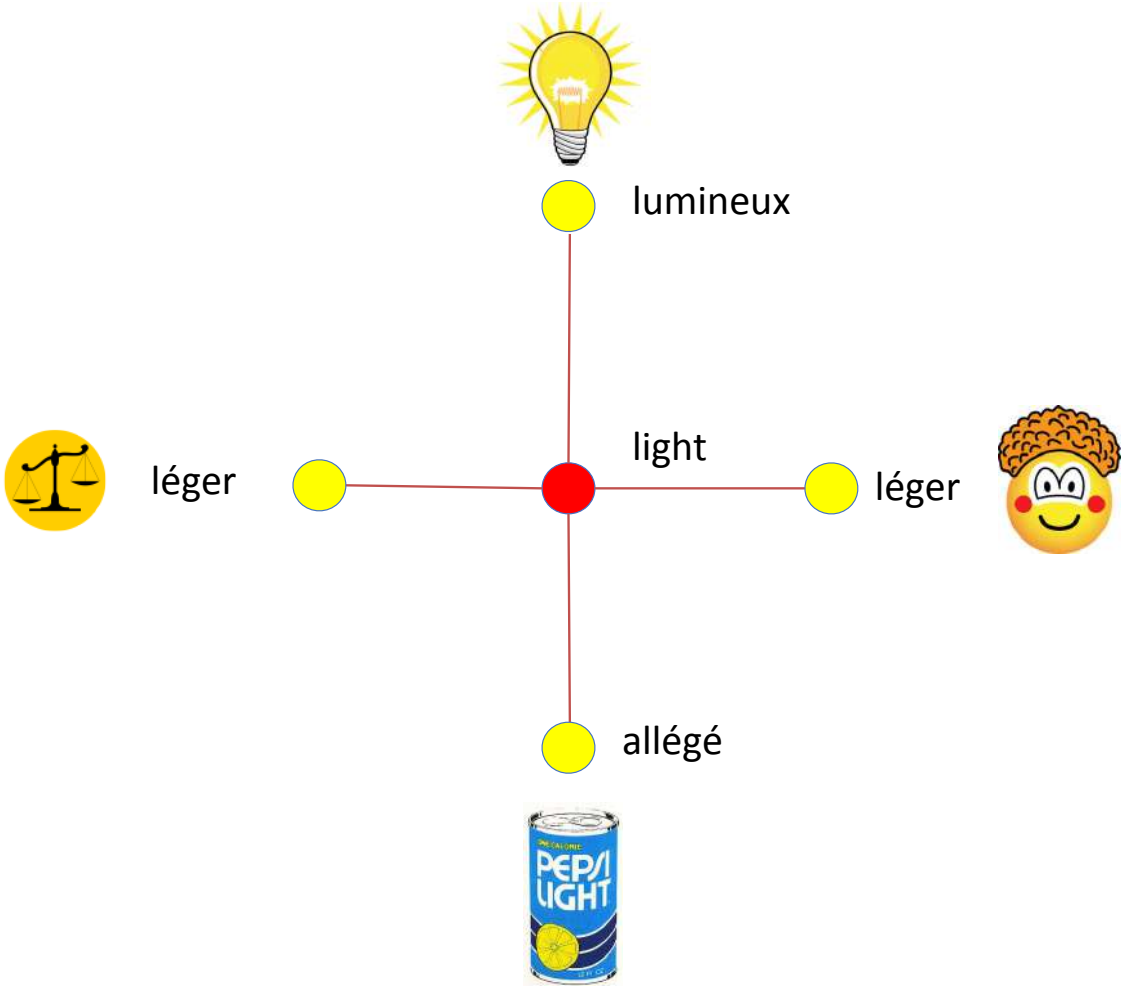


SMURF =
Spelling/ Meaning
Unit Reference

DUCKS = Data
Unified Concept
Knowledge Set

why multilingual dictionaries were impossible

kamusi.org

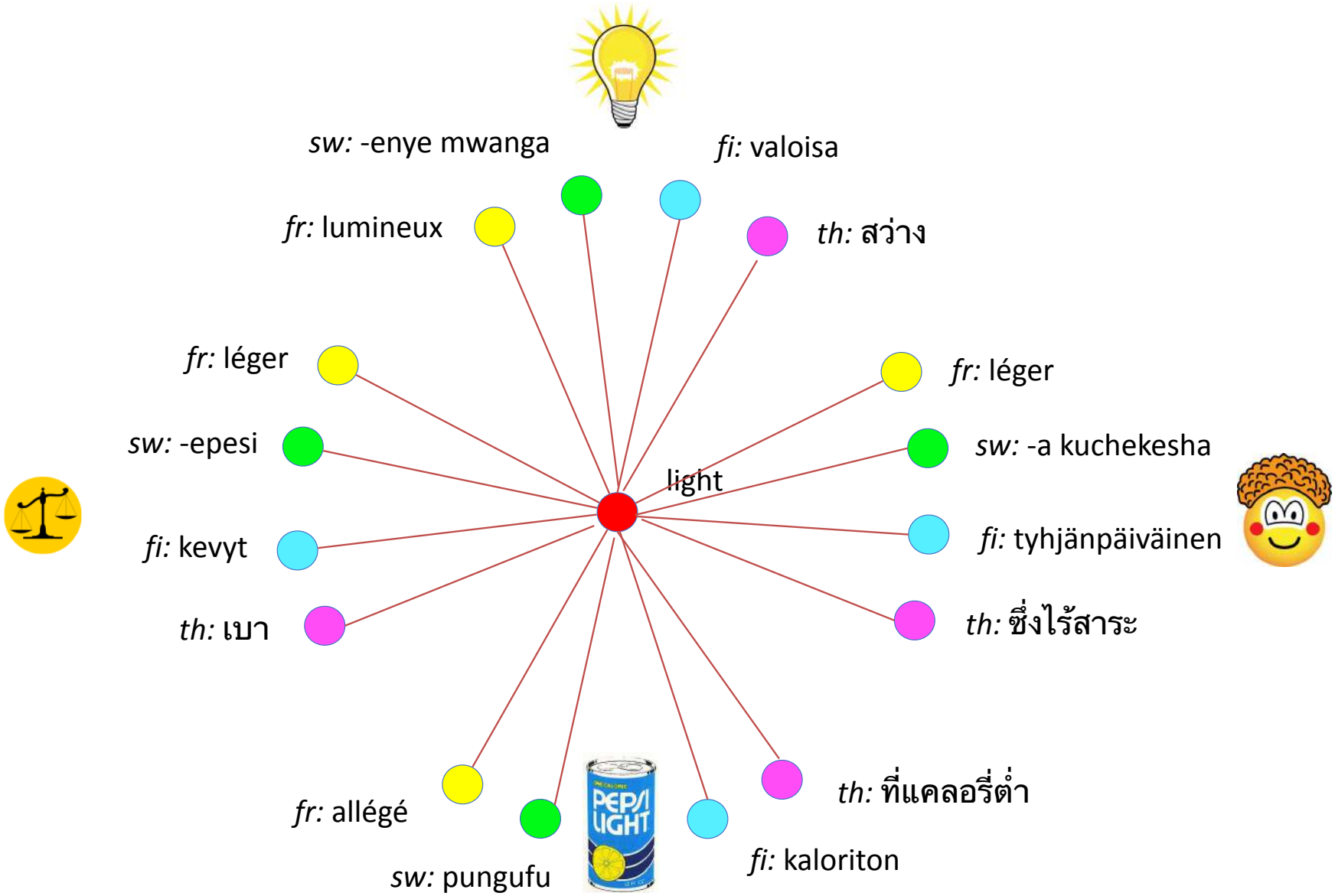


SMURF =
Spelling/ Meaning
Unit Reference

DUCKS = Data
Unified Concept
Knowledge Set

why multilingual dictionaries were impossible

kamusi.org



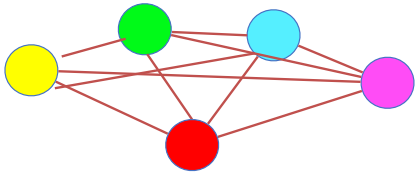
why multilingual dictionaries were impossible

kamusi.org

sw: -enye mwanga

fi: valoisa

fr: lumineux

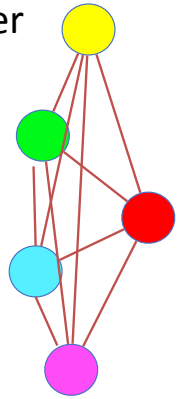


th: สว่าง

en: light

fr: léger

sw: -epesi

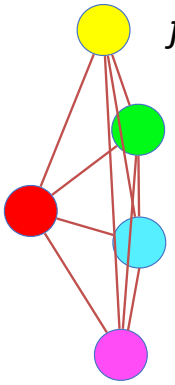


en: light

fi: kevyt

th: เบา

fr: léger



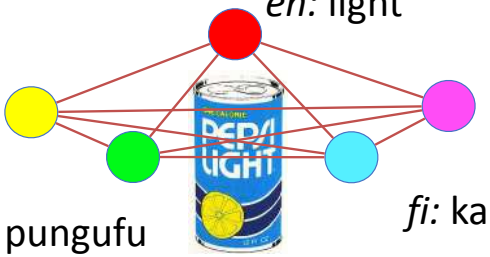
sw: -a kuchekesha

en: light

fi: tyhjänpäiväinen

th: ซึ่งไร้สาระ

en: light



th: ที่แคลอรี่ต่ำ

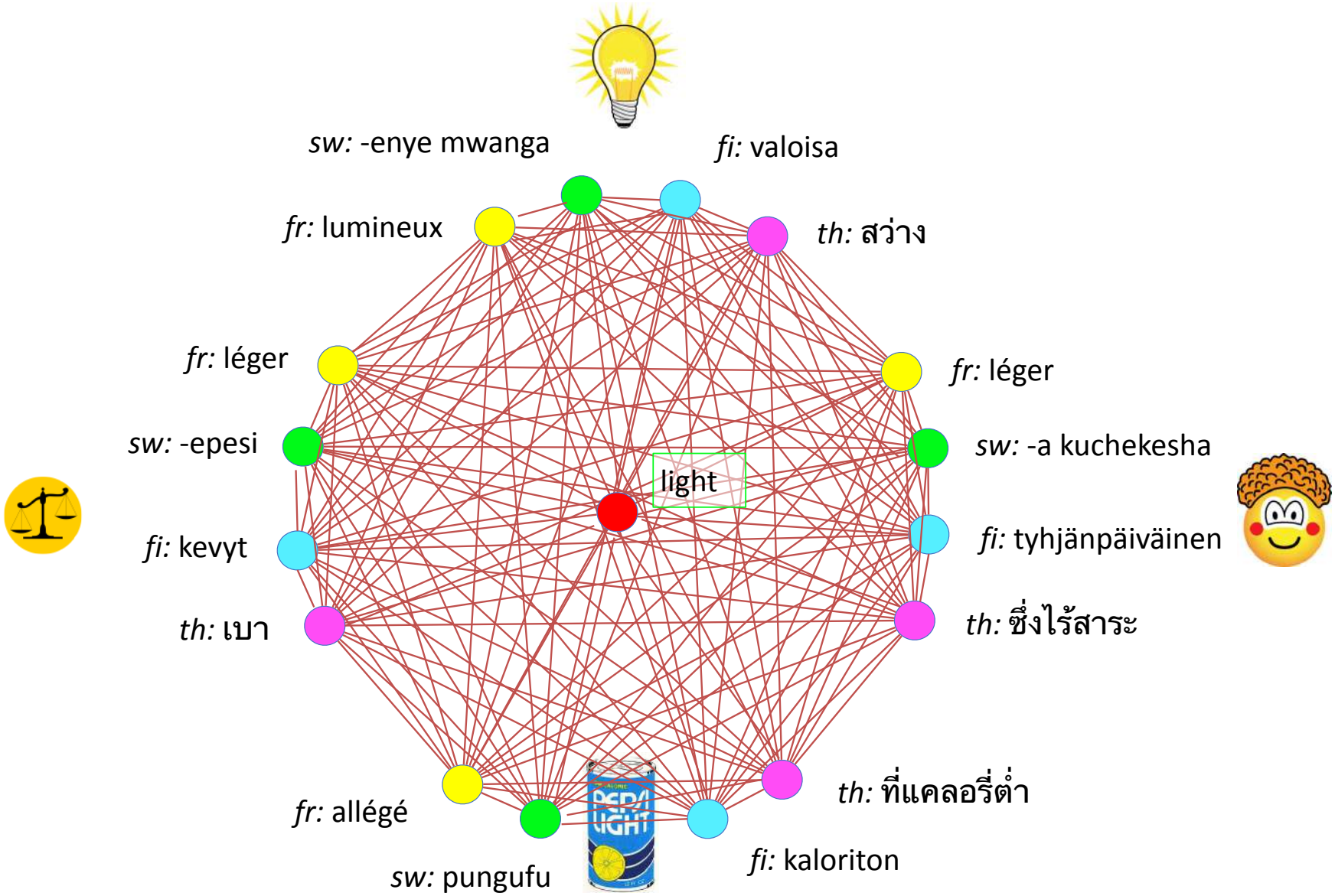
fr: allégé

sw: pungufu

fi: kaloriton

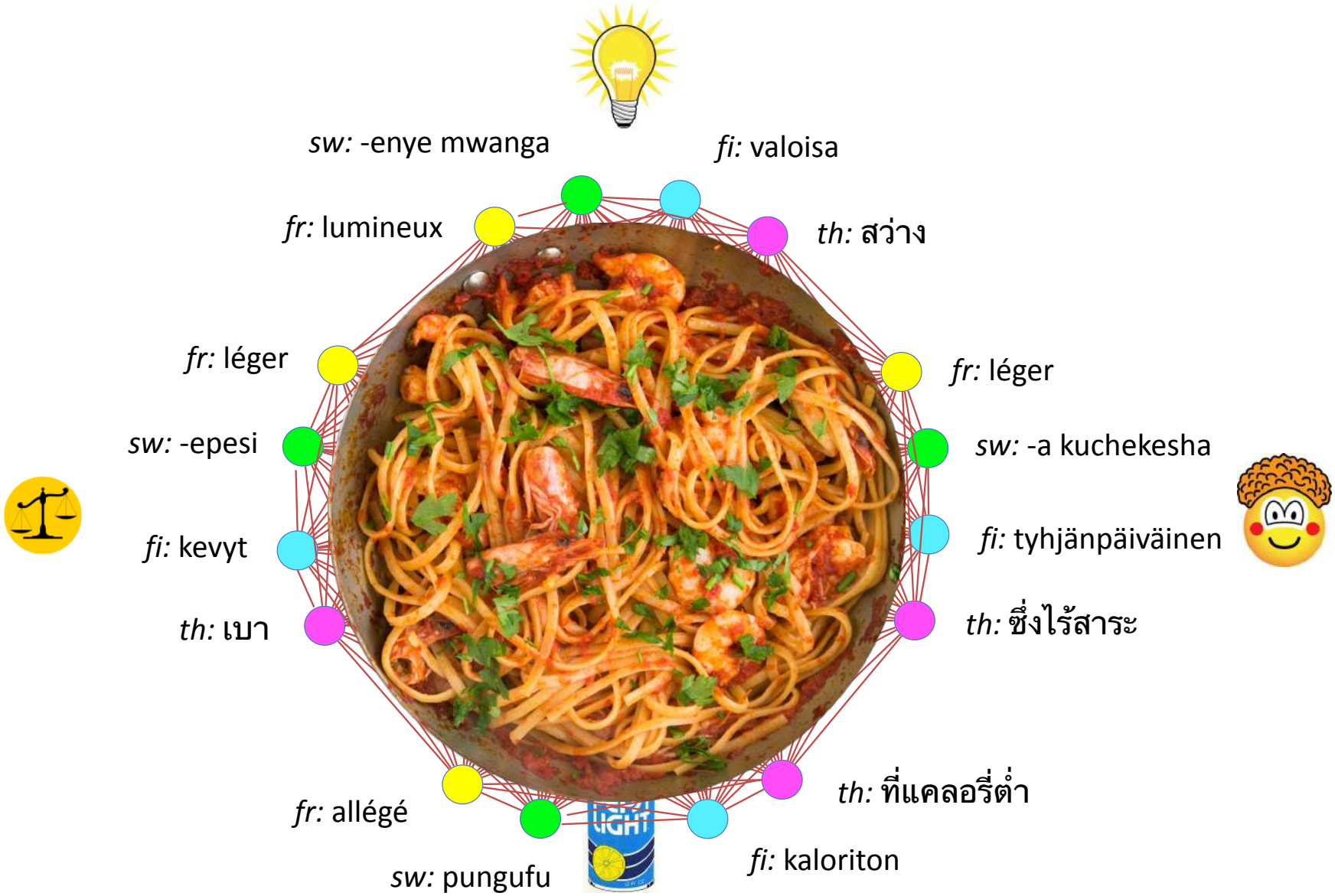
why multilingual dictionaries were impossible

kamusi.org



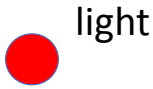
why multilingual dictionaries were impossible

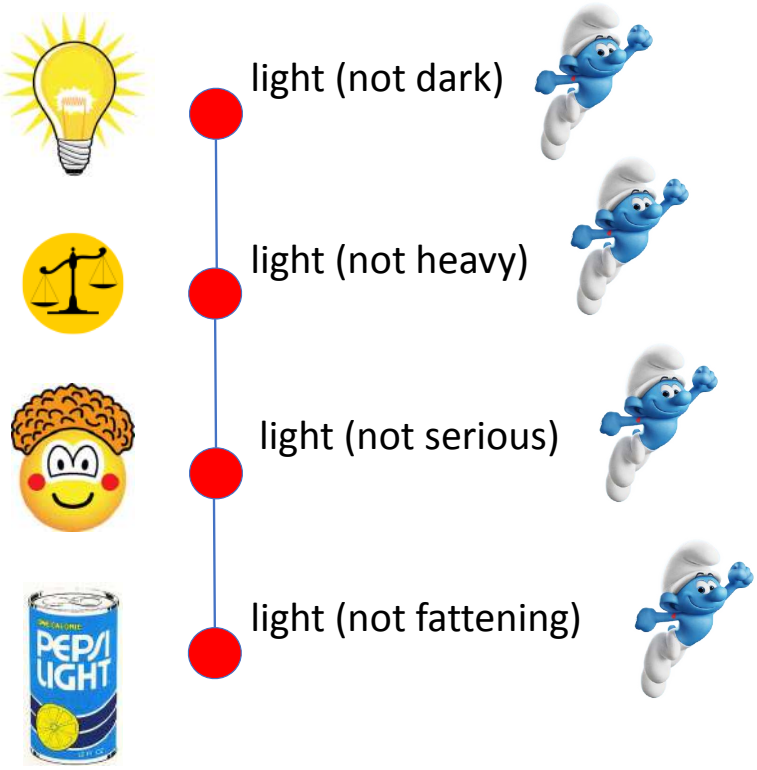
kamusi.org



why multilingual dictionaries were impossible

kamusi.org



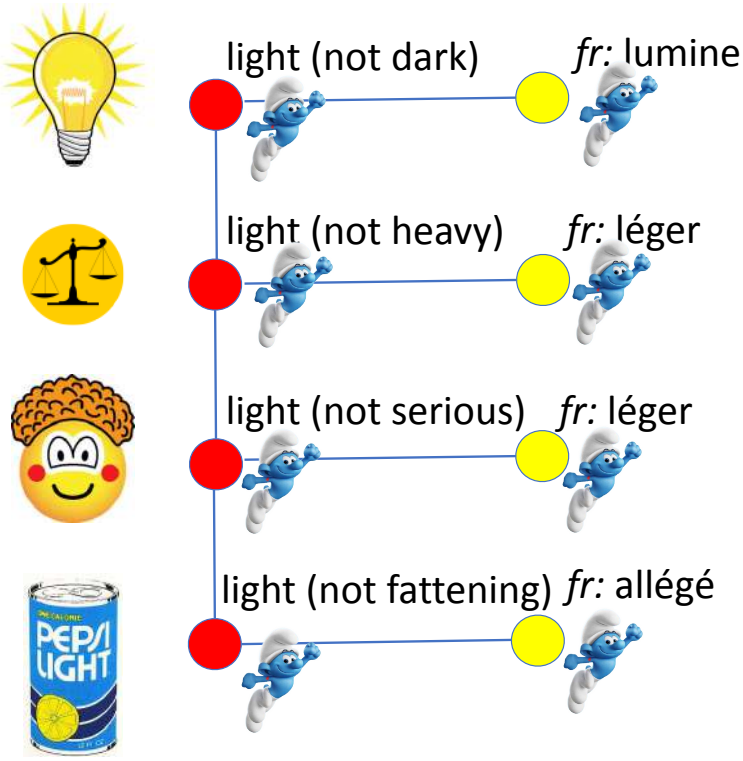


SMURF =
Spelling/ Meaning
Unit Reference

DUCKS = Data
Unified Concept
Knowledge Set

how Kamusi makes a multilingual dictionary possible

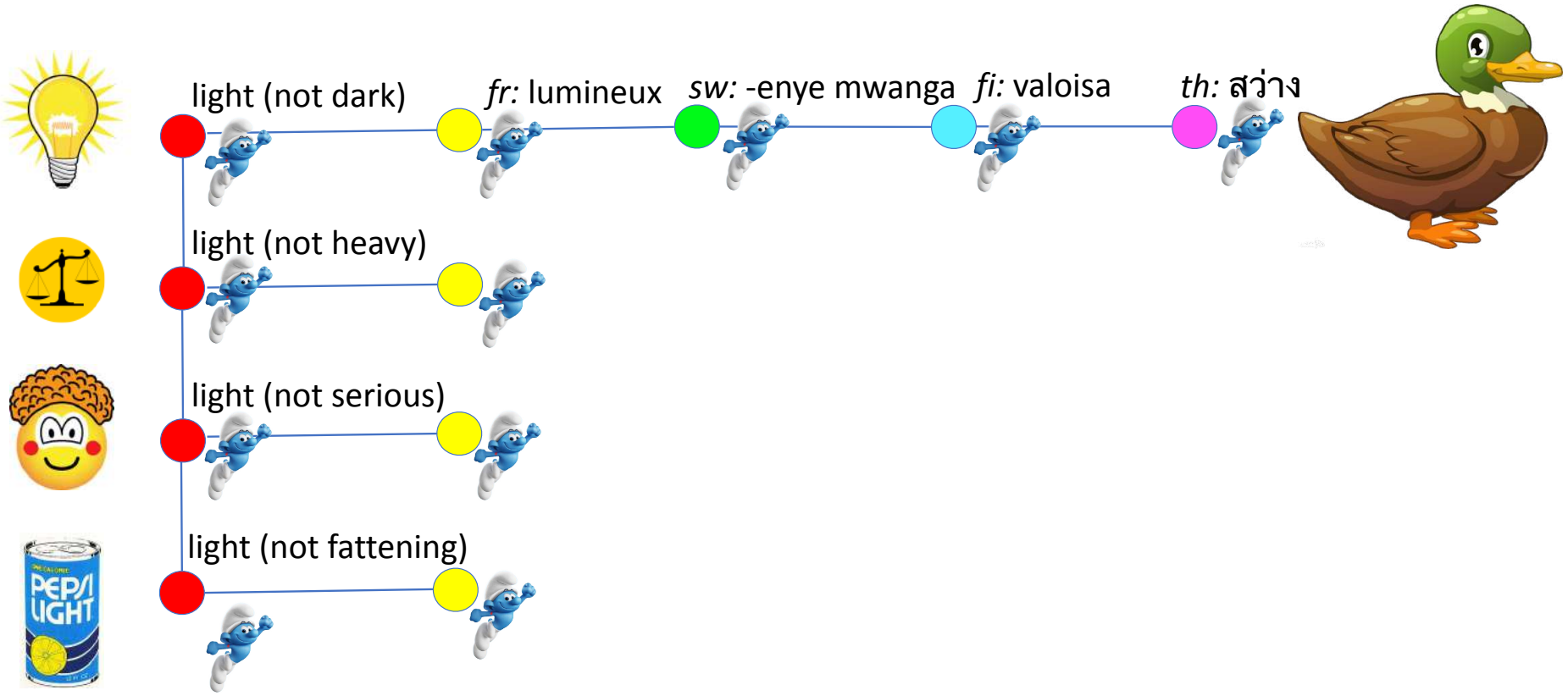
kamusi.org



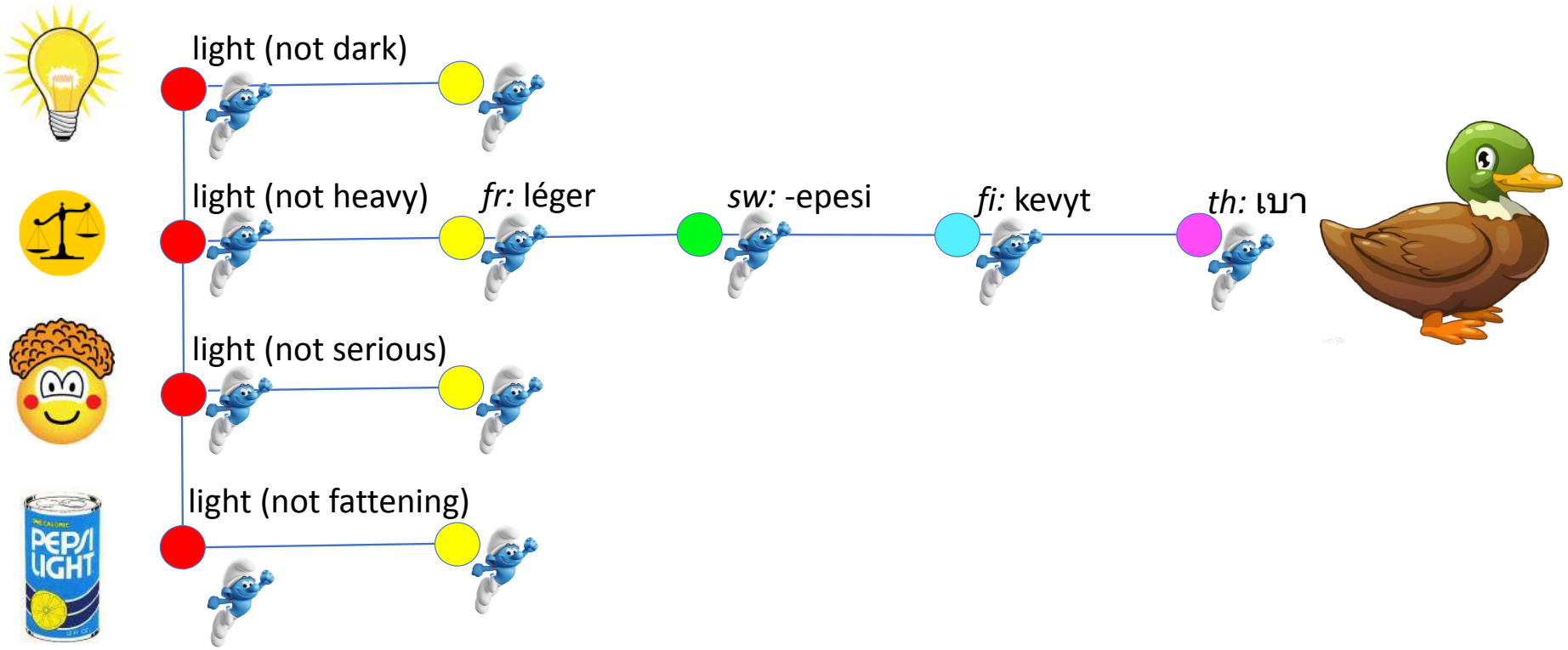
SMURF =
Spelling/ Meaning
Unit Reference

DUCKS = Data
Unified Concept
Knowledge Set

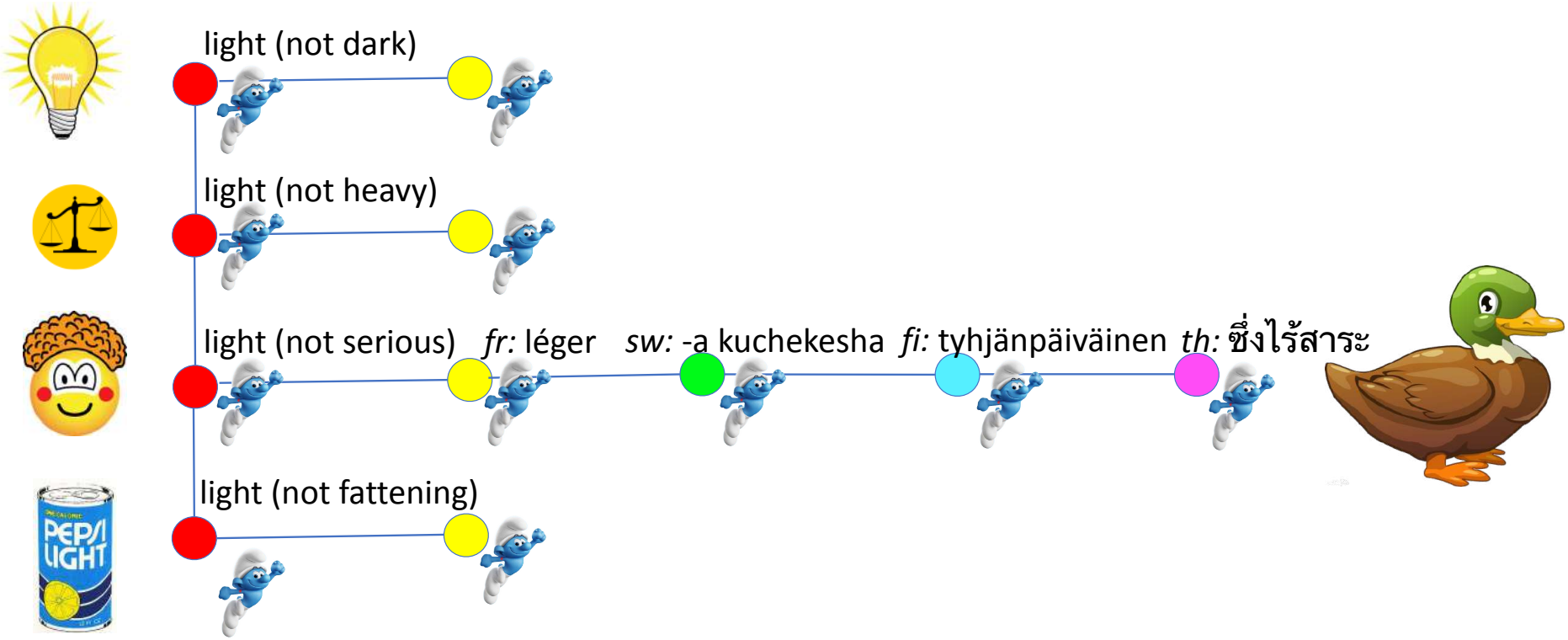
how Kamusi makes a multilingual dictionary possible



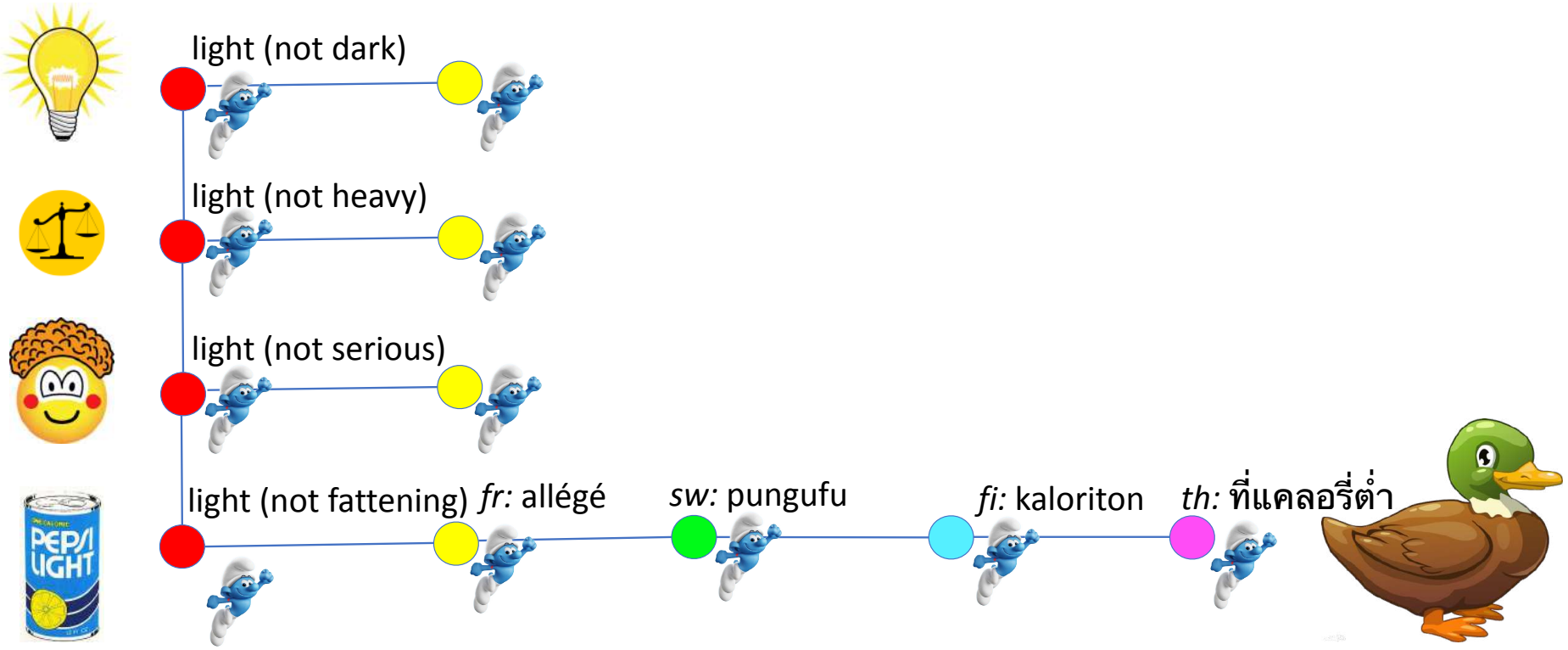
how Kamusi makes a multilingual dictionary possible



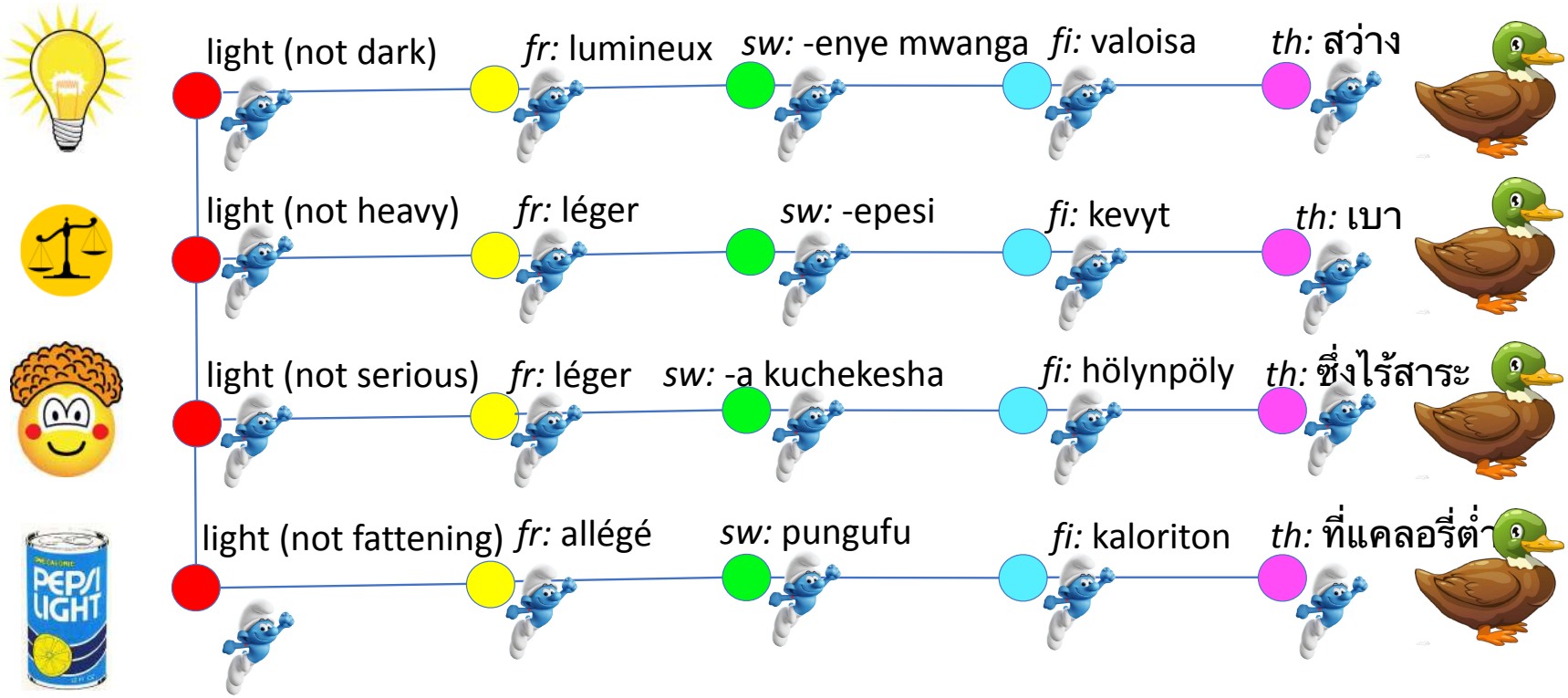
how Kamusi makes a multilingual dictionary possible



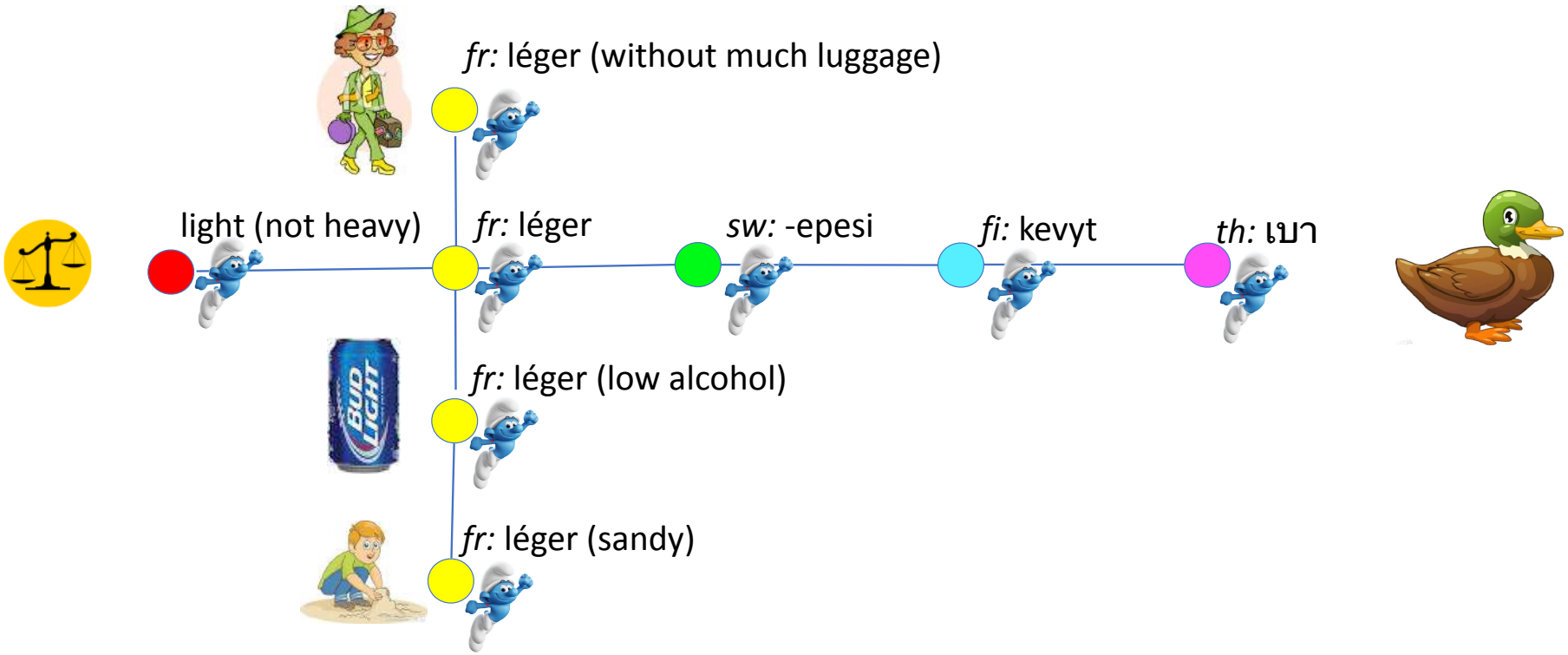
how Kamusi makes a multilingual dictionary possible



how Kamusi makes a multilingual dictionary possible



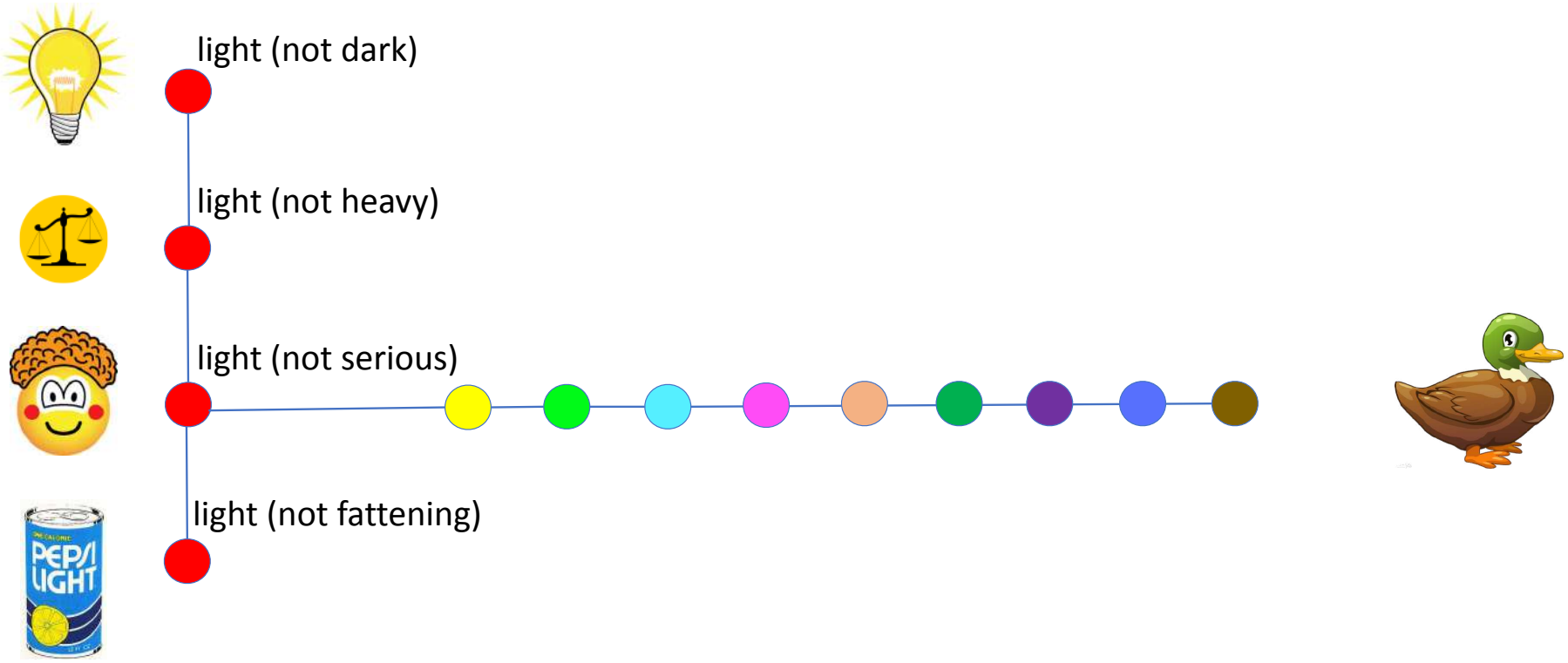
how Kamusi makes a multilingual dictionary possible



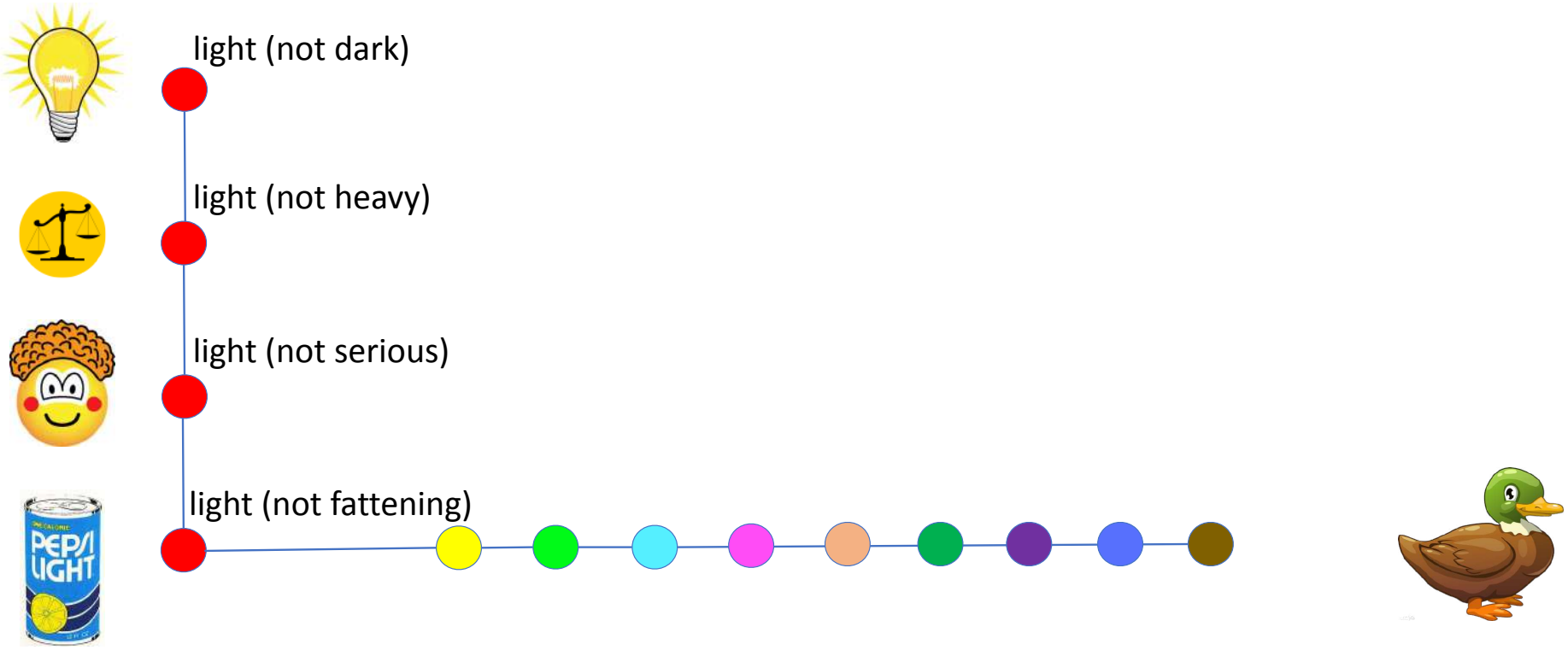
how Kamusi makes a multilingual dictionary possible

how Kamusi makes a multilingual dictionary possible

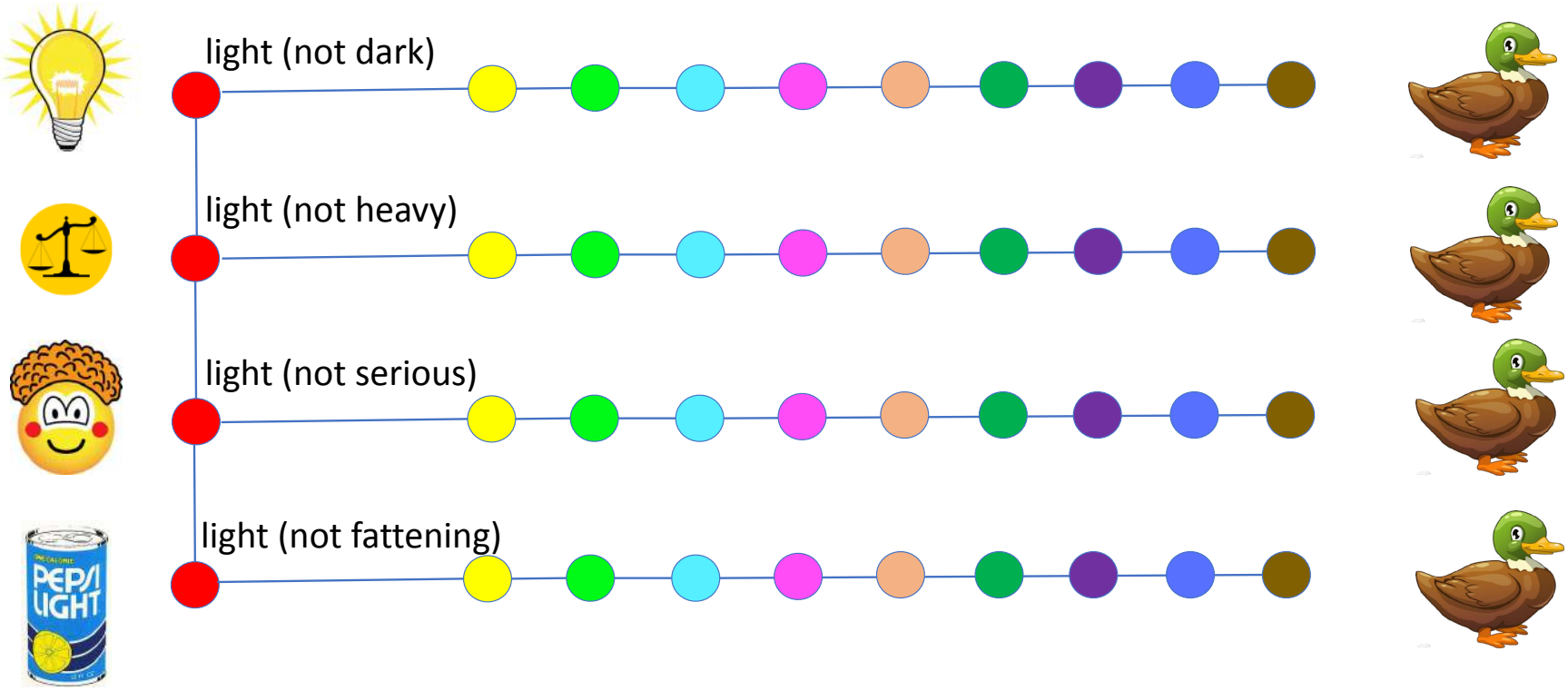
how Kamusi makes a multilingual dictionary possible



how Kamusi makes a multilingual dictionary possible



how Kamusi makes a multilingual dictionary possible


light (not dark)
light (not heavy)
light (not serious)
light (not fattening)
PEPSI
LIGHT

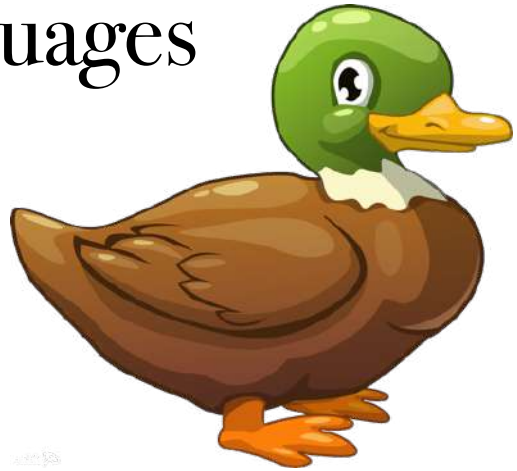
Realistic Fantasies about Computation and Translation

November 2020 census:

- 2,099,419 Smurfs
- 122 Languages

- Smurfs and Ducks
- Kam4D – kamu.si/kam4d
- SlowBrew

- ~138,000 Ducks
- 44 Languages



kamusi.org

Realistic Fantasies about Computation and Translation

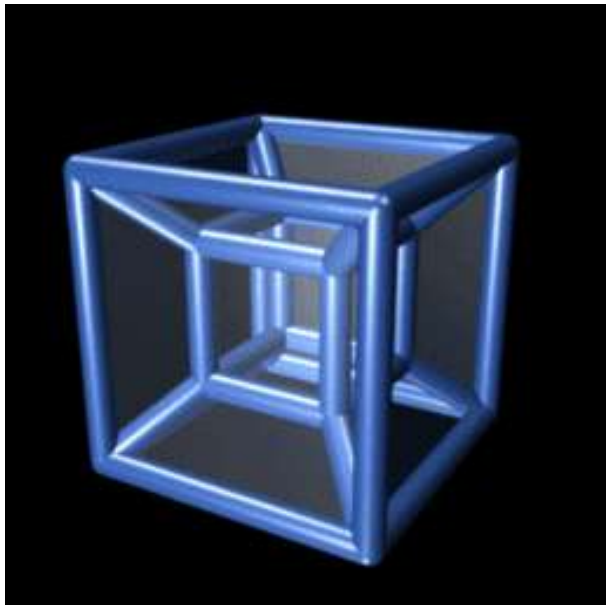
- Smurfs and Ducks
- Kam4D
- SlowBrew

kamusi.org

Realistic Fantasies about Computation and Translation

- 4D = Four Dimensional
  - Time is the fourth dimension - capacity to treat language change and historical languages
- Graph database structure for a complete matrix of human expression across time and space
  - the structure is realistic; the final goal is an impossible aspiration
- Molecular lexicography design

- Smurfs and Ducks
- Kam4D – kamu.si/kam4d
- SlowBrew

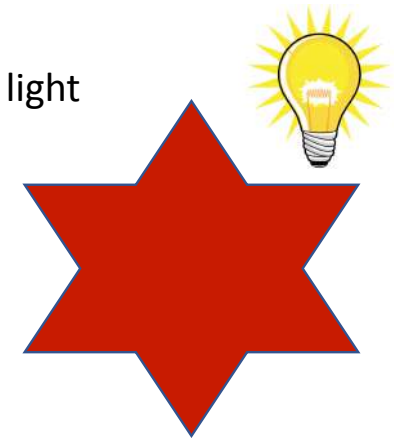


kamusi.org

- Smurfs and Ducks
- Kam4D – kamu.si/kam4d
- SlowBrew

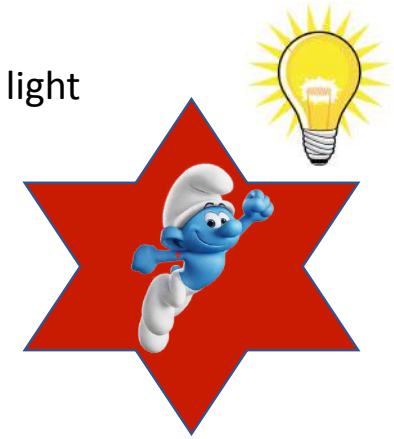
light

kamusi.org



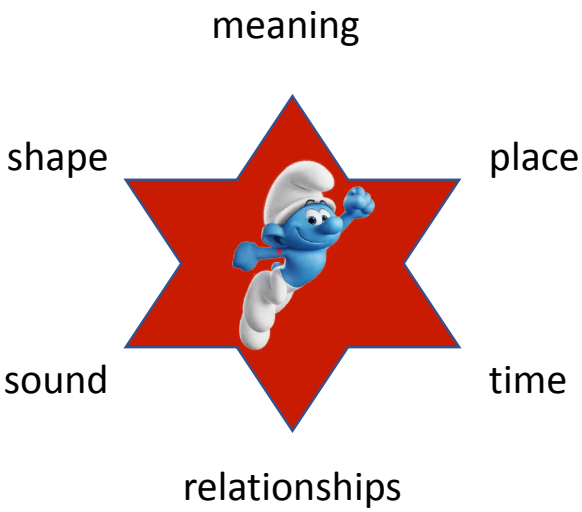
- Smurfs and Ducks
- Kam4D – kamu.si/kam4d
- SlowBrew

kamusi.org



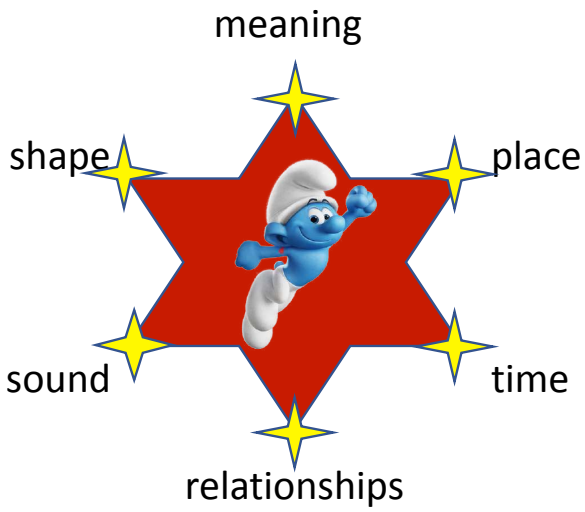
- Smurfs and Ducks
- Kam4D – kamu.si/kam4d
- SlowBrew

kamusi.org



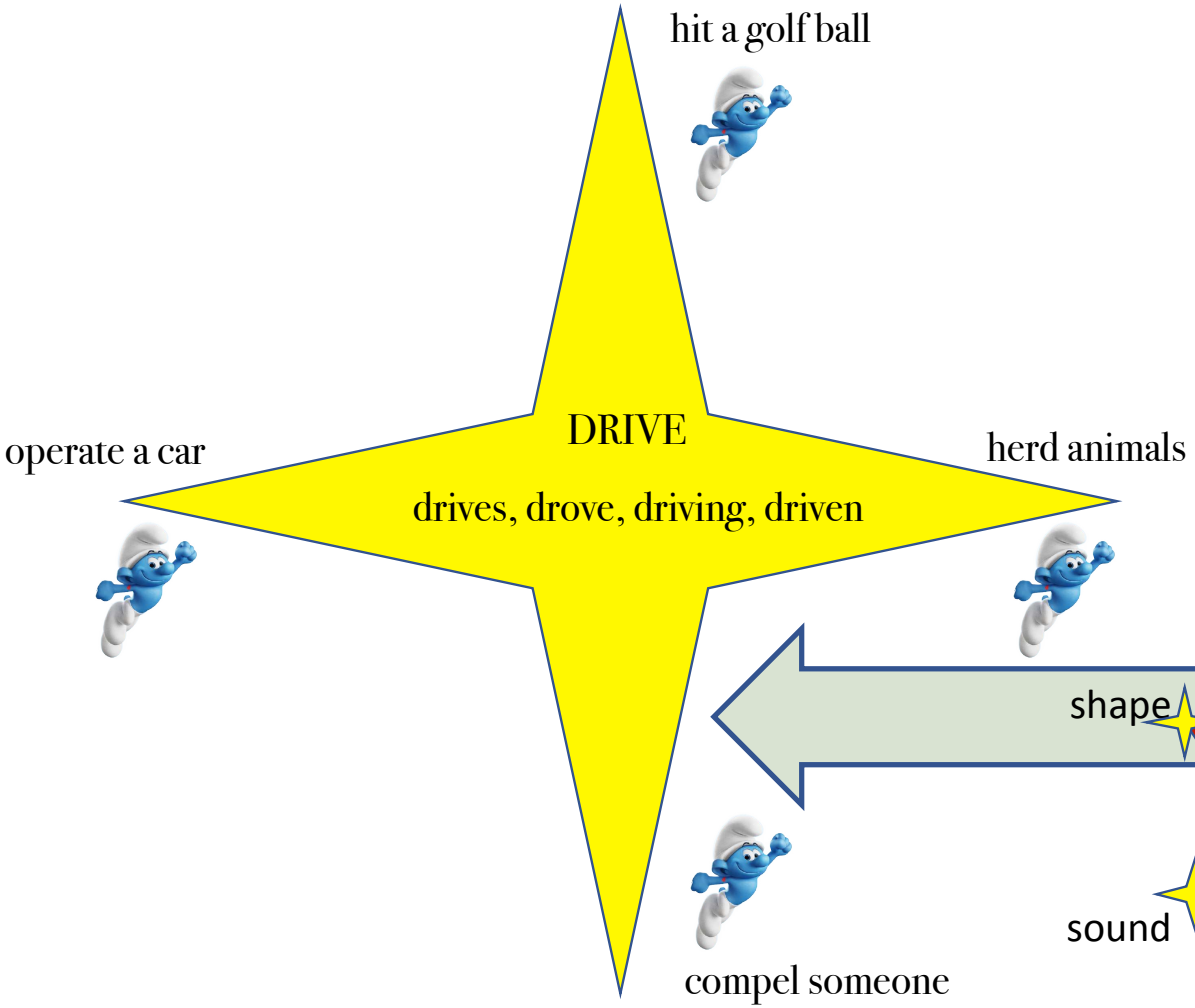
- Smurfs and Ducks
- Kam4D – kamu.si/kam4d
- SlowBrew

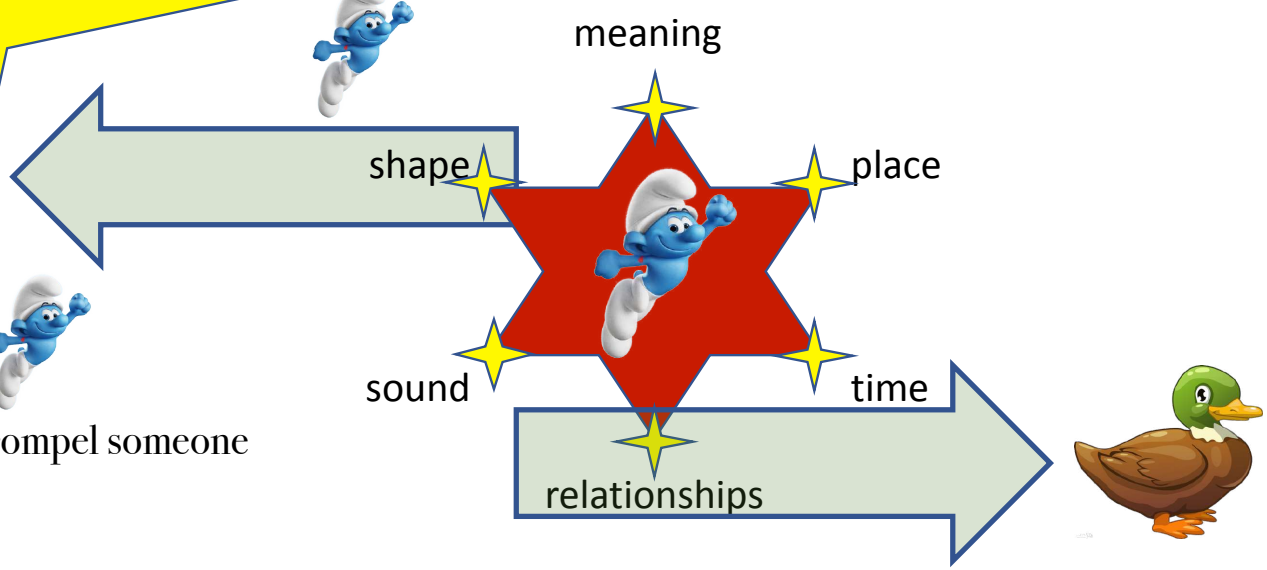
kamusi.org



- Smurfs and Ducks
- Kam4D – kamu.si/kam4d
- SlowBrew

kamusi.org



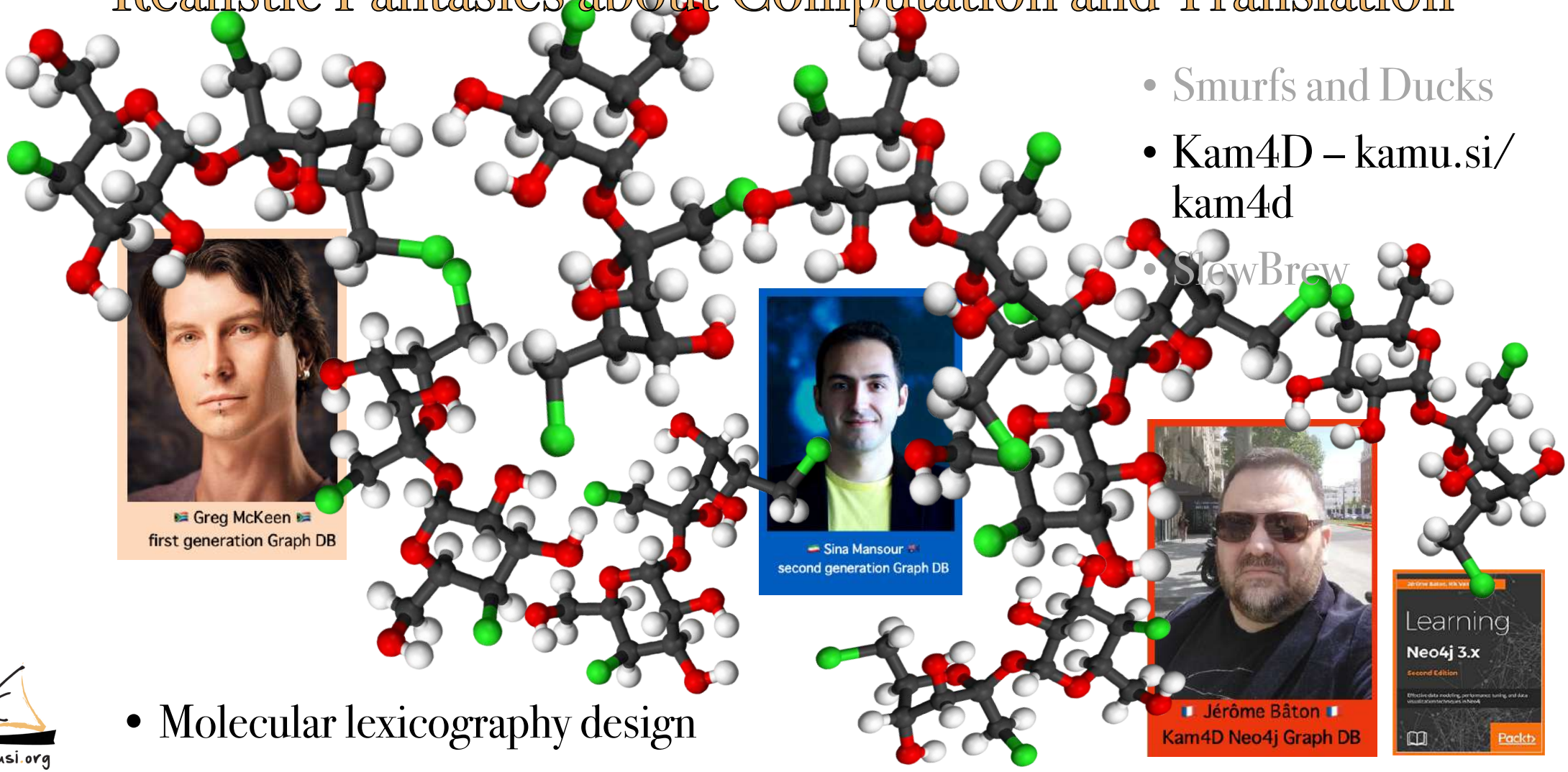


- Smurfs and Ducks
- Kam4D – kamu.si/kam4d
- SlowBrew

kamusi.org

Realistic Fantasies about Computation and Translation

- Smurfs and Ducks
- Kam4D – kamu.si/kam4d
- SlowBrew



- Molecular lexicography design

kamusi.org

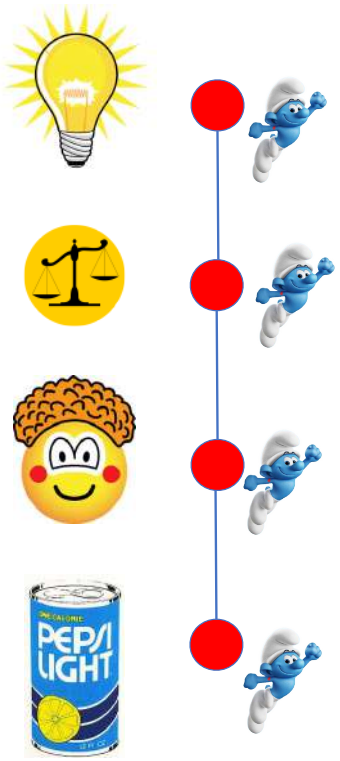
Realistic Fantasies about Computation and Translation

- Smurfs and Ducks
- Kam4D
- SlowBrew

kamusi.org

Realistic Fantasies about Computation and Translation

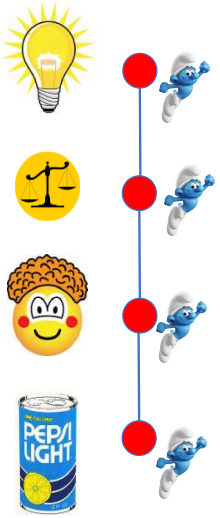
- User selects their meaning on the source side (predisambiguation)
  - Users can suggest missing senses



- Smurfs and Ducks
- Kam4D – kamu.si/kam4d
- SlowBrew

kamusi.org

Realistic Fantasies about Computation and Translation

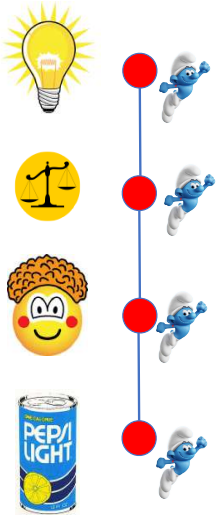


- User selects their meaning on the source side (predisambiguation)
  - Users can suggest missing senses
- SlowBrew suggests Party Terms (MWEs), or users can mark their own
  - Party Terms are treated as Smurfs in Kam4D
  - Separated expressions easily conjoined (unlike NMT)

- Smurfs and Ducks
- Kam4D – kamu.si/kam4d
- SlowBrew

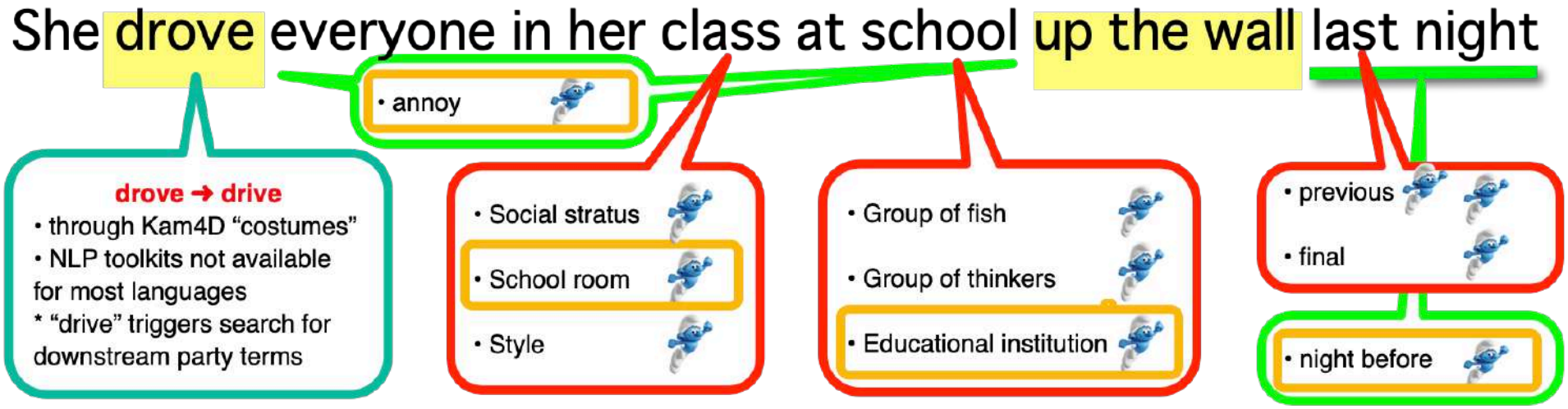
kamusi.org

Realistic Fantasies about Computation and Translation

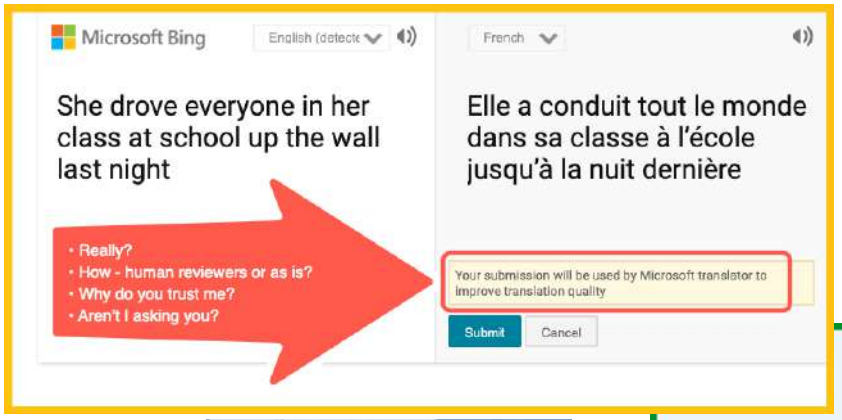


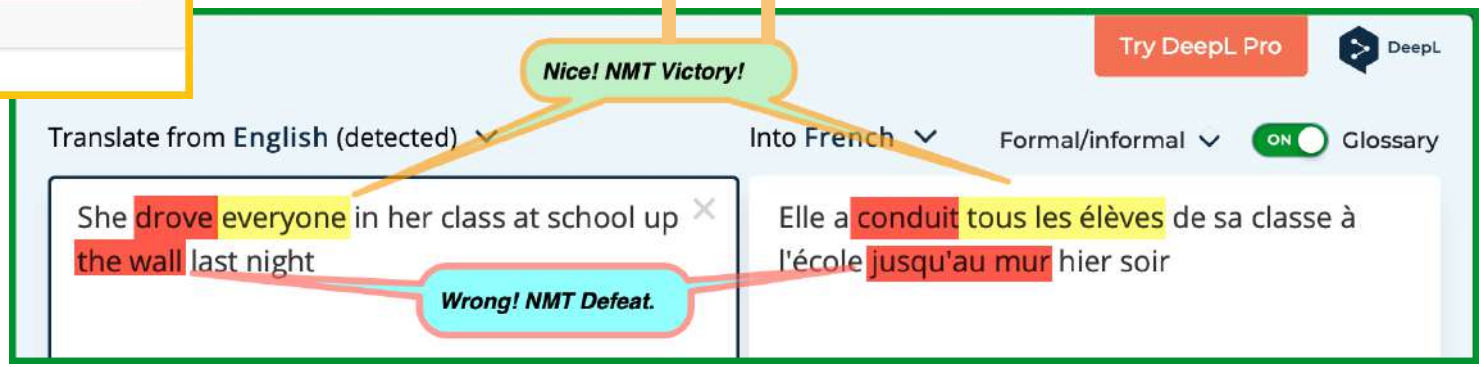
- User selects their meaning on the source side (predisambiguation)
  - Users can suggest missing senses
- SlowBrew suggests Party Terms (MWEs), or users can mark their own
  - Party Terms are treated as Smurfs in Kam4D
  - Separated expressions easily rejoined (unlike NMT)

- Smurfs and Ducks
- Kam4D – kamu.si/kam4d
- SlowBrew



kamusi.org





She drove everyone in her class at school up the wall last night

- annoy

drove → drive
- through Kam4D "costumes"
- NLP toolkits not available for most languages

* "drive" triggers search for downstream party terms

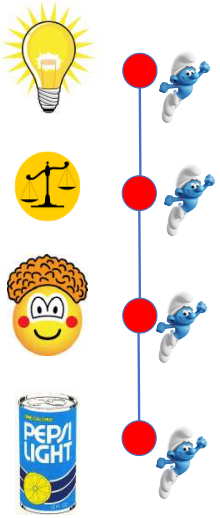
- Social stratus
- School room
- Style

- Group of fish
- Group of thinkers
- Educational institution

- previous
- final
- night before

kamusi.org

Realistic Fantasies about Computation and Translation

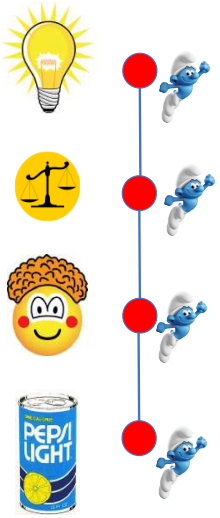


- User selects their meaning on the source side (predisambiguation)
  - Users can suggest missing senses
- SlowBrew suggests Party Terms (MWEs), or users can mark their own
  - Party Terms are treated as Smurfs in Kam4D
  - Separated expressions easily rejoined (unlike NMT)
- DUCKS finds equivalent term in Language B

- Smurfs and Ducks
- Kam4D – kamu.si/kam4d
- SlowBrew

kamusi.org

Realistic Fantasies about Computation and Translation



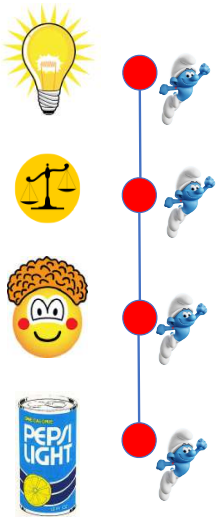
- User selects their meaning on the source side (predisambiguation)
  - Users can suggest missing senses
- SlowBrew suggests Party Terms (MWEs), or users can mark their own
  - Party Terms are treated as Smurfs in Kam4D
  - Separated expressions easily rejoined (unlike NMT)
- DUCKS finds equivalent term in Language B

- Machine learns from context-specific user selections
  - Crowdsourced dataset of spelling/meaning annotations
  - AI builds from human intelligence on the source-side

- Smurfs and Ducks
- Kam4D – kamu.si/kam4d
- SlowBrew

kamusi.org

Realistic Fantasies about Computation and Translation



Unanswered Questions:

- Will users take the time to predisambiguate?
  - People take time to choose images
  - People take time to spellchick
- Syntax on the target side?
  - Outside Kamusi wheelhouse – partners needed
- How to pay for it?

- Smurfs and Ducks
- Kam4D – kamu.si/kam4d
- SlowBrew

kamusi.org

How AI Cured Coronavirus and Delivered Universal Translation, and Other MT Myths and Magic

martin@kamusi.org

recommended reading:
- ➢ teachyoubackwards.com
- ➢ kamu.si/kam4d

Martin Benjamin

18 November 2020 Translating and the Computer
ASLING TC42 online
Keynote Address