

日英字幕付き音映像コーパスによる 英語ライティング学習

ミント音声教育研究所 田淵 龍二

tabuchiryuji@nifty.ne.jp



メールはこちら→



← 発表で使う対訳コーパス
コーポラはこちらから開く

<http://www.mintap.com/talkies/pac/corpora.html>

テーマ AI 翻訳の手法で 効率的に学習しよう！

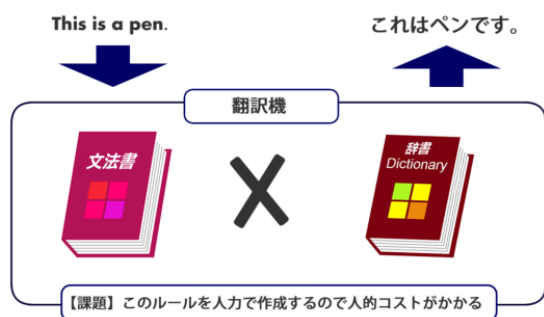
今日のお題

たかが香りで男が変わる

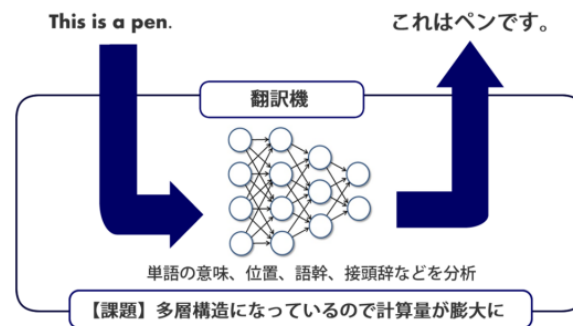
とあるテレビコマーシャルのキャッチフレーズより

はじめに AI 翻訳の歴史

1. ルールベース (1980 年代)

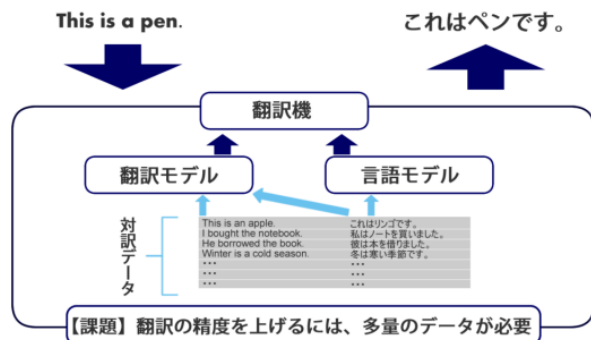


3. ニューラルベース (2010 年代)



*1

2. コーパスベース (1990 年代)



ニューラル機械翻訳

NMT (Neural Machine Translation) は「1 単語 1 単語の対応を陽に記憶する従来法と比べて、実数ベクトルで単語や文脈の類似性を考慮することで、より柔軟に文脈が表現できることが挙げられる。」 *2

*1 Google 翻訳の精度はなぜ上った? <https://gotcha.alc.co.jp/entry/20161220-it-translation1/13> *2 統計的機械翻訳を用いた自然言語からのソースコード生成 (2016, 札幌寛之, 小田悠介, Graham Neubig, 吉野幸一郎, 中村哲, 言語処理学会全国大会予稿集)

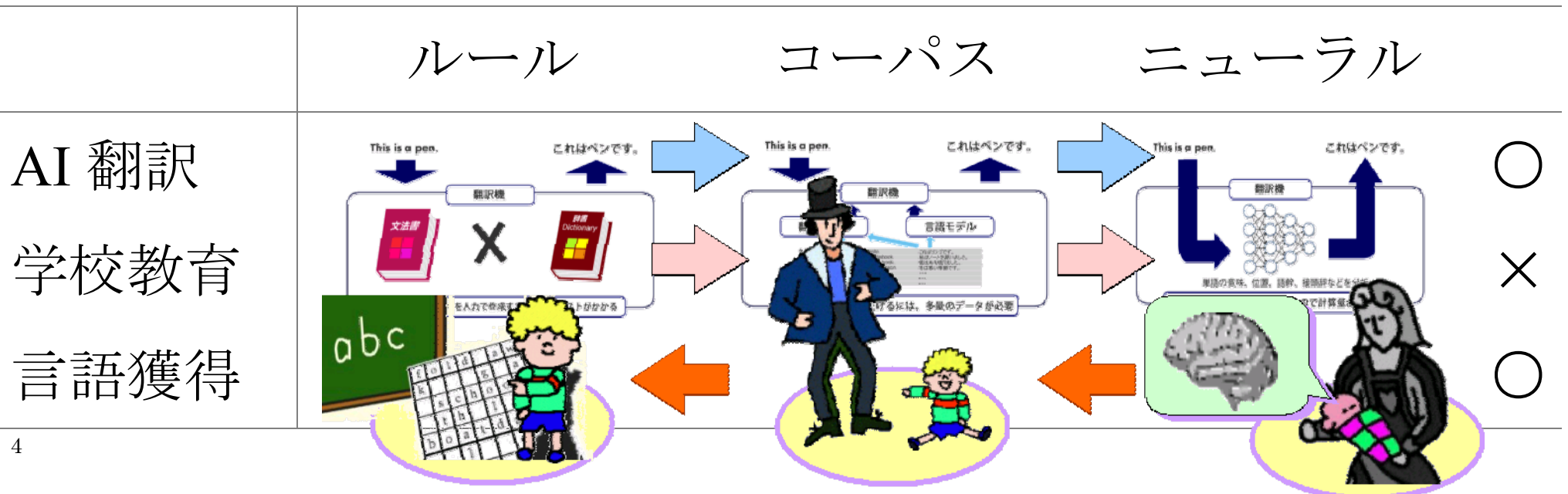
Q : 従来の学習法の効能と限界と突破口はどこにある？

A1 (効能) 英語が必要な (一部の) 人には効果的




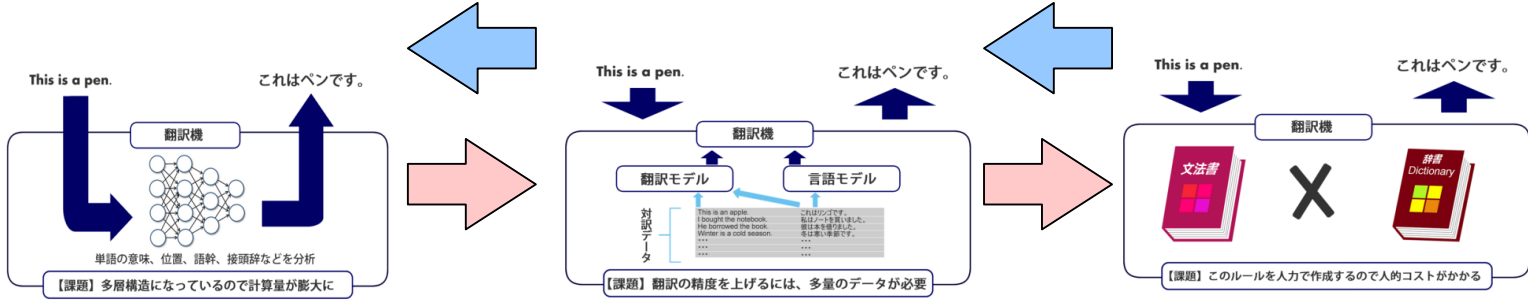
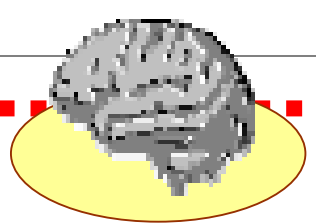


A2 (限界) (一部の) 成功体験は万人に通用しない

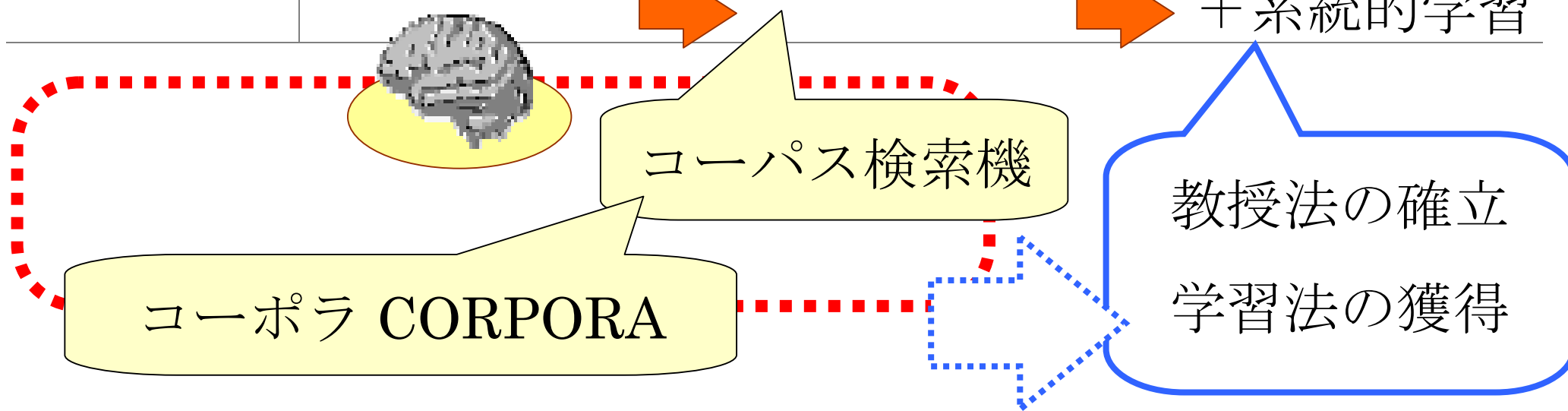
A3 (突破口) AI 翻訳の学習法 ≡ ヒトの言語獲得様式

A : AI 翻訳の手法を真似しよう！



Q : AI 翻訳の手法を真似するためには何が必要？

AI 翻訳	ニューラル  コーパス  ルール 
AI の歴史 言語獲得	 <p>The diagram illustrates the evolution of AI translation in three stages from right to left:</p> <ul style="list-style-type: none"> Stage 1 (Right): Rule-based translation. It shows a '翻訳機' (Translator) box connected to '文法書' (Grammar book) and '辞書' (Dictionary). A large 'X' is placed between them, indicating that manually creating rules is costly. A note below says: 【課題】このルールを人力で作成するので人的コストがかかる (Problem: It costs a lot of human resources to create these rules manually). Stage 2 (Middle): Neural Network-based translation. It shows a '翻訳機' box containing '翻訳モデル' (Translation model) and '言語モデル' (Language model). It also shows '対訳データ' (Parallel data) with examples like 'This is an apple.' and 'これはリンゴです。'. A note below says: 【課題】翻訳の精度を上げるには、多量のデータが必要 (Problem: To improve translation accuracy, a large amount of data is required). Stage 3 (Left): Deep learning-based translation. It shows a '翻訳機' box with a neural network diagram. A note below says: 【課題】多層構造になっているので計算量が膨大に (Problem: Because it has a multi-layer structure, the amount of calculation is enormous).
AI の真似	<p>持っている  コーパス体験  自律習得 + 系統的学習 </p>



日英対訳コーパス CORPORA でライティング

コーポラ CORPORA 基本情報

URL : <http://www.mintap.com/talkies/pac/corpora.html>



		Seleaf	TED Talks
1	作品数	25 films	2,439 talks
2	収録時間	46 時間	610 時間
3	字幕数	43,659	661,787
4	総単語数	0.3M	5M words
5	異なり語数	13K	67K tokens
6	見出し語数	6K	16K lemmas
7	対訳形式	フレーズ訳	文訳 (ズレあり)
8	適応学年	中 2~高 2	高 1~高 3

いずれも言語は英語、訳は日本語

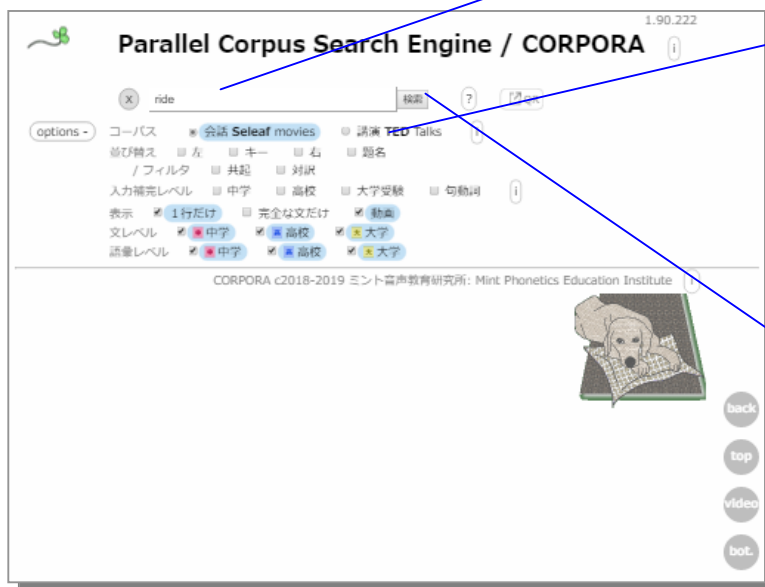
CORPORA 基本機能

1. 語彙検索 — 検索語彙を含むテキスト収集

2. フィルタ — 収集結果を解析し、絞り込む

2.1 対訳フィルタ

2.2 共起フィルタ



1. 検索語を入力

2. コーパスを選ぶ
選べるコーパスは

- Seleaf (映画)
- TED Talks (講演)

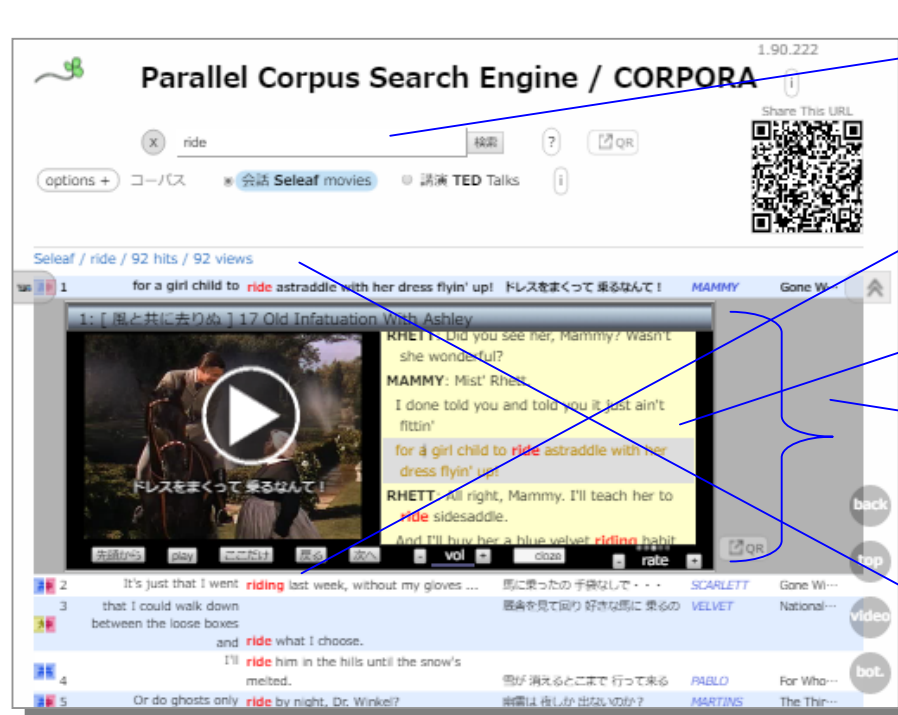
3. 検索ボタンか
Enter キーを押す

1. 語彙検索 — 検索語彙を含むテキスト収集

入力： 英語あるいは日本語

書式： 英語 (正規表現・見出し語検索対応)

結果： 検索語彙を含む字幕と前後 10 数秒の文脈シーン



ride で検索すると

rides や riding もヒット

ride を含む字幕を中心に

前後 10 数秒のシーンが出て
文脈理解を助ける

結果表示は more する

2. フィルタ — 収集結果を解析し、絞り込む

2.1 対訳フィルタ

検索語彙を含む対訳テキストに使われている語彙解析

The screenshot shows a search interface for the word 'ride'. The search results are filtered by '対訳' (Bilingual) and sorted by frequency of kanji characters. The results are as follows:

Rank	Frequency	Character	Source
1	22	乗	雲があって 乗れるんだ
2	11	馬	スペイン人が 馬に乗ってる
3	5	出	彼らが 行ったところ
4	4	行	わかったよ 横乗りにさせるさ
5	3	来	わかった 乗ればいい
6	3	勝	わかった 乗ればいい

1. option を開いて

2. 対訳を選ぶと

3. 対訳フィルタが開く

4. 訳語（漢字）の出現頻度の多い順に並ぶ

• ride の語義「乗」と利用場面が推測できる

2.2 共起フィルタ

検索語彙と関連 (共起) する語句解析

The screenshot shows the Seleaf search interface. The search term is 'interest'. The results are sorted by frequency. A '共起 フィルタ' (Co-occurrence Filter) overlay is visible, showing a list of words and their frequencies. The words are: BE (60), I (46), YOU (36), IN (35), TO (17), NOT (15), THE (14), A (12), OF (11). The word 'IN' is highlighted in blue, indicating it is selected.

1. 共起を選び
2. 共起フィルタを開く
3. 頻度の多い順に出現語彙が並ぶ
4. IN の頻度が高い
(機能語などは薄く表示される)
5. IN を押すと用例が収集される (次ページ)

テーマ AI 翻訳の手法で 効率的に学習しよう！

今日のお題

たかが香りで男が変わる

とあるテレビコマーシャルのキャッチフレーズより



YouTube

手順 1 : Voijjar でお題 (和文) を翻訳

The screenshot shows the Voijjar application interface. At the top, it displays "Cross Learning Notebook β' 1.90.306" and "Voijjar ボイジャー". There are buttons for "二人でチャット", "How to", "お手本を聞く", and "お手本を作る". A QR code is visible with the text "share this work" and "リンクをコピー". Below the header, there is a prompt "↓ 得意な言葉を選ぶ" and a central workspace with a pencil icon. The workspace contains three text boxes: a green box with the Japanese text "たかが香りで男が変わる", a blue box with the English translation "A man changes with a scent", and a white box with the Japanese text "人は香りで変わる". At the bottom left, there is a timestamp "2019-03-08 18:18".



Voijjar

手順 2 : CORPORA で「たかが」を検索して絞りこむ

日英対訳コーパス / コーポラ
1.90.308a

Parallel Corpus Search Engine / CORPORA

key words

options - コーパス 会話 Seleaf movies 講演 TED Talks

並び替え 左 キー 右 題名
/ フィルタ 共起 対訳

入力補完レベル 中学 高校 大学受験 句動詞

表示 1行だけ 完全な文だけ 動画

文レベル 中学(10) 高校(24) 大学(3)

語彙レベル 中学(21) 高校(9) 大学(7)

Share This URL



TED / たかが / 37 hits / 37 views (33.7 秒) (5.19 秒) (3.32 秒)

1	Because exploring a city on foot only gets you so far.	歩いて街を散策しても たかが 知れています	Aris Venetiki... Making ...
2	that in fact it's only a little 10 or 15 percent,	たかが 10-15%に過ぎず	Afra Raymond Three m...
3	it's just a news story.	たかが ニュース記事です	Molly Crockett Beware ...
4	They're just machines, after all.	たかが 機械です	Rodney Brooks Robots ...
5	It's just a bunch of cells. You understand your liver.	脳も、 たかが 細胞の固まり。肝臓なら分かるでしょ	Jeff Hawkins How bra...
7	It's a frog. And you think, frog.	カエル たかが カエルと思うでしょう	Michael Archer How we...
8	This is data, statistics. What does it tell us about these conflicts?	たかが データ、統計です それか戦争の何を教えてくれるのでしょうか	Sean Gourley The mat...
9	It may be television's not so important,	たかが テレビかもしれませんが	Sandi Toksvig A politic...
10	I don't feel like a bunch of cells, you know. I'm conscious.	たかが 細胞の固まりだなんて気がしない、という人もいます。意識があり、	Jeff Hawkins How br...
6	And I realized just what had happened to me at Swarthmore, た	すると、スワースモアで自分に何が起きたか たかが 理解できました	Patrick Awuah How to ...
11	that if you are someone	もし皆さんのどなた たかが	JD Schramm Break th...
12	and one of you is going to do this, right?	ここのだなた たかが 調べてくれるのでしょうか	Sajay Samuel How coll...

back

top

video

bot.

手順 3 : 「たかが」の英語表現を対訳フィルタで絞る

日英対訳コーパス / コーポラ 1.90.308a

Parallel Corpus Search Engine / CORPORA

key words 検索 ? QR

options - コーパス 会話 Seleaf movies **講演 TED Talks** i


並び替え 左 キー 右 題名
/ フィルタ 共起 対訳

入力補完レベル 中学 高校 大学受験 句動詞 i

表示 1行だけ 完全な文だけ 動画

文レベル 中学(10) 高校(24) 大学(3)

語彙レベル 中学(21) 高校(9) 大学(7)

Share This URL


TED / たかが / 37 hits / 37 views (33.7 秒) (5.19 秒) (3.32 秒)

1	Because exploring a city on foot only gets you so far.	歩いて街を散策しても たかが 知れています	Ar
2	that in fact it's only a little 10 or 15 percent,	たかが 10-15%に過ぎず	Afr
3	it's just a news story.	たかが ニュース記事です	Mo
4	They're just machines, after all.	たかが 機械です	Ro
5	It's just a bunch of cells. You understand your liver.	脳も、 たかが 細胞の固まり。肝臓なら分かるでしょ	Jef
6	And I realized just what had happened to me at Swarthmore,	すると、スワースモアで自分に何が起きた たかが 理解できました	Pat
7	It's a frog. And you think, frog.	カエル たかが カエルと思うでしょう	Mic
8	This is data, statistics. What does it tell us about these conflicts?	たかが データ、統計です。それが戦争の何を教えてくれるのでしょうか?	Se
9	It may be television's not so important,	たかが テレビかもしれないんですが	Sav
10	I don't feel like a bunch of cells, you know. I'm conscious.	たかが 細胞の固まりだなんて気がしない、という人もいます。意識があり、	Je
11	that if you are someone	もし皆さんのどな たかが	JL
12	and one of you is going to do this, right?	このどな たかが 調べてくれるでしょう?	S

対訳 フィルタ x

刈り込み +

絞り込みキー +

clear +

ONLY 2

JUST 4

THIS 4

KNOW 4

CAN 4

HAVE 4

ON 3

手順 4 : 案文 (英語) を Voijjar で日本語に翻訳する

The screenshot shows the Voijjar application interface. At the top, it displays "Cross Learning Notebook β 1.90.306" and "Voijjar ボイジャー". There are buttons for "二人でチャット", "How to", "お手本を聞く", and "お手本を作る". A QR code is visible with the text "share this work" and "リンクをコピー". Below the header, there is a prompt "↓ 得意な言葉を選ぶ". The main content area features three text boxes: a green box with the English sentence "A man just changes only with a scent", a blue box with the Japanese translation "男は香りだけで変わる", and a white box with the English sentence "A man changes only by its fragrance". A red circle highlights a pencil icon in the top right corner of the main content area. At the bottom, there is a green box with the Japanese sentence "たかが香りで男が変わる".

Mint Applications 1.90.306

i Cross Learning Notebook β 1.90.306

Voijjar ボイジャー

二人でチャット

How to

お手本を聞く

お手本を作る

share this work

リンクをコピー

↓ 得意な言葉を選ぶ

A man just changes only with a scent

男は香りだけで変わる

A man changes only by its fragrance

たかが香りで男が変わる

3. まとめ

- 悩んだら AI (Voijjar) に聞いてみよう！
- こだわりの表現で和文検索 (CORPORA) しよう！
- 案文ができたなら AI (Voijjar) で試してみよう！

限界：

- 和語の語感を再現する英語表現探索はヒトの仕事
- AI を使いこなすのもそのヒトの力量で決まる
- 隣に母語話者 (英語) がいたら最高！