# Advanced Contextualized Models

Pierluigi Cassotti, pierluigi.cassotti@gu.se

# TempoBERT

- Use time as additional context
- Exploit time masking

YEAR: 1800 ⟶ "<1800> The mountains have an awful majesty."

YEAR: 2020 ⟶ "<2020> You look awful today."

**(a) TempoBERT is trained on temporal corpora, where each sequence is prepended with temporal context information.**

**Time prediction:** "[MASK] Today's weather is awful." ⟶ <2020>

**Time-dependent MLM:** "<1800> He has an awful [MASK]." ⟶ presence

"<2020> He has an awful [MASK]." ⟶ temper

**(b) TempoBERT can be used for inference in two modes: (1) time prediction; (2) time-dependent mask filling.**

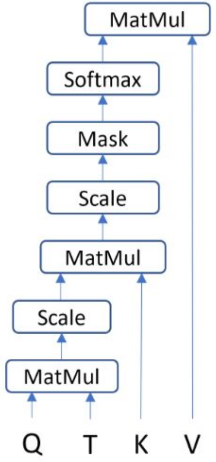
**Figure 1: Example of TempoBERT's time masking for training and inference. The word 'awful' changed its meaning in the last two centuries from marvelous to disgusting.**

Rosin, Guy D., Ido Guy, and Kira Radinsky. "Time masking for temporal language models." *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*. 2022.

# Temporal Attention

- Extends self-attention to include time dimension

$$TemporalAttention(Q, K, V, T) = \mathrm{softmax}\left(\frac{Q\frac{T^{\intercal}T}{\|T\|}K^{\intercal}}{\sqrt{d_k}}\right)V$$




**Time-specific weight matrix**

Guy D. Rosin and Kira Radinsky. 2022. Temporal Attention for Language Models. In *Findings of the Association for Computational Linguistics: NAACL 2022*, pages 1498–1508, Seattle, United States. Association for Computational Linguistics.

# XLM-RoBERTa

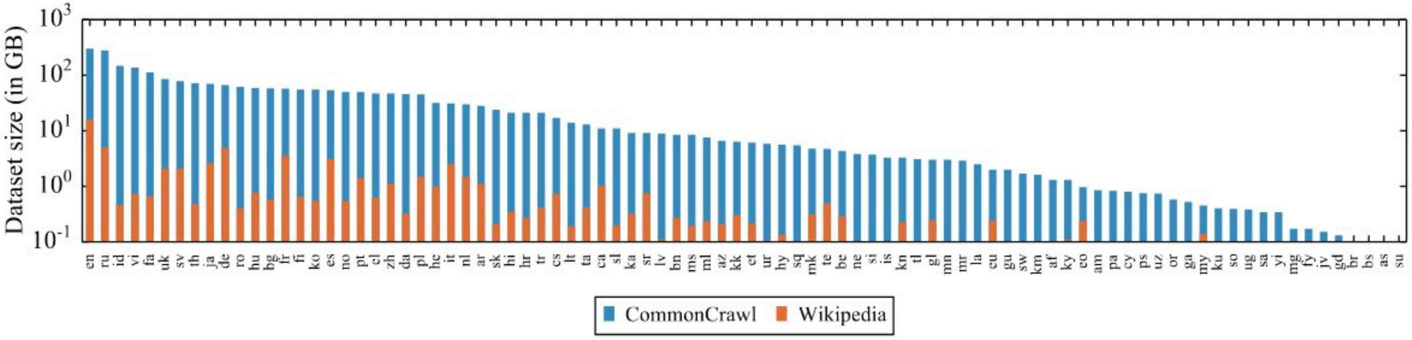



Figure 1: Amount of data in GiB (log-scale) for the 88 languages that appear in both the Wiki-100 corpus used for mBERT and XLM-100, and the CC-100 used for XLM-R. CC-100 increases the amount of data by several orders of magnitude, in particular for low-resource languages.

Sebastian Ruder, Anders Søgaard, and Ivan Vulić. 2019. Unsupervised Cross-Lingual Representation Learning. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Tutorial Abstracts*, pages 31–38, Florence, Italy. Association for Computational Linguistics.

# Gloss Reader

- Rely on XLM-RoBERTa and trained on a English Word Sense Disambiguation (WSD) dataset (SemCor)
- Zero-shot ability on other languages such as Russian

The bank can guarantee deposits will eventually cover future tuition costs because it invests in adjustable-rate mortgage securities.

**Context Encoder**

| | | |
|---|---|---|
| bank[1] | Gloss: | a financial institution that accepts deposits and channels the money into lending activities |

| | | |
|---|---|---|
| bank[2] | Gloss: | sloping land (especially the slope beside a body of water) |

**Gloss Encoder**

Rachinskiy, Maxim, and Nikolay Arefyev. "Zeroshot Crosslingual Transfer of a Gloss Language Model for Semantic Change Detection." *Computational linguistics and intellectual technologies: Papers from the annual conference Dialogue*. 2021.

# Deep Mistake

- Pretrained XLM-R fintuned on MCL-WiC task
- Not depends on fixed sense inventories

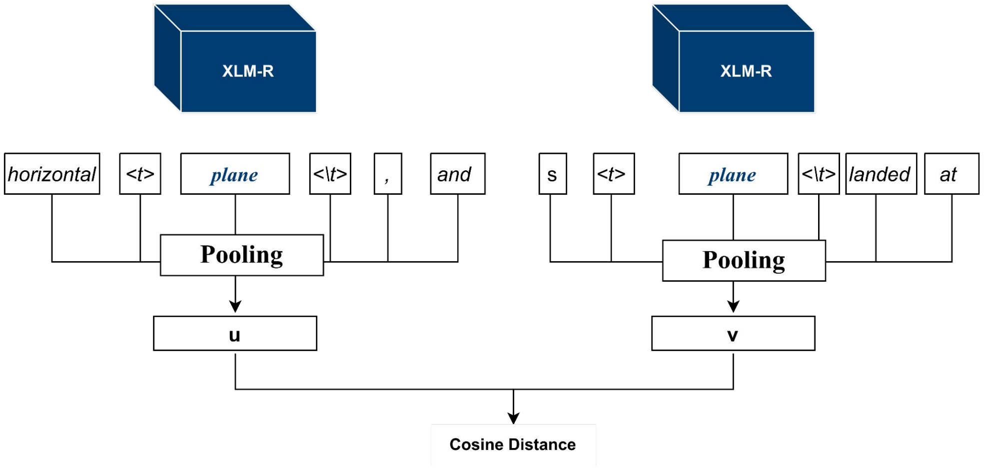
| Lang | Target | Context-1 | Context-2 | Label |
|---|---|---|---|---|
| EN | Beat | We beat the competition | Agassi beat Becker in the tennis championship. | True |
| DA | Tro | Jeg tror p˚a det, min mor fortalte. | Maria troede ikke sine egne øjne. | True |
| ET | Ruum | Uhel hetkel olin v̄ aljaspool aega ja ruumi. | Umberringi oli l̀ oputu t̀ uhi ruum. | True |
| FR | Causticité | Sa causticité lui a fait bien des ennemis. | La causticité des acides. | False |
| KO | 틀림 | 틀림이 있는지 없는지 세어 보시오. | 그 아이 하는 짓에 틀림이 있다면 모두 이 어미 죄이지요. | False |
| ZH | 發 | 建築師希望發大火燒掉城市的三分之一。 | 如果南美洲氣壓偏低，則印度可能發乾旱 | True |
| FA | صرف | صرف غذا نیم ساعت طول کشید | معلم صرف افعال ماضی عربی را آموزش داد | False |

Arefyev, Nikolay, et al. "DeepMistake: Which Senses are Hard to Distinguish for a WordinContext Model." *Computational linguistics and intellectual technologies: Papers from the annual conference Dialogue*. 2021.

# XL-LEXEME

*Provide a large table; this is a horizontal* ***plane****, and will represent the ground plane, viz.*

The President's ***plane*** *landed at Goose Bay at 9:03 p. m.*





Pierluigi Cassotti, Lucia Siciliani, Marco DeGemmis, Giovanni Semeraro, and Pierpaolo Basile. 2023. [XL-LEXEME: WiC Pretrained Model for Cross-Lingual LEXical sEMantic changE](). In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 1577–1585, Toronto, Canada. Association for Computational Linguistics.

# XL-LEXEME

| Dataset | Languages |
|---|---|
| **WiC**<br>Pilehvar et al., (2019 | **Monolingual**<br>EN |
| **XL-WiC**<br>(Raganato et al., 2020) | **Multilingual**<br>EN, BG, ZH, HR, DA, NL, ET, FA, FR, DE, IT, JA, KO |
| **MCL-WiC**<br>(Martelli et al., 2021) | **Multilingual**<br>**EN, AR, FR, RU, ZH** |
| | **Crosslingual**<br>**AR, FR, RU, ZH** |
| **AM²ICO**<br>(Liu et al., 2021) | **Crosslingual**<br>EN, DE, RU, JA, KO, ZH,AR, IN, FI, TR, EU, KA, UR, BN, KK |

Pierluigi Cassotti, Lucia Siciliani, Marco DeGemmis, Giovanni Semeraro, and Pierpaolo Basile. 2023. XL-LEXEME: WiC Pretrained Model for Cross-Lingual LEXical sEMantic changE. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 1577–1585, Toronto, Canada. Association for Computational Linguistics.

# XL-LEXEME

| | | | EN | LA | DE | SV | ES | RU | | | NO | | ZH | Avg$_w$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | $C_1 - C_2$ | $C_1 - C_2$ | $C_1 - C_2$ | $C_1 - C_2$ | $C_1 - C_2$ | $C_1 - C_2$ | $C_2 - C_3$ | $C_1 - C_3$ | $C_1 - C_2$ | $C_2 - C_3$ | $C_1 - C_2$ | $C_i - C_j$ |
| form-based | APD | BERT | .563 | - | .271 | .270 | .335 | .518 | .482 | .416 | .441 | .466 | .656 | .449 |
| | | mBERT | .363 | .102 | .398 | .389 | .341 | .368 | .345 | .386 | .279 | .488 | .689 | .371 |
| | | XLM-R | .444 | .151 | .264 | .257 | .386 | .290 | .287 | .318 | .195 | .379 | .500 | .316 |
| | | XL-LEXEME | **.886*** | **.231** | **.839*** | **.812*** | **.665*** | **.796*** | **.820*** | **.863*** | **.659** | **.640*** | **.731*** | **.751*** |
| | | *SOTA: sup.* | *.757* | *-.056* | *.877* | *.754* | *n.a.* | *.799* | *.833* | *.842* | *.757* | *.757* | *n.a.* | |
| | | *SOTA: uns.* | *.706* | *.443* | *.731* | *.602* | *n.a.* | *.372* | *.480* | *.457* | *.389* | *.387* | *n.a.* | |
| | PRT | BERT | .457 | - | .422 | .158 | .413 | .400 | .374 | .347 | .507 | .444 | **.712** | .406 |
| | | mBERT | .270 | .380 | .436 | .193 | .543 | .391 | .356 | .423 | .219 | .438 | .524 | .395 |
| | | XLM-R | .411 | .424 | .369 | .020 | .505 | .321 | .443 | .405 | .387 | .149 | .558 | .381 |
| | | XL-LEXEME | **.676** | **.506*** | **.824** | **.696** | **.632** | **.704** | **.750** | **.727** | **.764*** | **.519** | .699 | **.693** |
| | | *SOTA: sup.* | *.531* | *n.a.* | *n.a.* | *n.a.* | *n.a.* | *n.a.* | *n.a.* | *n.a.* | *n.a.* | *n.a.* | *n.a.* | |
| | | *SOTA: uns.* | *.467* | *.561* | *.755* | *.392* | *n.a.* | *.294* | *313* | *313* | *.378* | *.270* | *n.a.* | |
| sense-based | AP+JSD | BERT | .289 | - | .469 | - .090 | .225 | .069 | .279 | .094 | **.314** | .011 | .165 | .179 |
| | | mBERT | .181 | .277 | .280 | .023 | .067 | .017 | .086 | - .116 | .035 | - .090 | .465 | .077 |
| | | XLM-R | .278 | **.398** | .224 | -.076 | .224 | - .068 | **.209** | **.130** | - .100 | .030 | .448 | .142 |
| | | XL-LEXEME | **.493** | .033 | **.499** | **.118** | **.392** | **.106** | .053 | .117 | .297 | **.381** | **.308** | **.223** |
| | | *SOTA: sup.* | *n.a.* | *n.a.* | *n.a.* | *n.a.* | *n.a.* | *n.a.* | *n.a.* | *n.a.* | *n.a.* | *n.a.* | *n.a.* | |
| | | *SOTA: uns.* | *.436* | *.481* | *.583* | *.343* | *n.a.* | *n.a.* | *n.a.* | *n.a.* | *n.a.* | *n.a.* | *n.a.* | |
| | WiDiD | BERT | .385 | - | .355 | .106 | .383 | .135 | .102 | .243 | .233 | .087 | .533 | .239 |
| | | mBERT | .323 | - .039 | .312 | .195 | .343 | - .068 | .160 | .142 | .241 | .290 | .338 | .181 |
| | | XLM-R | .564 | - .064 | .499 | .129 | .459 | **.268** | .216 | .342 | .226 | .349 | .382 | .314 |
| | | XL-LEXEME | **.652** | **.236** | **.677** | **.475** | **.522** | .178 | **.354** | **.364** | **.561** | **.457** | **.563** | **.422** |
| | | *SOTA: sup.* | *n.a.* | *n.a.* | *n.a.* | *n.a.* | *n.a.* | *n.a.* | *n.a.* | *n.a.* | *n.a.* | *n.a.* | *n.a.* | |
| | | *SOTA: uns.* | *.651* | *-.096* | *.527* | *.499* | *.544* | *.273* | *.393* | *.407* | *n.a.* | *n.a.* | *n.a.* | |

Periti, F., & Tahmasebi, N. (2024). A Systematic Comparison of Contextualized Word Embeddings for Lexical Semantic Change.

# XL-LEXEME

| | | EN | DE | SV | ES | RU | | | NO | | ZH | Avg$_w$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $C_1 - C_2$ | $C_1 - C_2$ | $C_1 - C_2$ | $C_1 - C_2$ | $C_1 - C_2$ | $C_2 - C_3$ | $C_1 - C_3$ | $C_1 - C_2$ | $C_2 - C_3$ | $C_1 - C_2$ | $C_i - C_j$ |
| WiC | BERT | .503 | .350 | .221 | .319 | .314 | .344 | .350 | .429 | .406 | .516 | .358 |
| | mBERT | .332 | .344 | .284 | .289 | .280 | .273 | .293 | .283 | .333 | .413 | .301 |
| | XLM-R | .352 | .289 | .255 | .288 | .212 | .250 | .251 | .317 | .261 | .392 | .272 |
| | XL-LEXEME | **.626** | **.628** | **.631** | **.547** | **.549** | **.558** | **.564** | **.484** | **.521** | **.630** | **.568** |
| | GPT-4.0 | .606 | - | - | - | - | - | - | - | - | - | - |
| | *Agreement* | *.633* | *.666* | *.672* | *.531* | *.531* | *.567* | *.564* | *.761* | *.667* | *.602* | *.593* |
| WSI | BERT | .136 / .700 | .047 / .662 | .023 / .596 | .189 / .695 | - / - | - / - | - / - | .251 / .771 | .247 / .758 | .279 / .759 | .166 / .702 |
| | mBERT | .067 / .644 | .054 / .679 | .024 / .648 | .228 / .700 | - / - | - / - | - / - | .241 / .759 | .159 / .753 | .172 / .713 | .146 / .696 |
| | XLM-R | .068 / .737 | .024 / .725 | .031 / .680 | .164 / .755 | - / - | - / - | - / - | .179 / .775 | .183 / .715 | .279 / .806 | .133 / .743 |
| | XL-LEXEME | .273 / .834 | **.300 / .788** | **.249 / .766** | **.400 / .820** | - / - | - / - | - / - | **.337 / .806** | **.304 / .808** | **.448 / .836** | **.339 / .810** |
| | GPT-4.0 | **.340 / .877** | - / - | - / - | - / - | - / - | - / - | - / - | - / - | - / - | - / - | - / - |
| GCD | BERT | .425 | .116 | .148 | .284 | .487 | .452 | .469 | .571 | .521 | **.808** | .422 |
| | mBERT | .120 | .205 | .234 | .394 | .372 | .325 | .408 | .290 | .454 | .737 | .357 |
| | XLM-R | .219 | .069 | .143 | .464 | .284 | .301 | .375 | .395 | .345 | .557 | .324 |
| | XL-LEXEME | .801 | **.799** | **.721** | **.655** | **.780** | **.824** | **.851** | **.620** | **.567** | .716 | **.754** |
| | GPT-4.0 | **.818** | - | - | - | - | - | - | - | - | - | - |

Periti, F., & Tahmasebi, N. (2024). A Systematic Comparison of Contextualized Word Embeddings for Lexical Semantic Change.

# Definition Generation

FLAN T5

Mario Giulianelli, Iris Luden, Raquel Fernandez, and Andrey Kutuzov. 2023. [Interpretable Word Sense Representations via Definition Generation: The Case of Semantic Change Analysis](). In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3130–3148, Toronto, Canada. Association for Computational Linguistics.

# Definition Generation

| | | WordNet | | | Oxford | | |
|---|---|---|---|---|---|---|---|
| Model | Test | BLEU | ROUGE-L | BERT-F1 | BLEU | ROUGE-L | BERT-F1 |
| Huang et al. (2021) | *Unknown* | 32.72 | - | - | **26.52** | - | - |
| Flan-T5 XL | Zero-shot (task shift) | 2.70 | 12.72 | 86.72 | 2.88 | 16.20 | 86.52 |
| Flan-T5 XL | In-distribution | 11.49 | 28.96 | 88.90 | 16.61 | 36.27 | 89.40 |
| Flan-T5 XL | Hard domain shift | 29.55 | 48.17 | 91.39 | 8.37 | 25.06 | 87.56 |
| Flan-T5 XL | Soft domain shift | **32.81** | **52.21** | **92.16** | 18.69 | **38.72** | **89.75** |

Table 3: Results of the definition generation experiments.

Mario Giulianelli, Iris Luden, Raquel Fernandez, and Andrey Kutuzov. 2023. Interpretable Word Sense Representations via Definition Generation: The Case of Semantic Change Analysis. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3130–3148, Toronto, Canada. Association for Computational Linguistics.

# Definition Generation

| Usage example | Target word | Generated definition |
|---|---|---|
| ‘about half of the soldiers in our rifle platoons were **draftees** whom we had trained for about six weeks’ | **draftee** | ‘A PERSON WHO IS BEING ENLISTED IN THE ARMED FORCES’ |

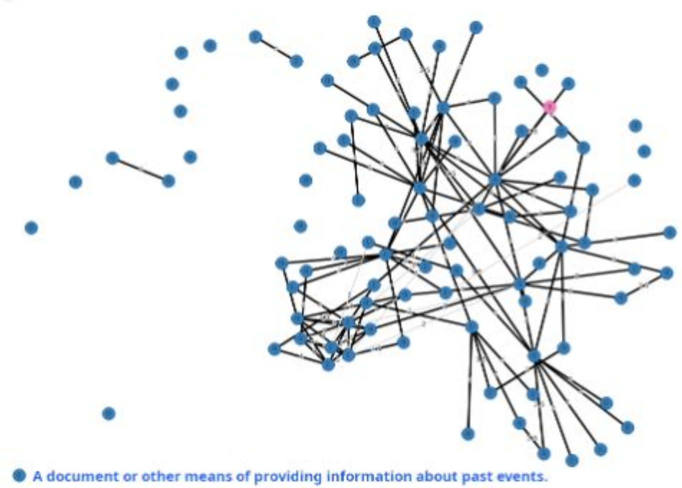
Table 1: An example of a definition generated by our fine-tuned Flan-T5 XL. The model is prompted with the usage example, post-fixed with the phrase *‘What is the definition of draftee?’*

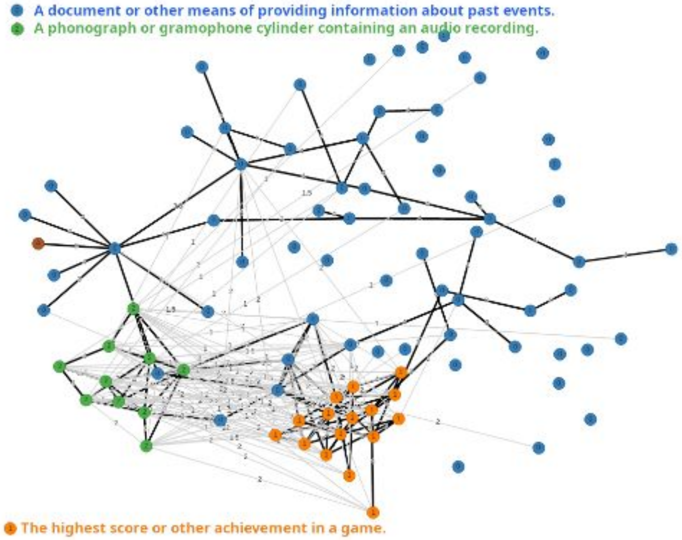
| Method | Cosine | SacreBLEU | METEOR |
|---|---|---|---|
| Token embeddings | 0.141 | - | - |
| Sentence embeddings | 0.114 | - | - |
| **Generated definitions** | | | |
| Flan-T5 XL Zero-shot | 0.188 | 0.041 | 0.083 |
| Flan-T5 XXL Zero-shot | 0.206 | 0.045 | 0.092 |
| Flan-T5 base FT | 0.221 | 0.078 | 0.077 |
| Flan-T5 XL FT | **0.264** | **0.108** | **0.117** |

Table 4: Correlations with pairwise similarity judgements by humans. ‘FT’ stands for ‘fine-tuned model’.

Mario Giulianelli, Iris Luden, Raquel Fernandez, and Andrey Kutuzov. 2023. Interpretable Word Sense Representations via Definition Generation: The Case of Semantic Change Analysis. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3130–3148, Toronto, Canada. Association for Computational Linguistics.

# Definition Generation




Mario Giulianelli, Iris Luden, Raquel Fernandez, and Andrey Kutuzov. 2023. Interpretable Word Sense Representations via Definition Generation: The Case of Semantic Change Analysis. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3130–3148, Toronto, Canada. Association for Computational Linguistics.

# Substitution-based

| | GEMS | SE Eng | SE Ger | SE Lat | SE Swe | Average | Average (weighted) |
|---|---|---|---|---|---|---|---|
| Number of words | 96* | 37 | 40 | 48 | 31 | | |
| *Static Embedding Methods* | | | | | | | |
| Pömsl and Lyapin (2020) | - | 0.422 | **0.725** | 0.412 | 0.547 | - | - |
| Montariol et al. (2021) [static] | 0.347 | 0.321 | 0.712 | 0.372 | **0.631** | 0.477 | 0.452 |
| *Contextual Embedding Methods* | | | | | | | |
| Martinc et al. (2020b) | 0.510 | 0.313 | 0.436 | 0.467 | -0.026 | 0.340 | 0.394 |
| Montariol et al. (2021) [contextual] | 0.352 | 0.437 | 0.561 | 0.488 | 0.321 | 0.432 | 0.422 |
| Scaled JSD | **0.535** | **0.547** | 0.563 | **0.533** | 0.310 | **0.498** | **0.514** |

Dallas Card. 2023. Substitution-based Semantic Change Detection using Contextual Embeddings. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 590–602, Toronto, Canada. Association for Computational Linguistics.

# Substitution-based

| Word | SE rating | SE rank | Scaled JSD | Scaled JSD rank | Corpus A substitutes (1810–1860) | Corpus B substitutes (1960–2010) |
|---|---|---|---|---|---|---|
| plane | 0.88 | 1 | 0.97 | 1 | plane line planes point surface lines | plane aircraft planes jet airplane car |
| graft | 0.55 | 4 | 0.97 | 2 | tree plant stock vine fruit wood | corruption bribery fraud crime violence |
| tip | 0.68 | 2 | 0.85 | 7 | tipped tip covered end filled tips give | tip tips end tipped edge point top ends |
| gas | 0.16 | 23 | 0.72 | 14 | gas gases vapor air fire water | gas gasoline oil gases fuel water air |
| head | 0.30 | 10 | 0.68 | 16 | head face hand heads hands eyes | head face heads hand body hands eyes |
| bit | 0.31 | 9 | 0.51 | 23 | bit piece sort little pieces bits kind | bit little lot touch tad piece bits pieces |
| fiction | 0.02 | 35 | 0.41 | 27 | fiction history literature art poetry | fiction fact fantasy story stories novels |
| tree | 0.07 | 33 | 0.22 | 33 | trees tree plants branches plant wood | trees tree plants woods branches bushes |
| ounce | 0.28 | 11 | 0.08 | 37 | ounce inch pounds hour acre dollars | ounce pounds inch inches cups pieces |

Dallas Card. 2023. Substitution-based Semantic Change Detection using Contextual Embeddings. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 590–602, Toronto, Canada. Association for Computational Linguistics.