

BERTje neemt een badje, but for how long?

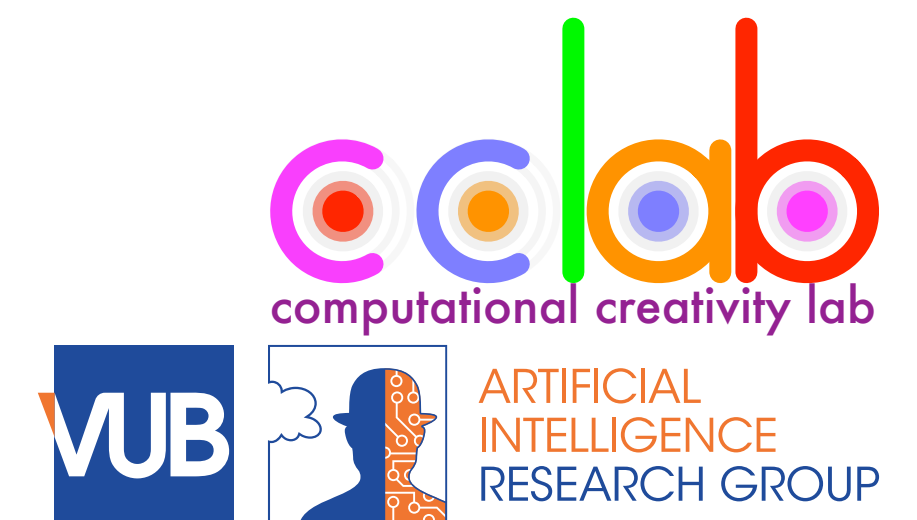
BERTje takes a little bath

Studying event duration of Dutch light verb constructions

CLIN 2026, Brussels (Belgium)

Lin de Huybrecht & Geraint A. Wiggins

Vrije Universiteit Brussel



Durative events in Light Verb Constructions* are perceived differently than their full verb counterparts (Wittenberg & Levy, 2017)

Light Verb Construction

Alice geeft Bob een por.

Alice gives Bob a poke.

<

Full Verb Construction

Alice port Bob.

Alice pokes Bob.

perceived as
taking less time

*Other names for Light Verb Construction (LVC) include support verb construction, stretched verb construction, complex predicate, expanded predicate and verbonominal construction (Gilquin, 2019; Ronan and Schneider, 2015).

Durative events in Light Verb Constructions are perceived differently than their full verb counterparts (Wittenberg & Levy, 2017)

Human experiments

By asking participants to estimate and categorise the duration of events (Wittenberg & Levy, 2017)

Computational experiments

By semantically projecting

- contextualised BERT embeddings (Liu & Chersoni, 2023)
- English co-occurrence count-based vectors (Huybrecht & Wiggins, 2026) onto a Duration scale

Durative events in Light Verb Constructions are perceived differently than their full verb counterparts (Wittenberg & Levy, 2017)

RQ: Can we find **the same effect** in Dutch?

- by using the same semantic projection technique (Grand et al., 2022) as Liu & Chersoni (2023) and Huybrecht & Wiggins (2026)
 - in a Flemish Dutch co-occurrence count-based vector space
 - and with BERTje embeddings
- using a dataset of 158 Dutch LVC-FVC pairs we built based on the English dataset (Huybrecht & Wiggins, 2026)

We use a broad definition of **Light Verb Constructions** (Gilquin, 2019)

LVC = eventative deverbal noun
+ delexical verb

onderzoek doen
(do research)

een **beslissing** maken
(take a decision)

moeite doen
(make an effort)

FVC

onderzoeken
(research)

beslissen
(decide)

proberen
(try)

Experiment 1:
studying event duration of Dutch LVC-FVC pairs
using a co-occurrence count-based vector space

We built our co-occurrence count-based vector space based on the Flemish part of Corpus Gesproken Nederlands

- a basis of 2000 content tokens:
 - excluding prepositions, pronouns, tokens consisting of only one character
 - excluding the 50 most frequently occurring lemmas
- a 5x5 tokens context window, minimum sentence length of 5
- Positive Pointwise Mutual Information (PPMI) as a weighting scheme

$$PPMI = \max \left(\log_2 \left(\frac{p(T, C)}{p(T)p(C)} \right), 0 \right)$$

We semantically project our vectors onto a Duration scale

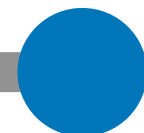
(Grand et al., 2022)

(Liu & Chersoni, 2023)

$$Proj = \frac{\overrightarrow{target} \cdot \overrightarrow{DURATION}}{||\overrightarrow{DURATION}||}$$

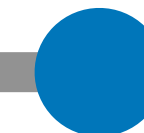
adj.: beknopt, onmiddellijk,
kortstondig, kort
noun: minuut, moment,
ogenblik, seconde

short



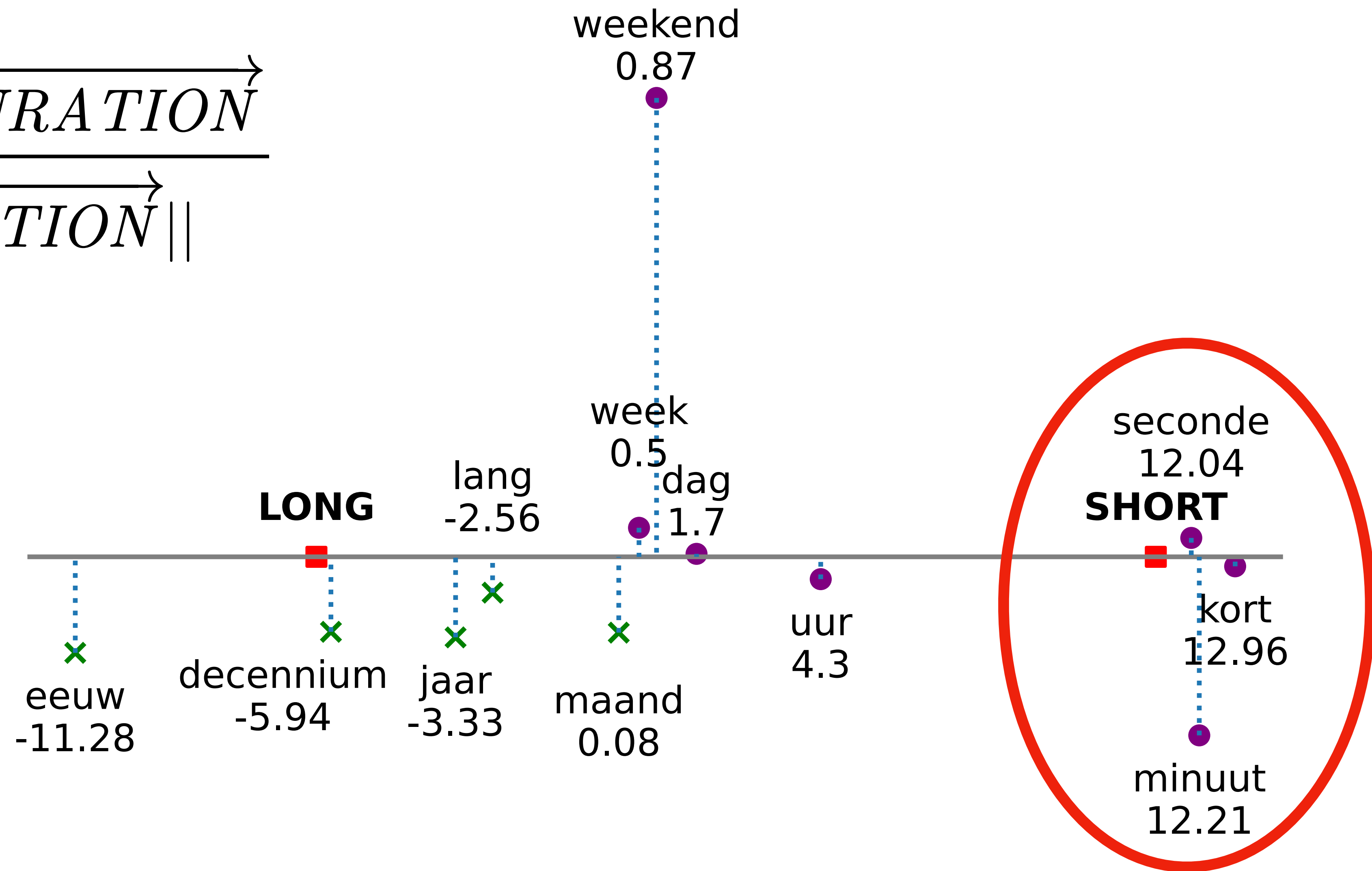
adj.: jarenlang, lang
langdurig, uitvoerig,
voortdurend
noun: decennium, eeuw,
eeuwigheid, jaar

long



We calibrate our Duration scale

$$Proj = \frac{\overrightarrow{target} \cdot \overrightarrow{DURATION}}{\|\overrightarrow{DURATION}\|}$$



We calibrate our Duration scale

shorter
duration

Lemma	EN translation	POS	Proj	Abs. Freq.
kort	short	adjective	12.960	520
minuut	minute		12.210	801
seconde	second		12.043	150
uur	hour		4.298	2,251
dag	day	noun	1.702	2,462
weekend	weekend		0.867	561
week	week		0.500	2,013
maand	month		0.077	914
lang	long	adjective	-2.560	1,754
jaar	year	noun	-3.333	4,642
decennium	decade		-5.939	16
eeuw	century		-11.282	221

longer
duration

EN translation
problem
sentence:

*you have been
talking for **10
min.**, the
connection will
be terminated
in **30 sec.***

We semantically project 134 Dutch LVC-FVC pairs in isolation onto our Duration scale and find a statistically significant difference

een por geven

give a poke

porren

poke

LVC = article + nominal + Light Verb

<

FVC = Full Verb

additive
compositionality

modelled as
taking less time

We semantically project 134 Dutch LVC-FVC pairs in isolation onto our Duration scale and find a statistically significant difference

LVC = article + nominal + Light Verb

additive compositionality

<

FVC = Full Verb

modelled as taking less time

	LVC proj.	FVC proj.
Min.	-1.633	-0.859
Q1	2.975	0.173
Median	4.086	0.892
Mean	4.001	1.048
Std.	1.473	1.194
Q3	4.864	1.633
Max.	7.521	4.634

Paired T-test where $\alpha = 0.05$

p-value $< 2.2 \times 10^{-16}$

$t(133) = 19.223$, CI $[2.698, +\infty)$

Cohen's d of 2.199, 95% CI $[1.783, 2.616]$

Exp. 2: studying event duration of Dutch
LVC-FVC pairs using BERTje's embeddings

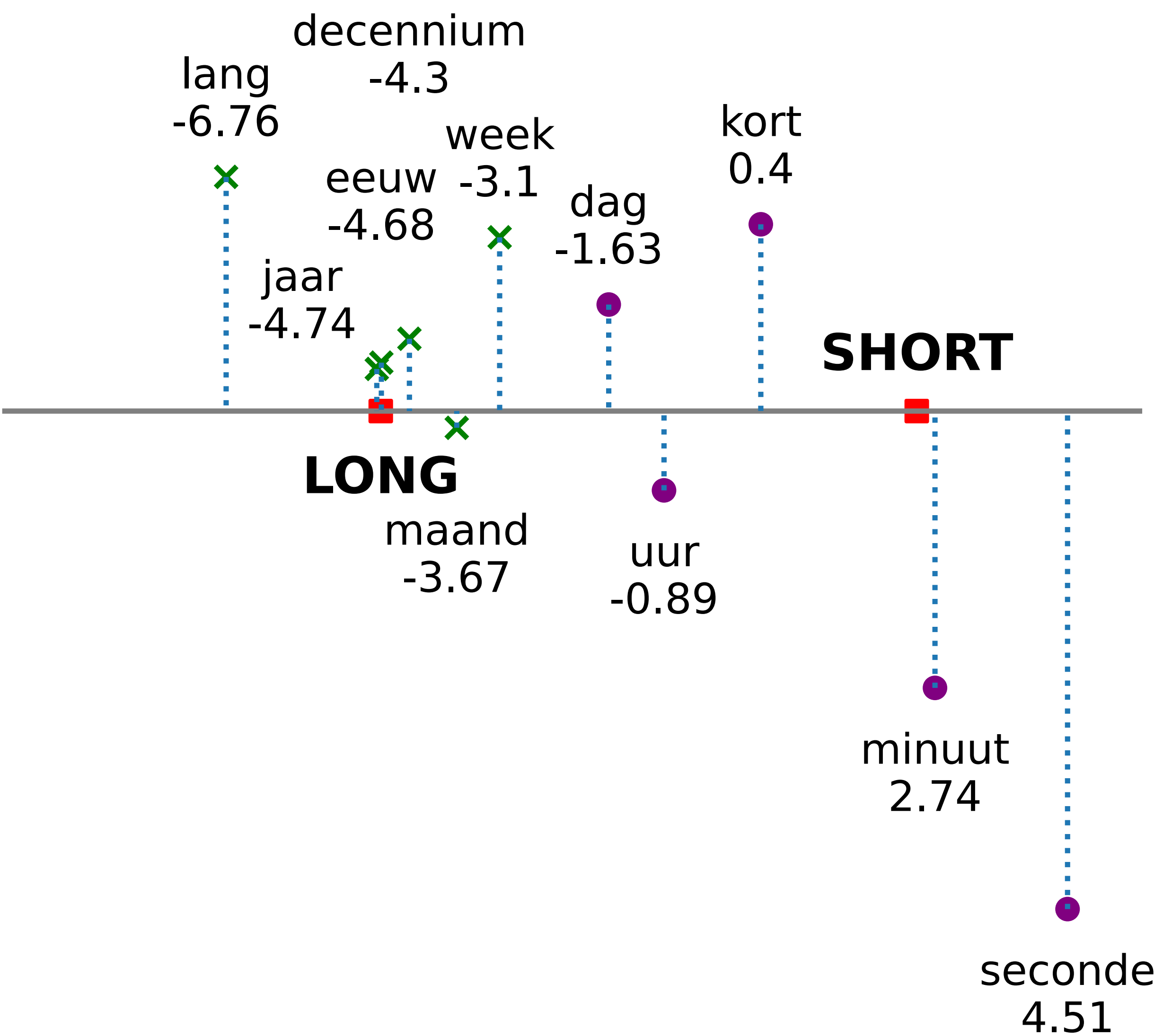
We used the base cased version of BERTje (de Vries et al., 2019) to study event duration of Dutch LVC-FVC pairs

We obtain the embeddings from BERTje by using the values of the last layer:

- LVCs: we use the contextualised embedding of the nominal of the LVC (Liu & Chersoni, 2023)
 - een por geven (give a poke)
- FVCs: porren (to poke)

When calibrating our projection scale for BERTje, the relative order of all lemmas makes sense, except for "jaar" (year)

Lemma	EN Translation	Proj.
seconde	second	4.512
minuut	minute	2.736
kort	short	0.403
uur	hour	-0.894
dag	day	-1.634
week	week	-3.096
maand	month	-3.671
decennium	decade	-4.305
eeuw	century	-4.681
jaar	year	-4.742
lang	long	-6.760



Using embeddings from BERTje, we semantically project
158 Dutch LVC-FVC pairs

een **por** geven (give a **poke**)

proj(contextualised LVC **nominal**)

porren (poke)

proj(FVC)

Using embeddings from BERTje, we semantically project 158 Dutch LVC-FVC pairs and find a statistically significant difference

$\text{proj}(\text{contextualised LVC nominal}) < \text{proj}(\text{FVC})$
modelled as
taking less time

construction	LVC nom.	FVC
Min.	-2.659	-2.392
Q1	-0.707	-1.432
Median	-0.110	-0.306
Mean	-0.109	-0.920
Std.	0.936	0.683
Q3	0.401	-0.452
Max.	2.692	1.051

Right tailed paired T-test where $\alpha = 0.05$

p-value < 2.2×10^{-16}

$t(157) = 11.530$, CI $[0.706, +\infty)$

Cohen's d of 0.990, 95% CI $[0.784, 1.197]$

For BERTje embeddings, we do not find a statistically significant difference in duration between 134 pairs of normal and diminutive LVC nominals

een **por** geven (give a **poke**)

proj(contextualised LVC nominal)

een **porretje** geven (give a little **poke**)

proj(contextualised LVC
diminutive nominal)

For BERTje embeddings, we do not find a statistically significant difference in duration between 134 pairs of normal and diminutive LVC nominals

proj(contextualised LVC nominal)

≈

proj(contextualised LVC diminutive nominal)

no statistically significant difference

proj	LVC nom.	LVC nom. dim.
Min.	-1.723	-1.725
Q1	-0.511	-0.494
Median	-0.016	-0.139
Mean	0.049	-0.072
Std.	0.867	0.671
Q3	0.474	0.416
Max.	2.692	1.975

Paired Wilcoxon signed-rank test

$\alpha = 0.05$

W= 3,321 and p-value = 0.098

eff. size $r = 0.147$, 95% CI [0.01, 0.33]

Studying event duration of Dutch Light Verb Constructions

Embeddings	compared constructions	statistically significant result
co-occurrence count-based	LVC & FVC	yes, LVC (nominal) shorter than FVC
BERTje	LVC nominal & FVC	
	LVC nominal & LVC diminutive nominal	no statistically significant difference

Studying event duration of Dutch Light Verb Constructions

- We find evidence that choice of syntactic construction influences modelled event duration, in both experiments using Dutch data.
- Further work includes
 - studying event duration of LVCs in context (within sentences)
 - studying the relationship between LVCs and FVCs in the semantic space

If this *sparks an interest*,
then *give* this QR code *a scan*
so we can *have a chat*!



We preprocessed the Flemish part of Corpus Gesproken Nederlands (CGN)

- by removing all metadata and punctuation
- by only keeping the lemmas and part-of-speech tags
 - keeping the original capitalisation: Geel (city) vs geel (colour yellow)

lemma₁_POS₁▯lemma₂_POS₂▯...▯lemma_N_POS_N\n

POS ∈ {ADJ, ADV, SUBST, VERB, PREP, PRON}

Geel_SUBST zijn_VERB een_ART stad_SUBST

Geel_SUBST be_VERB a_ART city_SUBST

We use a broad definition of **Light Verb Constructions** (Gilquin, 2019)

1. Eventative deverbal noun

+

2. Delexical verb

- identical to a verb:
onderzoek doen - onderzoeken
(do research)
- derived from a verb:
een beslissing maken - beslissen
(take a decision - decide)
- semantically related to a verb:
moeite doen - proberen
(make an effort - try)

- high frequency verbs:
*geven (give), maken (make),
nemen (take), hebben*
(have), *doen (do)*
- general verbs:
bieden (offer)
- other verbs:
plegen (commit)