



LATVIJAS
UNIVERSITĀTE

Loģikas "kļūdu" noteikšana

Autori:

Haralds Upītis hu21001

Paulis Kristofers Siks ps21049

Piemēri

- Ad hominem (atsaukšanās uz personību)

"Roberts Kīlis tagad stāstīs par izglītības politiku? Viņš taču uz debati ieradies šortos un sandalēs ar zeķēm!"

- Slidenā nogāze (slippery slope)

"Ja mēs tagad šim klientam iedosim atlaidi, rīt visi klienti par to zinās un aizparīt mēs varēsim bodīti klapēt ciet!"

- Pārsteidzīgs vispārinājums (faulty generalization)

"Latvijā nedzīvo neviens savvaļas zilonis, tāpēc ziloni nedzīvo savvaļā."

Mērķis

Sākotnējais mērķis:

Loģikas kļūdu un negodīgas pārliecināšanas paņēmieni noteikšana teikumā **latviešu valodā**.

Reālais mērķis:

Loģikas kļūdu un negodīgas pārliecināšanas paņēmieni noteikšana teikumā **angļu valodā**.

Galvenie izaicinājumi (latviešu valoda)

- Viens coreference resolution risinājums priekš latviešu valodas - "lvcoref" [4], likumbāzēts, prasa daudz informāciju ievadē, bez pilnvērtīgas dokumentācijas.
- Atradām vienu NLI zero-shot klasifikatoru [5], kurš apgalvoja, ka ir trenēts arī ar latviešu valodas datiem, bet nedeva labus rezultātus.

Esošie risinājumi

- Trīs publikācijas no 2022. [1] (ņēmām par pamatu mūsu risinājumam), 2023. [2] un 2024. [3] gada 18. aprīļa (nebijām atraduši risinājuma laikā).
- Jaunākajā risinājumā tiek lietoti few-shot klasifikatori, kuri dod daudz labākus rezultātus par zero-shot klasifikatoriem.

Esošie risinājumi

	F_1	P	R	Acc
Random	12.02	7.24	35.00	0.00
<i>Zero-shot classifiers directly tested on LOGIC</i>				
TARS	8.62	3.86	6.67	2.33
BART-MNLI	11.05	6.63	33.67	0.00
GPT3	12.20	12.00	12.00	12.00
RoBERTa-MNLI	12.22	7.51	36.00	0.33
GPT2	13.67	13.67	13.67	13.67
<i>Finetuned and tested on LOGIC</i>				
ALBERT	12.50	6.67	100.00	0.00
BigBird	15.02	8.61	90.00	0.33
DistilBERT	26.96	22.06	74.00	4.67
MobileBERT	35.68	29.05	71.00	7.33
BERT	45.80	40.73	73.67	18.00
DeBERTa	50.29	45.79	73.00	24.67
Electra	53.31	51.59	72.33	35.66
Electra- <i>StructAware</i>	58.77	55.25	63.67	47.67

2022. gada [1]

Model	Type	LOGIC			LOGIC Climate		
		P	R	F1	P	R	F1
Freq-based	baseline	0.094	0.094	0.093	0.120	0.079	0.080
Codex	few-shot	0.594	0.422	0.386	0.198	0.093	0.077
ELECTRA	baseline	0.614	0.602	0.599	0.276	0.229	0.217
	CBR	0.663	0.664	0.657	0.355	0.254	0.270
RoBERTa	baseline	0.577	0.561	0.560	0.237	0.211	0.200
	CBR	0.631	0.619	0.619	0.379	0.248	0.245
BERT	baseline	0.585	0.598	0.586	0.166	0.130	0.120
	CBR	0.613	0.616	0.611	0.359	0.204	0.200

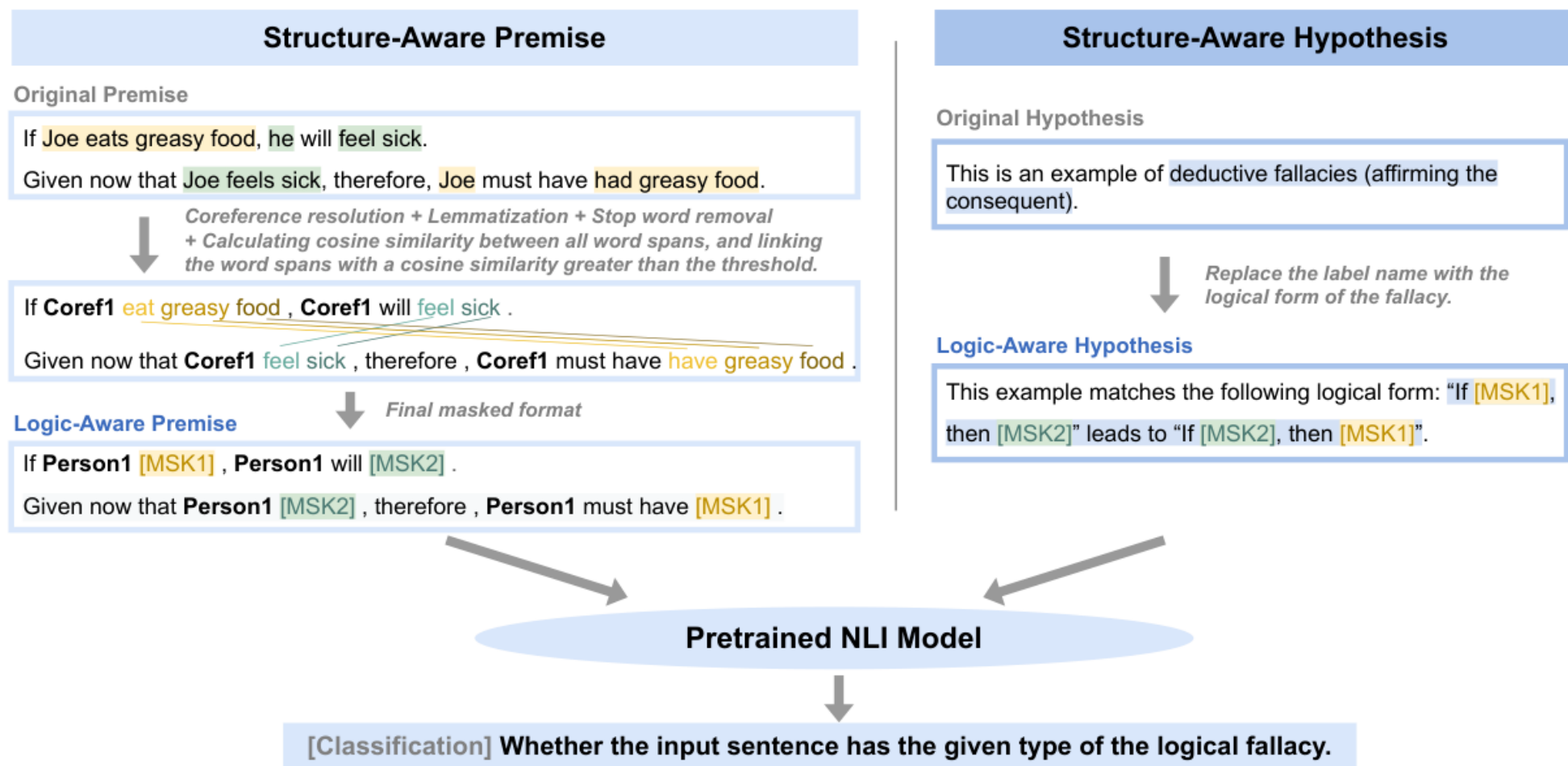
2023. gada [2]

4.3 Main Results

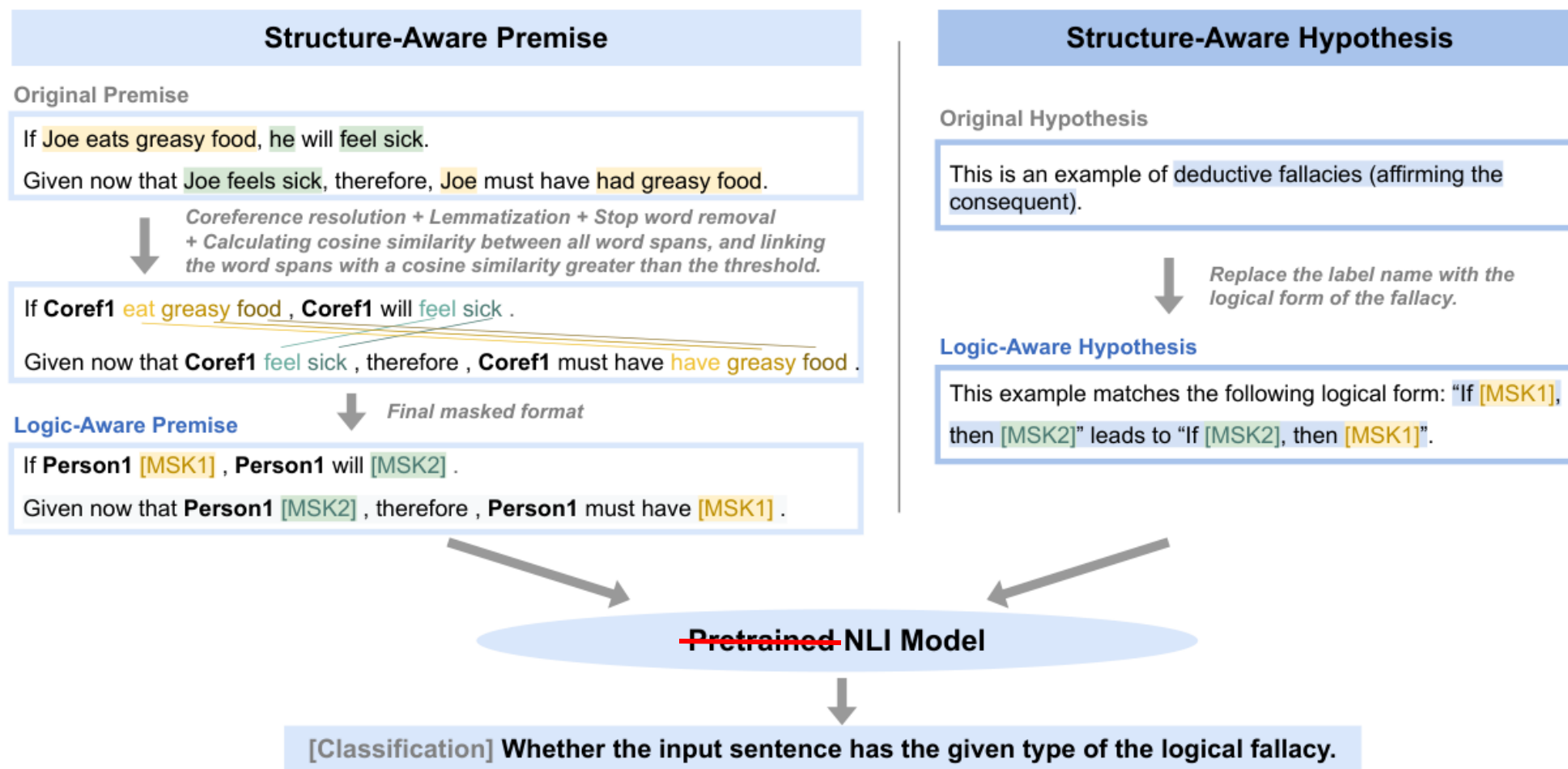
Model	Acc	P	R	F1
BART-MNLI (Zero Shot)	0.58	1	0.15	0.26
Llama-7b (Few Shot)	0.41	0.45	0.82	0.58
Mistral-7b-Instruct (Few Shot)	0.85	0.85	0.86	0.85
GPT3.5 (Few Shot)	0.88	0.86	0.91	0.89
GPT4 (Few Shot)	0.95	0.97	0.94	0.95
GPT4 (Few Shot with COT)	0.94	0.95	0.94	0.94
Claude 3 Opus (Few Shot)	0.97	0.96	0.98	0.97
NL2FOL (Few Shot)	0.63	0.58	0.92	0.71

2024. gada [3]

Mūsu risinājuma pamats



Mūsū risinājums



Premisas maskēšana

Orģinālā premisa

If Joe eats greasy food, he will feel sick.
Given now that Joe feels sick, therefore, Joe must have had greasy food

Koreference + lemmatizācija

If [MSK1] eat greasy food, [MSK1] will feel sick.
Give now that [MSK1] feel sick, therefore, [MSK1] must have have greasy food

Visi vardu sprīži

If;eat; eat greasy; eat greasy food;
greasy; greasy food; ...

Vārdu sprīžu pāri
ar cos-sim > .9

{greasy; have greasy; 0.91}
{greasy; have greasy; 0.92}
{greasy food; have greasy food; 0.92}
...

sprīžu grupas

[eat greasy; have eat greasy]

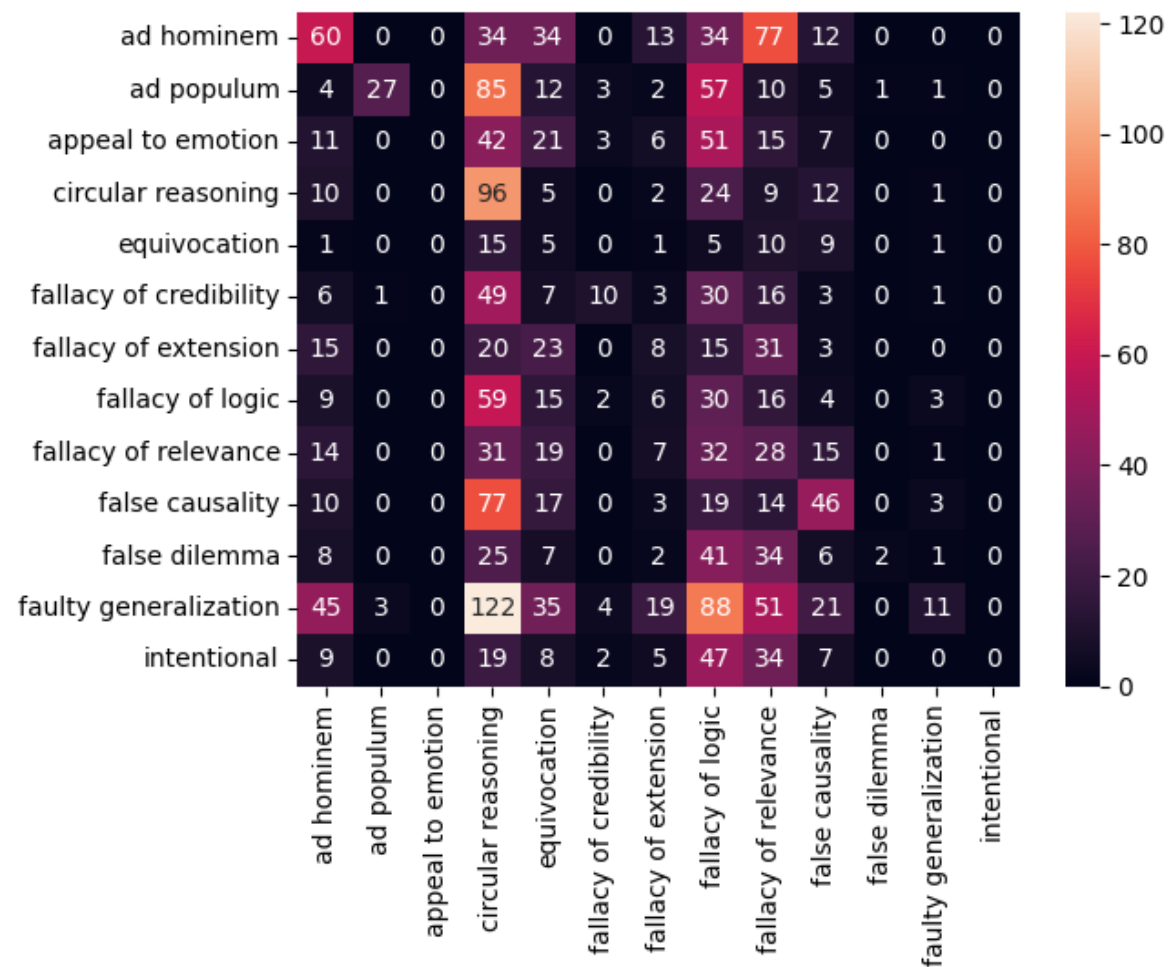
Aizvieto sprīžus ar maksu

If [MSK1] eats [MSK2], [MSK1] will [MSK3].
Given now that [MSK1] [MSK3], therefore, [MSK1] must have [MSK2]

Izmantotie modeļi

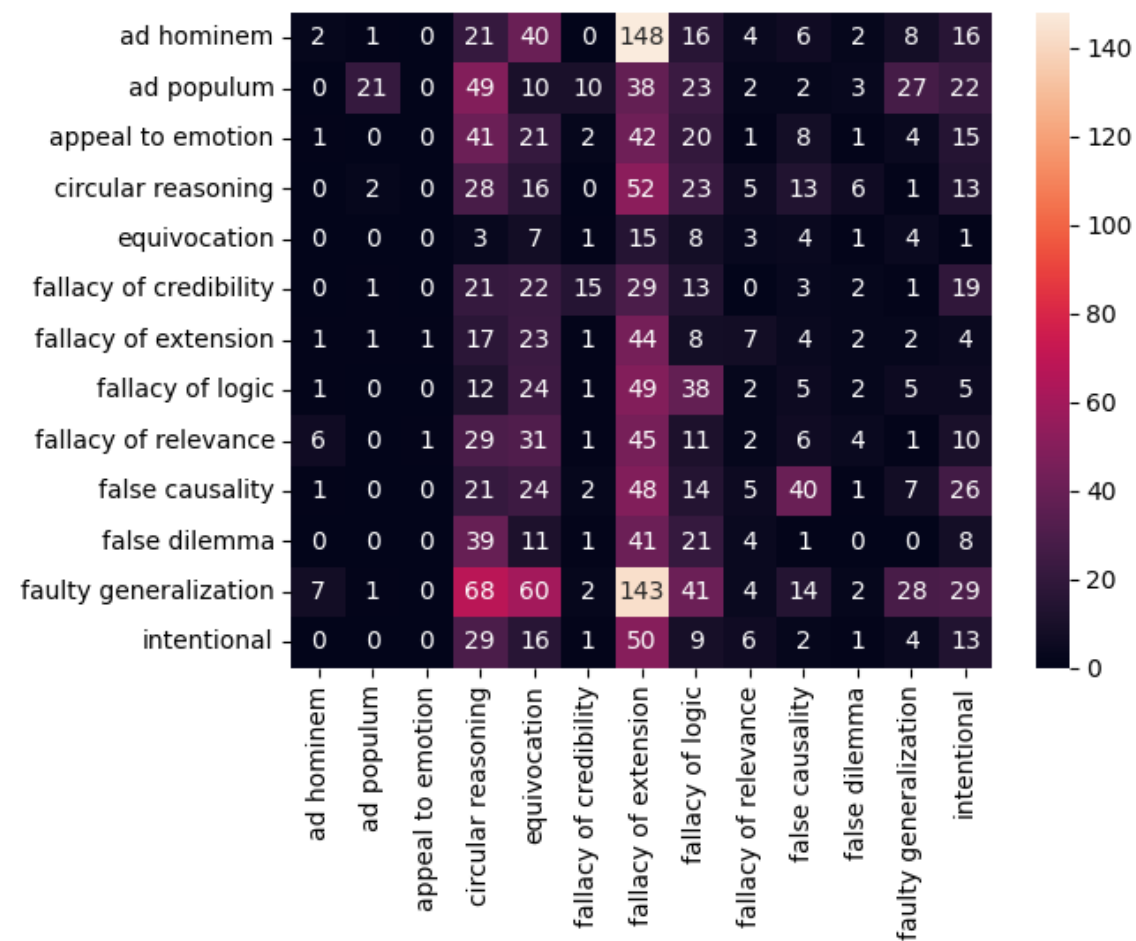
- Koreference, lematizācija - stanza, CoreNLP
- Kosinusa līdzība - sentence-transformers/bert-base-nli-mean-tokens
- Klasifikācija - nli-distilroberta-base, distilbert-base-uncased-mnli, facebook/bart-large-mnli

legūtie rezultāti



nli-distilroberta-base

legūtie rezultāti



distilbert-base-uncased-mnli

Galvenie izaicinājumi (maskēšana un klasifikatori)

- Oriģinālajā rakstā nebija minēts, kā tiek veikta maskēšana, tikai elementi
- Datu kopas maskēšana un klasificēšana paņem vairākas stundas
- Laika ierobežojums - nepaspējām paveikt visu, ko vēlējāmies (modeļu pielāgošana, maskēšanas uzlabojumi)

Atsauces

- [1] <https://arxiv.org/pdf/2202.13758>
- [2] <https://arxiv.org/pdf/2301.11879>
- [3] <https://arxiv.org/pdf/2405.02318>
- [4] <https://github.com/aznotins/LVCoref>
- [5] https://huggingface.co/knowledgator/comprehend_it-multilingual-t5-base

Paldies par uzmanību!

