



## **Equal goes it loose**

Hilft es Transformern beim Übersetzen von Redewendungen, sie mit einem Idiom-Tagger zu einem Multitasking-Lernsystem zu verbinden?

# Inhalt

- Einführung
- Dankers, Lucas and Titov 2022s Beobachtungen
- Daten
- Getestete Multitask-Learning Architekturen
  - Einfache MTL-Modelle, die keine Attention-Muster modifizieren
  - Modelle, die Cross-Attention modifizieren
  - Modelle, die die Self-Attention im Encoder modifizieren
- Evaluation
  - Haupttask: Übersetzung vom Deutschen ins Englische
  - Hilfstask: Detektieren von Redewendungen in den Quellsätzen
- Ideen zur Verbesserung

# Einführung

- Für maschinelle Übersetzungsmodelle ist das Übersetzen von Redewendungen immer noch eine Herausforderung
- Sie tendieren dazu, die Redewendungen kompositionell und damit zu wörtlich zu übersetzen
- Dankers, Lucas, und Titov 2022 kritisieren daher die Tendenz, Kompositionelle Generalisierung von Übersetzungsmodellen noch zu fördern

German source sentence

Ich werde mit der Tür ins Haus fallen .

English gold translation

I will come straight to the point .

Transformer's translation

i 'm going to drop in the house with the door .



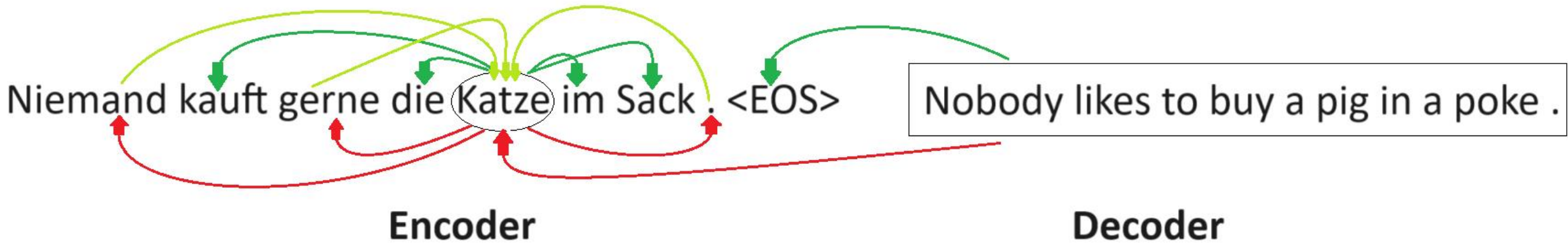
# Einführung

- Dankers, Lucas, und Titov 2022 schlagen stattdessen einen Multitask Learning (MTL) Ansatz vor, bei dem Idiomerkenkung trainiert wird
- In meiner Arbeit habe ich dafür mehrere MTL-Modelle trainiert, die parallel zum Übersetzen das Erkennen von Idiomen in den Quellsätzen lernt
- Einige Modelle nutzen die Ausgabe ihrer Idiom-Tagging-Module , um die Attention-Muster des Transformers zu modifizieren
- Die Modifizierungen basieren auf Beobachtungen, die Dankers, Lucas, und Titov 2022 für Fälle gemacht haben, in denen der Transformer Redewendungen richtig übersetzt

# Dankers, Lucas and Titov 2022s Beobachtungen



- stark erhöhte Attention
- leicht erhöhte Attention
- verringerte Attention



# Daten

- Parallel-Korpus für den Übersetzungstask:
  - IWSLT14 Deutsch → Englisch
  - „Examining the Tip of the Iceberg: A Data Set for Idiom Translation “ (Fadaee, Bisazza, und Monz 2018)
- Idiom Label für die deutschen Quellsätze:
  - Automatisch generierte Silbertrainingsdaten
  - Dazu verwendeter Idiom-tagger: ID10M (Tedeschi et al. 2022)
  - Grund für diese Wahl:
    - Multi-lingual
    - Sequence-Labeling anstatt binärer Klassifizierung von Phrasen
  - Probleme:
    - Hat in seinen Trainingsdate viele Kollokationen mit wörtlichem Sinn z.b. «Die Herausforderung annehmen» oder «Die Kontrolle verlieren»
    - Daher sind in meinem Trainingsdaten leider viele solcher Phrasen als Idiome gekennzeichnet und viele richtige Idiome wurden übersehen



# Daten

Wer ist nun schuld , dass diese Situation so außer Kontrolle gerät und was wird als nächstes passieren . False Positive

O O O O O O O O B-IDIOM I-IDIOM I-IDIOM O O O O O O O

Schwerwiegendere Fehlwahrnehmungen liegen ähnlich klar auf der Hand. **False Negative**

Der Mann nahm den Hut ab und setzte Klein Forti in ein Saatsfeld am Wegesrand . True Negative

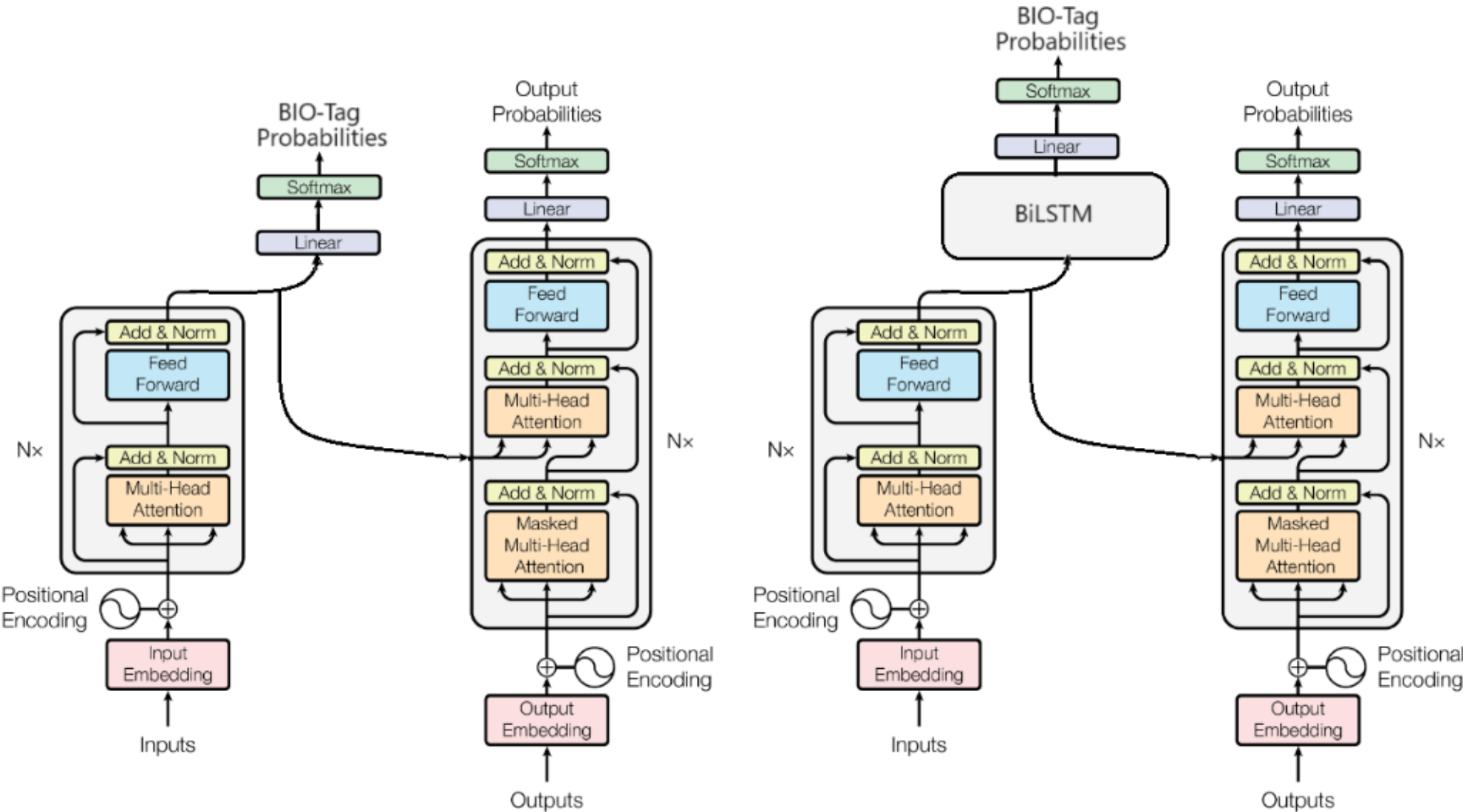
International gesehen steckt der Panna-Fussball allerdings noch in den Kinderschuhen.

O O O O O O B-IDIOM I-IDIOM I-IDIOM I-IDIOM O

True Positive

# Getestete Multitask-Learning Architekturen

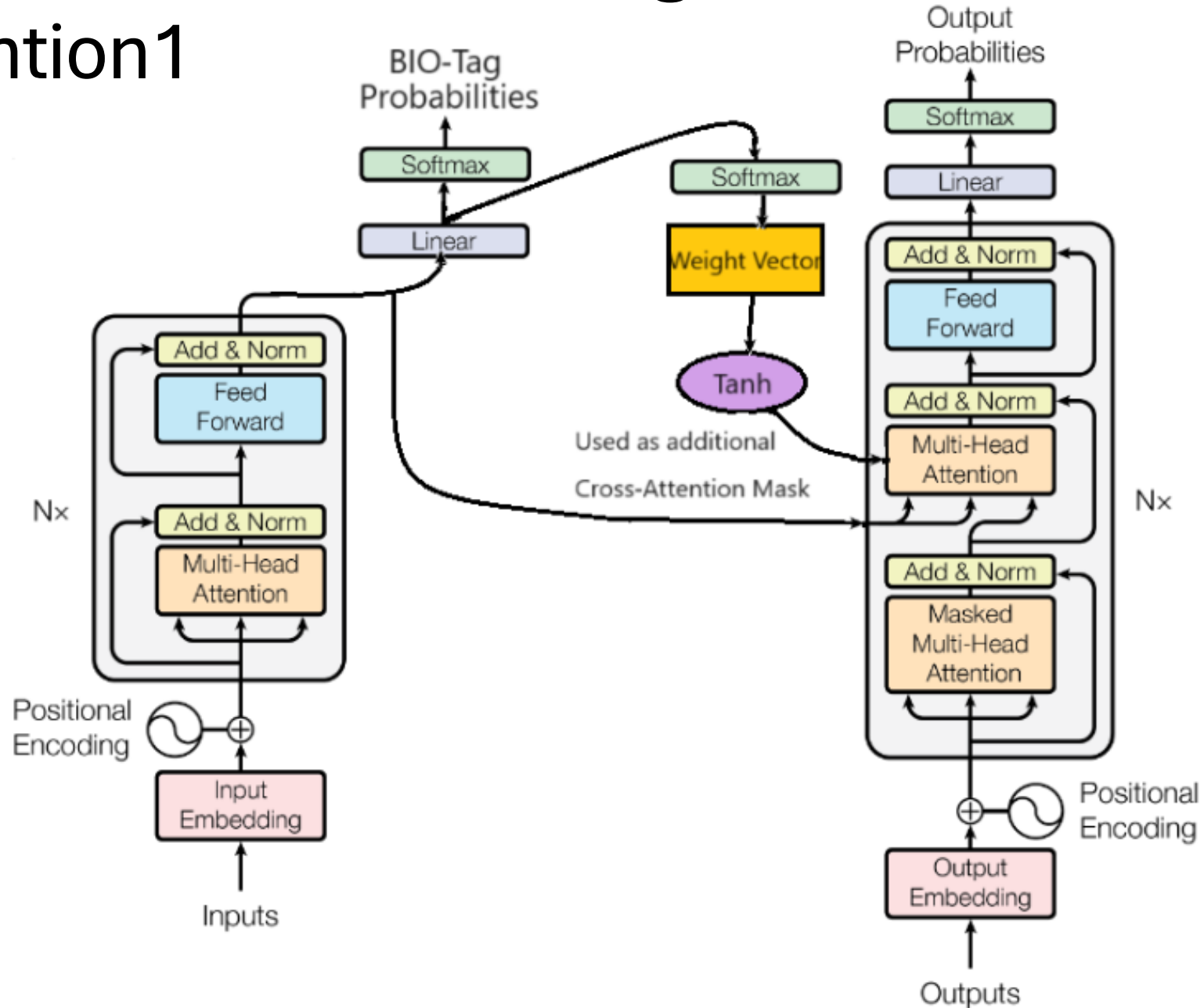
## Ohne Attention Modifizierung





# Getestete Multitask-Learning Architekturen

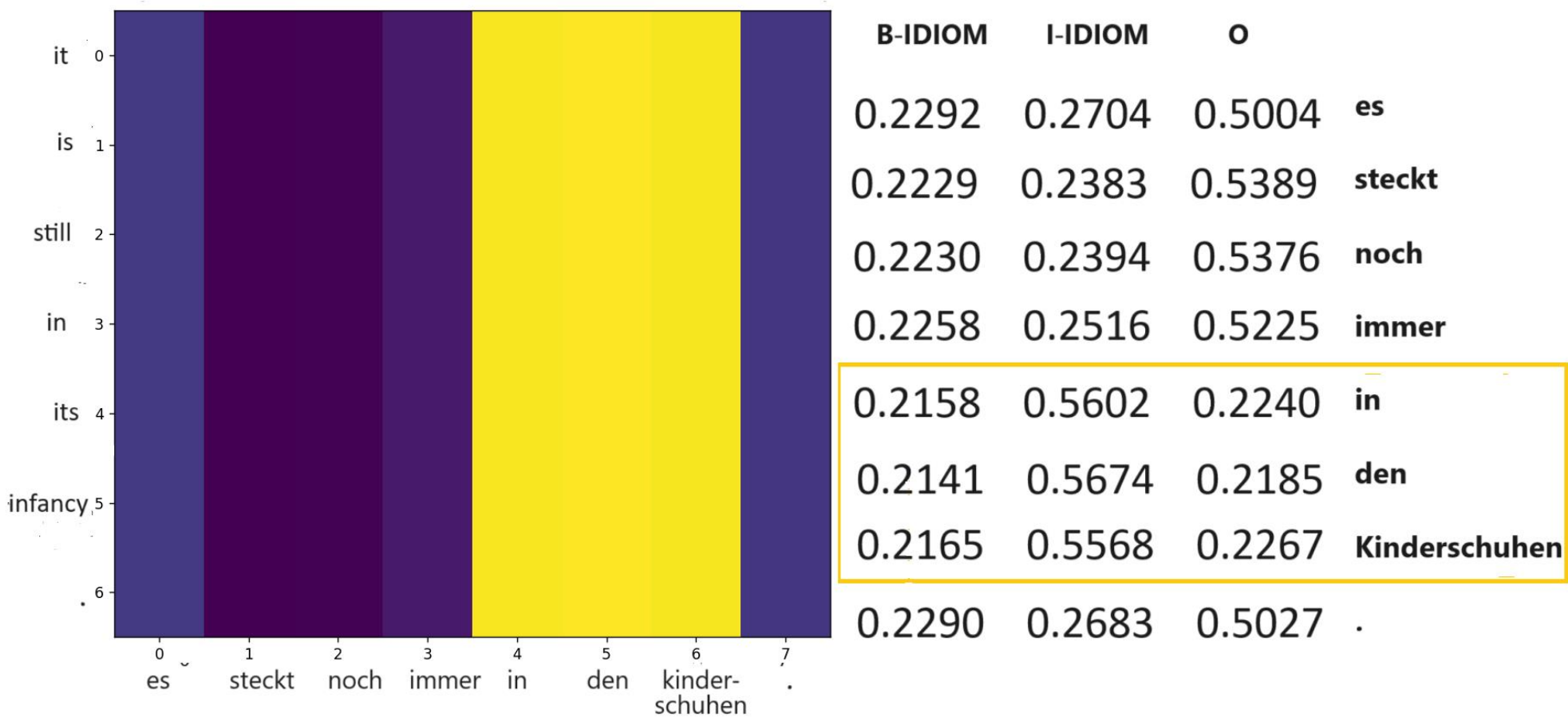
## Cross-Attention1



# Getestete Multitask-Learning Architekturen

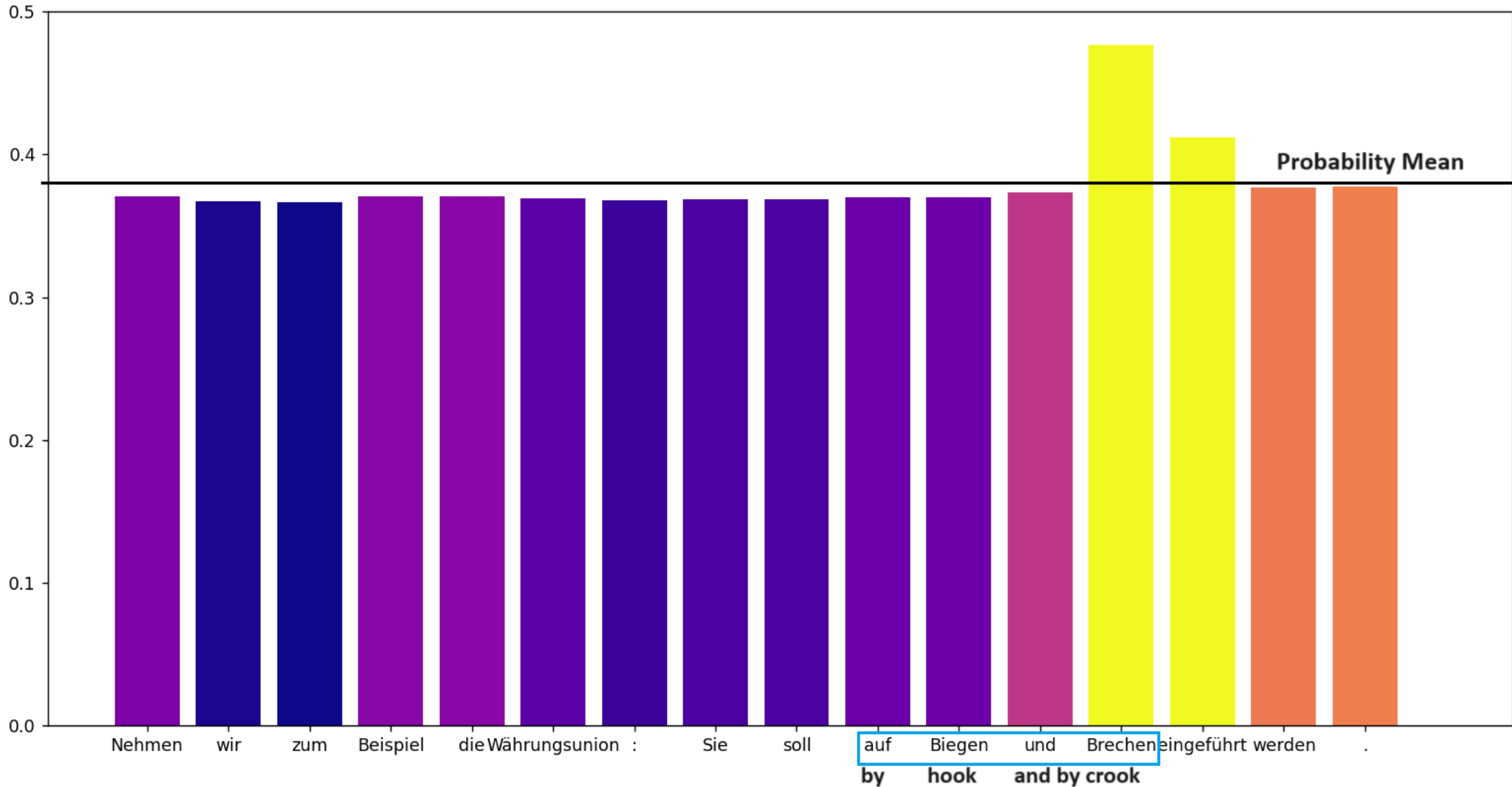
## Cross-Attention 1

$$\text{weightvector} = \begin{bmatrix} 1.0839 \\ 0.2406 \\ -0.2202 \end{bmatrix} \quad \begin{bmatrix} B - IDIOM \\ I - IDIOM \\ O \end{bmatrix}$$



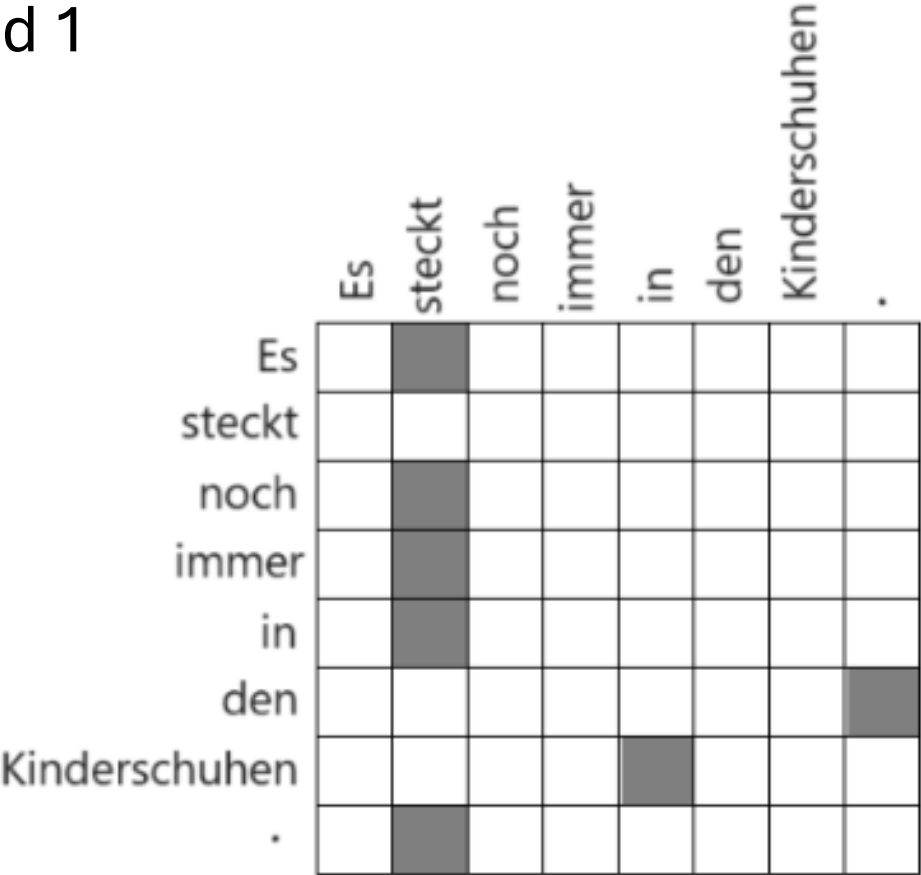
# Getestete Multitask-Learning Architekturen

## Cross-Attention 2

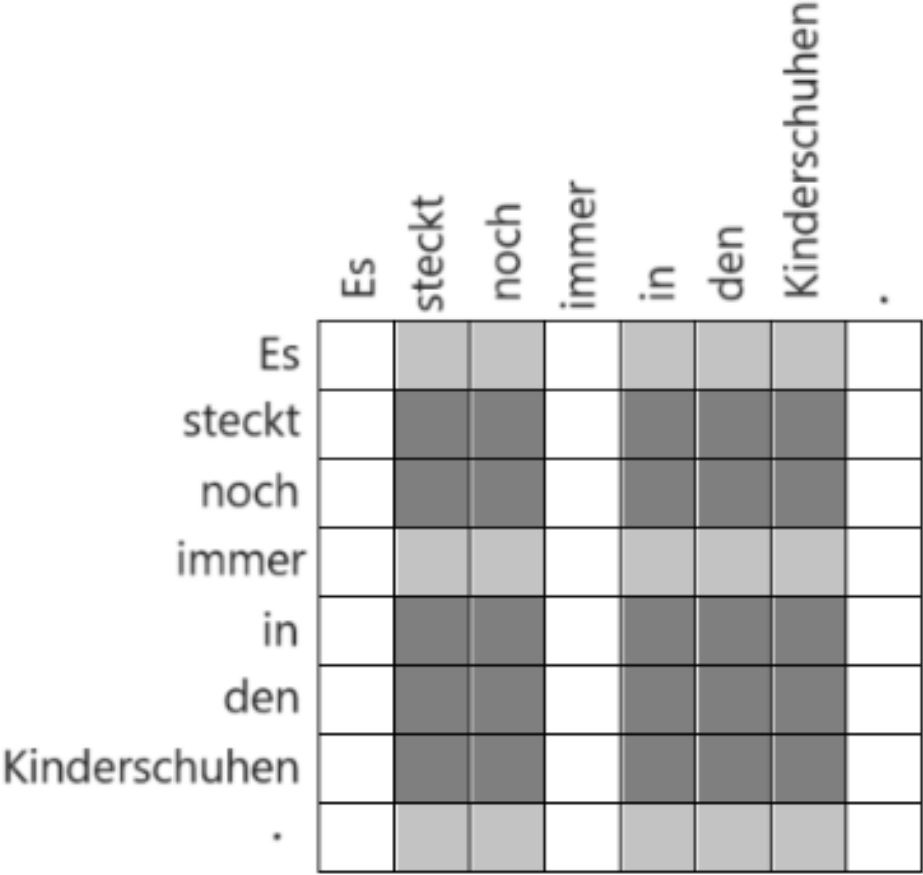


# Getestete Multitask-Learning Architekturen

## Special Head 1

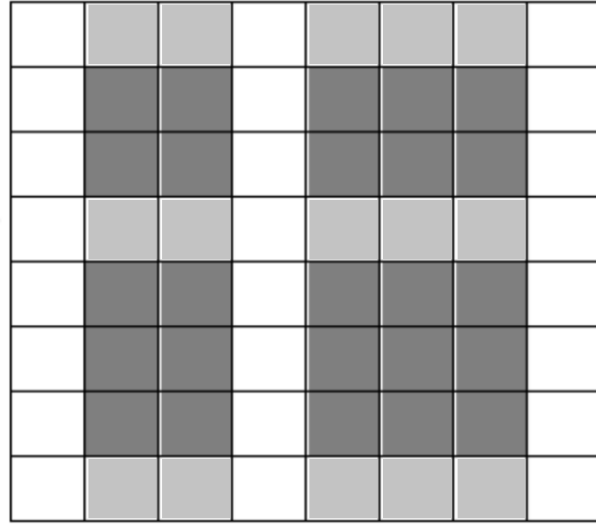


Self-Attention-Matrix nach Deguchi, Tamura und Ninomiya 2019 : Die Token sollen die höchste Aufmerksamkeit auf ihren Dependenzkopf richten



Self-Attention-Matrix für denselben Satz, die die Token innerhalb der Idiome gemäß Dankers, Lucas, und Titov 2022 gruppiert

# Getestete Multitask-Learning Architekturen

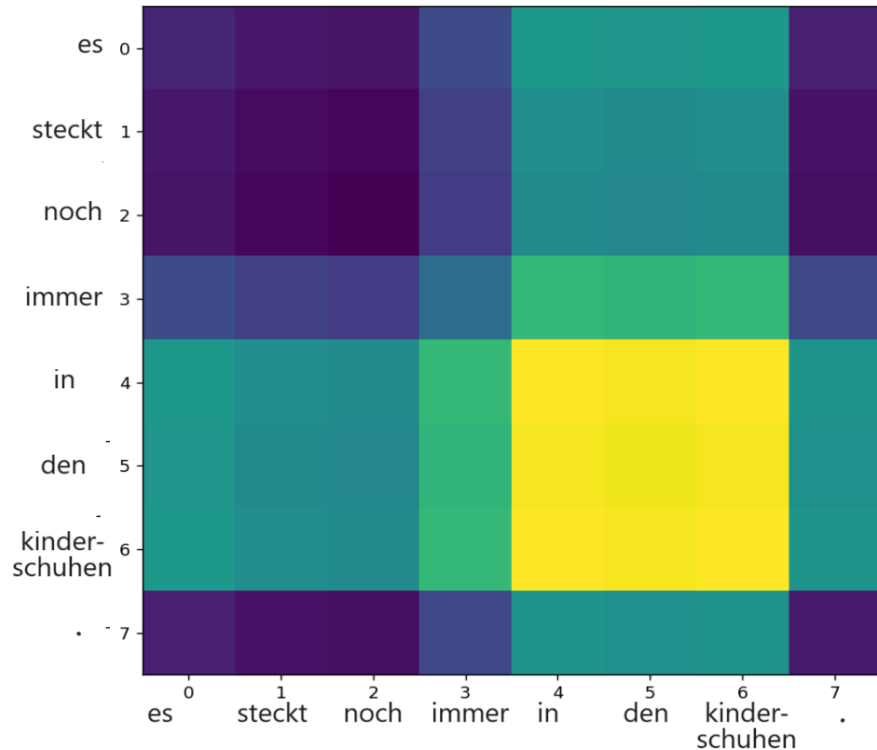


## Special Head 1

$$\text{weightvector} = \begin{bmatrix} 0.3653 \\ -0.0405 \\ -0.8752 \end{bmatrix} \quad \begin{bmatrix} B - IDIOM \\ I - IDIOM \\ O \end{bmatrix}$$

$$\text{bias} = 0.1249$$

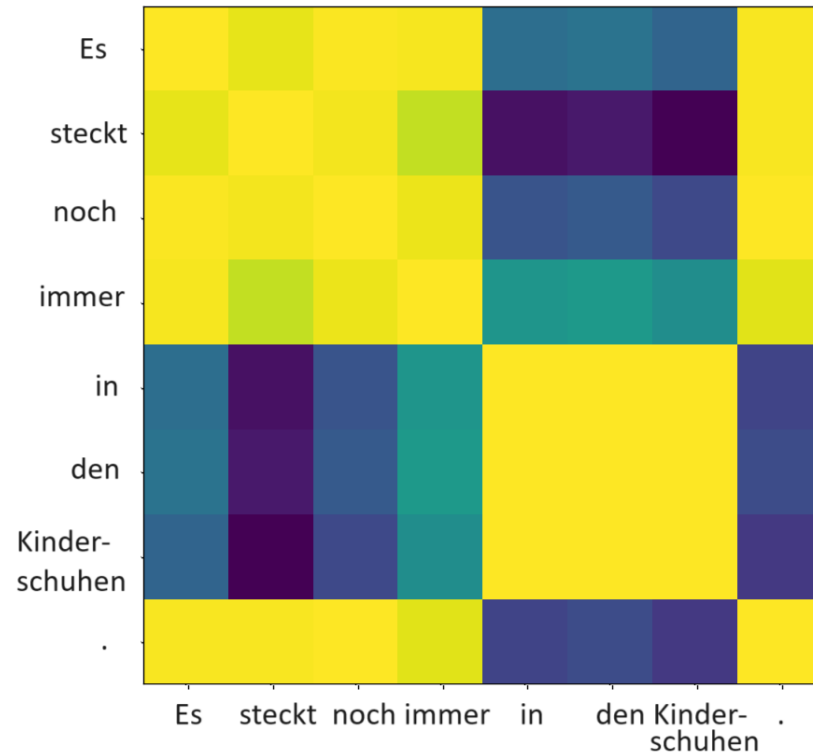
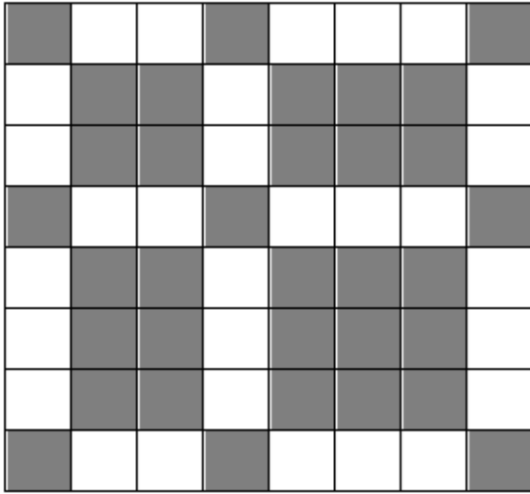
$$\text{scale-parameter} = 0.6576$$



B-IDIOM	I-IDIOM	O	
0.2283	0.2639	0.5078	es
0.2248	0.2463	0.5289	steckt
0.2232	0.2404	0.5364	noch
0.2360	0.3271	0.4369	immer
0.2167	0.5570	0.2263	in
0.2195	0.5439	0.2366	den
0.2166	0.5571	0.2263	Kinderschuhen
0.2268	0.2571	0.5161	.

# Getestete Multitask-Learning Architekturen

## Special Head 2



B-IDIOM	I-IDIOM	O	
0.2338	0.3025	0.4637	Es
0.2252	0.2495	0.5253	steckt
0.2315	0.2850	0.4835	noch
0.2361	0.3339	0.4300	immer
0.2187	0.5478	0.2335	in
0.2200	0.5424	0.2376	den
0.2167	0.5566	0.2267	Kinder-schuhen
0.2300	0.2758	0.4942	.



## Evaluation

Haupttask: Übersetzung vom Deutschen ins Englische

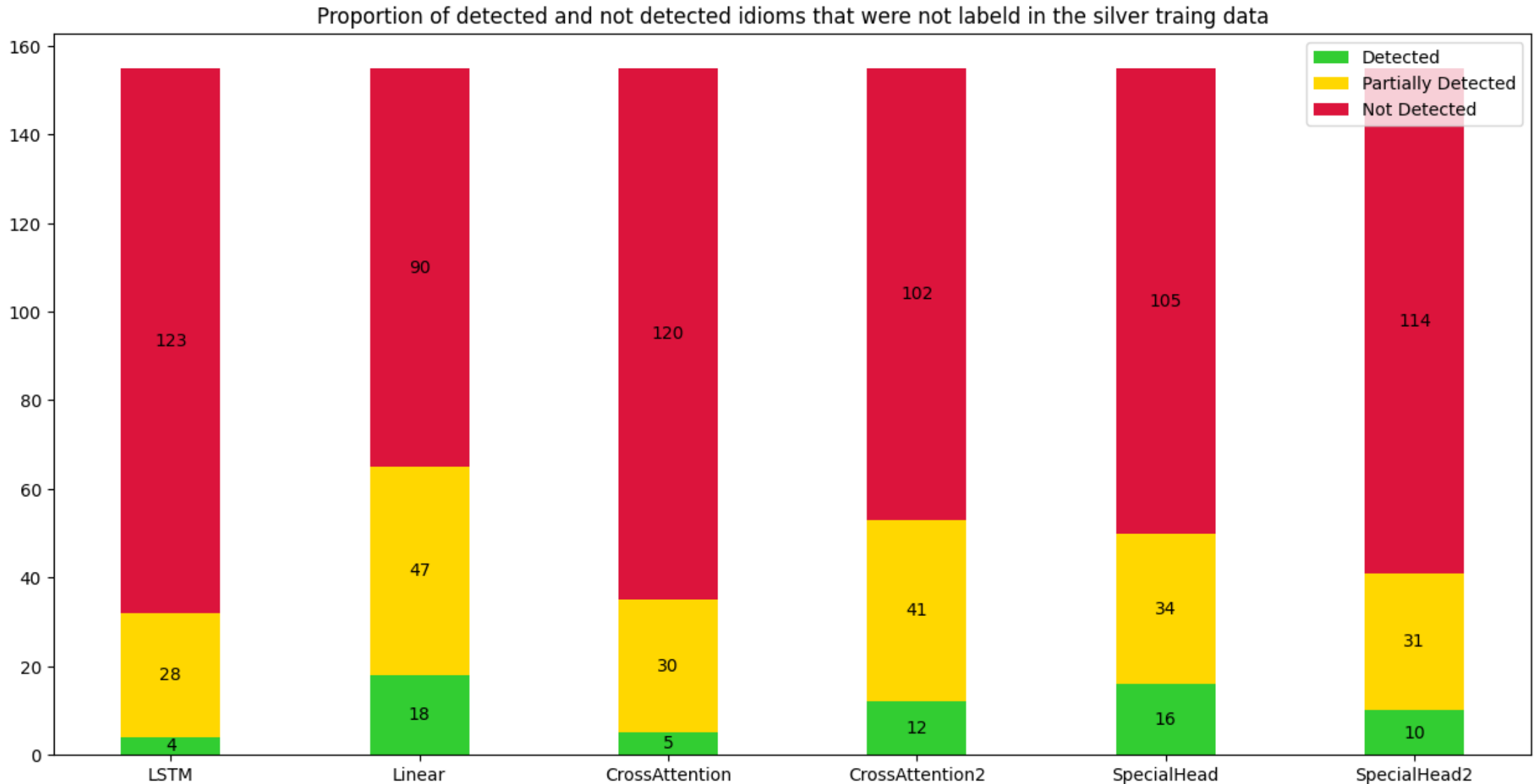
	BLEU	accuracy all idioms	accuracy I-IDIOM F1 > 0	accuracy I-/B-IDIOM F1 > 0
Vanilla	30.44	46.63	56.31	64.66
BiLSTM	30.44	<b>48.31</b>	60.36	69.17
Linear	30.57	48.03	<b>62.16</b>	<b>72.18</b>
Cross-Attention	<b>30.65</b>	47.47	59.01	70.68
Cross-Attention2	<b>30.65</b>	44.94	57.21	63.16
Special Head	30.46	44.38	56.76	68.42
Special Head 2	30.58	44.94	58.11	69.93

Evaluation  
Hilfstask: Detektieren von Redewendungen in den Quellsätzen

Model	Cross Entropy Loss	Precision	Recall	F1-Score
LSTM	0.8950	12.25	7.4 1.43	9.23
Linear	0.8670	61.29	9.07 1.67	15.8
CrossAttention	0.8652	55.56	4.77 0.48	8.79
CrossAttention2	0.8642	62.5	3.582.15	6.77
Special Head	0.8659	52.17	5.73 1.43	10.32
Special Head 2	0.8659	62.86	5.25 1.91	9.69

# Evaluation

## Hilfstask: Detektieren von Redewendungen in den Quellsätzen



# Interpretation der unterschiedlichen Modellergebnisse

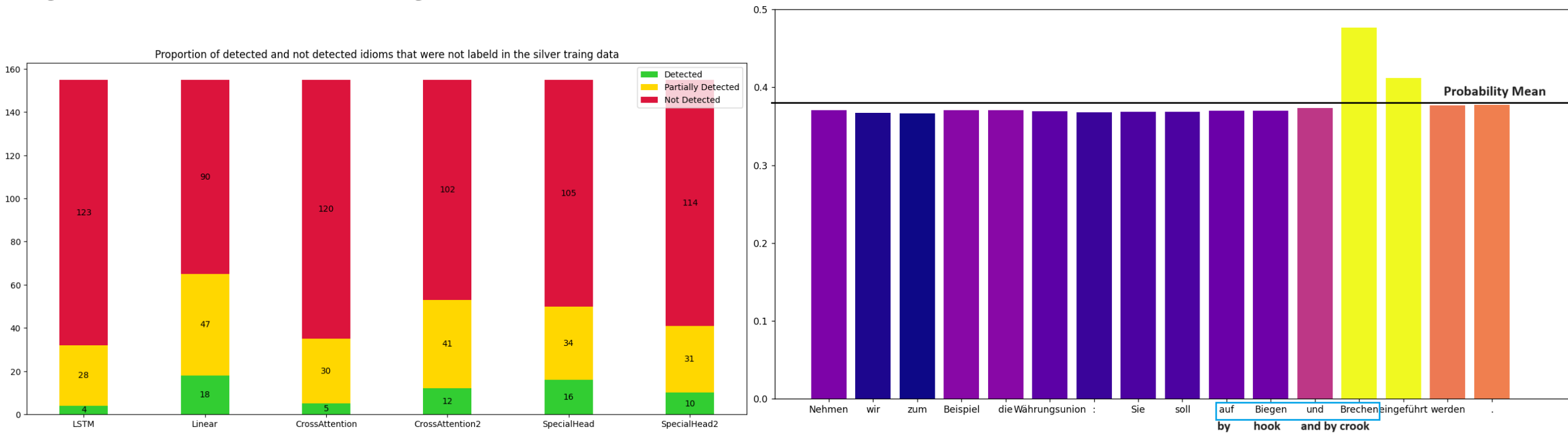
Modelle, die die Attentionmuster nicht modifizieren:

- Linear-Layer Version übersetzt Idiome am besten von allen Modellen und der Idiomtagger hat den höchsten Recall und beste Generalisierungsfähigkeit
- BiLSTM Version scheint Wissen über Idiome nicht wirklich zu verwenden. Verbesserte Idiomübersetzungs-Accuracy wahrscheinlich durch erhöhte Anzahl an Parametern
- Mit Probing Information über idiomatizität In den Embeddings untersuchen spannend an dieser Stelle  
(Vergleich: Klublička, Nedumpozhimana, and Kelleher 2023)

# Interpretation der unterschiedlichen Modellergebnisse

## Cross-Attention modifizierende Modelle

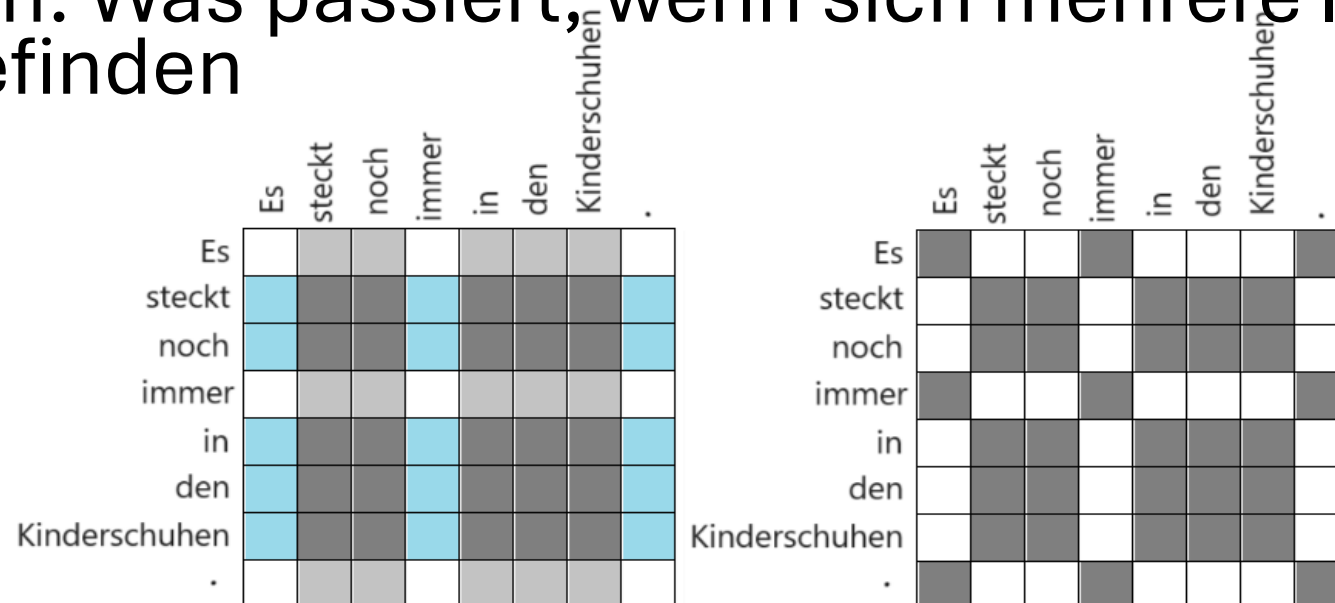
- Beide haben besten BLEU-score
- Von den Attentionmustern modifizierenden Modellen ist „Cross-Attention1“ das beste
- Die Idiomtranslationaccuracy von „Cross-Attention2“ ist deutlich geringer
- Beide haben einen sehr niedrigen Recall, aber „Cross-Attention2“ kann besser generalisieren → Probing auch hier spannend



# Interpretation der unterschiedlichen Modellergebnisse

Self-Attention modifizierende Modelle mit spezialisiertem Attention-Head:

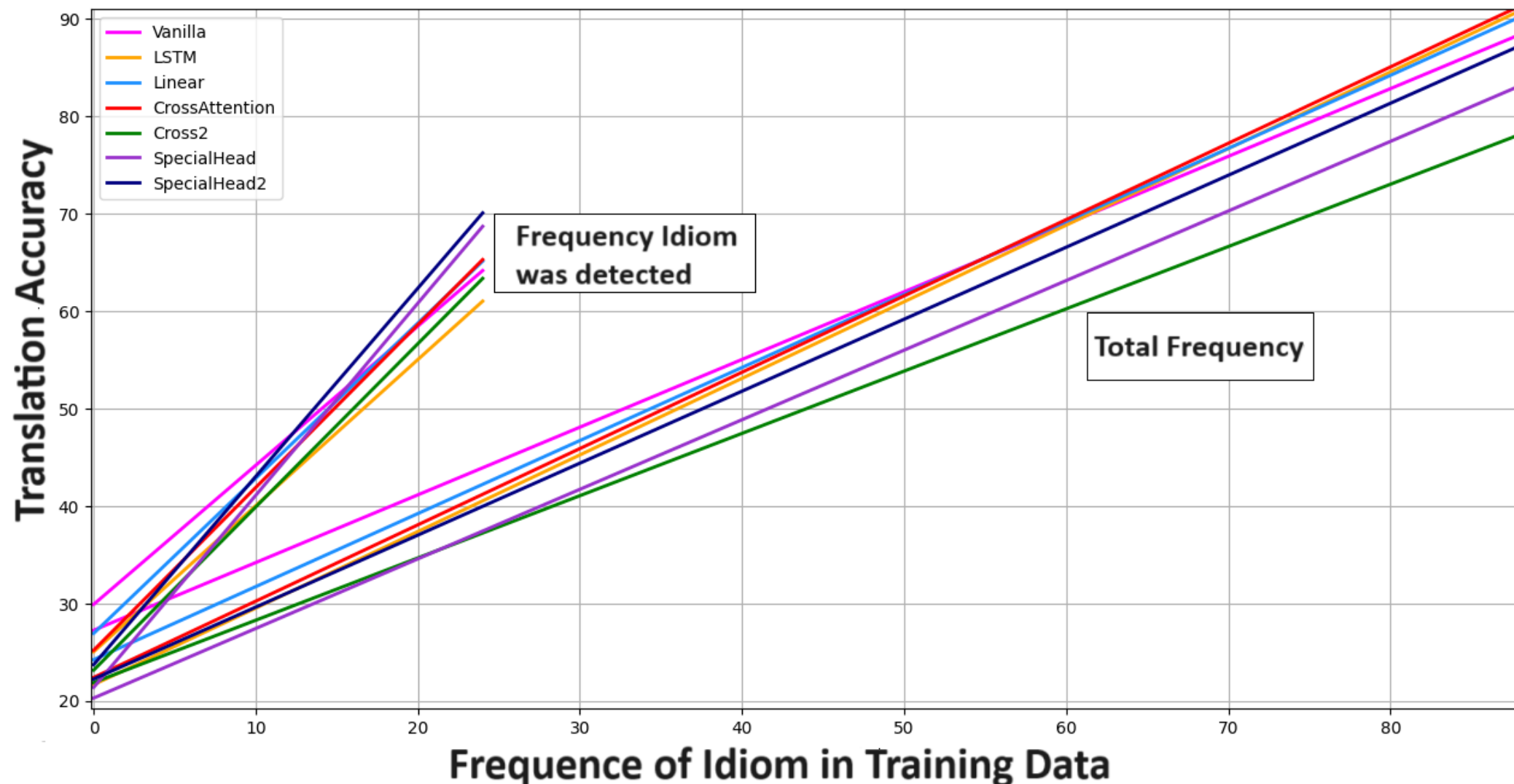
- Profitieren am meisten von richtig gelabelten Idiominstanzen in den Trainingsdaten
- Zweite Version ohne Gewichte und Maske, die Attention zwischen Idiomem und Kontext in beide Richtungen reduziert, funktioniert besser
- Noch zu testen: Was passiert, wenn sich mehrere Idiome in einem Satz befinden





# Interpretation der unterschiedlichen Modellergebnisse

## Self-Attention modifizierende Modelle mit spezialisiertem Attention-Head:



# Ideen zur Verbesserung für alle Modelle

- Qualität der Trainingsdaten für die Idiomerkennung verbessern
  - Mehr Daten vorhanden für Englisch → Deutsch
  - Idiom-tagging Module könnten dann beispielsweise mit PIE corpus (Zhou et al. 2021) fine-tuned werden
- Fähigkeit der Idiomtagging-Module zur Generalisierung verbessern:
  - Zusätzlicher Hilfstagsk Pos-tagging
  - Ausweitung zu einem multi-lingualen Modell
  - Idiome paraphrasieren lernen (auch mit PIE corpus möglich)
- Probing mit MAGPIE (wie bei Klublička et al. 2023), um kontextualisierte Embeddings auf Information über Idiomazität zu überprüfen
- Code einsehbar unter:  
[https://github.com/LeHarter/joeynmt\\_MTL\\_Idiomtranslation](https://github.com/LeHarter/joeynmt_MTL_Idiomtranslation)

Gibt es Fragen?



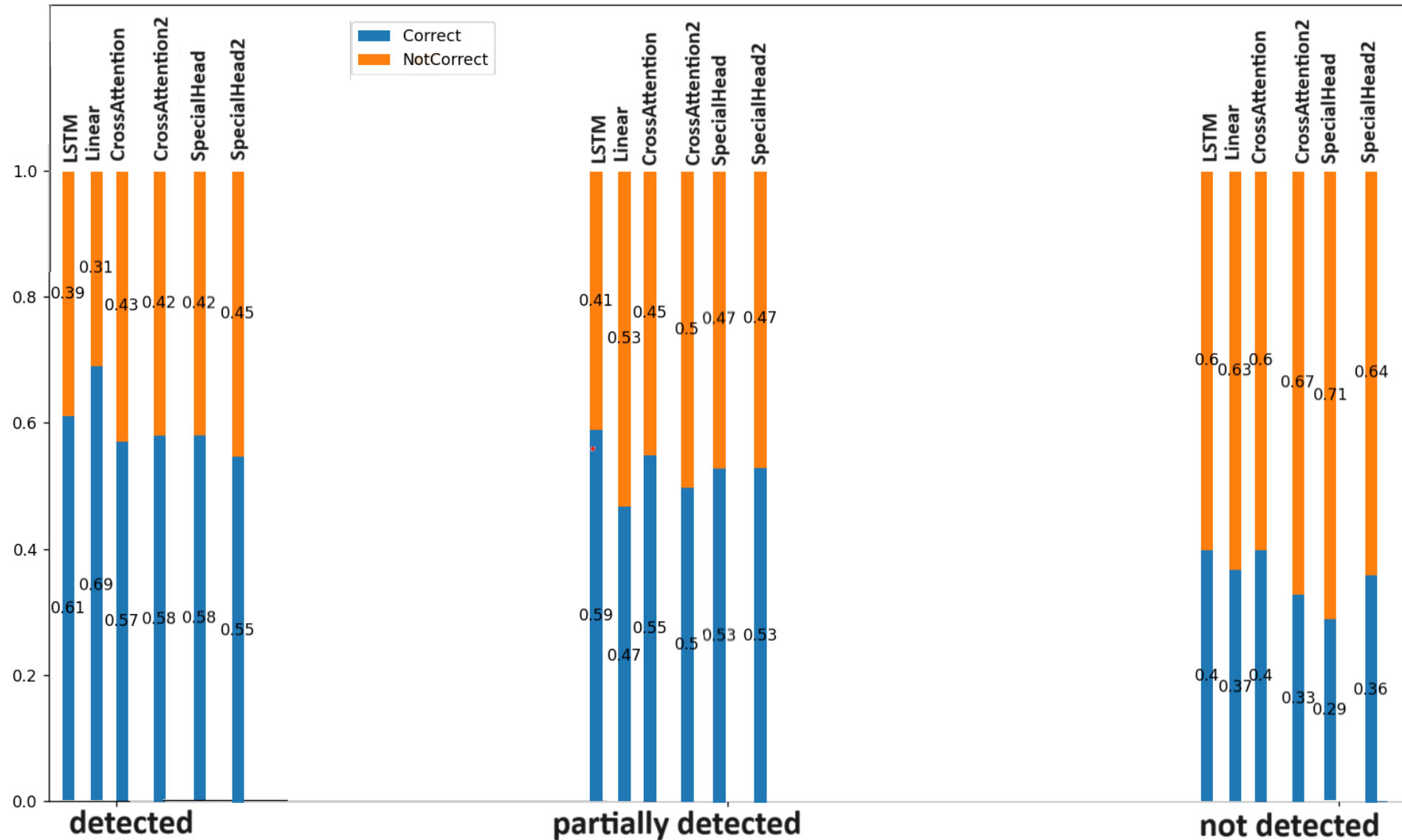
# Quellen

- **Dankers, Verna, Christopher G Lucas, and Ivan Titov (2022).** “Can transformer be too compositional? analysing idiom processing in neural machine trans-lation”. In: arXiv preprint arXiv:2205.15301
- **Deguchi, Hiroyuki, Akihiro Tamura, and Takashi Ninomiya (2019).** “Dependency-Based Self-Attention for Transformer NMT”. In: Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019). Varna, Bulgaria: INCOMA Ltd., pp. 239–24
- **Fadaee, Marzieh, Arianna Bisazza, and Christof Monz (2018).** “Examining the Tip of the Iceberg: A Data Set for Idiom Translation”. In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018). Miyazaki, Japan: European Language Resources Association (ELRA)
- **Klubička, Filip, Vasudevan Nedumpozhimana, and John D Kelleher (2023).** “Idioms, Probing and Dangerous Things: Towards Structural Probing for Idiomaticity in Vector Space”. In: arXiv preprint arXiv:2304.14
- **Tedeschi, Simone, Federico Martelli, and Roberto Navigli (2022).** “ID10M: Idiom identification in 10 languages”. In: Findings of the Association for Computational Linguistics: NAACL 2022, pp. 2715–272
- **Zhou, Jianing, Hongyu Gong, and Suma Bhat (2021).** “PIE: A Parallel Idiomatic Expression Corpus for Idiomatic Sentence Generation and Paraphrasing”. In: Proceedings of the 17th Workshop on Multiword Expressions (MWE 2021). Ed. by Paul Cook, Jelena Mitrović, Carla Parra Escartín, 43 Ashwini Vaidya, Petya Osenova, Shiva Taslimipoor, and Carlos Ramisch. Online: Association for Computational Linguistics, pp. 33–48

# Evaluation

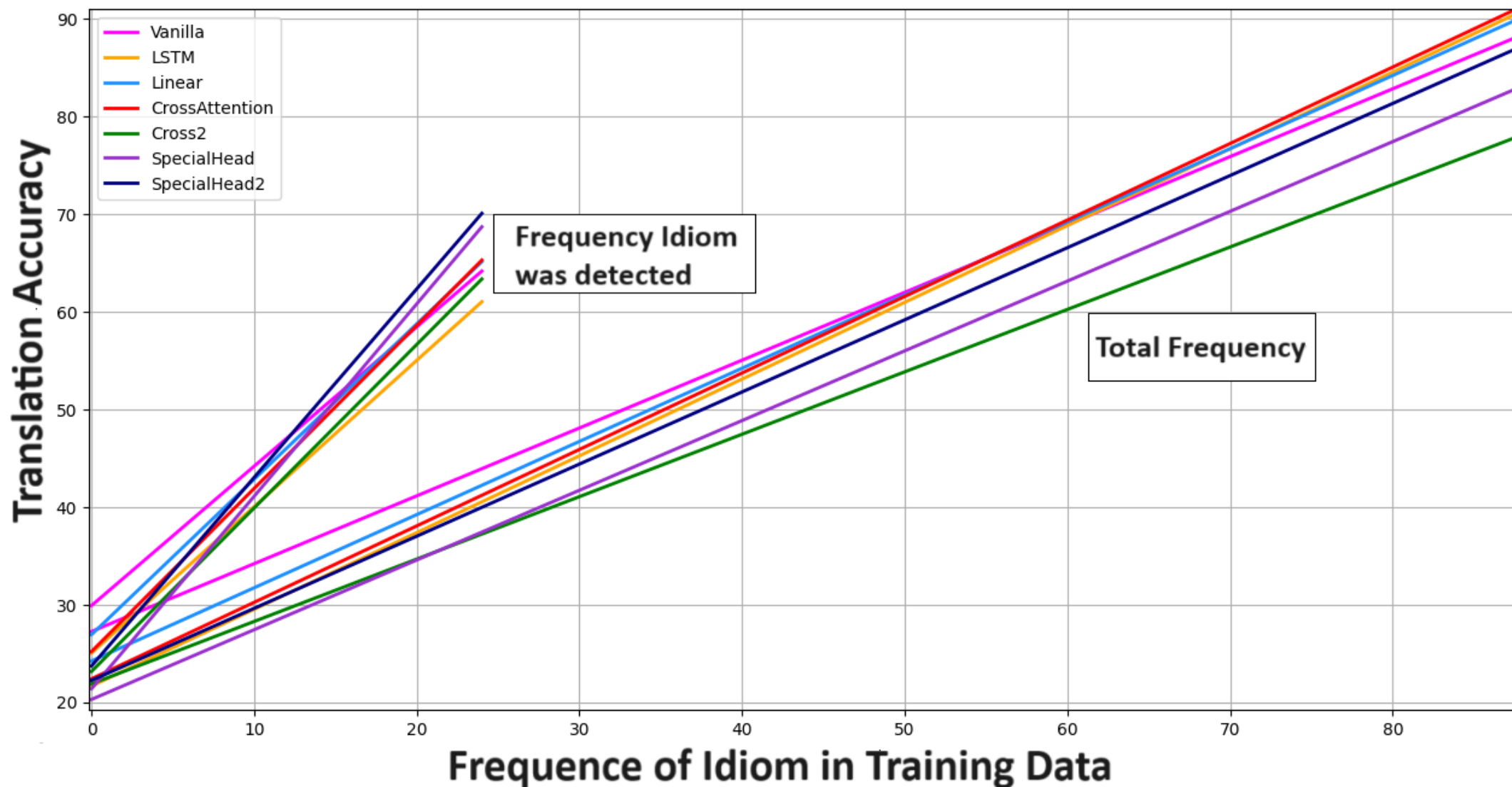
## Haupttask: Übersetzung vom Deutschen ins Englische

Likelihood models translate an Idiom correctly given it detected the idiom



# Evaluation

Haupttask: Übersetzung vom Deutschen ins Englische





# Getestete Multitask-Learning Architekturen

## Special Head 1

