

WiC-TSV-de: German Word-in-Context Target-Sense-Verification Dataset and Cross-Lingual Transfer Analysis



Anna Breit, Artem Revenko, Narayani Blaschke

<https://github.com/semantic-web-company/wic-tsv>

Target Sense Verification (TSV)

Traditional WSD: Find most suitable sense from a given sense inventory

Target Sense Verification: Verify a word in context against *one* target sense:

Alphabet ?

"I had a very interesting interview with Terno Schwab, the CEO of **Alphabet**."

+ no need to model the entire sense inventory
+ no assumption of complete data
→ more suitable for enterprise and domain-specific use cases

Dataset Creation

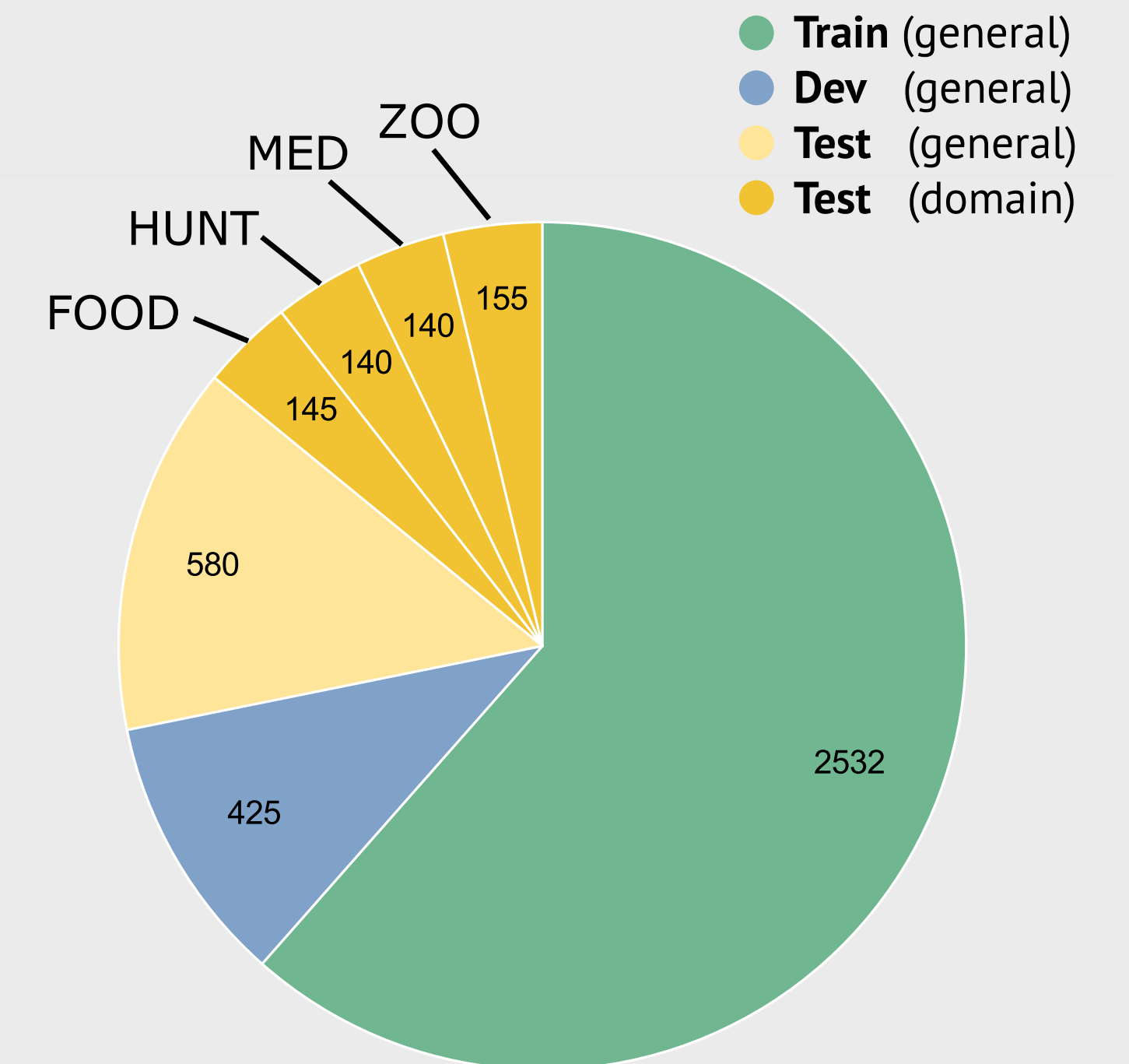
- **Domain-Independent Instances** are from German Wiktionary
- **Domain-Bound Instances**
 - contexts are manually collected
 - **hypernyms and definitions** - domain-specific knowledge sources (e.g. MeSH)

Creating challenging instances

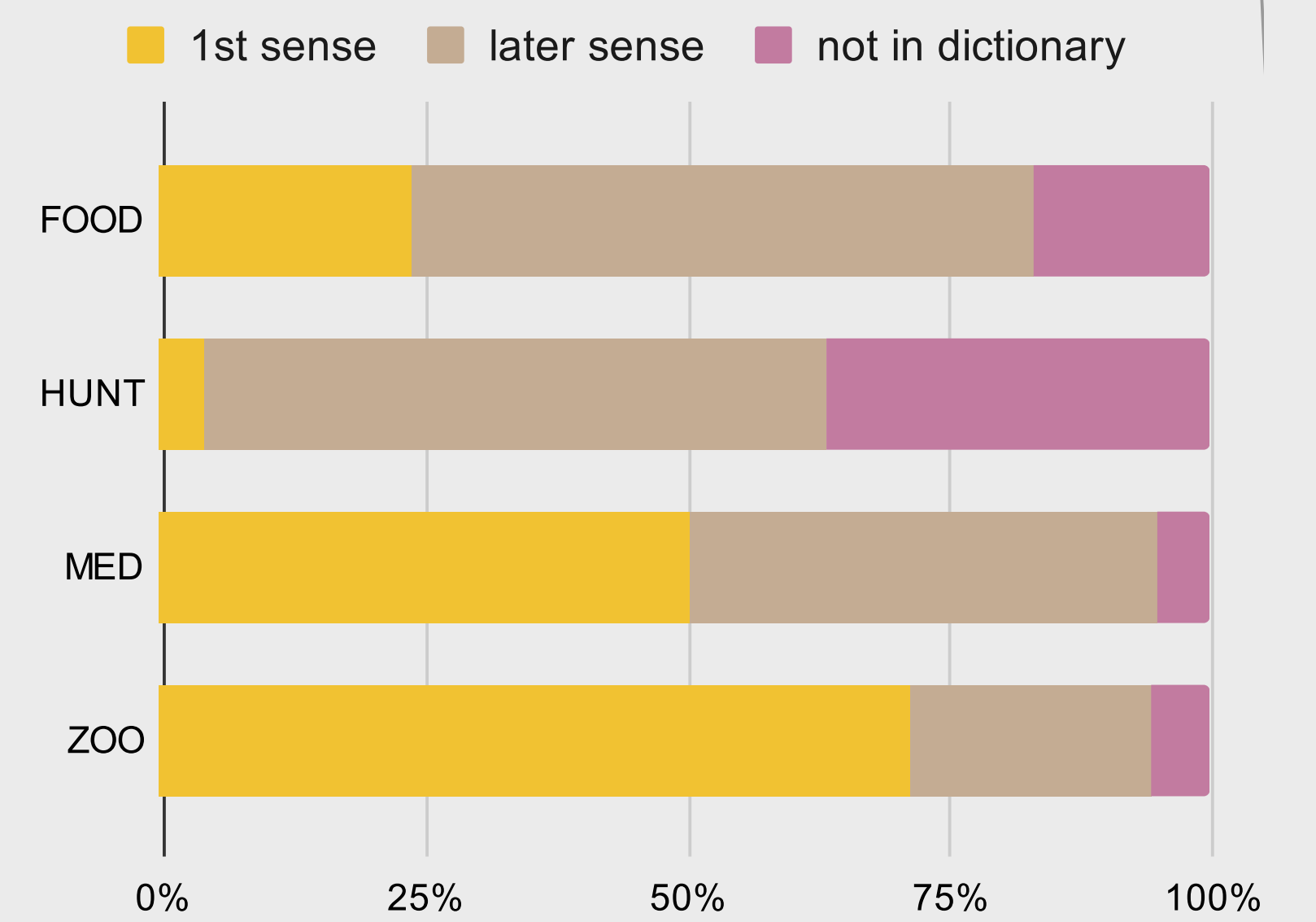
- **Mixed commonality** of target senses
- **In-Domain Ambiguities**, e.g. "Bär" both referring to *bear* as well as to *male marmot* in hunting domain.
- **Mixed Contexts**, e.g. "The virus gets its name from a distinctive *corona* of sugary-proteins that projects from its surface."
- **Trigger Word**, e.g. "The *jaguar* was run over by a sports *car*."

Dataset Statistics

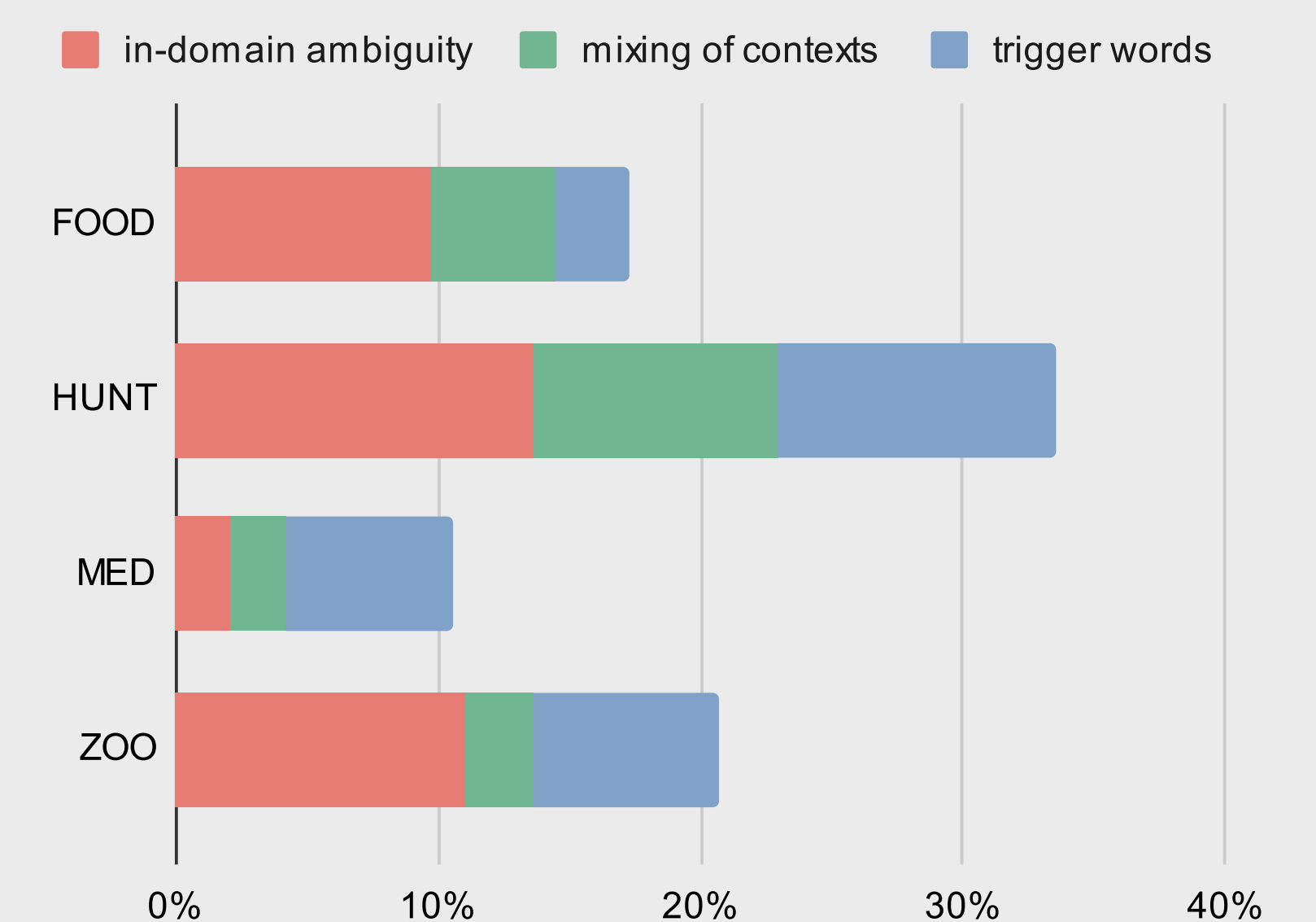
Distribution



Target Sense Commonality



Hard Instance Strategies

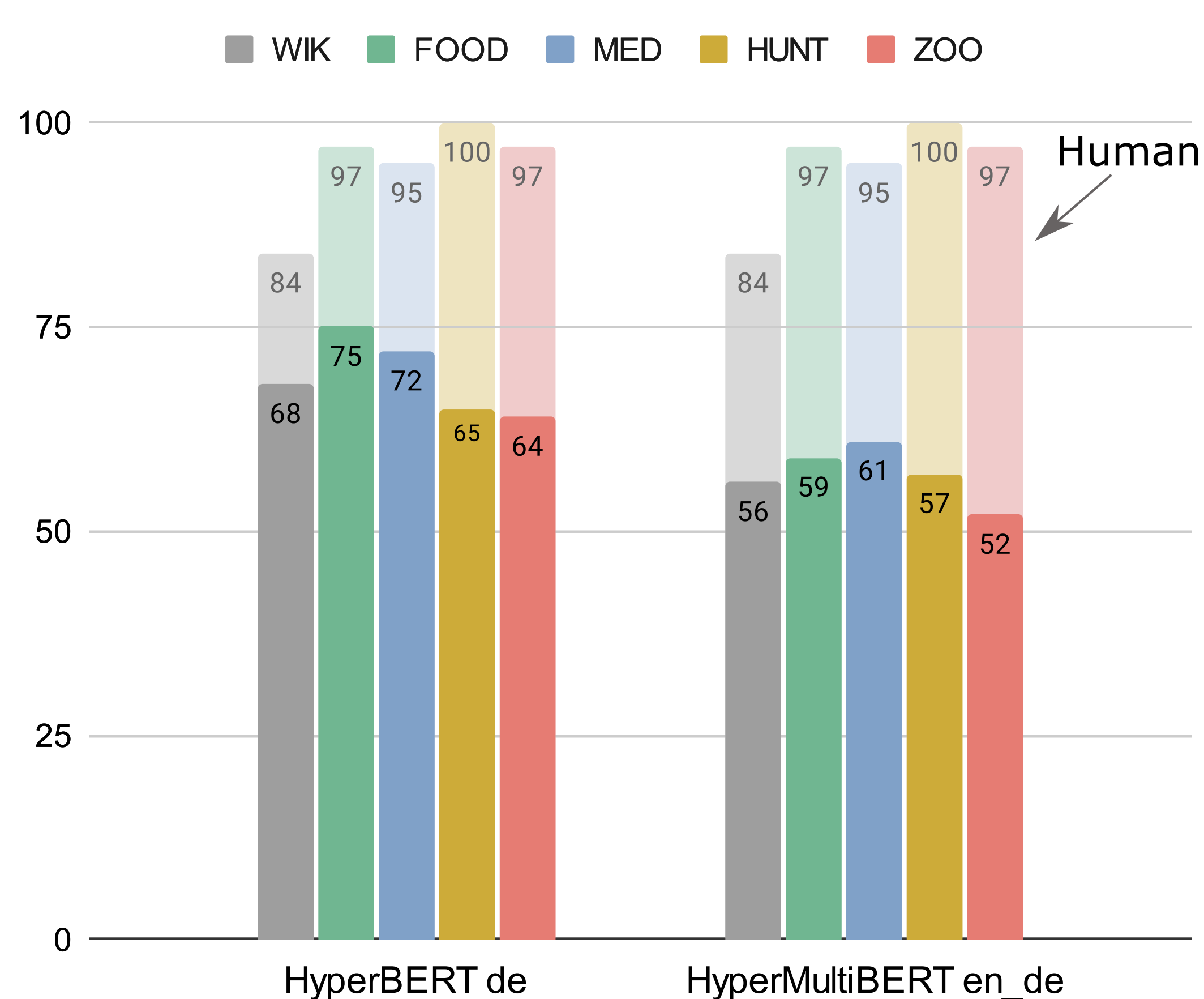


Examples of Instances

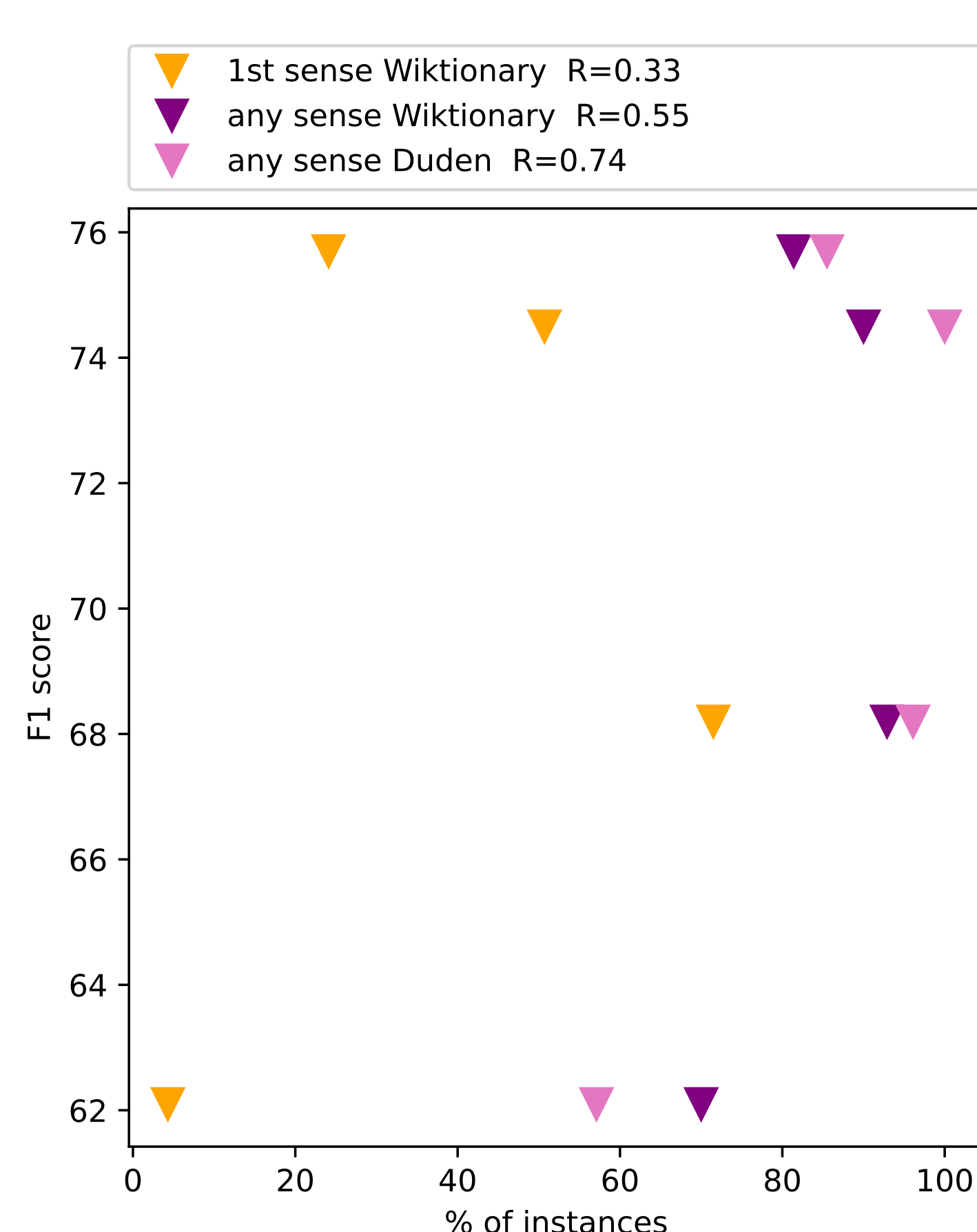
Label	Context	Definition	Hypernyms
<i>Domain-independent (WIK)</i>			
F	Der Ruf "Feuer!" hallte durch das Haus. <small>The call "Fire!" echoed through the house.</small>	das Ansehen, das jemand bei anderen hat <small>the reputation that someone has among others</small>	Ansehen, Status <small>reputation, status</small>
<i>Gastronomy (FOOD)</i>			
T	Betrachten wir nun ein Objekt, beispielsweise einen Apfel, treffen die von diesem Apfel reflektierten Lichtstrahlen auf unsere Hornhaut. <small>If we now look at an object, for example an apple, the light rays reflected from this apple hit our cornea.</small>	Frucht des Apfelbaums <small>fruit of the apple tree</small>	Kernobst <small>Pome fruit</small>
<i>Hunting (HUNT)</i>			
T	Bemerken möchte ich dazu, dass die Katzen und vorjährigen Murren viel mehr schreien als die älteren Bären . <small>I would like to add to this that the female marmots and one-year-old marmots scream much more than the older male marmots.</small>	Die Bezeichnung für das männliche Murmeltier <small>The name for the male marmot</small>	Niederhaarwild <small>small wild game</small>
<i>Medicine (MED)</i>			
F	Mit diesem Beitrag sollen die Fakten aufgezeigt werden, die das Einmessen mit der ägyptischen Elle bezeugen. <small>This article intends to show the facts documenting the calibration using the Egyptian cubit</small>	Kleinfingerseitig gelegener länglicher Röhrenknochen des Unterarms <small>Elongated tubular bone of the forearm on the small finger side</small>	Armknöchel <small>arm bone</small>
<i>Zoology (ZOO)</i>			
F	Doch die Ente mauserte sich mit der Zeit zum geliebten Kultauto. <small>But over time, the Citroën 2CV "moulded" (note: as in "transformed") into a beloved cult car.</small>	Enten, Anatinae, heterogene Unterfamilie der Entenvögel mit etwa 35 Gattungen und 115 Arten. <small>Ducks, Anatinae, heterogeneous subfamily of duck birds with about 35 genera and 115 species</small>	Entenvögel; Anatidae <small>Duck Birds; Anatidae</small>

Results

Performance on the Test-Subsets (mono- and cross-lingual)



Influence of Target Sense Commonality (monolingual)



Influence of Hard Example Strategy (monolingual)

