



UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

UniHD at TSAR-2022 Shared Task: Is Compute All We Need for Lexical Simplification?

Dennis Aumiller and Michael Gertz

Heidelberg University

Institute of Computer Science

Database Systems Research Group

aumiller@informatik.uni-heidelberg.de

Current Simplification Systems



UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

Complex Word

*: requires external knowledge/tooling

Source: Qiang et al., 2020: LSBert: A Simple Framework for Lexical Simplification

Aumiller and Gertz: "UniHD at TSAR-2022: Is Compute All We Need for Lexical Simplification?"

Current Simplification Systems



UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

Complex Word



Stemming*

*: requires external knowledge/tooling

Source: Qiang et al., 2020: LSBert: A Simple Framework for Lexical Simplification

Aumiller and Gertz: "UniHD at TSAR-2022: Is Compute All We Need for Lexical Simplification?"

Current Simplification Systems



UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

Complex Word



Stemming*



Substitute
Generation*

*: requires external knowledge/tooling

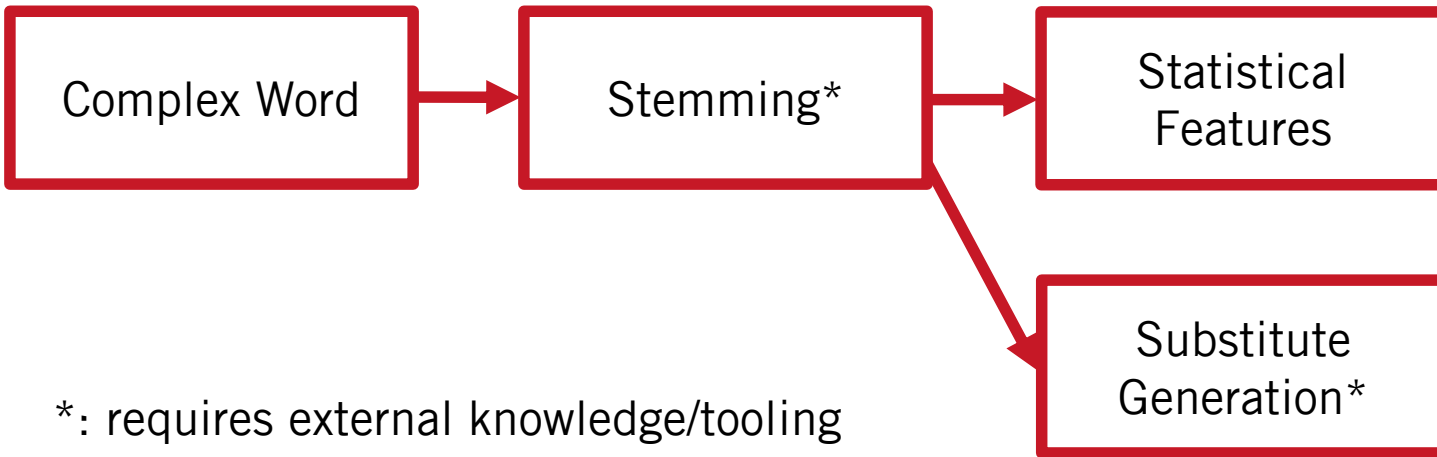
Source: Qiang et al., 2020: LSBert: A Simple Framework for Lexical Simplification

Aumiller and Gertz: "UniHD at TSAR-2022: Is Compute All We Need for Lexical Simplification?"

Current Simplification Systems



UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

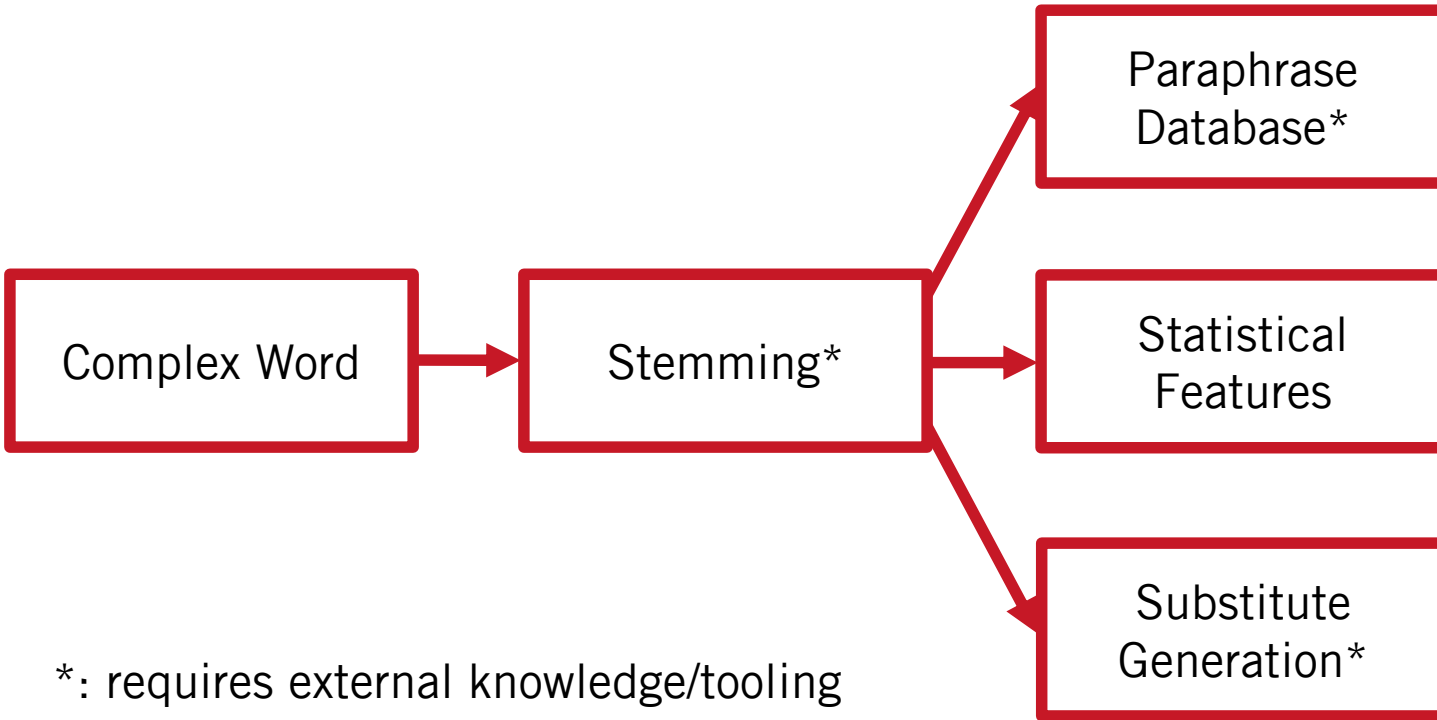


*: requires external knowledge/tooling

Source: Qiang et al., 2020: LSBert: A Simple Framework for Lexical Simplification

Aumiller and Gertz: "UniHD at TSAR-2022: Is Compute All We Need for Lexical Simplification?"

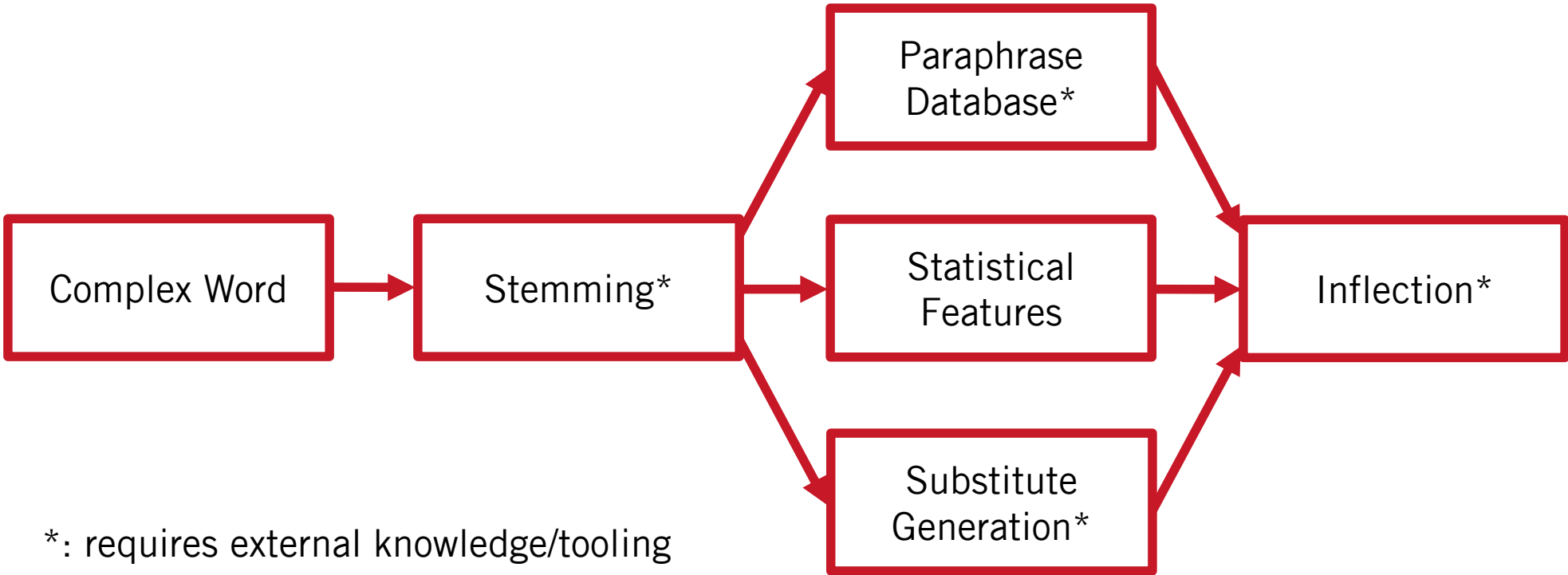
Current Simplification Systems



*: requires external knowledge/tooling

Source: Qiang et al., 2020: LSBert: A Simple Framework for Lexical Simplification

Current Simplification Systems



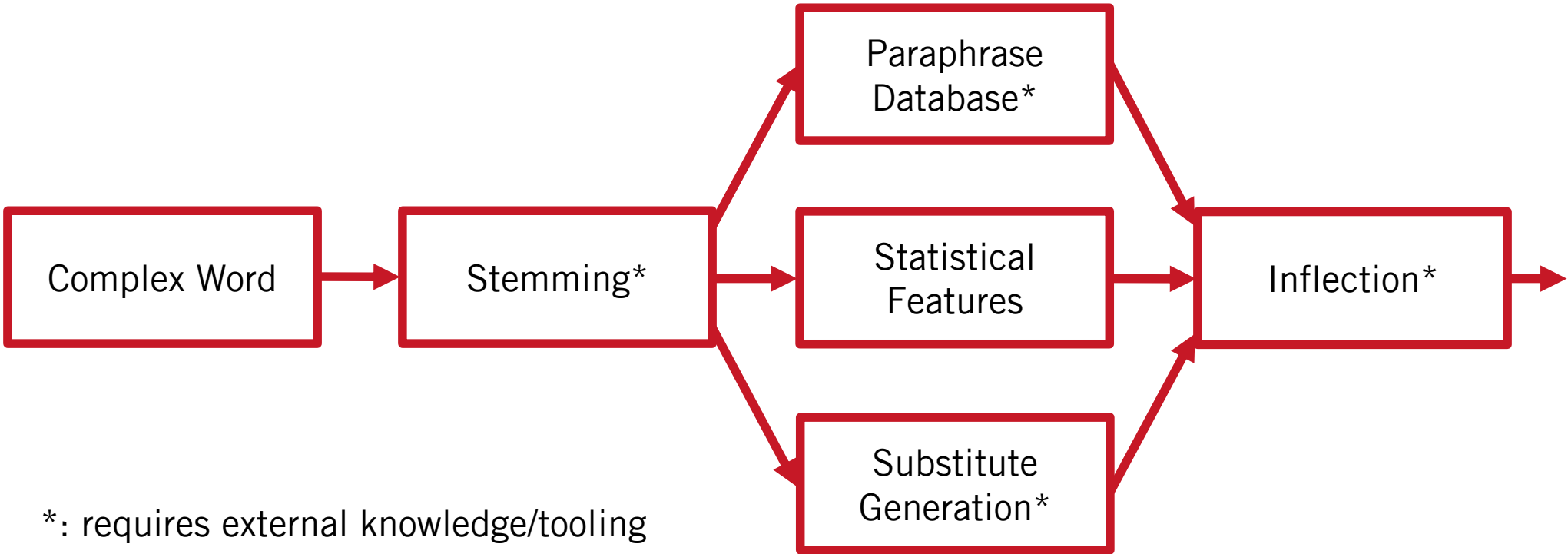
*: requires external knowledge/tooling

Source: Qiang et al., 2020: LSBert: A Simple Framework for Lexical Simplification

Current Simplification Systems



UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386



*: requires external knowledge/tooling

Source: Qiang et al., 2020: LSBert: A Simple Framework for Lexical Simplification

Aumiller and Gertz: "UniHD at TSAR-2022: Is Compute All We Need for Lexical Simplification?"

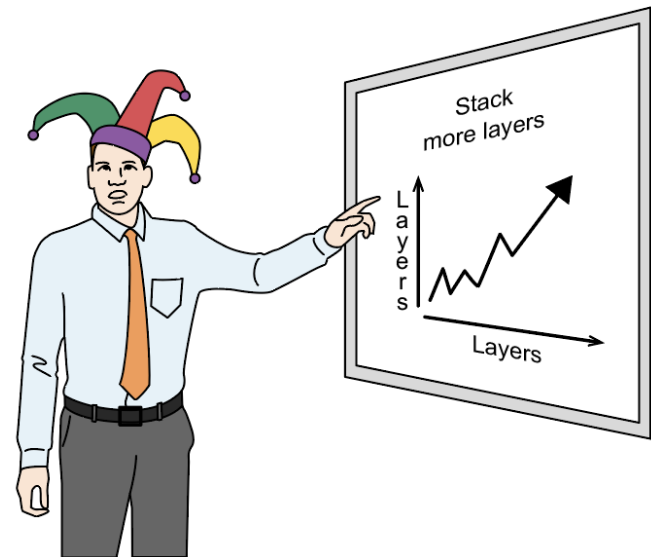


Wishful Thinking

- A **simple pipeline**, that could do all of that (but better)?
 - No domain-specific resources required
 - No tedious adaptation to new domains/languages

Wishful Thinking

- A **simple pipeline**, that could do all of that (but better?)
 - No domain-specific resources required
 - No tedious adaptation to new domains/languages



Source: <https://freecontent.manning.com/neural-network-architectures/>



Contributions

- Consider very large LMs as zero-shot recommenders for few-shot (contextualized) Lexical Simplification
- Improve prediction diversity and coverage by sampling from multiple system prompts



UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

The Magical World of Prompt Design

Task Setup



- Shared task trial data: only 10 labeled samples
 - Require great zero-shot performance
- Post-filtering as quality control

Task Setup



- Shared task trial data: only 10 labeled samples
 - Require great zero-shot performance
- Post-filtering as quality control





Model Considerations

- Large enough to exhibit emergent behavior
 - >60B parameters
- Provide reasonable API or inference endpoints
- Explicitly trained on (at least) one shared task lang
 - Consider “parameters per language” in train data



Model Considerations

- Large enough to exhibit emergent behavior
 - >60B parameters
- Provide reasonable API or inference endpoints
- Explicitly trained on (at least) one shared task lang
 - Consider “parameters per language” in train data
- Went with GPT-3 (*text-davinci-002*)



Play around, Find out

- Finding the “best” prompt is difficult
 - No observable gradients
 - Inference costs
- Bigger models -> higher prompt tolerance



Play around, Find out

- Finding the “best” prompt is difficult
 - No observable gradients
 - Inference costs
- Bigger models -> higher prompt tolerance

Give me five easier to understand synonyms for the following word: instilled

A: imparted, injected, infused, permeated, steeped



Play around, Find out

- Finding the “best” prompt is difficult
 - No observable gradients
 - Inference costs
- Bigger models -> higher prompt tolerance

Give me five easier to understand synonyms for the following word: instilled

A: imparted, injected, infused, permeated, steeped

Give me five easier to understand synonyms for the following word: instilled

1. implanted
2. inserted
3. infused
4. permeated
5. saturated



Play around, Find out

- Finding the “best” prompt is difficult
 - No observable gradients
 - Inference costs
- Bigger models -> higher prompt tolerance

Give me five easier to understand synonyms for the following word: instilled

A: imparted, injected, infused, permeated, steeped

Give me five easier to understand synonyms for the following word: instilled

1. implanted
2. inserted
3. infused
4. permeated
5. saturated

Ensuring Structured Outputs



- “Exam question trick”
 - Adding explicit cues improves consistency
 - Still no guarantee for consistency!

Question: List ten alternatives for the word "instilled" that are easier to understand.

Answer:

1. Taught
2. Trained
3. Educated
4. instructed
5. imparted
6. conveyed
7. communicated
8. imparted
9. introduced
10. implanted



Incorporating Context

Context: Rajoy's conservative government had instilled markets with a brief dose of confidence by stepping into Bankia, performing a U-turn on its refusal to spend public money to rescue banks.

Question: Given the above context, list ten alternative words for "instilled" that are easier to understand.

Answer:

1. planted
2. sowed
3. introduced
4. infused
5. instilled
6. imparted
7. injected
8. communicated
9. conveyed
10. transferred

- Downside: More context -> higher inference cost



Filtering

- Merge different output formats
- Filter identity predictions & repetitions
- Remove verbose substitutes

Question: List ten alternatives for the word "maniacs" that are easier to understand.

Answer:

1. people who are out of control
2. people who are crazy
3. people who are insane
4. people who are demented
5. people who are deranged
6. people who are disturbed
7. people who are erratic
8. people who are wild
9. people who are uncontrolled
10. people who are unhinged

Official Results



- Particularly impressive @1 & @3 performance
- @10 performance drops of

Run	ACC@1	Acc@k@Top1			MAP@k			Potential@k		
		$k = 1$	$k = 2$	$k = 3$	$k = 3$	$k = 5$	$k = 10$	$k = 3$	$k = 5$	$k = 10$
Single (Ours)	0.7721	0.4262	0.5335	0.5710	0.5090	0.3653	0.2092	0.8900	0.9302	0.9436
MANTIS-1	0.6568	0.319	0.4504	0.5388	0.473	0.3599	0.2193	0.8766	0.9463	0.9785
UoM&MMU-1	0.6353	0.2895	0.4530	0.5308	0.4244	0.3173	0.1951	0.8739	0.9115	0.9490
LSBert	0.5978	0.3029	0.4450	0.5308	0.4079	0.2957	0.1755	0.8230	0.8766	0.9463
TUNER	0.3404	0.1420	0.1689	0.1823	0.1706	0.1087	0.0546	0.4343	0.4450	0.4450



UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

Improving Coverage of Predictions



Improving Coverage

- Filtering removes “too many” candidates
- Candidate sets can be homogenous
- Order of candidates suboptimal



Improving Coverage

- Filtering removes “too many” candidates
- Candidate sets can be homogenous
- Order of candidates suboptimal

Idea: Query with multiple (different) prompts!

- Add few-shot examples
- With/Without context sentence

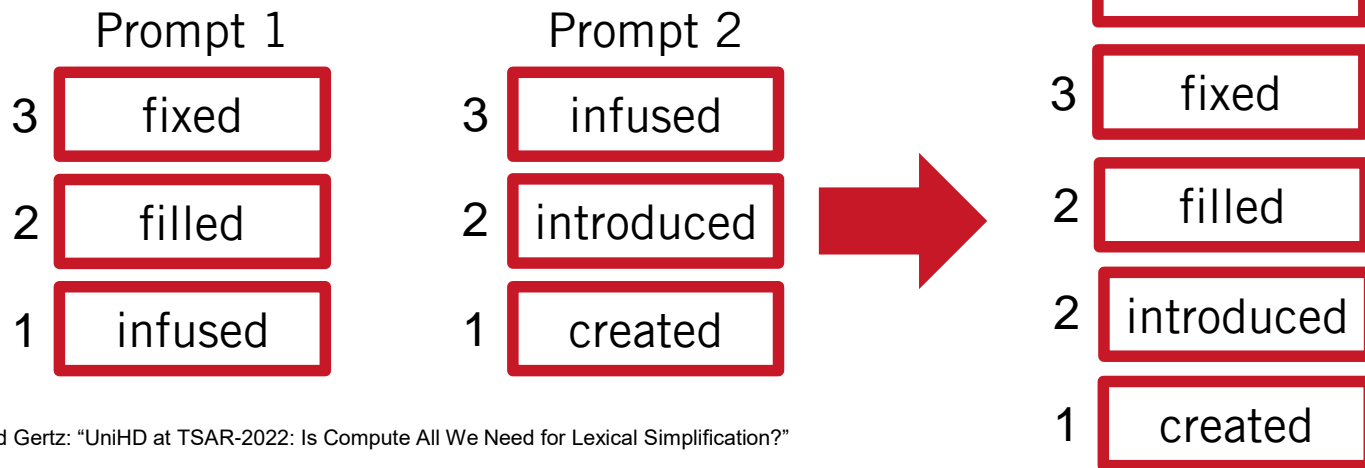
Merging Predictions

- Assign scores to each prompt-specific result set
- Aggregate & re-rank based on total score
- Cut down to k predictions

Prompt 1		Prompt 2	
3	fixed	3	infused
2	filled	2	introduced
1	infused	1	created

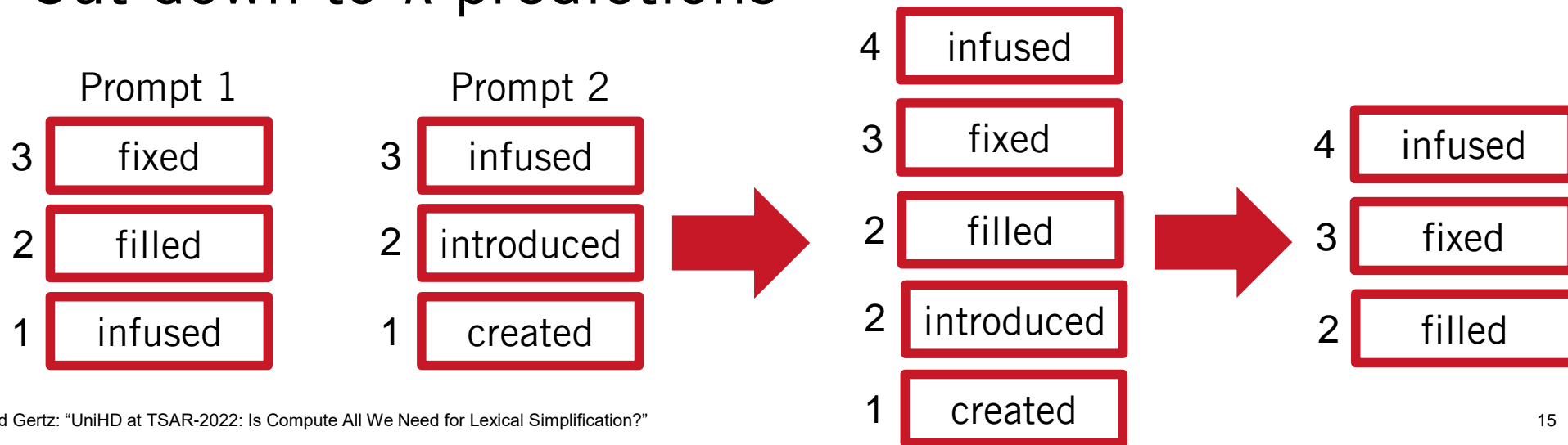
Merging Predictions

- Assign scores to each prompt-specific result set
- Aggregate & re-rank based on total score
- Cut down to k predictions



Merging Predictions

- Assign scores to each prompt-specific result set
- Aggregate & re-rank based on total score
- Cut down to k predictions





Updated Results

- Slight improvements @1
- Drastically improved @10 performance

Run	ACC@1	Acc@k@Top1			MAP@k			Potential@k		
		$k = 1$	$k = 2$	$k = 3$	$k = 3$	$k = 5$	$k = 10$	$k = 3$	$k = 5$	$k = 10$
Ensemble (Ours)	0.8096	0.4289	0.6112	0.6863	0.5834	0.4491	0.2812	0.9624	0.9812	0.9946
Single (Ours)	0.7721	0.4262	0.5335	0.5710	0.5090	0.3653	0.2092	0.8900	0.9302	0.9436
MANTIS-1	0.6568	0.319	0.4504	0.5388	0.473	0.3599	0.2193	0.8766	0.9463	0.9785
UoM&MMU-1	0.6353	0.2895	0.4530	0.5308	0.4244	0.3173	0.1951	0.8739	0.9115	0.9490
LSBert	0.5978	0.3029	0.4450	0.5308	0.4079	0.2957	0.1755	0.8230	0.8766	0.9463
TUNER	0.3404	0.1420	0.1689	0.1823	0.1706	0.1087	0.0546	0.4343	0.4450	0.4450



UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

Bonus: Multilingual Extension

Language Transfer with LLMs



UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

- What if we “ask nicely”?

Language Transfer with LLMs



UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

- What if we “ask nicely”?

Context: Además de partidos de fútbol americano, el estadio ha sido utilizado para una gran variedad de eventos, entre los que se destacan varios partidos de la selección nacional de fútbol de los Estados Unidos, y fue el hogar del ahora difunto club de la MLS, el Tampa Bay Mutiny.

Question: Given the above context, list ten alternative words for "difunto" that are easier to understand.

Answer:

1. deceased
2. late
3. departed
4. gone
5. lost
6. forgotten
7. unlamented
8. dead
9. defunct
10. extinct

Language Transfer with LLMs



UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

- What if we “ask nicely”?

Context: Además de partidos de fútbol americano, el estadio ha sido utilizado para una gran variedad de eventos, entre los que se destacan varios partidos de la selección nacional de fútbol de los Estados Unidos, y fue el hogar del ahora difunto club de la MLS, el Tampa Bay Mutiny.

Question: Given the above context, list ten alternative words for "difunto" that are easier to understand.

Answer:

1. deceased
2. late
3. departed
4. gone
5. lost
6. forgotten
7. unlamented
8. dead
9. defunct
10. extinct

Context: Además de partidos de fútbol americano, el estadio ha sido utilizado para una gran variedad de eventos, entre los que se destacan varios partidos de la selección nacional de fútbol de los Estados Unidos, y fue el hogar del ahora difunto club de la MLS, el Tampa Bay Mutiny.

Question: Given the above context, list ten alternative Spanish words for "difunto" that are easier to understand.

Answer: 1. Fallecido

2. Difunto
3. Muerto
4. Falecido
5. Difunto
6. Fallecido
7. muerto
8. desaparecido
9. extinto
10. finado

Spanish and Portuguese Results



Spanish:

Run	ACC@1	Acc@k@Top1			MAP@k			Potential@k		
		$k = 1$	$k = 2$	$k = 3$	$k = 3$	$k = 5$	$k = 10$	$k = 3$	$k = 5$	$k = 10$
Ensemble (Ours)	0.6521	0.3505	0.5108	0.5788	0.4281	0.3239	0.1967	0.8206	0.8885	0.9402
Single (Ours)	0.5706	0.3070	0.3967	0.4510	0.3526	0.2449	0.1376	0.6902	0.7146	0.7445
PresiUniv-1	0.3695	0.2038	0.2771	0.3288	0.2145	0.1499	0.0832	0.5842	0.6467	0.7255
UoM&MMU-3	0.3668	0.1603	0.2282	0.269	0.2128	0.1506	0.0899	0.5326	0.6005	0.6929
LSBert	0.2880	0.0951	0.1440	0.1820	0.1868	0.1346	0.0795	0.4945	0.6114	0.7472
TUNER	0.1195	0.0625	0.0788	0.0842	0.0575	0.0356	0.0184	0.144	0.1467	0.1494

Spanish and Portuguese Results



Spanish:

Run	ACC@1	Acc@k@Top1			MAP@k			Potential@k		
		$k = 1$	$k = 2$	$k = 3$	$k = 3$	$k = 5$	$k = 10$	$k = 3$	$k = 5$	$k = 10$
Ensemble (Ours)	0.6521	0.3505	0.5108	0.5788	0.4281	0.3239	0.1967	0.8206	0.8885	0.9402
Single (Ours)	0.5706	0.3070	0.3967	0.4510	0.3526	0.2449	0.1376	0.6902	0.7146	0.7445
PresiUniv-1	0.3695	0.2038	0.2771	0.3288	0.2145	0.1499	0.0832	0.5842	0.6467	0.7255
UoM&MMU-3	0.3668	0.1603	0.2282	0.269	0.2128	0.1506	0.0899	0.5326	0.6005	0.6929
LSBert	0.2880	0.0951	0.1440	0.1820	0.1868	0.1346	0.0795	0.4945	0.6114	0.7472
TUNER	0.1195	0.0625	0.0788	0.0842	0.0575	0.0356	0.0184	0.144	0.1467	0.1494

Portuguese:

Run	ACC@1	Acc@k@Top1			MAP@k			Potential@k		
		$k = 1$	$k = 2$	$k = 3$	$k = 3$	$k = 5$	$k = 10$	$k = 3$	$k = 5$	$k = 10$
Ensemble (Ours)	0.7700	0.4358	0.5347	0.6229	0.5014	0.3620	0.2167	0.9171	0.9491	0.9786
Single (Ours)	0.6363	0.3716	0.4625	0.5160	0.4105	0.2889	0.1615	0.7860	0.8181	0.8422
GMU-WLV-1	0.4812	0.2540	0.3716	0.3957	0.2816	0.1966	0.1153	0.6871	0.7566	0.8395
Cental-1	0.3689	0.1737	0.2433	0.2673	0.1983	0.1344	0.0766	0.524	0.5641	0.6096
LSBert	0.3262	0.1577	0.2326	0.286	0.1904	0.1313	0.0775	0.4946	0.5802	0.6737
TUNER	0.2219	0.1336	0.1604	0.1604	0.1005	0.0623	0.0311	0.2673	0.2673	0.2673



UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

Take Away and Limitations

Is Compute All We Need?



UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

- **Maybe?**
 - Great performance
 - Expensive inference
 - Hard to steer generation
- Useful for silver labels?

Thank you for your attention!



UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

Code: <https://github.com/dennlinger/TSAR-2022-Shared-Task>

Any questions?



@d_aumiller

aumiller@informatik.uni-heidelberg.de



Alternative Models:

- Smaller models struggle with coherence

Name ten alternative words for "instilled" that are easier to understand.

instructed, instilled, ingrained, ingrained, ingrained, implanted, implanted, implanted,
implanted, planted

All Prompts:



UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

- TODO