Laboratoire
Linguistique, **Langues**, **Parole** | LILPA | UR 1339
Université de Strasbourg

Language, Inclusion, Remediation, Interculturality & Communication | LiRiC
The **interdisciplinary thematic** institutes
of the University of Strasbourg & CNRS & Inserm

ACL 2025 VIENNA
JULY 27 – AUGUST 1

GeNRe: A French Gender-Neutral Rewriting System Using Collective Nouns

Enzo Doyen & Amalia Todirascu (LiLPa / University of Strasbourg, France)

1. Introduction

- Masculine generics (MG) in gender-marked languages (e.g., French, German, Dutch): use of the masculine form as a default/neutral form to refer to mixed group of men/women. Examples in French:

  (a) « Les **étudiants**[MG] sont partis. » (**Students**[MG] left.)

- Psycholinguistics studies show that **MG induce gender bias** and **amplify male-centric mental representations**[1,2]
- Task of **gender rewriting**: propose alternatives to gendered sentences to balance datasets and reduce use of MG and related biases

2. Task of Gender Rewriting

- **Three types** of existing gender rewriting systems: **neutral** rewriting (English[3,4]), **inclusive** rewriting (French[5], Portuguese[6], German[7]) and **gender-switching** rewriting (Arabic[8])
- **No neutral rewriting system for French**, and **collective nouns (gender-fixed in French) not tested** as potentially good candidates for gender-neutralization
- Automatically switching from MG to collective nouns in French requires **number and gender changes, re-inflection**

Masculine generics	Collective noun
Tous les anciens employés[MG] sont invités au programme.	L'ancien personnel est invité au programme.
(All former employees[MG] are invited to the program.) →	(Former staff is invited to the program.)

3. Methodology

a. Create a French member noun-collective noun dictionary based on the work of **Lecolle (2019)**[9], corpus search and scraping (**315** entries total)

Member noun	Collective noun
employés (*employees*)	personnel (*staff*)
jurés (*jury members*)	jury
policiers (*police officers*)	police
soldats (*soldiers*)	armée (*army*)
…	…

b. Use 3 approaches: RBS (**spaCy**[10] for syntactic dependency detection; **inflecteur**[11] for re-inflecting); fine-tuned models (M2M100/T5) and instruct-based model (Claude 3 Opus)
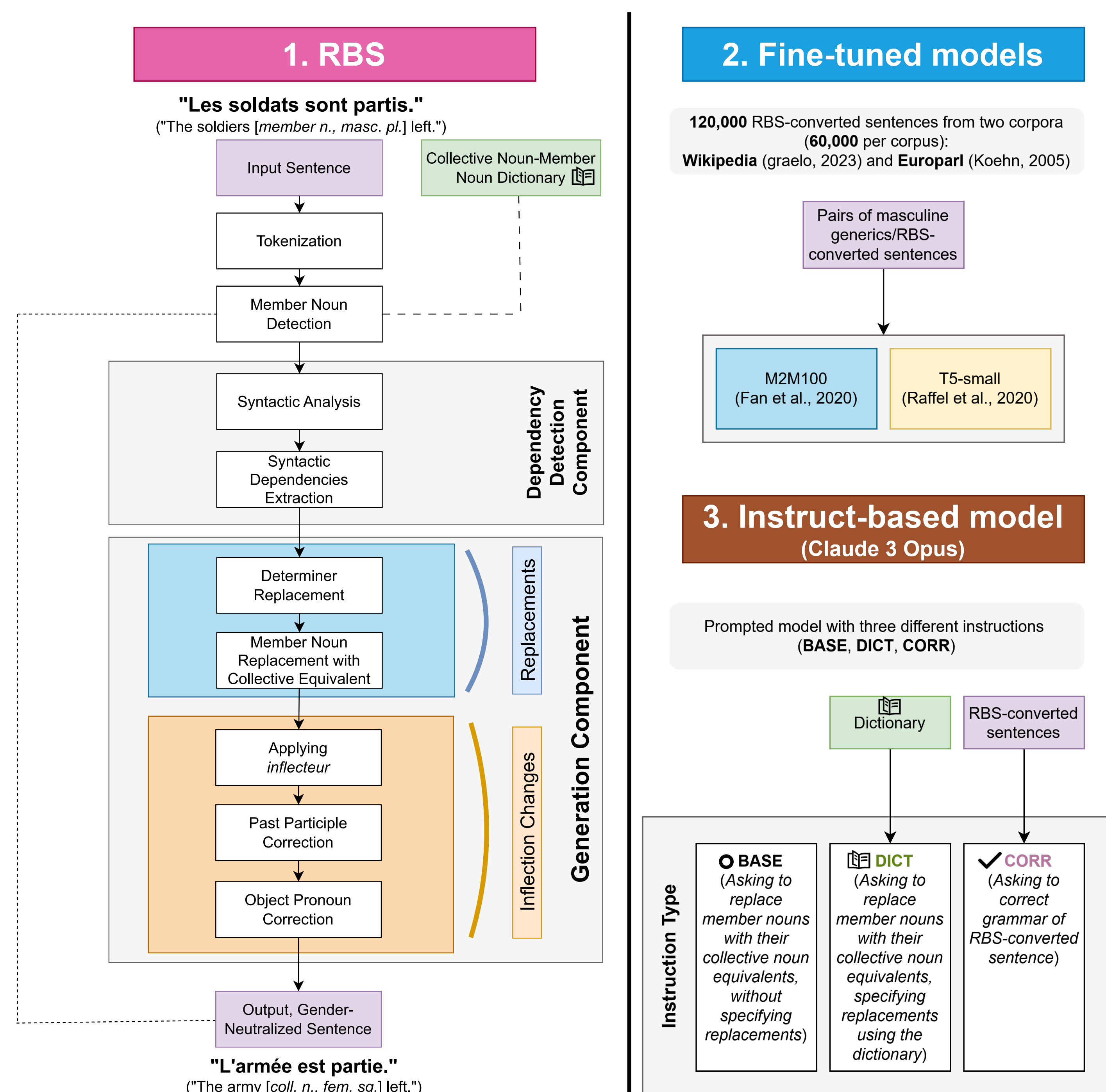
4. Evaluation and Results

- Evaluation dataset of 500 sentences (250 Wikipedia/250 Europarl)
- **RBS Dependency Detection Component**: Manually annotated syntactic dependencies in evaluation sentences
  - ⊙ Objective: evaluate correct detection of syntactic dependencies of the member noun to be modified
  - ⊙ Baseline: default spaCy-detected member noun's dependencies (excluding punctuation)

|	Wikipedia			Europarl			Avg.		
	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
Baseline	0.096	0.723	0.169	0.115	0.689	0.197	0.1055	0.706	0.183
GeNRe-RBS	0.773	0.855	0.812	0.758	0.813	0.785	0.7655	0.834	0.7985

RBS Dependency Detection Component Results

- **Generation (All Models)**: Manually gender-neutralized evaluation sentences using collective nouns
  - ⊙ Objective: evaluate correct correct conversion of sentences
  - ⊙ Baseline: original, unconverted sentence (as per previous works)

	Wikipedia			Europarl			Avg.		
	WER (↓)	BLEU (↑)	Cosine (↑)	WER (↓)	BLEU (↑)	Cosine (↑)	WER (↓)	BLEU (↑)	Cosine (↑)
Baseline	12.611%	80.688	97.436	12.446%	82.871	97.008	12.529%	81.779	97.222
GeNRe-RBS	**4.03%**	92.096	**98.88**	**3.814%**	93.707	99.22	**3.81%**	92.887	**99.05**
GeNRe-T5	6.726%	87.358	98.508	4.259%	93.111	99.1	5.492%	90.234	98.804
GeNRe-M2M100	6.566%	88.186	97.232	4.247%	93.197	98.992	5.406%	90.692	98.112
Claude-BASE	13.87%	80.205	96.532	10.713%	85.313	97.128	12.291%	82.759	96.83
Claude-DICT	4.702%	**92.79**	98.812	4.197%	**94.247**	99.264	4.45%	**93.519**	99.038
Claude-CORR	11.282%	84.954	98.092	8.992%	85.257	98.056	10.137%	85.25	98.074

Global Generation Results

- **Generation Errors**: For RBS and fine-tuned models, two annotators manually annotated with subcategories (italicized in plot)
- For Claude 3 Opus, used LLM-as-a-judge to automatically annotate in a two-step process: 1. ask to describe the error; 2. assign a subcategory

Error Counts by Category and Model
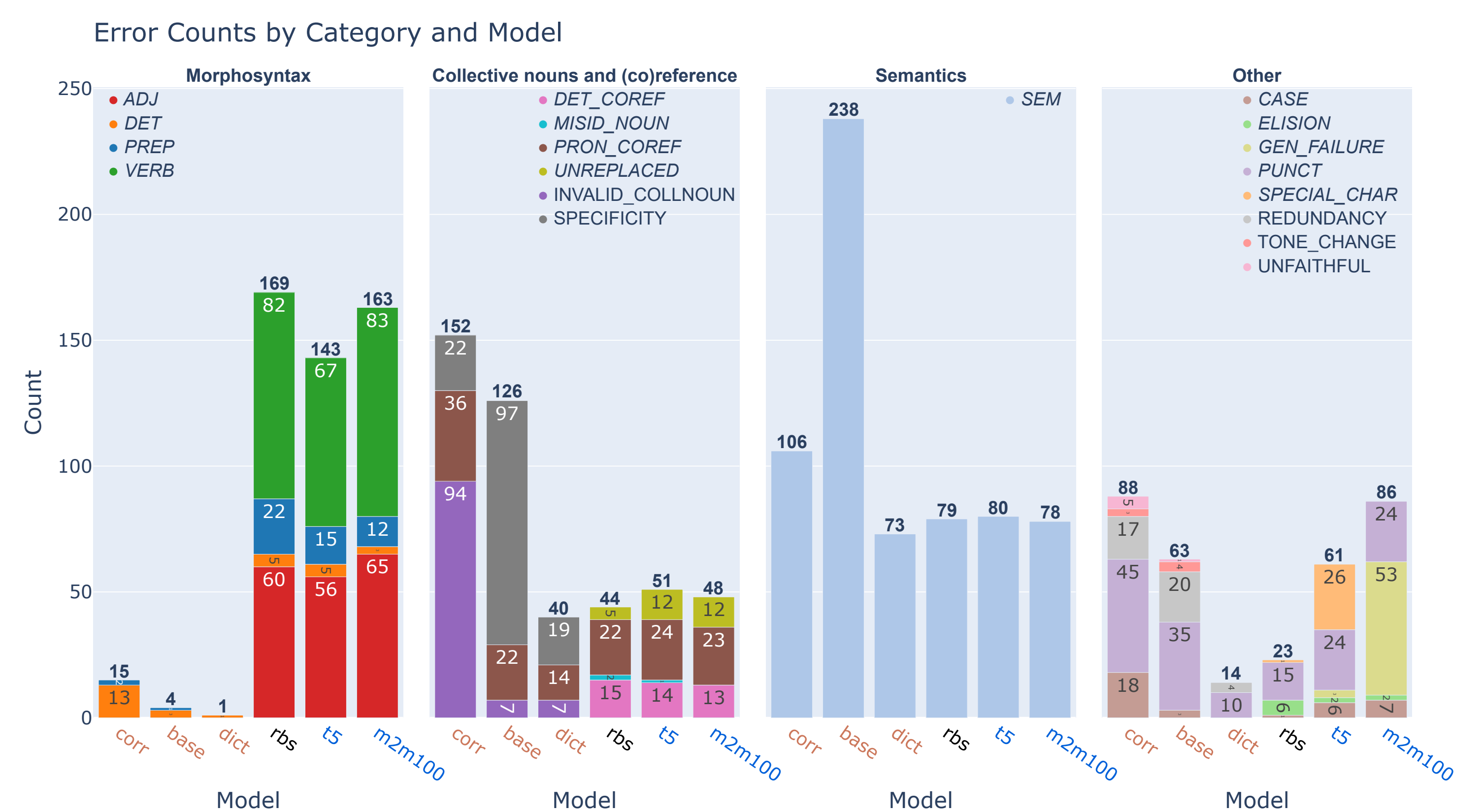
5. Takeaways

- RBS and instruct-based model combined with pre-created resources achieve similar results; lower morphosyntactic errors for Claude 3 Opus
- Collective nouns useful for gender neutralization, but limited by restrictive semantics

References

[1] Braun, F. et al. (2005). Cognitive Effects of Masculine Generics in German. *Communications*, 30(1).

[2] Gygax, P. et al. (2012). The Masculine Form and Its Competing Interpretations in French. *Journal of Cognitive Psychology*, 24(4).

[3] Vanmassenhove, E. et al. (2021). NeuTral Rewriter.

[4] Sun, T. et al. (2021). They, Them, Theirs. arXiv.

[5] Lerner, P. et al. (2024). INCLURE: A Dataset and Toolkit for Inclusive French Translation.

[6] Veloso, L. et al. (2023). A Rewriting Approach for Gender Inclusivity in Portuguese. In *Findings of the Association for Computational Linguistics: EMNLP 2023* (pp. 8747–8759). Singapore: Association for Computational Linguistics.

[7] Pomerenke, D. (2022). INCLUSIFY.

[8] Habash, N. et al. (2019). Automatic Gender Identification and Reinflection in Arabic. In *Proceedings of the First Workshop on Gender Bias in Natural Language Processing* (pp. 155–165). Florence, Italy: Association for Computational Linguistics.

[9] Lecolle, M. (2019). *Les Noms Collectifs Humains En Français. Enjeux Sémantiques, Lexicaux et Discursifs.* (C. Jacquet-Pfau, Ed.). Université de Lorraine: Lambert-Lucas.

[10] Montani, I. et al. (2024). Explosion/spaCy.

[11] Chuttarsing, A. (2021). Inflecteur.