INCLURE: a Dataset and Toolkit for Inclusive French Translation
Paul Lerner, Cyril Grouin

To cite this version:
Paul Lerner, Cyril Grouin. INCLURE: a Dataset and Toolkit for Inclusive French Translation. The 17th Workshop on Building and Using Comparable Corpora (BUCC @ LREC 2024), 2024, Turin, Italy. hal-04531938

HAL Id: hal-04531938
https://hal.science/hal-04531938
Submitted on 5 Apr 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Inclure: a Dataset and Toolkit for Inclusive French Translation

Paul Lerner¹, Cyril Grouin²
Sorbonne Université, CNRS, ISIR¹, Université Paris-Saclay, CNRS, LISN²
75005, Paris, France¹, 91400, Orsay, France²
lerner@isir.upmc.fr, cyril.grouin@liris.upsaclay.fr

Abstract
Inclusive French (gender-neutral language) is a variety of French that is used to highlight awareness of gender and identity against Standard French, which enforces the use of masculine for generic usage or plural. Although widely used and challenging to a set of NLP tools, Inclusive French was very little studied in NLP. Detractors of Inclusive French argue that it is difficult to read, while its supporters argue that it provides a fairer representation of women and gender minorities. We provide Inclure, the first large-scale parallel corpus for Standard to Inclusive French translation, and vice-versa, thus providing a “bilingual” access to French, for both detractors and supporters of Inclusive French. This corpus comes with a toolkit that can be readily applied to larger French corpora and could be extended to other languages, for which the number of inclusive varieties is growing. We also provide Fabien.ne BARTHez, a sequence-to-sequence model trained on Inclure. Apart from its direct application to translation, this model could also be used in most NLP pipelines, either as a pre-processing step to improve downstream processing or as a post-processing according to the user’s preference.

Keywords: Inclusive French, Gender-neutral Language, Parallel Corpus, Neural Machine Translation

1. Introduction

Inclusive French (gender-neutral language) is a variety of French used to highlight awareness of gender and identity against Standard French, which enforces the use of masculine for generic usage (e.g., un doctorant se doit de publier¹) or plural (e.g., mon frère et ma sœur sont des doctorants²). Inclusive French would include women in these speeches mainly in two different manners (Grouin, 2022) (see Figure 1):

1. coordination of feminine and masculine forms: un doctorant ou une doctorante;
2. morphological combination of masculine and feminine flectional endings (colloquially known as inclusive writing or écriture inclusive in French): un.e doctorant.e.

Although Inclusive French is prone to controversy³, several studies have found that Standard French shadows women and impacts the mental representations of the speakers (Sczesny et al., 2016). To avoid this issue, Touraille and Allassonnière-Tang (2023) argued generalizing gender-neutral words in French by proposing a new non-binary inflexional ending⁴. Other studies focus on the perception of sentences written in inclusive French, highlighting that feminization and coordination of feminine and masculine forms are better accepted than other processes (Delaborde et al., 2021). We choose not to choose. With the Inclure dataset and toolkit, anyone should be able to translate from Standard to Inclusive French, and vice-versa, thus providing “bilingual” access to French.

Inclusive French was very little studied in the NLP community. To our knowledge, this is only the second study of Inclusive French, after the exploratory study of Grouin (2022), and the first for Inclusive French Translation. We propose:

• Inclure, a dataset of 69K aligned sentences (bitext)⁶;
• Fabien.ne BARTHez, a sequence-to-sequence model trained on Inclure, able to translate from Standard to Inclusive French, and vice-versa⁷.

¹Meaning “a PhD Student must publish”. The feminine form of un doctorant is une doctorante.
²Meaning “My brother and sister are PhD students”.
³The Académie Française considers that Inclusive French puts the French language “in mortal peril” and wishes to ban its usage (Grouin, 2022). The Rassemblement National of Marine Le Pen shares this opinion and proposed another law to ban Inclusive French on October 12th, 2023 https://www.assemblee-nationale.fr/dyn/16/textes/l16b0777_proposition-loi.
⁴The authors proposed to use the final vowel “-i” to produce non-binary words: il doctoranti est heureusi meaning “the Ph.D. student is happy”.
⁵We use the term translate for lack of a better one, but the problem is much simpler than translating from French to any other language. Standard and Inclusive French are but varieties of the same language, the grammar is identical. This will be further demonstrated in Section 5.
⁶https://huggingface.co/datasets/PaulLerner/oscar_inclure
⁷https://huggingface.co/PaulLerner/fabien.ne_barthez
2. Related Work

Translating from Inclusive to Standard French as pre-processing in an NLP pipeline would broadly relate our work to other studies that tackle out-of-vocabulary words (Spriet et al., 1996; Maurel, 2004; Cartoni, 2008; Stouten et al., 2010; Rabary et al., 2015) or user-generated content (Baranes and Sagot, 2014; Farzindar and Roche, 2013; Benamara et al., 2018).

As for NLP studies of gender-neutral languages, Lauscher et al. (2022) focuses on coreference resolution to find that new gender-inclusive pronouns in English are challenging to state-of-the-art models. Grouin (2022) is the first NLP study of Inclusive French. Based on a very small corpus made of political speeches and French government publications (Inclusive French Corpus – IFC), they found that Inclusive French was challenging for two standard NLP tools, namely TreeTagger (Schmid, 1994) and spaCy9 (Montani et al., 2023). They study POS tagging, lemmatization, and Named Entity Recognition. They find that Inclusive French is much more challenging to these tools than Standard French.

However, their IFC corpus is too small to train a translation model (we identify 72 parallel sentences). We bridge this gap by proposing INCLURE, as described in the next section.

Other resources for Inclusive French, which have not made the object of a scientific publication, are available online10. However, they are limited to a bilingual dictionary (i.e., single-word translation) and only available through their GUI. In contrast, we propose open-source resource and models.

3. The Inclure Corpus

3.1. Methods

To build a corpus of parallel sentences (bitext) of Inclusive/Standard French, we seek to detect sentences in Inclusive French, and automatically translate them to Standard French using a rule-based system. We argue that such a system can easily be built for the Inclusive to Standard direction, but not the opposite (see Figure 1). In this regard, our strategy is similar to back-translation (Sennrich et al., 2016; Burlot and Yvon, 2018). Indeed, translating from Standard to Inclusive French is a difficult task, which requires solving the following semantic challenges:

1. knowing which nouns refer to people: e.g., “un doctorant” should be translated “un.e doctorant.e” because it refers to a person (PhD student) but “un marteau” should be kept “un marteau” because it refers to an object (a hammer);

2. resolving co-references: e.g., “Un doctorant se doit de publier. Il doit aussi...” where the pronoun il should be made inclusive vs. “Un marteau sert à planter des clous. Il sert aussi...” where the pronoun il should stay masculine.

---

8Our code is available at https://github.com/PaulLerner/inclure
9In particular, the fr_core_news_sm model.
10https://incluzor.org/ and https://eninclusif.fr/
A PhD Student must publish

Un.e doctorant ou doctorante se doit de publier

Syntactic parsing (spaCy)

Un.e doctorant ou doctorante se doit de publier

COORDINATION

Un.e doctorant ou doctorante se doit de publier

MORPHOLOGICAL COMBINATION

Regular expression

Un doctorant se doit de publier

Output sentence (Standard French)

Figure 2: Simplified diagram of our rule-based system for Inclusive to Standard French translation, used to generate the Inclure parallel corpus.

This task is best learned automatically from data, as described in Section 4.

More precisely, we focus on the two main processes of Inclusive French, which are easily detected automatically (see Figure 2):

1. coordination: e.g., un doctorant ou une doctorante is detected through a syntactic analysis: the head of doctorante is doctorant, but both share the same lemma;

2. morphological combination: e.g., les doctorant.e.s is detected through a regular expression.

The regular expression is built around common French feminine suffixes: {esse|sse|e|euse|se|ienne|enne|nne|ne|ère|ere|re|trice|rice|ice). Because Inclusive French is yet unstandardized, we see several variants of the same suffix, e.g., trice|rice|ice. These might occur in auteur.trice, auteur.rice, or auteur.ice (all meaning “author”). Likewise, the ordering of the masculine and feminine suffix is variable; both auteur.trice and auteur.ice are acceptable. Therefore, the core of our regex substitution method lies in two regexes:

• <FEM>s?\.(esse|sse|e|euse|se|ienne|enne|nne|ne|ère|ere|re|trice|rice|ice)

Where <FEM> stands for the feminine suffixes listed above. Parenthesis shows the captured sections of the string that are substituted back (e.g., teur in auteur.teur to obtain auteur, the masculine form). s marks the plural. Instead of [a-z], we use all lowercase French letters, including accents and diacritics ([a-zàâéèêëîïôùûüÿçæœ]). These might occur in auteur.trice, auteur.rice, or auteur.ice (all meaning “author”). Likewise, the ordering of the masculine and feminine suffix is variable; both auteur.trice and auteur.ice are acceptable. Therefore, the core of our regex substitution method lies in two regexes:

• {{a-z}+}<FEM>{s?}\b, when the feminine suffix comes after.

Table 1: Average sentence length, vocabulary size, and type-to-token ratio (TTR) of Inclure and the Inclusive French Corpus (IFC), in the Inclusive (I) or Standard (S) version.

<table>
<thead>
<tr>
<th></th>
<th>Inclure</th>
<th>IFC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length</td>
<td>33.0</td>
<td>29.4</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>70,200</td>
<td>66,500</td>
</tr>
<tr>
<td>TTR</td>
<td>0.93</td>
<td>0.90</td>
</tr>
</tbody>
</table>

Note that the interpunct (“.”, U+00B7) is frequently used as a separating sign instead of the dot (“.”, U+002E). However, the interpunct is absent in BARThez vocabulary (Eddine et al., 2021), which we use as a foundation model for our translation model (Section 4). Therefore, all interpuncts between two lowercase letters are replaced by dots in preprocessing.

3.2. Implementation

Syntactic dependency parsing, lemmatization, and morphological analysis are done using spaCy, more precisely the fr_dep_news_trf model, based on CamemBERT (Martin et al., 2020), which is pretrained on OSCAR 2019 (Suárez et al., 2019) and fine-tuned on the Sequoia Corpus (Candito et al., 2014). We use a single NVIDIA V100 GPU with 32GB of memory to process a subset of OSCAR 22.01 in 20 hours.

Our code is available so that Inclure can be easily extended to larger corpora and other languages.

3.3. Processing OSCAR

A random 1.3% of French OSCAR 22.01 was processed, that is 681K documents of a total 2.29M sentences. Our system estimates that 0.3% of these sentences are Inclusive French, yielding 69K aligned sentences (bitext) in Standard and Inclusive French. We denote the resulting dataset Inclure.

The dataset has a total vocabulary of 70,200 different words in its original Inclusive French and a smaller 66,500 words in the translated Standard French, as words have fewer inflected forms in Standard French. Likewise, we find Standard French
sentences to be shorter and with a smaller type-to-token ratio. These statistics are summarized in Table 1.

The dataset is split randomly into three subsets: train (90%), validation (5%), and test (5%). We show two random examples of the test set, for each Inclusive French process:

1. coordination: Toutes les informations utiles sur la sécurité des données et les éventuels risques pour la sécurité, sur le type d’enregistrement des données, leur étendue et leur conservation, et sur les droits des clientes et clients, doivent être communiquées. ⇐⇒ Toutes les informations utiles sur la sécurité des données et les éventuels risques pour la sécurité, sur le type d’enregistrement des données, leur étendue et leur conservation, et sur les droits des clients, doivent être communiquées.11

2. morphological combination: Le message est clair : ces organisations et personnalités sont accusé.e.s de complicité dans les attentats commis ces dernières semaines. ⇐⇒ Le message est clair : ces organisations et personnalités sont accusés de complicité dans les attentats commis ces dernières semaines12

4. Inclusive French Translation with Fabien.ne BARThez

4.1. Method

We adopt the now-standard learning method to translate end-to-end with a sequence-to-sequence model (Cho et al., 2014; Sutskever et al., 2014; Bahdanau et al., 2015), in either translation direction, while our main interest lies in the Standard to Inclusive direction.

The Transformer architecture, now more widely known for large language models, was originally proposed for translation and is well-suited for the task (Vaswani et al., 2017). We leverage the BARThez model of Eddine et al. (2021), a sequence-to-sequence model of 139M parameters13 pre-trained to reconstruct a corrupted input, in the manner of BART (Lewis et al., 2020), but for French instead of English. BARThez was pre-trained on 66 GB of French raw text from diverse sources, mostly from CommonCrawl. It uses the SentencePiece tokenizer (Kudo and Richardson, 2018) trained on a 10 GB random sample from their pre-training corpus. We leave studies on the impact of the vocabulary and tokenizer for future work.

Although the training data differs, we fine-tune BARThez using the same loss function as for its pre-training, i.e., minimizing the cross-entropy between the predicted output and the ground truth. Each prediction is conditioned on the whole input and the preceding output tokens, using teacher forcing as systematically done with Transformers. We note this fine-tuned model Fabien.ne BARThez.

4.2. Implementation and Hyperparameters

We use the same hyperparameters for both translation directions. The model is trained using the Adam optimizer (Kingma and Ba, 2015) with an initial learning rate of $5 \times 10^{-5}$ linearly decreasing for a maximum of 10K steps if training is not interrupted before, according to the validation loss. At inference, decoding is done using greedy search as we have found that beam search decreased BLEU on the validation set.

We use a single NVIDIA V100 GPU with 32GB of memory holding a batch of 128 aligned sentences. In both translation directions, models start overfitting, and training is interrupted after 3K steps (≈ 6 epochs), after about an hour of training.

Our implementation is based upon Transformers (Wolf et al., 2020), itself built upon PyTorch (Paszke et al., 2019). Our code is freely available to ensure the reproducibility of our results.

5. Results

5.1. Evaluation Data and Metric

In addition to the IID test set of Inclure, we evaluate the out-of-domain (OOD) performance of Fabien.ne BARThez using the Inclusive French Corpus of Grouin (2022). Indeed, this corpus mostly contains transcripts of political speeches, whose oral style differs from the text typically found in OSCAR/CommonCrawl. Exceptions are six examples used to illustrate the use of the inclusive neutralization process described by Alpheratz (2019). These six examples were written by Grouin (2022) to complete the coverage of their corpus, as they could not find the natural occurrence of this process, which hints at its rareness. We will return to these examples in Section 6.

As for Inclure, all separating signs of Inclusive French are normalized to use a standard dot (".",

---

11 Meaning “All relevant information on data security and possible security risks, on the type of data storage, its scope and retention, and on customer rights, must be provided.”

12 Meaning “The message is clear: these organizations and personalities are accused of complicity in the attacks of recent weeks.”

13 Eddine et al. (2021) report 165M parameters but we find 139M in their released model. The embedding layer of 38M parameters is tied to the output layer, counting it twice would result in 178M parameters.
We remove them from the Standard version of the which refers to Martinicans) in the Standard version (Post, 2018). We leave the study of other metrics and the OOD corpus, although no fine-tuning or hyperparameter-tuning on the Inclusive French Corpus.

Table 2: Main results: BLEU scores from Standard to Inclusive French. IID: results on the test set of INCLURE, after training and tuning hyperparameters on the dedicated IID subsets. OOD: out-of-domain results, without fine-tuning or hyperparameter-tuning on the Inclusive French Corpus.

<table>
<thead>
<tr>
<th>Model</th>
<th>IID</th>
<th>OOD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identity (baseline)</td>
<td>76.30</td>
<td>79.74</td>
</tr>
<tr>
<td>Fabien.ne BARTHez</td>
<td><strong>92.83</strong></td>
<td><strong>83.05</strong></td>
</tr>
</tbody>
</table>

Table 3: Additional results: BLEU scores from Inclusive to Standard French.

<table>
<thead>
<tr>
<th>Model</th>
<th>IID</th>
<th>OOD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identity (baseline)</td>
<td>77.12</td>
<td>79.89</td>
</tr>
<tr>
<td>Rule-based</td>
<td>–</td>
<td>86.63</td>
</tr>
<tr>
<td>Fabien.ne BARTHez</td>
<td><strong>96.07</strong></td>
<td><strong>94.60</strong></td>
</tr>
</tbody>
</table>

We will see in the next section that BLEU is better suited to evaluate Standard French outputs, where our model achieves nearly perfect BLEU scores on both the IID and OOD evaluation sets.

5.3. From Inclusive to Standard French

Although our main research interest lies in the Standard to Inclusive direction, we study in this section the opposite direction, both for completeness but also to demonstrate that our model generalizes beyond learning the inverse function of our rule-based system, which generated the training data (cf. Section 3.1). BLEU scores are reported in Table 3. In addition to the Identity baseline, we also report the performance of our rule-based system, which generated the INCLURE corpus. This system is, therefore, not evaluated on the IID subset where

14`refs:1|case:mixed|eff:no|tok:13a|smooth:exp|version:2.3.1`

15Meaning “Elected representatives’ allowances capped at median salary.”

16Meaning “Revocability of elected representatives.”

17Meaning “All those I didn’t get to see during this brief visit.”
it should get 100 BLEU. Because it was designed to be precise, sometimes at the expense of recall, it does not systematically detect Inclusive French in the OOD evaluation set. In this case, we fall back to the Identity baseline (i.e., compute the BLEU between the Inclusive French input and the Standard French ground truth).

The rule-based system outperforms the Identity baseline but is largely inferior to Fabien.ne BARTHez, which achieves near-perfect BLEU scores on both the IID and OOD evaluation sets, thus demonstrating its generalization capacities. Unlike the Standard to Inclusive direction, BLEU is reasonably well-suited to compare Standard French outputs to the ground truth. Coming back to our earlier examples, our model correctly predicts indemnités d’élu plafonnées au salaire médian and Révocabilité des élus, which perfectly match the ground truth.

Again, in the Inclusive to Standard direction, the irregularities of Inclusive French are smoothed out. For example, Tous ceux que je n’ai pu voir au cours de cette brève visite […] is correctly predicted, which explains the high BLEU scores.

### 6. Discussion

**Language fixation** Since the inclusive French language is constantly evolving, offering a variety of processes, we have not yet observed a language fixation of phrases produced by coordinating feminine and masculine words. In the Inclure corpus, we found about as many female-male coordinations as male-female coordinations (see Table 4). Nevertheless, we observed a majority of female-male coordinations (93%) in the IFC corpus. Despite its low number of examples, we hypothesize that political discourse mainly uses female-male coordination to highlight women for political reasons, fixing de facto those phrases. Adopting a linguistic point of view, we may consider that using female words first makes it more distinctive from standard French which uses male words to encompass both men and women (bonjour à toutes et à tous vs. bonjour à tous18).

**Inferring Feminization** We have focused on the two main phenomena of Inclusive French, coordination and morphological combination, which counteract Standard French’s use of masculine for generic usage or plural. However, another aspect of Inclusive French is the feminization of nouns that refer to women, particularly job titles. The IFC corpus contains a few of these examples, where feminization must be inferred from the gender of the name, e.g., Giorgia Marras, illustrateur et auteure de bande dessinée, est née à Gênes en Italie, en 198819 must be translated to Giorgia Marras, illustratrice et auteure de bande dessinée, est née à Gênes en Italie, en 1988 because Giorgia Marras is a woman, which may be inferred from her name.

Our model cannot infer this, because such examples are absent from Inclure. We leave this for future work. Wikidata may be a useful resource for this, as it currently holds 52K entities that have different feminine and masculine labels in French, e.g., Q644687 illustrateur or illustratrice20.

**Morphological Neutralization** As mentioned in Section 5.1, the IFC corpus of Grouin (2022) contains six synthetic examples, based on the work of Alpheratz (2019), to cover another rare process of Inclusive French: morphological neutralization. It consists in creating new neutral lexical units (e.g., træur, which means both frère or sœur) or new inflected forms (e.g. députæs instead of député.es). Our model did not learn those processes either, as they are absent from Inclure. However, we believe it may be addressed as a post-processing step according to the user’s preference (e.g., replacing é.es with æs). The same could be said about non-binary markers (e.g. député.e.x21).

**Rare words** Another limitation of our model, which we have observed on the OOD evaluation set, is its brittleness to rare words. For example, a speech beginning with Martiniquais [...] (addressing to Martinicans) is automatically translated to Martiniquais, Martiniciennes [...] instead of Martiniquaises, as ienne is a common feminine suffix. This paper tackles the translation from Inclusive French to Standard French, and vice-versa. Inclusive French is a gender-neutral language used to highlight an awareness of gender and identity against the generic use of masculine in Standard French.

<table>
<thead>
<tr>
<th>Inclure</th>
<th>IFC</th>
</tr>
</thead>
<tbody>
<tr>
<td>F M (toutes et tous)</td>
<td>51%</td>
</tr>
<tr>
<td>M F (tous et toutes)</td>
<td>49%</td>
</tr>
</tbody>
</table>

Table 4: Statistics of the gender ordering in coordinations, on both Inclure and the Inclusive French Corpus (IFC).

---

18Respectively “Good morning to all (women) and to all (men)” vs. “Good morning to all (men, including women)”

19Meaning “Giorgia Marras, illustrator and comic strip author, was born in Genoa, Italy, in 1988”

20https://w.wiki/7k3d

21According to https://eninclusif.fr/. The corpus of Grouin (2022) does not contain such examples.
from inclure.x import exclure
import spacy

model = spacy.load("fr_dep_news_trf")
# exclure yields aligned sentences for each sentence in the input text
list(exclure(model("Bonjour à toutes et tous")))

Listing 1: Generating parallel sentences using the inclure toolkit python interface

from transformers import pipeline, AutoModelForSeq2SeqLM

inclure = pipeline("text2text-generation", model="PaulLerner/fabien.ne_barthez")
# high-level pipeline to get the output directly
inclure("Bonjour à tous")

Listing 2: Translating from Standard to Inclusive French using Fabien.ne BARThez via the Transformers library

French. Inclusive French was shown to provide fairer representations to the speakers but is also criticized for being difficult to read. With inclure, we sought to provide a “bilingual” access to Standard and Inclusive French.

Despite being widely used and challenging to NLP tools, Inclusive French has been very little studied in NLP. We present the second study and the first for Inclusive French translation. We provide inclure, a dataset of 69K aligned sentences (bitext) as well as Fabien.ne BARThez, a model able to translate from Standard to Inclusive French, and vice-versa. This model generalizes very well to out-of-domain data, through experiments on the Inclusive French Corpus (IFC) of Grouin (2022).

INCLURE comes with a toolkit for automatic annotation, which can readily be applied to larger corpora and may be extended to languages other than French, as discussed in the next section. INCLURE comes with a CLI, which can generate new training data as python -m inclure.x <input> <output>, where <input> should contain JSONL files formatted as OSCAR. Listing 1 shows how to use the Python interface. The Fabien.ne BARThez translation models can be accessed directly through the Hugging Face prediction GUI or via the Transformers library, see Listing 2.

We discuss our perspectives for future work in the next section.

8. Future Work

8.1. Vocabulary and Tokenization

We adopted BARThez as the foundation model in this work and kept its SentencePiece tokenizer. This is, however, likely suboptimal because inclusive words (e.g., député.e.s) are over-tokenized (e.g. _député . e . s). We assume that morphological tokenization (e.g., _député + <inclusive plural>) would be beneficial. A first step would be training the SentencePiece tokenizer on an Inclusive French corpus such as inclure. Remember that the BARThez tokenizer does not contain the interpunct, which hints at how little Inclusive French it was trained on (e.g., _député e s is tokenized into _député <unk> e <unk> s).

However, switching tokenizers would imply retraining the model from scratch, which would allow studying two additional factors:

- the model size: do we need 139M parameters?
- its pre-training: is BARThez’ pretraining (corrupted input reconstruction) beneficial to Inclusive French Translation?

8.2. More Processes for Inclusive French

In this work, we focused on two main processes used in Inclusive French, the coordination of feminine and masculine forms, and the combination of feminine and masculine flectional endings. We plan to add other existing processes to produce Inclusive French, such as feminization of job titles and neutralization of gendered forms in producing
new morphological forms (such as the controversial *il* personal pronoun including both masculine *il* and feminine *elle* pronouns). Another emerging process is proximity agreement, where the adjective agrees with the closest noun instead of keeping the generic masculine (e.g., *les garçons et les filles sont belles* instead of *beaux*23; Riban and Gerin, 2017). Such syntactic rules could be detected using a dependency parser, similarly to what is described in Section 3.1.

### 8.3. Beyond French

French is far from the only language with inclusive varieties (Sczesny et al., 2016). Spanish, for example, uses similar processes, e.g., using @ or x to mark neutral gender instead of o (masculine) and a (feminine), for example *latinx* (Lomotey, 2015). Our work could be easily extended to other inclusive languages, such as Inclusive Spanish.

### 8.4. Beyond BLEU

We found in Section 5.2 that BLEU was not always suited to evaluate Inclusive French generation, due to the irregularities of Inclusive French, and the semantic equivalence between its two main processes (coordination and morphological combination). The machine translation community is gradually moving away from surface metrics like BLEU in favor of neural metrics (Nakhlé, 2023), such as COMET (Rei et al., 2020) or BLEURT (Sellam et al., 2020). We should, however, be careful before using these metrics on Inclusive French, which may be out-of-domain of the underlying language model. We should first assess the correlation between these metrics and human judgments, which would need to be collected, e.g., for the corpus of Grouin (2022).

### 9. Acknowledgements

We thank the reviewers for their helpful feedbacks. This work was partly funded by the French Agence Nationale de la Recherche (ANR) under grant ANR-22-CE23-0033 / MaTOS.

### 10. Bibliographical References


---

23Meaning “the boys and girls are pretty”. *filles* is the closest noun to the adjective *belles*, which therefore agrees with the feminine.


11. Language Resource References

