Inclure: a Dataset and Toolkit for Inclusive French Translation

Paul Lerner[1], Cyril Grouin[2]
Sorbonne Université, CNRS, ISIR[1], Université Paris-Saclay, CNRS, LISN[2]
75005, Paris, France[1], 91400, Orsay, France[2]
lerner@isir.upmc.fr, cyril.grouin@lisn.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 (Alpheratz, 2018, 2019). Indeed, Standard French, as other languages (Hellinger and Bußmann, 2015), enforces the use of masculine for generic usage (e.g., *un doctorant se doit de publier*[1]) or plural (e.g., *mon frère et ma sœur sont des doctorants*[2]). 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 controversy[3], 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 ending[4]. 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[5] 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-versa[7].

[1] Meaning “a PhD Student must publish”. The feminine form of *un doctorant* is *une doctorante*.

[2] Meaning “My brother and sister are PhD students”.

[3] The *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.

[4] The authors proposed to use the final vowel “-i” to produce non-binary words: *li doctoranti est heureusi* meaning “the Ph.D. student is happy”.

[5] We 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.

[6] https://huggingface.co/datasets/PaulLerner/oscar_inclure

[7] https://huggingface.co/PaulLerner/fabien.ne_barthez

BUCC 2024: The 17th Workshop on Building and Using Comparable Corpora, pages 59–68

Standard French	Inclusive French
co-reference **Un doctorant** se doit de publier. **Il** doit aussi... PERSON	MORPHOLOGICAL COMBINATION **Un.e doctorant ou doctorante** se doit de publier. **Il ou elle** doit aussi... COORDINATION COORDINATION
A PhD Student must publish. They must also...	
co-reference **Un marteau** sert à planter des clous. **Il** sert aussi... OBJECT	**Un marteau** sert à planter des clous. **Il** sert aussi... *(identical)*
A hammer is used to drive nails. It is also used...	

Figure 1: Overview of our research problem and proposed solution: while translating from Standard to Inclusive French requires semantically-heavy capacities, such as knowing which nouns refers to a person or an object, or resolving co-references, the two main processes of Inclusive French, morphological combination and coordination, can be easily detected with a regular expression and a syntactic parser, respectively. Gender-marking words are printed in bold.

This reasonably large dataset enables further studies on Inclusive French translation, e.g., on the importance of vocabulary and tokenization, and comes with a rule-based system that can be readily applied to larger French corpora and could be extended to other languages[8]. Fabien.ne BARThez could be used directly by interested users. In NLP, it could also be applied either as pre-processing (e.g., when translating "*un.e doctorant.e se doit de publier*" to English, or post-processing (e.g., "French Ph.D. students are under-payed" may be translated either to Standard or Inclusive French depending on the user's preference).

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 spaCy[9] (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.

[8]Our code is available at `https://github.com/PaulLerner/inclure`

[9]In particular, the `fr_core_news_sm` model.

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 online[10]. 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.

[10]`https://incluzor.org/` and `https://eninclusif.fr/`

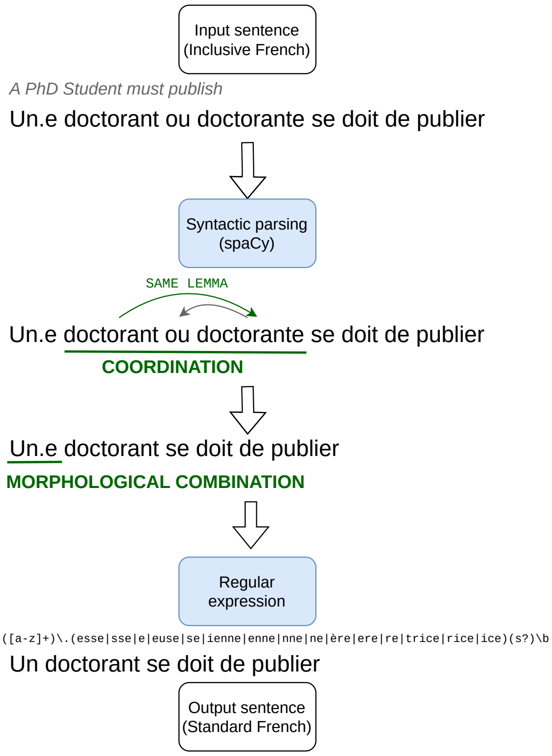

Figure 2: Simplified diagram of our rule-based system for Inclusive to Standard French translation, used to generate the INCLURE parallel corpus.

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 autrice.teur are acceptable. Therefore, the core of our regex substitution method lies in two regexes:

- <FEM>s?\.([a-z]+)\b, when the feminine suffix comes before the separating dot;
- ([a-z]+)\.<FEM>(s?)\b, when the feminine suffix comes after.

	INCLURE		IFC	
	I	S	I	S
Length	33.0	29.4	33.9	32.1
Vocabulary	70,200	66,500	899	860
TTR	0.93	0.90	0.91	0.87

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.

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 autrice.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]), but left them out above to improve readability.

Note that the interpunct ("·", U+00B7) is frequently used as a separating sign instead of the dot (".", U+002E). However, the interpunct is absent of 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.* $\iff$ *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.* $\iff$ *Le message est clair : ces organisations et personnalités sont* accusés *de complicité dans les attentats commis ces dernières semaines*[12]

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 parameters[13] 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 5×10^{-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.

Model	IID	OOD
Identity (baseline)	76.30	79.74
Fabien.ne BARThez	**92.83**	**83.05**

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.

U+002E), to ease evaluation. Note that the corpus of Grouin (2022) originally contained various separating signs in addition to the dot and interpunct, such as the slash, dash, and parenthesis. Moreover, Grouin (2022) kept the demonyms coordination (e.g. *les Martiniquaises et les Martiniquais*, which refers to Martinicans) in the Standard version of the corpus, as they are a kind of named entity. We remove them from the Standard version of the corpus as we are more interested in translation than named entity recognition. Additionally, we segment the corpus in sentences. This is easily done automatically as there is a 1-1 mapping between Standard and Inclusive French sentences, in the same order. We filtered out identical sentences in both varieties (as some documents contained mixed varieties) to arrive at 72 aligned sentences.

The dataset has a total vocabulary of 899 different words in its original Inclusive French and a smaller 860 words in the translated Standard French, similarly to INCLURE. Again, Standard French sentences are shorter and have a smaller type-to-token ratio. These statistics are summarized in Table 1.

Quantitative evaluation is done using BLEU (Papineni et al., 2002) implemented with SacreBLEU[14] (Post, 2018). We leave the study of other metrics for translating Inclusive French to future work, as they would require collecting human judgments.

5.2. From Standard to Inclusive French

Our main results, translating from Standard to Inclusive French, are reported in Table 2. As both varieties of French are close, we use as a baseline the identity function, i.e., simply computing the BLEU score between the Standard French input and Inclusive French ground truth. This baseline, or lower bound, gives very high BLEU scores, between 76 and 80, depending on the evaluation corpus.

Fabien.ne BARThez nevertheless largely outperforms the baseline, on both the IID test set and the OOD corpus, although no fine-tuning or hyperparameter-tuning was done on the latter. We find, however, a 10 absolute BLEU point gap between the two corpora, which would suggest a poorer performance of our model on the OOD corpus. Our qualitative analysis reveals, however, that most OOD examples with relatively modest BLEU scores are semantically equivalent, because of the limitations of the surface metric that is BLEU. Take for example the ground-truth *Indemnités d'élu.e plafonnées au salaire médian.*[15], for which our model provided *Indemnités d'élu et d'élue plafonnées au salaire médian.*, preferring the coordination process over the morphological combination process, and scoring only 51 BLEU. It is even worse for *Révocabilité des élu.e.s.*[16] vs. *Révocabilité des élues et élus.*, which scores only 13 BLEU, despite being equivalent. Likewise, while the ordering of the feminine (*élues*) and masculine (*élus*) does not matter, *Révocabilité des élues et élus.* vs. *Révocabilité des* élus et élues*.* would only score 21 BLEU.

Furthermore, Inclusive French is sometime inconsistent, especially in its oral form present in the Inclusive French Corpus. For example, one speech begins with Tous ceux *que je n'ai pu voir au-cours de cette brève visite*[17] while our model correctly predicts Tous ceux et celles *que je n'ai pu voir au-cours de cette brève visite*.

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.

Model	IID	OOD
Identity (baseline)	77.12	79.89
Rule-based	–	86.63
Fabien.ne BARThez	**96.07**	**94.60**

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

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] `nrefs:1|case:mixed|eff:no|tok:13a|smooth:exp|version:2.3.1`

[15] Meaning "Elected representatives' allowances capped at median salary."

[16] Meaning "Revocability of elected representatives."

[17] Meaning "All those I didn't get to see during this brief visit".

	Inclure	IFC
F M (*toutes et tous*)	51%	93%
M F (*tous et toutes*)	49%	7%

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

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 tous*[18]).

[18] Respectively "Good morning to all (women) and to all (men)" vs. "Good morning to all (men, including women)"

Inferring Feminization We have focused on the two main phenomenons 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 auteur *de bande dessinée, est née à Gênes en Italie, en 1988*[19] 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 *illustratrice*[20].

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. *frœur*, 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.x*[21]).

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.

7. Conclusion

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

[19] Meaning "Giorgia Marras, illustrator and comic strip author, was born in Genoa, Italy, in 1988"

[20] https://w.wiki/7k3d

[21] According 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")))
[('Bonjour à toutes et tous', 'Bonjour à 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")
[{'generated_text': 'Bonjour à toutes et à tous'}]
# or load model for complete control
>>> model = AutoModelForSeq2SeqLM.from_pretrained("PaulLerner/fabien.ne_barthez")
```

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[22] or via the Transformers library, see Listing 2.

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

[22] Upon acceptance of the paper, similarly to https://hf.co/moussaKam/barthez.

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 *iel* 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

My Alpheratz. 2018. Français inclusif : conceptualisation et analyse linguistique. *SHS Web Conf.*, 46:13003.

My Alpheratz. 2019. Français inclusif : du discours à la langue ? *Le Discours et la Langue Revue de linguistique française et d'analyse du discours*, (111):53–74.

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural Machine Translation by Jointly Learning to Align and Translate. In *ICLR 2015*.

Marion Baranes and Benoît Sagot. 2014. Normalisation de textes par analogie: le cas des mots inconnus. In *TALN-Traitement Automatique du Langage Naturel*, pages 137–148.

Farah Benamara, Diana Inkpen, and Maite Taboada. 2018. Introduction to the special issue on language in social media: exploiting discourse and other contextual information. *Computational Linguistics*, 44(4):663–681.

Franck Burlot and François Yvon. 2018. Using monolingual data in neural machine translation: a systematic study. In *Proceedings of the Third Conference on Machine Translation: Research Papers*, pages 144–155, Brussels, Belgium. Association for Computational Linguistics.

Bruno Cartoni. 2008. De l'incomplétude lexicale en traduction automatique: vers une approche morphosémantique multilingue (université de genève).

Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1724–1734, Doha, Qatar. Association for Computational Linguistics.

Marine Delaborde, Auphélie Ferreira, Loïc Grobol, Gabrielle Le Tallec, Benjamin Fagard, and Olga Seminck. 2021. Usages et perception du langage inclusif : des pratiques langagières clivantes ? Colloque Entre féminin et masculin – langue(s) et société, Lisbonne, Portugal.

Moussa Kamal Eddine, Antoine Tixier, and Michalis Vazirgiannis. 2021. Barthez: a skilled pretrained french sequence-to-sequence model. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 9369–9390.

Atefeh Farzindar and Mathieu Roche. 2013. Les défis du traitement automatique du langage pour l'analyse des réseaux sociaux. *Revue TAL–Traitement Automatique des langues*, 54(3):7–16.

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

Marlis Hellinger and Hadumod Bußmann. 2015. Gender across languages: The linguistic representation of women and men. *Gender across languages*, pages 1–26.

Diederik P. Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *ICLR (Poster)*.

Taku Kudo and John Richardson. 2018. SentencePiece: A simple and language independent subword tokenizer and detokenizer for Neural Text Processing. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 66–71, Brussels, Belgium. Association for Computational Linguistics.

Anne Lauscher, Archie Crowley, and Dirk Hovy. 2022. Welcome to the modern world of pronouns: Identity-inclusive natural language processing beyond gender. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 1221–1232, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7871–7880, Online. Association for Computational Linguistics.

Benedicta Adokarley Lomotey. 2015. On sexism in language and language change–the case of peninsular spanish. *Linguistik online*, 70(1).

Louis Martin, Benjamin Muller, Pedro Javier Ortiz Suárez, Yoann Dupont, Laurent Romary, Éric de la Clergerie, Djamé Seddah, and Benoît Sagot. 2020. CamemBERT: a Tasty French Language Model. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7203–7219, Online. Association for Computational Linguistics.

Denis Maurel. 2004. Les mots inconnus sont-ils des noms propres. *Actes des JADT*.

Ines Montani, Matthew Honnibal, Matthew Honnibal, Adriane Boyd, Sofie Van Landeghem, Henning Peters, Paul O'Leary McCann, jim geovedi, Jim O'Regan, Maxim Samsonov, Daniël de Kok, György Orosz, Marcus Blättermann, Madeesh Kannan, Duygu Altinok, Raphael Mitsch, Søren Lind Kristiansen, Edward, Lj Miranda, Peter Baumgartner, Raphaël Bournhonesque, Richard Hudson, Explosion Bot, Roman, Leander Fiedler, Ryn Daniels, kadarakos, Wannaphong Phatthiyaphaibun, and Schero1994. 2023. explosion/spaCy: v3.7.1: Bug fix for 'spacy.cli' module loading.

Mariam Nakhlé. 2023. L'évaluation de la traduction automatique du caractère au document: un état de l'art. *Actes de CORIA-TALN 2023. Actes des 16e Rencontres Jeunes Chercheurs en RI (RJCRI) et 25e Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL)*, page 143.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pages 311–318.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library. *Advances in Neural Information Processing Systems*, 32.

Matt Post. 2018. A call for clarity in reporting BLEU scores. In *Proceedings of the Third Conference on Machine Translation: Research Papers*, pages 186–191, Belgium, Brussels. Association for Computational Linguistics.

Christelle Rabary, Thomas Lavergne, and Aurélie Névéol. 2015. Etiquetage morpho-syntaxique en domaine de spécialité: le domaine médical. In *Actes de la 22e conférence sur le Traitement Automatique des Langues Naturelles. Articles courts*, pages 192–198.

Ricardo Rei, Craig Stewart, Ana C Farinha, and Alon Lavie. 2020. Comet: A neural framework for mt evaluation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2685–2702.

Chloé Riban and Murielle Gerin. 2017. Les garçons et les filles sont belles.

Helmut Schmid. 1994. Probabilistic part-of-speech tagging using decision trees. In *Proceedings of International Conference on New Methods in Language Processing, Manchester, 1994*.

Sabine Sczesny, Magda Formanowicz, and Franziska Moser. 2016. Can gender-fair language reduce gender stereotyping and discrimination? *Frontiers in psychology*, 7:25.

Thibault Sellam, Dipanjan Das, and Ankur Parikh. 2020. BLEURT: Learning robust metrics for text generation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7881–7892, Online. Association for Computational Linguistics.

Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016. Improving neural machine translation models with monolingual data. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 86–96, Berlin, Germany. Association for Computational Linguistics.

Thierry Spriet, Frédéric Béchet, Marc El-Bèze, Claude de Loupy, and Liliane Khouri. 1996. Traitement automatique des mots inconnus. In *Proceedings of TALN*, volume 96, pages 170–179.

Frederik Stouten, Irina Illina, and Dominique Fohr. 2010. Regroupement des occurrences des mots hors-vocabulaire répétés en vue de leur modélisation pour la transcription d'émissions radio. *Mons, Belgique*, page 173.

Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 2014. Sequence to Sequence Learning with Neural Networks. In *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc.

Priscille Touraille and Marc Allassonnière-Tang. 2023. Idéer une catégorie épicène et la matérialiser cohéremment dans la langue. Une nécessité épistémologique autant que politique. In Patricia Lemarchand, editor, *Qu'est-ce qu'une femme ? Catégories homme/femme : débats contemporains*, Essais, chapter 8, pages 167–233. Editions Matériologiques, Paris.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M. Rush. 2020. HuggingFace's Transformers: State-of-the-art Natural Language Processing. *arXiv:1910.03771 [cs]*. ArXiv: 1910.03771.

11. Language Resource References

Marie Candito, Guy Perrier, Bruno Guillaume, Corentin Ribeyre, Karën Fort, Djamé Seddah, and Éric Villemonte de La Clergerie. 2014. Deep syntax annotation of the sequoia french treebank. In *International Conference on Language Resources and Evaluation (LREC)*.

Cyril Grouin. 2022. Impact du français inclusif sur les outils du TAL (Impact of French Inclusive Language on NLP Tools). In *Actes de la 29e Conférence sur le Traitement Automatique des Langues Naturelles. Volume 1 : conférence principale*, pages 126–135, Avignon, France. ATALA.

Pedro Javier Ortiz Suárez, Benoît Sagot, and Laurent Romary. 2019. Asynchronous pipeline for processing huge corpora on medium to low resource infrastructures. In *7th Workshop on the Challenges in the Management of Large Corpora (CMLC-7)*. Leibniz-Institut für Deutsche Sprache.