

# Comment l'éthique de l'IA parle d'équité : étude linguistique de l'usage du terme “*fairness*” dans un corpus de chartes anglophone

Nicolas Allègre, Josephine Bernard, Dan Hayoun, Laury Magne,  
Tiphaine Viard☆, Maria Boritchev★

☆ i<sup>3</sup>, Télécom Paris, Institut Polytechnique de Paris, France

★ LTCl, Télécom Paris, Institut Polytechnique de Paris, France

tiphaine.viard@telecom-paris.fr

maria.boritchev@telecom-paris.fr

MOTS-CLÉS : Sémantique, contexte, syntaxe, graphes, AMR.

KEYWORDS: Semantics, context, syntax, graphs, AMR.

La base de données du Conseil de l'Europe liste des thèmes et mots clé communs à travers les initiatives portant sur l'éthique de l'intelligence artificielle<sup>1</sup>. Parmi ces mots clé, nous nous sommes particulièrement intéressées aux usages du terme “*fairness*”. Dans la liste de thèmes citée ci-dessus, ce terme a une place particulière, notamment dans un cadre de recherche francophone ; l'usage (qui fait la langue) est de parler de « la *fairness* » plutôt que d'utiliser une des traductions possibles (« impartialité », « équité ») lorsque l'on parle d'intelligence artificielle en français. Nous nous sommes donc attachées à explorer l'usage de “*fairness*” dans un contexte anglophone dans un premier temps, afin de pouvoir comparer ensuite cet usage à celui rencontré dans des contextes francophones spécialisés.

Le corpus MapAIE<sup>2</sup> est constitué de 436 chartes et manifestes de l'éthique de l'intelligence artificielle (IA), en anglais, provenant de communautés industrielles, d'entités législatives ou politiques (ex : Commission Européenne), mais aussi de communautés scientifiques et médiatiques (Gornet *et al.*, 2024). La sélection des documents compris dans MapAIE a été faite notamment par rapport à leur mention dans des études sociologiques précédentes ; dans l'étude présentée ici, nous nous sommes appuyées sur une version non-restreinte du corpus MapAIE, ci-après dénommée SourceAIE. Cette version contient 553 documents.

Nous présentons ici le résultat de notre exploration de l'usage du terme “*fairness*” dans le corpus SourceAIE : comment le terme “*fairness*” est-il utilisé dans les textes portant sur l'éthique de l'intelligence artificielle ? Notre contribution est double :

- Notre étude montre que le terme “*fairness*” apparaît principalement comme élément accessoire des phrases du corpus SourceAIE.

1. <https://www.coe.int/en/web/artificial-intelligence/national-initiatives>

2. <https://mapaie.telecom-paris.fr/>

- Dans le cadre de notre étude, nous avons utilisé une méthode à notre connaissance peu ou pas utilisée en linguistique computationnelle : l’étude de la forme des graphes AMR (Abstract Meaning Representation, (Banarescu *et al.*, 2013)). Nous présentons ici notre méthodologie.

## 1 Analyse morphosyntaxique

Nous commençons par une analyse morphosyntaxique en dépendances du corpus `SourceAIE`. Nous identifions ainsi les divers rôles syntaxiques que prend le lexème “*fairness*”, en utilisant `SpaCy`<sup>3</sup>. Le but de cette analyse est d’interroger l’importance grammaticale du terme, s’il apparaît plutôt en tant que sujet ou objet.

| Étiquette             | Fréquence dans le corpus |
|-----------------------|--------------------------|
| <code>pobj</code>     | 731                      |
| <code>compound</code> | 411                      |
| <code>conj</code>     | 246                      |
| <code>dobj</code>     | 294                      |
| <code>nsubj</code>    | 63                       |

TABLE 1 – Étiquettes associées à “*fairness*” dans le corpus `SourceAIE`, avec les fréquences relatives.

Le résultat de cette analyse (table 1) montre que “*fairness*” est principalement utilisé en tant qu’objet de préposition `pobj`, et deuxièmement en tant que partie d’un *compound*<sup>4</sup> – d’une combinaison de lexèmes qui se comportent morphosyntaxiquement comme un seul mot. “*fairness*” est 10 fois moins utilisé en tant que sujet de phrase `nsubj` que objet de préposition.

Cette première analyse nous permet de formuler les observations suivantes :

- “*fairness*” semble rarement être le sujet principal d’une phrase du corpus `SourceAIE`.
- “*fairness*” a une importance indirecte : en tant qu’objet de préposition, et aussi en association avec d’autres concepts.

À la suite de ces observations, nous proposons une analyse sémantique de l’usage de “*fairness*” dans le corpus `SourceAIE`. Pour ce faire, nous avons effectué un *parsing* sémantique automatique des 1201 phrases du corpus comprenant le lexème “*fairness*” en graphes AMR (Abstract Meaning Representation, (Banarescu *et al.*, 2013)), en utilisant `AMRlib`<sup>5</sup>, afin de pouvoir ensuite analyser les structures des graphes AMR obtenus.

3. <https://spacy.io/models/en>

4. <https://universaldependencies.org/u/dep/compound.html>

5. <https://amrlib.readthedocs.io/en/latest/>

## 2 Analyse des graphes AMR

Les graphes AMR permettent d'accéder à une abstraction de la structure sémantique des phrases d'origine. Nous nous intéressons ici au terme *“fairness”*, à ses relations sémantiques et à son rôle sémantique dans des contextes différents. En termes de structure AMR, nous avons donc analysé les éléments suivants :

- la place dans le graphe AMR (racine, noeud, feuille) du terme *“fairness”* ;
- les parents (concepts supérieurs) du terme *“fairness”* ;
- les adelphe (concepts ayant le même parent) du terme *“fairness”* ;
- les enfants (concepts inférieurs) du terme *“fairness”* ;
- les relations liant *“fairness”* à ses parents et ses enfants.

Nous avons observé deux concepts AMR se rattachant au terme *“fairness”* : *fair-01* (concept PropBank<sup>6</sup>) et *fairness* (concept résultant du *parsing* automatique).

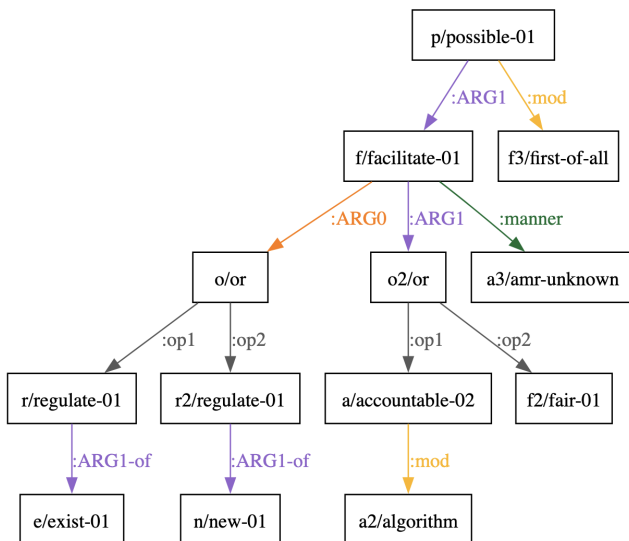


FIGURE 1 – Visualisation<sup>8</sup> du *parsing* automatique en AMR de la phrase “First of all, how can existing regulations or new ones facilitate algorithmic accountability or **fairness**?”.

**“fairness”, racine** : les concepts se rattachant à *“fairness”* n’apparaissent que 13 fois en tant que racine (concept principal) du graphe AMR. Ces occurrences n’ont lieu que dans des titres de documents ou de paragraphes, parfois comprenant le lexème *“fairness”* seul.

**Parents de “fairness”** : les trois concepts parents apparaissant le plus sont :

- *and* (476 fois) : cas où *“fairness”* est utilisé en conjonction avec d’autres concepts.

6. <https://propbank.github.io/v3.4.0/frames/fair.html#fair.01>

8. Via l’éditeur graphique metAMoRphosED <https://github.com/Orange-OpenSource/metamorphosed>

— *define-01* (57 fois) : la branche se lit comme “*define fairness*”.

— *principle* (51 fois) : la branche se lit comme “*principle of fairness*”.

*and* est parent de “*fairness*” 5 fois plus que *define-01*, cette forme de graphe AMR se retrouve dans 36.25% des concepts parents de “*fairness*”. Nous observons donc que “*fairness*” semble souvent être présenté comme un élément parmi d’autres dans les énoncés de SourceAIE, à travers des conjonctions.

**Adelphes de “*fairness*”** : les 5 concepts partageant le parent *and* avec “*fairness*” :

- *transparency* : 131,
- *accountable-02* : 111,
- *safe-01* : 50,
- *explain-01* : 37,
- *privacy* : 36.

“*transparency*”, “*accountability*”, “*safe*”, “*explainability*” et “*privacy*” font partie de la liste de thèmes et mots clé du Conseil de l’Europe au même titre que “*fairness*”. Nous les retrouvons ici comme éléments d’énumérations conjonctives.

**Relations liant “*fairness*” à son parent** : les relations les plus fréquentes sont :

- *:ARG1* : 407
- *:op1* : 263
- *:op2* : 175
- *:topic* : 113
- *:op3* : 61

La relation *:ARG1* désigne “*fairness*” comme un argument de son concept parent. Les relations *:opX* concernent les opérateurs, *and* et *or*. La relation la plus fréquente liant “*fairness*” à son parent est *:opX*, totalisant 499 occurrences. Cette observation nous permet de conclure que “*fairness*” est majoritairement utilisé comme partie d’une énumération conjonctive dans SourceAIE.

Cette étude linguistique de l’usage de “*fairness*” dans le corpus SourceAIE nous a permis plusieurs observations. Le lexème “*fairness*” est très présent dans le corpus, est principalement utilisé en tant qu’objet, et le terme est fréquemment utilisé comme élément d’une conjonction. Nous formulons donc l’hypothèse que l’usage de “*fairness*” dans le corpus SourceAIE est celui d’un terme vide/mot *buzz*, et non d’un terme central. Dans la suite de nos explorations linguistiques du corpus, nous souhaitons investiguer davantage la sémantique de “*fairness*” (vide ou non, significative ou non, dans quel contexte ?), ainsi que les différentes lexicalisations possibles des concepts linguistiques liés au terme “*fairness*”, tels que “*equity*”, “*integrity*”, ou “*justness*”. Nous avons conscience des limitations techniques de notre étude, notamment liés à l’usage de *parsing* automatique pour la production des graphes AMR : le graphe présenté en figure 1 peut être discuté (le rattachement de *fair-01* à *or* plutôt qu’à *algorithm*?), et c’est nécessairement le cas d’autres graphes parmi les 1201 de cette étude. Nous espérons dans un premier temps produire une méthodologie d’exploration de graphes AMR reproductible et évaluable, pour pouvoir ensuite l’appliquer

sur des graphes d'une meilleure qualité (au minimum relus et corrigés par des personnes expertes).

## Références

BANARESCU L., BONIAL C., CAI S., GEORGESCU M., GRIFFITT K., HERMJAKOB U., KNIGHT K., KOEHN P., PALMER M. & SCHNEIDER N. (2013). Abstract meaning representation for sembanking. In *Proceedings of the 7th linguistic annotation workshop and interoperability with discourse*, p. 178–186.

GORNET M., DELARUE S., BORITCHEV M. & VIARD T. (2024). Mapping ai ethics : a meso-scale analysis of its charters and manifestos. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, p. 127–140.