

Guide d’annotation

19 mai 2022

1 Introduction

Les documents sont des grilles FERA portant sur des récits émotionnels. Chaque grille est divisée en 5 parties : Fait extérieur, Fait extérieur, Émotion, Raison de l’émotion et Action. Nous ne prenons pas en compte la dernière partie Action. Chaque partie est composée d’une ou plusieurs phrases. Le nom des parties reprend la terminologie d’Alain Finkel. Les Faits extérieurs sont ceux que toute personne, qui aurait été témoin de la scène, aurait pu observer. Les Faits intérieurs sont les pensées et les ressentis vécus par l’AUTEUR qui ne sont pas directement visibles de l’extérieur.

Le but est d’annoter des groupes. Les groupes identifient dans le texte les mots ou groupes de mots (empan de texte) faisant référence aux entités impliquées dans la situation décrite par l’AUTEUR de la grille. Les différents types de groupe sont : INDICE, AUTEUR, SOURCE, ASPECT, CAUSE, MODIFIEUR, TERRITOIRE et NEGATION. Une partie de la grille peut n’avoir aucun groupe identifié. Les groupes sont reliés entre eux par des relations. L’INDICE et la CAUSE peuvent être reliés par aucune relation, le reste des groupes possèdent obligatoirement une relation. Un groupe ne peut avoir plus d’une relation. Le croisement de frontière est autorisé seulement si un groupe ne peut être relié avec un autre au sein du même phrase.

2 Règles

2.1 Règles générales

Ces règles s’appliquent à tous les groupes.

- Règle des référents différents : s’il existe plusieurs entités qui possèdent un même référent, alors on annote l’entité la plus proche de l’INDICE selon la *règle de la distance minimale*. Sinon, chaque entité doit être annotée séparément. Exemple : « Je **m**’énervé » et non « Je **Je m**’énervé ».
- Règle du chaînage des entités : une coréférence doit être résolue si le référent à un plus haut niveau de spécification que la coréférence. Ainsi, on ne relie pas deux pronoms (ex. il et lui) et deux noms propres (ex. Gustave et Gustave). [INSERER EXEMPLE NIVEAU SPECIFICATION DIFFERENT].

2.2 Règles particulières

Ces règles s’appliquent à certains groupes seulement (cf. Description des groupes).

- Règle de la distance minimale : on annote le groupe le plus proche de l’INDICE selon la distance de mots à gauche ou à droite. Même si une personne est désignée par un

pronom, il faut le sélectionner si celui-ci est le plus proche de l'INDICE et résoudre la corréférence si possible. Ainsi, si plusieurs entités sont présentes dans le texte, mais désignent le même référent, alors c'est seulement l'entité la plus proche de l'INDICE qui sera annotée. Le reste des entités est résolu en annotant les relations de corréférence.

- Stratégie minimaliste : on considère le plus petit empan de texte pour l'annotation du groupe. De telle sorte que si on retire un mot de l'empan, alors l'annotation n'a plus de sens pour le groupe sélectionné. Inclure les déterminants s'ils sont présents. (ex. on annoté « les collègues » et non « collègues »). Ainsi, on considère seulement les mots sémantiquement porteurs de sens pour le groupe sélectionné.
- Stratégie maximaliste : on considère le plus grand empan de texte pour l'annotation du groupe. De telle sorte que si on retire un mot de l'empan, alors l'annotation garde son sens pour le groupe, même s'il est désormais atrophié. !Cela consiste donc à sélectionner la séquence de mots la plus étendue, incluant au besoin, les modifieurs, appositions, subordonnées relatives, voire conjonctions de modifieurs de toutes sortes y compris de relatives, jusqu'au discours rapporté ou aux citations !.

2.3 Règles linguistiques

- Les mots composites avec « - » sont considérés comme un seul mot. Exemple : « attrape-tout » est annoté ensemble.
- Ne pas annoter les ponctuations (« »', ; : . ? !) en début ou fin d'empan de texte.

3 Description des groupes

L'INDICE est l'expression d'une émotion. Cela peut être un mot ou un groupe de mots correspondant à un événement indiquant la présence d'une émotion. Toutes les relations partent de l'INDICE. L'orientation de ces dernières n'est pas importante. L'INDICE peut avoir plusieurs CAUSEs, ASPECTs et TERRITOIRES si ces derniers sont des entités ou des événements différents. Un INDICE ne peut avoir qu'une SOURCE et/ou qu'un seul AUTEUR. L'annotation d'un INDICE est une condition nécessaire pour l'annotation des autres groupes, c'est ce que l'on cherche à annoter en premier. Toutefois, il peut exister des CAUSEs qui ne sont pas reliées directement à un INDICE. *Stratégie minimaliste*. Exemple : « J'ai **peur** d'être viré ».

L'AUTEUR est l'auteur de la grille, celui qui ressent le(s) émotion(s). On annoté les adjectifs possessifs comme « ma », « mon », et « mes », les pronoms personnels comme « je », « j' », « m' » et « me » et les pronoms possessifs comme « la mienne ». Il faut prendre en compte l'apostrophe lors de l'annotation. Si l'auteur est inclus avec d'autres personnes, comme dans « la nôtre » ou « nous », on rajoute un attribut qui indique la cardinalité (un ou plusieurs). *Stratégie minimaliste, règle de la distance minimale*. Exemple : « **J'**ai peur d'être viré ».

La CAUSE est un événement ou une entité (humaine ou non) qui provoque une émotion chez l'AUTEUR. Si la CAUSE est une entité humaine, on rajoute un attribut « humain ». Les conjonctions comme « parce que », « car », « mais », « et » et « ou » sont inclus. Généralement, la partie Fait extérieur contient les CAUSEs de l'émotion. *Stratégie maximaliste*. Exemple : « J'ai peur **d'être viré** ».

L'ASPECT représente un point de vue sur une entité ou une situation. Une émotion peut être tournée vers plusieurs aspects.

Le MODIFIEUR est un mot ou groupe de mots indiquant l'intensité de l'émotion comme « pas du tout », « beaucoup », « très », etc. *Stratégie minimaliste*. Exemple : « J'ai **très** peur d'être viré ».

La NEGATION est un mot ou groupe de mots indiquant la négation. Les termes indiquant la négation ne sont annotés que lorsqu'ils concernent un INDICE. Les négations indiquent alors l'absence (avec les mots « sans », « aucun », « personne », etc.) ou l'inverse (avec les mots « n'/ne ... pas », « non », etc.) d'une émotion exprimée. *Stratégie minimaliste*. Exemple : « Je **n'**ai **pas** peur d'être viré » et « J'ai une **absence** de colère ».

Le TERRITOIRE est concept créé par Alain Finkel. Il désigne quelque chose dont l'AUTEUR a le contrôle ou le pouvoir. Les TERRITOIRES sont attaqués (colère et peur), perdus (tristesse), retrouvés ou conquis (joie). Généralement, les TERRITOIRES sont annotés dans la partie Raison. *Stratégie maximaliste*. Exemple : « Il attaque **ma capacité à être respecté en classe** et **ma reconnaissance** ».

4 Éléments difficiles

- Les mots avec comme lemme « attaque » sont annotés comme CAUSE. Faire de même pour les synonymes d' « attaquer ».
- Un empan de texte peut appartenir à plusieurs groupes différents. [EXEMPLE]
- Certains groupes peuvent être imbriqués partiellement ou totalement. [EXEMPLE]

5 Aides

Voici quelques questions que vous pouvez vous poser pour l'annotation des groupes :

- Est-ce qu'il y a des termes qui indiquent qu'une émotion est ressentie ? (INDICE)
- Est-ce que le texte spécifie que l'auteur de la grille ressent une émotion ? (AUTEUR)
- Quels sont les séquences de mots qui décrivent un évènement, une personne ou une circonstance qui provoque une émotion ? (CAUSE)
- Quels sont les aspects de l'AUTEUR qui sont attaqués, perdus, retrouvés ou conquis (TERRITOIRE) ?

Il est également possible de reformuler une phrase pour mieux l'annoter : une seule personne (AUTEUR) ressent une émotion exprimée dans le texte à travers un ou plusieurs termes (INDICE) à cause de quelqu'un (CAUSE humaine) et/ou de quelque chose (CAUSE non-humaine). Certains aspects de l'AUTEUR sont attaqués, perdus, retrouvés ou conquis (TERRITOIRE).

6 Exemples

[INSÉRER EXEMPLE INCEPTION]