# **Expérimentation**

**Corpus**

Afin de procéder à notre expérimentation qui consiste à analyser l’outil ALP , nous utilisons des articles issus de site d’informations comme *ALJazeera, France24 Arabia et ALArabia*. Nous choisissons des articles qui parleront de politique et d’économie.

**Méthode**

Notre méthode de travail consistera à analyser les articles sur l’interface web de l’outil ALP, ce dernier est un outil gratuit et accessible sur le site internet *arabicnlp.pro/alp[[1]](#footnote-2).*  Notre objectif sera d’évaluer la reconnaissance des EN sur ALP et d’analyser les erreurs de l’outil, qui seront présentées dans des tableaux pour chaque article. Chaque tableau contiendra cinq colonnes, la première colonne sera appelée *Lexèmes*, nous écrirons dans cette colonne les mots qui portent une ambigüité ou une erreur dans l’outil ALP.

La deuxième colonne quant à elle sera le *Tag ALP* proposé par l’outil ALP. La troisième colonne aura pour nom *Tag Correct*, comme son nom l’indique ,il sera question de corriger les tags proposés par ALP. Dans la quatrième colonne nous proposerons l’analyse et la description de l’ambigüité et de l’erreur, cette colonne s’appellera donc *Analyse.* La dernière colonne s’appellera *Contexte* et aura pour tâche d’étiqueter toute la phrase dans laquelle nous avons identifié les *Lexèmes*.

Cette étape nous permettra de réfléchir en vue de proposer des algorithmes qui améliorerons les résultats obtenus.

**Les articles**

Les articles seront analysés en *mode normalisé[[2]](#footnote-3)* et *non normalisé,* un mode que propose l’outil ALP*.*

**Tableau XXX : des articles expérimentés**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Numéro de l’article** | **Titre de l’article** | **Sources** | **Type d’articles** | **Nombre de lexèmes** | **Date** |
| **1** | *»شغور منصب الرئيس... ماذا يقول دستور الجزائر «* | AlJazeera | Politique | 147 | 26/03/2019 |
| **2** | *» تقرير "مضلل" لبنك أميركي عملاق. هل جاء للتأثير على توجهات الناخبين الأتراك؟ «* | AlJazeera | Politique/ Economie | 804 | 26/03/2019 |
| **3** | وصواريخ باتريوت للشرق » الأوسط البنتاغون يرسل سفينة « حربية | Al Arabiya | Politique | 475 | 11/05/2019 |
| **4** | »فرنسا "مصممة" على فرض ضريبة على شركات الإنترنت العملاقة « | France 24 Arabic | Economie | 515 | 05/04/2019 |
| **5** | » ا ليابان: محكمة طوكيو تقرر تمديد توقيف كارلوس غصن حتى 14 أبريل « | France 24 Arabic | Economie | 160 | 05/04/2019 |
| **6** | » فرنسا تفرج عن التونسي بلحسن الطرابلسي شقيق زوجة بن علي « | France 24 Arabic | Politique | 419 | 12/05/2019 |
| **7** | »الجزائر: توجيه تهمتي "المساس بسلطة الجيش" و"المؤامرة ضد سلطة الدولة" للويزة حنون « | France 24 Arabic | Politique | 323 | 13/05/2019 |
| **8** | » الصين تعامل الولايات المتحدة بالمثل وتزيد الرسوم الجمركية بقيمة 60 مليار دولار« | France 24 Arabic | Economie | 251 | 13/05/2019 |
| **9** | »الجولاني يدعو أنصاره إلى "حمل السلاح" دفاعا عن إدلب من القصف السوري الروسي « | France 24 Arabic | Politique | 456 | 13/05/2019 |
| **10** | »إيطاليا: مسرح لاسكالا يلغي اتفاقا ماليا مع السعودية ويعيد إليها 3 ملايين يورو« | France 24 Arabic | Economie | 230 | 21/03/2019 |

Tableau 1: articles expérimentés

**Résultats**

Nous allons analyser dans cette partie les articles sur l’interface web de l’outil ALP.

ALP propose sur son interface le nombre de lexèmes dans le texte (figure 1), a droite nous avons la position du mot dans le texte. En deuxième position nous avons les mots du texte, et en troisième position le mot découpé (DET+NOM…)[[3]](#footnote-4).

L’interface contient aussi à gauche du tableau le tag proposé par ALP, nous notons que les EN sont étiquetées en couleur (personne, organisation, lieu, mois, évènements).

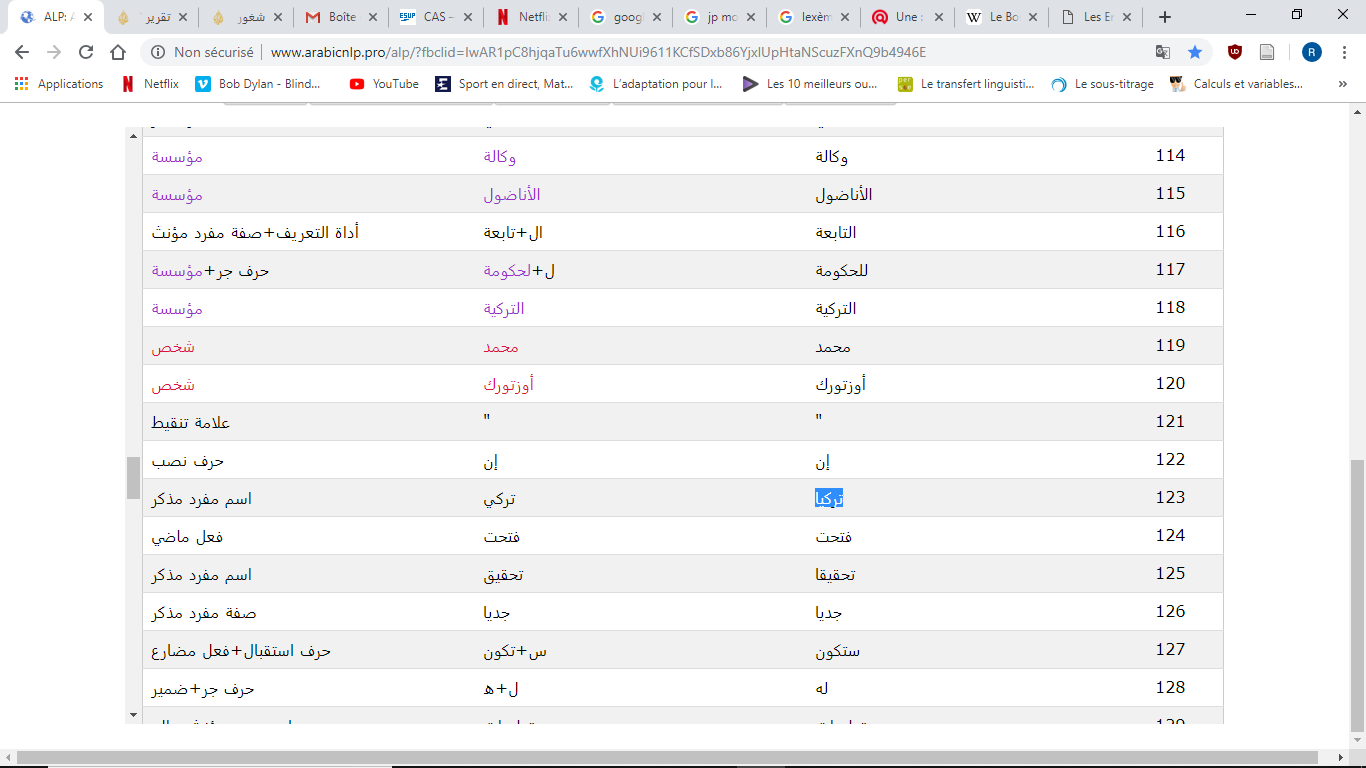


Figure 1: ALP tag

## **Artcile 1**

## Article analysé en texte *normalisé* et *non normalisé*.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Lexèmes** | **Tag ALP** | **Tag**  **correct** | **Analyse** | **Contexte** |
| *شغور* (*Vacance)*  *Mode normalisé* | PERS | NOM | Le mot se trouve juste avant la l’entité nommée : *présidence de la république*.Cette proximité. | يُعلَن شغور منصب رئاسة الجمهورية  (…*Déclare la vacance du poste de président de la République…*)  VERBE+PERS+NOM+NOM+  NOM |
| *المجلس الدستوري*, (*conseil constitutionnel*)  Mode *normalisé* | DET+NOM  +ADJ | ORG | Longueur du texte. | يجتمع المجلس الدستوري وجوبا.  *(Le conseil constitutionnel se reunit immediatement…)*  VERBE+NOM+ADJ+NOM  + ADV |
| و يكلف  *(Et sera chargé)*  Mode *normalisé* | PERS | PREP+VERBE | Présence de la préposition و (et) qui vient s’ajouter au verbe *يكلف* et qui donne une forte ressemblance avec le prenom : *Wyclef*. Cela dit l’erreur n’est pas presente quand le texte est en mode *non normalisé.* | ويُكلِّف بتولي رئاسة الدولة بالنيابة  *(…Et sera chargé du poste de president de la republiqu…).*  PERS+PERS+NOM+NOM  +NOM+ADV+NOM |
| بتولي *(se charger de)*  Mode *normalisé* | PERS | PREP+VERBE | ALP le marque comme PERS car le mot se trouve déjà après une entité nommée Pers (و يكلف) qui est aussi fausse. | ويُكلِّف بتولي رئاسة الدولة بالنيابة  *(…Et sera chargé du poste de president de la république…*).  PERS+PERS+NOM+NOM  +NOM+ADV+NOM |

Tableau 2: article 1

## **Artcile 2**

## Article analysé en texte *non normalisé* et *normalisé*.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Lexèmes** | **Tag ALP** | **Tag**  **Correct** | **Analyse** | **Contexte** |
| بينما يطوف الرئيس رجب طيب أردوغان بين الولايات التركية  (*Alors que le président Recep Tayyip Erdoğan fait le tour des villes turcs*)  Mode *normalisé* | CONJ+VERBE+DET  +NOM+PERS+PERS  +PERS+PERS+PERS  +PERS | CONJ+VERBE  +DET+NOM  +PERS+PERS  +PERS+PREP  +DET+NOM  + ADJ | L’ambigüité se situe au niveau des mots : *Bayn, Wilayat, Tourkiya.*  Le mot *Bayn (entre)* se place dans cette phrase entre le nom de personne *Erdoğan (EN)* et le mot *Wilayat (villes)* qui est décrit sur ALP comme PERS (erreur). Les mots *wilayat* et *tourkiya sont* aussi décrit comme PERS. | بينما يطوف الرئيس رجب طيب أردوغان بين الولايات التركية لحث شعبه.  (*Alors que le président Recep Tayyip Erdoğan fait le tour des villes turcs pour inciter son peuple.)*  CONJ+VERBE+DET  +NOM+PERS+PERS  +PERS+PERS+PERS  +PERS+VERBE+  PRON+NOM |
| جي.بي.مورغان  *(J.P.Morgan)*  Mode *non normalisé* | PERS+ORG+ORG | ORG | ALP découpe *G.P. MORGAN* en unité, ainsi *G* devient PERSquand *p et MORGAN* sont *ORG* et non comme une seule unité ORG. Le nom Morgan induit l’outil en erreur. | *وتزامن تقرير "جي.بي مورغان" مع تقرير لوكالة*.  *(Le rapport de JP MORGAN coïncide avec le rapport de l’agence.)*  PREP+PERS+ ORG+ORG |
| الليرة *(La Lira)*  Mode *normalisé* et *non normalisé* | NOM | MISC | ALP ne reconnait pas les monnaies et devises. | محاولات ضرب الليرة.  (…*Tenter d’enfoncer la Lira.)*  NOM+VERBE+NOM |
| *ديلي صباح*  *(Deli Sabah)*  Mode *normalisé* | PREP+NOM | ORG | ALP préconise l’utilisation des textes en mode N*ormalisé.* Cela dit, nous constatons beaucoup d’erreur quand nous utilisons le mode *normalisé*. Dans cet exemple il est question du nom d’un journal (*Deli Sabah)* que l’outil n’arrive pas à détecter. | *قال الصحفي الاقتصادي في صحيفة "ديلي صباح."*  *(Le journalise a dit dans le journal Deli Sabah…)*  DET+NOM+VERBE  +PREP+DET+NOM  PREP+NOM |
| *البنك الأميركي*  *(La Banque Mondial)*  Mode *normalisé* | *DET+NOM+*  *DET+NOM* | ORG | ALP préconise l’utilisation des textes en mode N*ormalisé.* Cela dit, nous constatons beaucoup d’erreur quand nous utilisons le mode *normalisé*. Dans cet exemple il est question du nom d’organisation (La Banque Américaine) que l’outil n’arrive pas à détecter. | *تقرير البنك الأميركي أثر سلبا.*  *(Le rapport de la Banque Américaine a influencé négativement…)*  NOM+DET+NOM+  DET+NOM+VERBE +ADV |
| *تركيا (Turquie)*  Mode *normalisé et non normalisé* | NOM | LOC | ALP n’arrive pas à détecter certains noms de lieux quand le texte est très long. | *إن تركيا فتحت تحقيقا.*  *(…que la Turquie avait ouvert une enquête…)*  ADV+NOM+VERBE  +ADJ+NOM |
| *(Le Golan)الجولان*  Mode *non normalisé* | DET+NOM | LOC | Longueur du texte. | *واعتراف ترامب بالسيادة الإسرائيلية على الجولان*  PREP+VERBE+PERS  +PREP+NOM+ADJ+  PREP**+**DET+NOM |
| البنك المركزي (*La Banque Centrale*)  Mode *normalisé* et *non normalisé* | DET+NOM+DET+ADJ | ORG | Texte analysé en mode normalisé. | *علق البنك المركزي بسعر الفائدة*  VERBE+ DET+NOM+  DET+ADJ+PREP+DET  +NOM+DET+NOM |

Tableau 3: article 2

**Article 3**

Article analysé en texte *non normalisé* et *normalisé*.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Lexèmes** | **Tag ALP** | **Tag**  **Correct** | **Analyse** | **Contexte** |
| (*Patriote*)باتريوت  Mode *normalisé* et *non normalisé* | PERS | ORG | ALP tag ce mot comme le nom d’une personne | إرسال صواريخ باتريوت إلى القيادة  VERBE+NOM+PERS  +PREP+DET+NOM |
| "يو اس اس آرلنغتون"  *(USS Arlington)*  Mode *non normalisé* | NOM+NULL+NULL  +ADJ | ORG | ALP ne reconnait pas cette ORG qui concerne un porte avion et la marque comme NULL | وافق على إرسال البارجة "يو اس اس آرلنغتون"  VERBE+PREP+NOM  NOM+NULL+NULL  +ADJ |
| "يو إس إس أبراهام لينكولن"  *(USS Abraham Lincoln)*  Mode *non normalisé* | PERS+PERS+PERS  +PERS+PERS | ORG | La présence du nom de personne Abraham Lincoln a induit l’outil ALP en erreur. | للانضمام إلى حاملة الطائرات "يو إس إس أبراهام لينكولن"  NOM+PREP+ADJ  +NOM+PERS+PERS  +PERS+PERS+PERS |
| (*Déclaration*)بيان  Mode *normalisé* | PERS | NOM | Le mot ressemble certainement à un autre nom de personne. | وقال بيان صادر عن وزير الدفاع  PREP+VERBE+PERS+  VERBE+PREP+NOM+  DET+NOM |
| (*Déclaration*)بيان  Mode *normalisé* | PERS | NOM | Le mot ressemble certainement à un autre nom de personne. | اضاف البيان أن وزارة الدفاع تراقب.  VERBE+PERS+  CONJ+DET+NOM+  PREP+DET+NOM+  VERBE |
| *(B 52)"*بي 52 "  Mode *non normalisé* | NOM+NULL | ORG | ALP ne reconnait pas les objets et les marques. L’exemple que nous avons ici concerne un type d’arme. | أن قاذفات من طراز "بي 52" وصلت إلى قاعدة.  CONJ+NOM+PREP+  NOM+NOM+NULL  +VERBE+PREP  +NOM. |
| (Iran)إيران  Mode *normalisé* | NOM | LOC | Longueur du texte et le mode utilisé (normalisé). | استعداد إيران لشن هجوم.  NOM+NOM+PREP  +VERBE+NOM. |
| الإدارة الأميركية للملاحة البحرية.  *(*[*United*](https://fr.wikipedia.org/wiki/United_States_Navy) *States Navy)*  Mode *normalisé* | NOM+ADJ+NOM+ADJ | ORG | Mode utilisé (normalisé),erreur que nous ne détectons pas quand nous utilisons le mode non normalisé. | وكانت الإدارة الأميركية للملاحة البحرية قالت.  PREP+VERBE+NOM+  ADJ+NOM+ADJ  +VERBE |
| جيم مالوي.  *(Jim Malloy)*  Mode *normalisé* | NOM+ADJ | PERS+PERS | Longueur du texte et le mode utilisé (normalisé), l’outil n’a pu reconnaitre cette entité nommée alors qu’il en a reconnu d’autre (Donald Trump) | وقال نائب الأدميرال جيم مالوي.  VERBE+NOM+NOM  +NOM+ADJ. |
| البحرين.  *(Le Bahreïn)*  Mode *normalisé* | DET+NOM | LOC | Longueur du texte et le mode utilisé (normalisé). | قائد الأسطول ……المتمركز في البحرين.  NOM+DET+NOM+  ADJ+PREP+DET  +NOM. |
| لوكالة "رويترز"  *(L’agence Reuters)*  Mode *normalisé* | ORG | PREP+NOM+  ORG | ALP ne distingue pas le mot لوكالة ,  qui est un nom, du mot رويترز qui est une ORG et le fait qu’il soit collé cela induit l’outil en erreur. | لوكالة "رويترز" الخميس  ORG+NOM. |

Tableau 4; article 3

**Article 4**

Article analysé en texte *non normalisé* et *normalisé*.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Lexèmes** | **Tag ALP** | **Tag**  **Correct** | **Analyse** | **contexte** |
| *(Bucarest)بوخارست*  Mode *normalisé et non normalisé* | PERS | LOC | La présence d’un nom PERS **(**لومير **)** avant le mot **(**بوخارست) a induit l’outil en erreur. | وقال لومير من بوخارست  PREP+PERS+PREP  +PERS |
| ***-***  Mode *normalisé* | PERS | PONCT | ALP a étiqueté le signe de ponctuation (-) comme PERS est cela est dû a la position de ce dernier entre deux noms PERS (جان-ايف) | جان-ايف لودريان  PERS+PERS+PERS+  PERS |
| *(Etats-Unis)الولايات المتحدة.*  Mode *normalisé* | NOM+ADJ | LOC | Le mode normalisé n’offre pas toujours des garanties de détection des entités nommées. Dans ce texte, l’entité nommée Etats-Unis est détectée une fois en tant que NOM+ADJ et plus loin dans le texte en tant que LOC. | تعبر فيها الولايات المتحدة عن استيائها  VERBE+PREP+NOM  +ADJ+PREP+NOM  +PRON |
| *(Union Européenne)*  *الاتحاد الأوروبي*  Mode *normalisé* | NOM+ADJ | ORG | Mode normalisé et longueur du texte. | تتناقض مع قواعد الاتحاد الأوروبي في مجال  VERBE+PREP+NOM  +NOM+ADJ+PREP+  NOM |
| *(PARIS)باريس*  Mode *normalisé* | PERS | LOC | ALP a confondu avec le nom PERS Paris . | لكن باريس مصممة على الاستمرار  CONJ+PERS+VERBE  PREP+NOM |
| *والسويد والدانمارك وفنلندا*  *(Et la Suède, le Danemark, et la Finlande.)*  Mode *normalisé* | PREP+DET+  NOM+PREP+DET+  NOM+PREP+NOM | PREP+LOC  +PREP+LOC  +PREP+LOC | ALP n’a pu détecter un nombre important de nom de lieux en utilisant le mode normalisé. | بسبب معارضة إيرلندا والسويد والدانمارك وفنلندا  LOCUTION+NOM  LOC PREP+NOM+  PREP+NOM +PREP+NOM |
| *بريطانيا* والنمسا  *(Grande-Bretagne et l’Autriche)*  Mode *normalisé ا* | NOM**+**PREP**+**DET  +ADJ | LOC+PREP+LOC | ALP n’a pu détecter un nombre important de nom de lieux en utilisant le mode normalisé. | تسير في الاتجاه نفسه بينها بريطانيا والنمسا  **VERBE+PREP+DET**  **+NOM+ADV+**NOM  **+**PREP**+**DET+ADJ |
| *بيار موسكوفيسي*  ***(Pierre Moscovici)***  Mode *normalisé* | NOM+ADJ | PERS+PERS | Longueur du texte ou mode normalisé. | أعرب …. بيار موسكوفيسي عن دعمه  **VERBE……+NOM+ADJ**  **PREP+NOM+PRON** |
| *موسكوفيسي*  ***(Moscovici)***  Mode *normalisé* | PERS | PERS | Tag correct. Le même mot (موسكوفيسي )est étiquette correctement alors qu’il ne l’est pas deux phrases au-dessus. | وأوضح موسكوفيسي  **PREP+VERBE+PERS.** |

Tableau 5: article 4

**Article 5**

Article analysé en texte *non normalisé* et *normalisé*.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Lexèmes** | **Tag ALP** | **Tag**  **Correct** | **Analyse** | **contexte** |
| *(Nissan)*نيسان  Mode *normalisé et non normalisé* | MOIS | ORG | Le mot *Nissan* est le 1er mois dans le calendrier Juif. ALP a dû confondre entre l’entreprise Nissan et le mois Nissan. | توقيف رئيس مجلس إدارة مجموعة نيسان.  VERBE+NOM+NOM  +NOM+NOM+MOIS |
| غصن (Ghosn)  Mode *normalisé et non normalisé* | NOM | PERS | ALP confond le nom de personne *Ghosn* qui s’écrit de la même façon que le nom arabe *Ghosn* (*Branche*). | كان غصن ينوي.  VERBE+NOM+VERBE. |
| رينو (*Renault*)  Mode *normalisé et non normalisé* | PERS | ORG | ALP confond le nom d’entreprise *Renault* avec le nom de personne *Renaud*. | كان رئيس مجلس إدارة رينو-نيسان  VERBE+NOM+NOM  +NOM+PERS+VERBE. |
| نيسان *(Nissan)*  Mode *normalisé et non normalisé* | VERBE | ORG | ALP a étiqueté le mot qui précédé Nissan comme PERS(*Renault*). Le mot *Nissan* a donc été étiqueté comme verbe a cause de sa position dans la phrase, derrière le nom de personne. | كان رئيس مجلس إدارة رينو-نيسان  VERBE+NOM+NOM  +NOM+PERS+VERBE. |
| دولار (Dollars)  Mode *normalisé et non normalisé* | NOM | MISC | ALP ne prend pas en considération les monnaies. | باستخدام خمسة ملايين دولار  VERBE+ADJ+NOM  + NOM |

Tableau 6: article 5

**Article 6**

Article analysé en texte *non normalisé* et *normalisé*.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Lexèmes** | **Tag ALP** | **Tag**  **Correct** | **Analyse** | **contexte** |
| كزافييه نوغيراس ومارسيل سيكالدي.  *(Xavier Nougueras et Marcel Ceccaldi)*  Mode *non normalisé et normalisé* | NOM+NOM+  PREP+ORG+  VERBE | PERS+PERS+  PREP+PERS+  PERS | ALP ne reconnait pas ces lexèmes comme entités nommées mais comme des noms. | وأعلن وكيلا الدفاع عنه كزافييه نوغيراس ومارسيل سيكالدي.  PREP+NOM+ADJ  DET+NOM+NOM+  PREP+ORG+VERBE |
| الطرابلسي *(Trabelsi)* | VERBE | PERS | ALP ne reconnait pas cette entité nommée quand elle est dissociée de sa première entité nommée Personne *(بلحسن).* | ويواجه الطرابلسي في فرنسا.  PREP+VERBE+VERBE  +PREP+NOM |
| الطرابلسي *(Trabelsi)*  Mode *non normalisé et normalisé* | NOM | PERS | ALP ne reconnait pas cette entité nommée quand elle est dissociée de sa première entité nommée Personne *(بلحسن).* | الذي سجن الطرابلسي  PREP+VERBE+NOM |
| (*Aix En Provence*)إيكس أن بروفانس  Mode *non normalisé et normalisé* | ORG+PREP+NOM | LOC | ALP ne reconnait pas l’entité  إيكس أن بروفانس car elle est précédé par les mot محكمة الاستئناف (Cours de cassation) qui introduit le nom de cette organisation. | في إيكس أن بروفانس  PREP+ORG+PREP+  NOM |
| الطرابلسي بنهب   *(Trabelsi de dérober)*  Mode *non normalisé et normalisé* | ORG+ORG | PERS+PREP+VERBE | L’entité nommée *Trabelsi* est ici précédé par une autre entité nommée Organisation, cela a induit ALP en erreur. | اتهمت النيابة العامة الطرابلسي بنهب  VERBE+ORG+ORG  +ORG+ORG |
| علي (Ali)  Mode *normalisé* | PREP | PERS | Ressemblance du mot *Ali* avec la préposition على (*ALA*) ,qui s’écrit à la fin avec un *i*, et que ALP transforme en lettre *A.* | زوجة الرئيس التونسي السابق زين العابدين بن علي.  NOM+DET+NOM  ADJ+ADJ+PERS+  PERS+PERS+PREP |
| قضائي (*Judiciaire*)  Mode *normalisé* | PERS | ADJ | Le mot est précédé par un autre mot (إشراف ) qui ressemble à un prénom arabe. | تحت إشراف قضائي.  PREP+NOM+PERS |
| مرسيليا (*Marseille*)  Mode *normalisé* | ADJ | LOC | Le mode normalisé n’arrive pas à détecter cette entité. | كان محتجزا في مدينة مرسيليا.  VERBE+ADJ+PREP  +NOM+ADJ |

Tableau 7: article 6

**Article 7**

Article analysé en texte *non normalisé* et *normalisé*.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Lexèmes** | **Tag ALP** | **Tag**  **Correct** | **Analyse** | **contexte** |
| مقران(*Mokrane*)  Mode *non normalisé et normalisé* | NOM | PERS | ALP ne reconnait pas cette entité nommée car *Mokrane* est un nom de Personne berbère. | أوضح مقران آيت العربي  VERBE+NOM+PERS  +PERS |
| بسعيد (*avec Saïd)*  Mode *non normalisé et normalisé* | PREP+NOM | PREP+PERS | La présence de la préposition avec a induit l’outil en erreur. | التقت فعلا بسعيد بوتفليقة  VERBE+ADV+PREP  +NOM+PERS |
| العربي (*Alarbi*)  Mode *non normalisé et normalisé* | NOM | PERS | Ressemblance avec le nom العربي (l’arabe). | أشار العربي إلى…  VERBE+NOM+PREP |

Tableau 8; article 7

**Article 8**

Article analysé en texte *non normalisé* et *normalisé*.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Lexèmes** | **Tag ALP** | **Tag**  **Correct** | **Analyse** | **contexte** |
| بكين *(Pékin)*  Mode *normalisé* | PREP+NOM | LOC | ALP découpe l’entité nommée Pékin en deux mots *ب* (à) comme PREP ( dans la langue arabe) , et *كين* en tant que NOM. | أن بكين ستفرض  CONJ+PREP+NOM  +VERBE |
| إن *(In)*  Mode *normalisé* | CONJ | PERS | ALP n’a pas reconnu cette entité nommée en mode normalisé. | قال ….. غينغ شوانغ إن  VERBE+PERS+PERS  +CONJ |
| الصين *(La Chine)*  Mode *normalisé* | DET+NOM | LOC | Entité non reconnue en mode normalisé. | الصين لن ترضخ  DET+NOM+ADV  VERBE |

Tableau 9: article 8

**Article 9**

Article analysé en texte *non normalisé* et *normalisé*.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Lexèmes** | **Tag ALP** | **Tag**  **Correct** | **Analyse** | **contexte** |
| هيئة تحرير [الشام](https://fr.wikipedia.org/wiki/Hayat_Tahrir_al-Cham) [(Hayat Tahrir al-Cham)](https://fr.wikipedia.org/wiki/Hayat_Tahrir_al-Cham) | NOM+VERBE+LOC | ORG | La présence d’un nom de lieu *Cham,* induit ALP en erreur. | [دعاالقائدالعام **لهيئة تحرير الشام**](https://www.france24.com/ar/20190110-سوريا-هيئة-تحرير-الشام-إدلب-اتفاق-سيطرة-فصائل-جهادية-معارضة)**.**  VERBE+DET+NOM  ADJ+PREP+NOM  +VERBE+LOC |
| تلغرام *(Telegram)* | NOM | ORG | ALP en mode normalisé n’a pas reconnu l’application *Telegram.* | في تطبيق تلغرام.  PREP+NOM+NOM |
| ريف حماة الشمالي  *(Rif Hamma du Nord)* | NOM+NOM+NOM | NOM+LOC | ALP ne reconnait pas ce lieu en mode normalisé. | أجريت في ريف حماة الشمالي.  VERBE+PREP+NOM  +NOM+NOM |
| ريف إدلب  *(Rif Idleb)* | PERS+PERS | NOM+LOC | Les deux mots sont précédés d’un verbe (témoigner) qui est souvent utilisé avant les noms propres , cela a induit ALP en erreur. | ويشهد ريف إدلب.  PREP+VERBE+PERS  +PERS |
| سوتشي *(Sotchi)* | NOM | LOC | ALP ne reconnait pas cette entité nommée nom de lieu. | من أستانا الى سوتشي.  PREP+LOC+PREP  +NOM |

Tableau 10: article 9

**Article 10**

Article analysé en texte *non normalisé* et *normalisé*.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Lexèmes** | **Tag ALP** | **Tag**  **Correct** | **Analyse** | **contexte** |
| لاسكالا *(La Scala)* | ADJ | LOC | Le mode normalisé ne reconnait pas cette entité nommée. | إدارة مسرح لاسكالا في  ميلانو.  NOM+NOM+ADJ  +PREP+LOC |
| بميلانو *(A Milan)* | PREP+NOM | PREP+LOC | Le mode normalisé ne reconnait pas cette entité nommée. | بميلانو الإيطالية الاثنين.  PREP+NOM+LOC  +SUBS |
| الإيطالية *(Italienne)* | LOC | ADJ | ALP confond entre l’adjectif *Italienne* et le pays *Italie.* | بميلانو الإيطالية الاثنين  PREP+NOM+LOC  +SUBS |
| الاثنين *(Lundi)* | LOC | SUBS | Le mot est précédé par une autre entité nommée الإيطالية (LOC). | بميلانو الإيطالية الاثنين  PREP+NOM+LOC  +SUBS |

Tableau 11: article 10

**Synthèses des erreurs**

Notre expérimentation avec ces 10 articles sur l’interface de l’outil ALP, nous a permis de détecter beaucoup d’erreurs dans la détection des entités nommées. Nos articles ont été analysés en utilisant le mode normalisé et non normalisé, que ALP propose pour les textes sans vocalisation ( la langue arabe est une langue qui peut se lire et s’écrire sans la transcription des voyelles).

Le mode Normalisé est le mode que l’outil suggère pour tout types de textes, ce mode est donc supposé être fiable. Cela dit, le mode normalisé est celui dans lequel nous trouvons le plus d’erreurs au niveau de la détection des entités nommées.

**Les différentes erreurs de l’outil ALP**

**Erreurs d’ordre grammaticale**

Nous avons détecté beaucoup d’erreurs d’ordre grammaticale, comme l’erreur trouvée dans l’article 1 où la présence d’une préposition collée au verbe a été détectée par ALP comme étant une EN PERS. ALP a aussi beaucoup de difficulté avec les textes sans voyelles, même s’il préconise le recours au mode normalisé pour tout type de texte.

**Erreurs d’ordre sémantique**

ALP fait des erreurs au niveau sémantique, nous avons des exemples de mots analysés comme le mot *ALARABI* ( L’ARABE) qui est une entité nommée dans notre texte et, que ALP étiquette en tant que nom , la priorité est donnée ici à l’aspect sémantique du mot.

**Position du mot**

La présence d’une suite d’EN (PERS) dans une seul phrase induit ALP aussi en erreur, car l’outil étiquète un mot ( verbe, nom…) en EN (PERS ) , même si cette dernière ne l’est pas. La présence aussi des EN noms de personnes dans une EN nom d’organisation pose des problèmes à l’outil.

**Reconnaissance des entités nommée**

ALP ne reconnait pas tous les noms de personnes non arabe, nous avons détecté quelques erreurs si le nom de personne est Berbère ou étranger. La même erreur intervient si le nom de personne fait partie d’un nom d’organisation, dans ce cas, ALP donne l’avantage aux noms de personnes et non aux noms d’organisation.

ALP ne reconnait pas les Monnaies, l’article souligne que ces dernières seront ajoutées plus tard.

**Longueur du texte**

ALP n’offre pas un traitement fiable pour les longs textes, nous sommes limité par le nombre de mots et, les erreurs à ce niveau sont très importantes. ALP est capable de détecter et d’étiqueter un mot en entité nommée dans le premier paragraphe et, de l’étiqueter différemment (verbe, nom…) dans le paragraphe suivant.

1. « <http://www.arabicnlp.pro/alp/?fbclid=IwAR1pC8hjqaTu6wwfXhNUi9611KCfSDxb86YjxIUpHtaNScuzFXnQ9b4946E> ». [↑](#footnote-ref-2)
2. Pour les textes sans voyelles. [↑](#footnote-ref-3)
3. La langue arabe étant une langue ou le déterminant est collé au nom. [↑](#footnote-ref-4)