

Classifications et grammaires des invariants lexicaux arabes en prévision d'un traitement informatique de cette langue: Construction d'un modèle théorique de l'arabe : la grammaire des invariants lexicaux temporels

Dhaou GHOUL

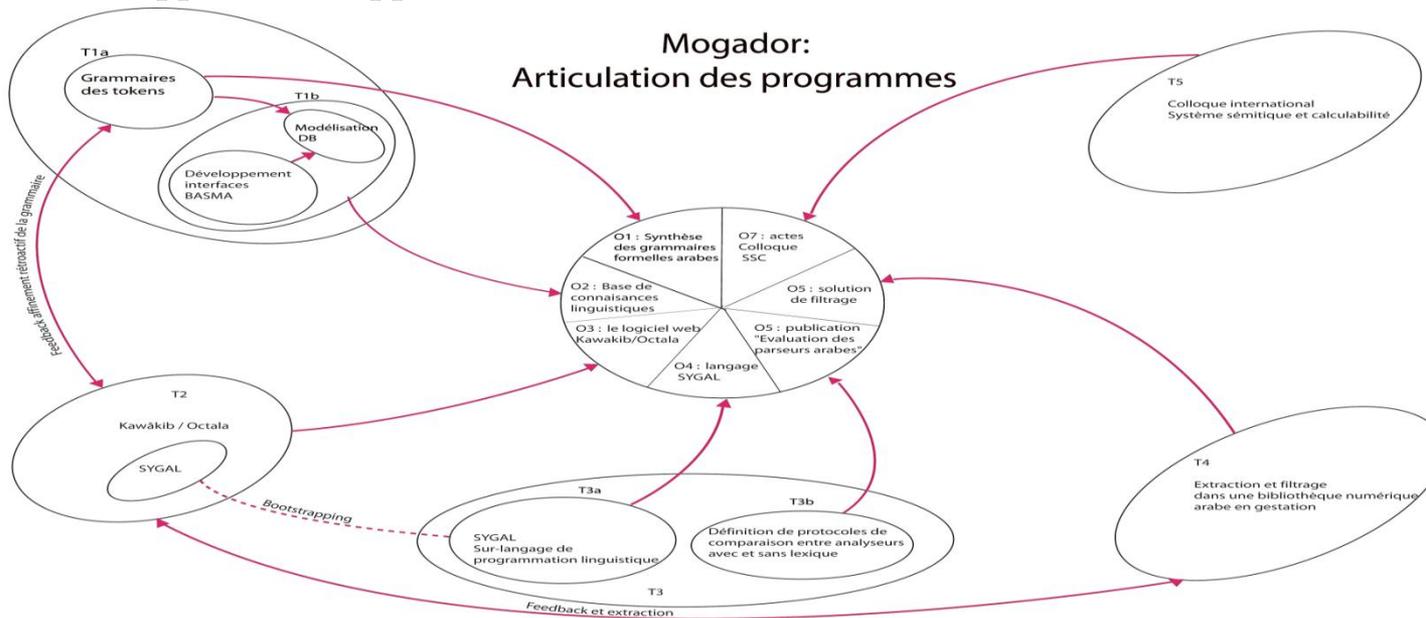
dhaou.ghoul@sorbonne-universite.fr

Le 18 /04/2019

Al-zuġġāġi : « al-ḥarf mā dalla 'alā ma'nan fi ġayrihi naḥwu min wa'ilā waṭumma » (La particule est ce qui dénote un sens dans un autre mot comme min, 'ilā, ṭumma).

Cadre général

- Projet Mogador: Modélisation des grammaires arabes, des données et des outils de recherche (TALA-MMSH, TALA-IFAO).
- Traitement par automates de la langue arabe (programme TALA).
 - ✓ L'approche algorithmique de la grammaire arabe.
 - ✓ Développement d'applications.



Al-zuġġāġi : « al-ḥarf mā dalla ‘alā ma‘nan fi ġayrihi naḥwu min wa’ilā waṭumma » (La particule est ce qui dénote un sens dans un autre mot comme min, ‘ilā, ṭumma).

I
Contexte et
objectifs.

Plan

II
Les invariants
lexicaux en
arabe.

III
Notre étude
des invariants
lexicaux
temporels.

IV
L'application
Kawâkib.

V
Conclusion et
perspectives.

1. Contexte et objectifs.
2. Les invariants lexicaux en arabe.
 - Utilité des invariants lexicaux dans le domaine TAL.
 - Classifications des invariants lexicaux: état de l'art et différents niveaux d'invariance.
3. Notre étude des invariants lexicaux temporels.
 - Classification des invariants étudiés.
 - Construction de corpus.
 - Méthodologie et modélisation par des schémas de grammaires
 - Les invariants lexicaux temporels simples.
 - Les invariants lexicaux temporels complexes.
4. L'application Kawâkib
5. Conclusions et perspectives.

Al-zuġġāġi : « al-ḥarf mā dalla 'alā ma'nān fi ġayrihi naḥwu min wa'ilā waṭumma » (La particule est ce qui dénote un sens dans un autre mot comme min, 'ilā, ṭumma).

I
**Contexte et
objectifs.**

I. Contexte et objectifs

II
**Les invariants
lexicaux en
arabe.**

✓ L'importance des invariants lexicaux dans la langue arabe (Audebert et Jaccarini, 1986).

✓ Les invariants lexicaux garde leurs propres formes dans le langage quotient.

III
**Notre étude
des invariants
lexicaux
temporels.**

✓ Traitement de la langue arabe avec des règles minimales (réduction du recours au lexique) .

✓ Étude Linguistique des invariants lexicaux temporels de la langue arabe.

✓ Grammaires des attentes: étude d'environnement syntaxique des invariants lexicaux (20 invariants lexicaux).

IV
**L'application
Kawâkib.**

✓ Fournir des schémas d'automates qui permettent de reconnaître les invariants lexicaux qui ont des marqueurs de temporalité et d'aspect.

✓ L'analyse morphosyntaxique automatisée de l'arabe afin d'améliorer quelques applications dans le domaine de traitement automatique de l'arabe indépendamment de la couverture des ressources.

V
**Conclusion et
perspectives.**

Al-zuġġāġi : « al-ḥarf mā dalla 'alā ma'nan fi ġayrihi naḥwu min wa'ilā waṭumma » (La particule est ce qui dénote un sens dans un autre mot comme min, 'ilā, ṭumma).

I

Contexte et
objectifs.

II

Les invariants
lexicaux en
arabe.

III

Notre étude
des invariants
lexicaux
temporels.

IV

L'application
Kawâkib.

V

Conclusion et
perspectives.

II. Les invariants lexicaux en arabe.

1. Utilité des invariants lexicaux arabes dans le domaine TAL:

- Rôle très important dans l'interprétation de la phrase et la cohérence d'un texte.
- Mots grammaticaux → prédire la structure de la phrase (Audebert et Jaccarini, 1988).
- Actualisateurs du prédicat: opérateur selon l'analyse matricielle définitoire (Ibrahim, 2010).
- Étude linguistique détaillées des invariants lexicaux → amélioration des applications TAL.

Al-zuġġāġi : « al-ḥarf mā dalla 'alā ma'nān fī ġayrihi naḥwu min wa'ilā waṭumma » (La particule est ce qui dénote un sens dans un autre mot comme min, 'ilā, ṭumma).

I
Contexte et
objectifs.

II. Les invariants lexicaux en arabe.

II

Les invariants
lexicaux en
arabe.

1. Utilité des invariants lexicaux arabes dans le domaine TAL:

❖ Quelques exemples:

○ système de traduction automatique: (SYSTRAN)

سأطوف به انحاء البلاد حتى يراه الجميع

Traduction Arabe / Anglais:

Will circulate in him across the country **even** sees him the all. ✗

Traduction référence:

I'll take him around the world **until** everyone sees him. ✓

III

Notre étude
des invariants
lexicaux
temporels.

➔ Mauvaise étude linguistique de l'invariant « ḥattā ».

➔ Attribution incorrecte d'étiquette syntaxique de « ḥattā ».

IV

L'application
Kawâkib.

➔ SYSTRAN prend «ḥattā» toujours comme un adverbe « même».

V
Conclusion et
perspectives.

Al-zuġġāġi : « al-ḥarf mā dalla 'alā ma'nān fi ġayrihi naḥwu min wa'ilā waṭumma » (La particule est ce qui dénote un sens dans un autre mot comme min, 'ilā, tumma).

I
Contexte et
objectifs.

II. Les invariants lexicaux en arabe.

1. Utilité des invariants lexicaux arabes dans le domaine TAL:

❖ Quelques exemples:

- Synthèse de parole (Transformer un texte écrit en un message oral):

Exemple d'un synthétiseur en ligne: Acapela

III

Notre étude
des invariants
lexicaux
temporels.

لَمَّا طَرَقْتُ الْبَابَ فَتَحَ لِي « *lammā ʔaraqtu albāba fataḥa lī* » (Quand je frappais à la porte il m'a ouvert). ✓

إِنَّهُ تَعْبِيرٌ جَمِيلٌ لَمَّا يَحْدُثُ هَذِهِ الْأَيَّامَ « *'innahu ta 'bīrun ḡamīl limā yaḥduṭu ḥadihi al'āyām* » (C'est une bonne expression de ce qui se passe pendant ces jours). ✓

IV

L'application
Kawâkib.

لَوْ لَمْ تَعْجَبْهَا لَمَّا اسْتَطَاعَتْ الْكِتَابَةَ عَنْهَا « *law lam ta 'ḡabahā lamā 'istaṭā 'at alkitābat 'anhā* » (Si elle n'avait pas aimé, elle n'aurait pas pu écrire à son sujet.) ✗
(*limā*)

V

Conclusion et
perspectives.

بَلْ لَمَّا يَذُوقُوا عَذَابَ « *bal lammā yaḍūqū 'adāba* » (Plutôt ils n'ont pas encore goûté à Mon châtement) (Sourate 38, verset 8). ✗ (*limmā*).

Al-zuḡḡāḡi : « al-ḥarf mā dalla 'alā ma 'nan fi ḡayrihi naḥwu min wa'ilā waṭumma » (La particule est ce qui dénote un sens dans un autre mot comme min, 'ilā, ṭumma).

I
Contexte et
objectifs.

II. Les invariants lexicaux en arabe.

II

Les invariants
lexicaux en
arabe.

III

Notre étude
des invariants
lexicaux
temporels.

IV

L'application
Kawâkib.

V

Conclusion et
perspectives.

1. Utilité des invariants lexicaux arabes dans le domaine TAL:

❖ Quelques exemples:

- Synthèse de parole (Transformer un texte écrit en un message oral):

Exemple d'un synthétiseur en ligne: Acapela

- ➔ L'absence des voyelles ou des signes diacritiques entraîne la mauvaise qualité de la synthèse vocale.
- ➔ La mauvaise catégorie grammaticale entraîne la mauvaise qualité de la synthèse vocale: la transcription correcte en phonème dépend essentiellement de la catégorie grammaticale.

Exemple: « *faqad* » (فقد): CC+ ADV (fa+ qad) ou VAC (faqada).

✓ فقد شارك في هذه الأمسية

✗ من يأسف على كذاب فقد لسانه في حادث

I

Contexte et
objectifs.

II

Les invariants
lexicaux en
arabe.

III

Notre étude
des invariants
lexicaux
temporels.

IV

L'application
Kawâkib.

V

Conclusion et
perspectives.

II. Les invariants lexicaux en arabe.

2. Classifications des invariants lexicaux: état de l'art et différents niveaux d'invariance:

2.1. État de l'art:

- ❖ Sîbawayh : « *al-ḥarf mā ḡā' lima 'na laysa bi'ism wa lā fi 'l* » (un mot ayant un sens qui n'est ni celui d'un nom ni celui d'un verbe).
- ❖ Al-Zuḡḡāḡī: « *al-ḥarf mā dalla 'alā ma 'na fi ḡayrihi naḥwu min wa 'ilā waṭumma* » (La particule est ce qui dénote un sens dans un autre mot comme *min* et *'ilā* et *tumma*).
- ❖ Ibn al-Sarrāḡ : « *al-ḥurūfu mā lā yaḡūzu 'an yuxbara 'anhā walā 'an takūna xabaran naḥwa min, wa, 'ilā* ».

I

Contexte et
objectifs.

II

Les invariants
lexicaux en
arabe.

III

Notre étude
des invariants
lexicaux
temporels.

IV

L'application
Kawâkib.

V

Conclusion et
perspectives.

II. Les invariants lexicaux en arabe.

2. Classifications des invariants lexicaux: état de l'art et différents niveaux d'invariance:

2.1. État de l'art:

- Définition littéraire des invariants lexicaux: 2 tendances
 - 1^{er} tendance: « *al-ḥarf* » dénote un sens dans un autre mot. (majorité des grammairiens arabes):

الفقاص عصفور في « *fī alqafasi 'sfūr* » (**Dans** la cage, il y a un oiseau).

- 2^{ème} tendance: il dénote un sens en lui-même comme le verbe et le nom:

وضعت الكرة تحت الطاولة « *waḍa 'tu alkurata taḥta alṭāwilat* » (j'ai mis le ballon **au-dessous** de la table).

➔ Les invariants lexicaux arabes appartiennent à deux classes : ceux qui ont leur propre sens indépendamment du mot qui les suit et les autres n'ont de sens que s'ils apparaissent dans un énoncé complet.

I

Contexte et objectifs.

II

Les invariants lexicaux en arabe.

III

Notre étude des invariants lexicaux temporels.

IV

L'application Kawâkib.

V

Conclusion et perspectives.

II. Les invariants lexicaux en arabe.

2. Classifications des invariants lexicaux: état de l'art et différents niveaux d'invariance:

2.1. État de l'art:

- Définition formelle des invariants lexicaux:
 - Ils se situent au même niveau que les schèmes et échappent aux règles de dérivation morphologique.
 - $INL = LEX \text{ inter } \{x ; SC(x) = \{x\}\} \{x ; x \in LEX \text{ et } SC(x) = \{x\}\}$

Exemple : $SC(\text{فإنهم} / fa'inahum) = \text{ف}$ $SC(\text{إن}) = \text{هم}$ $\text{هم} = \text{إن}$. Avec $\text{إن} \in LEX$,

- ✓ INL est un *lexème du langage quotient* de l'arabe L/RAC
- ✓ $LEX \subset L_{AR} \subset A^*$, LEX est un sous-ensemble particulier de L_{AR} .
- ✓ A^* : l'ensemble des concaténations des graphèmes arabes.
- ✓ L_{AR} : l'ensemble des formes graphiques arabes licites (mots graphiques séparés par deux blancs).
- ✓ L: l'ensemble des phrases grammaticales - non forcément douées de sens

I
Contexte et
objectifs.

II
Les invariants
lexicaux en
arabe.

III
Notre étude
des invariants
lexicaux
temporels.

IV
L'application
Kawâkib.

V
Conclusion et
perspectives.

II. Les invariants lexicaux en arabe.

2. Classifications des invariants lexicaux: état de l'art et différents niveaux d'invariance:

2.1. État de l'art:

- ✓ Relativement au langage L_{AR} , il est possible de considérer deux types de partition de LEX :
 - ✓ Une partition en catégories grammaticales définies selon le principe distributionnel de Bloomfield: $\forall x \in A^*, \forall y \in A^*, x.u.y \in L_{AR} \Leftrightarrow x.v.y \in L_{AR}$; ce que l'on note $u \sim_g v$ (« g », en indice, pour grammatical), avec $u, v \in LEX$.
 - ✓ Une partition en classes disjointes d'éléments recouvrant entièrement LEX dont chacune représente un paradigme lexical (ou schème, *wazn*). La relation associée à cette partition est également une relation de congruence notée $u \sim_{sc} v$ (« sc », en indice, pour schème)

$$\begin{array}{ccc} L_{AR}^* & \xrightarrow{\Pi} & L_{AR}^*/_{\sim_L} \\ \downarrow SC & & \downarrow SC \\ L_{AR}^*/_{\sim_{SC}} & \xrightarrow{\Pi} & (L_{AR}^*/_{\sim_L})/_{\sim_{SC}} = (L_{AR}^*/_{\sim_{SC}})/_{\sim_{L_{AR}}} \end{array}$$

I

Contexte et
objectifs.

II

Les invariants
lexicaux en
arabe.

III

Notre étude
des invariants
lexicaux
temporels.

IV

L'application
Kawâkib.

V

Conclusion et
perspectives.

II. Les invariants lexicaux en arabe.

2. Classifications des invariants lexicaux: état de l'art et différents niveaux d'invariance:

2.1. État de l'art: Classifications

- ✓ Classification en fonction de leurs rôle syntaxiques:(Wright, Robertson Smith, et de Goeje, 1896):
 - ✓ pronoms.
 - ✓ Prépositions,
 - ✓ Adverbes,
 - ✓ Conjonctions
 - ✓ Interjections
- ✓ Classification en famille: 24 familles (Gaubert, 2001):
 - ✓ 2 auxiliaires verbaux: kâna et ses conjugaisons, laysa et ses conjugaisons.
 - ✓ 12 invariants pré-verbaux.
 - ✓ 10 invariants pro-nominaux.

I

Contexte et
objectifs.

II

Les invariants
lexicaux en
arabe.

III

Notre étude
des invariants
lexicaux
temporels.

IV

L'application
Kawâkib.

V

Conclusion et
perspectives.

II. Les invariants lexicaux en arabe.

2. Classifications des invariants lexicaux: état de l'art et différents niveaux d'invariance:

2.2. Différents niveaux d'invariance :

- ✓ Les invariants qui ont leur propre projection par rapport au SC et se situent donc au même niveau que les classes d'équivalence (congruence) définies par les schèmes.
- ✓ Les invariants qui sont permutable avec un nom, un verbe, des syntagmes nominaux ou verbaux, etc.,
- ✓ Les invariants dont la sémantique est plus ou moins vide mais qui ne sont pas suppressibles.
- ✓ Les invariants qui sont suppressibles.
- ✓ Les invariants qui sont des opérateurs casuels
- ✓ Les invariants qui sont actualisateurs du prédicat.

III. Notre étude des invariants lexicaux temporels.

1. Classification des invariants temporels étudiés:

- ❑ 20 invariants temporels: : *l.m.ā* ; *'indamā* ; *'aṭnā'a* ; *'idā* ; *'id* ; *baynamā* ; *ḥattā*, *ḥīn/ ḥīna* ; *lam* ; *lan* ; *munḍu* ; *qad* ; *tumma* ; *ba'da* ; *ba'damā* ; *ḥīnamā* ; *kullamā* ; *sawfa* ; *ḥīnadāk* ; *haytūmā*.
- ❑ Distribution des invariants temporels par rapport à la totalité des invariants lexicaux en arabe:

	Nb phrases	Nb mots	Nb invariants lexicaux	Nb invariants temporels	% invariants temporels
Texte1	142	3460	1174	173	14.73
Texte2	63	1311	507	80	15.77
Texte3	123	2366	752	109	14.49
Texte4	11	344	108	10	9.25

III. Notre étude des invariants lexicaux temporels.

I
Contexte et
objectifs.

II
Les invariants
lexicaux en
arabe.

III
Notre étude
des invariants
lexicaux
temporels.

IV
L'application
Kawâkib.

V
Conclusion et
perspectives.

1. Classification des invariants temporels étudiés:

1.1. Classification en fonction des agglutations: 7 classes

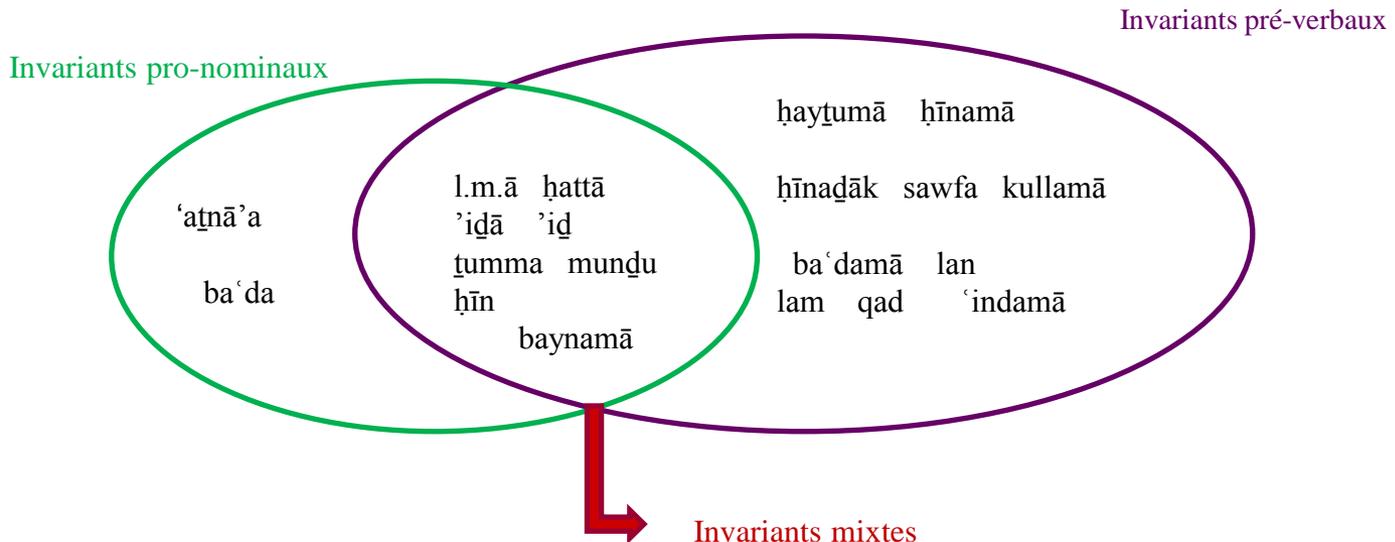
- Classe1: agglutination seulement à des coordonnants : *l.m.ā*, *'indamā*...
- Classe2: agglutination à une conjonction de coordination et/ou à une conjonction d'interrogation : *'idā*, *lam*, *lan*.
- Classe3: agglutination à une conjonction de coordination et/ou à une conjonction de corroboration (*la*) : *qad*.
- Classe4: agglutination à une conjonction de coordination et/ou à une préposition et à un pronom personnel attaché : *'atnā*'a.
- Classe5: agglutination à une conjonction de coordination et/ou à une préposition et à un pronom personnel attaché et à un pronom relatif : *hīn*.
- Classe6: agglutination à une conjonction de coordination et/ou à un pronom personnel attaché et à un pronom relatif : *ba*'da.
- Classe7: agglutination à une conjonction de coordination et/ou à une conjonction d'interrogation et à une conjonction de corroboration : *sawfa*.

III. Notre étude des invariants lexicaux temporels.

1. Classification des invariants temporels étudiés:

1.2. Classification en fonction du mot qui les suit:

- Les invariants pré-verbaux.
- Les invariants pro-nominaux.
- Les invariants mixtes.



Al-zuġġāġi : « al-ḥarf mā dalla 'alā ma'nan fi ġayrihi naḥwu min wa'ilā waṭumma » (La particule est ce qui dénote un sens dans un autre mot comme min, 'ilā, ṭumma).

III. Notre étude des invariants lexicaux temporels.

I
Contexte et
objectifs.

II
Les invariants
lexicaux en
arabe.

III
Notre étude
des invariants
lexicaux
temporels.

IV
L'application
Kawâkib.

V
Conclusion et
perspectives.

1. Classification des invariants temporels étudiés:

1.3. Classification en fonction de leurs rôles syntaxiques: 9 classes

- Classe1: Conjonction de subordination: *lammā*, *'indamā*, *baynamā*
- Classe2: Conjonction de coordination: *tumma*
- Classe3: Particule de négation: *lan*, *lam*
- Classe4: Conjonction d'interrogation: *limā*
- Classe5: Préposition: *hattā*,
- Classe6: ADVT: *lan*, *qad*, *lam*
- Classe7: CCOND: *'idā*, *'id*, *ḥaytumā*
- Classe8: Particule de surprise: *'id*, *'idā*
- Classe9: Particule de futur: *sawfa*

III. Notre étude des invariants lexicaux temporels.

1. Classification des invariants temporels étudiés:

1.4. Classification en fonction de sa fonction de rection dans la phrase: 2 classes



I
Contexte et
objectifs.

II
Les invariants
lexicaux en
arabe.

III
Notre étude
des invariants
lexicaux
temporels.

IV
L'application
Kawâkib.

V
Conclusion et
perspectives.

III. Notre étude des invariants lexicaux temporels.

I
Contexte et
objectifs.

II
Les invariants
lexicaux en
arabe.

III
Notre étude
des invariants
lexicaux
temporels.

IV
L'application
Kawâkib.

V
Conclusion et
perspectives.

1. Classification des invariants temporels étudiés:

1.5. Classification en fonction de leur ambiguïté:

- Ambiguïté graphique.
- Ambiguïté grammaticale.
- Ambiguïté sémantique.

Invariant	Ambiguïté graphique	Ambiguïté grammaticale	Ambiguïté sémantique
l.m.ā	lammā	CS, PNE, ADVT	Correspondance Temporelle, négation
	limā	CS	Conséquence, interrogation
	lamā	PNE	Négation
‘indamā	—	CS, ADVT	Correspondance Temporelle
‘aṭnā’a	‘aṭnā’a, aṭnā’a	CS, ADVT, PREP	Synchronisation temporelle
’idā	’idā, ādā, āḍan, ’idān	CS, CCOND, ADVT, PS	Condition, surprise
’id	’id, āḍ	CS, COND, ADVT, PS	Surprise, condition
baynamā	—	CS, ADVT	Simultanéité de deux phénomènes.
ḥattā	—	CS, CC, PREP	Coordination, intensité, justification, direction...
ḥīn	ḥīn, ḥīna	CS, ADVT, PREP	Correspondance temporelle, un moment
lam	—	ADVT, PNE	Négation
lan	—	ADVT, PNE	Négation
munḍu	—	CS, ADVT, PREP	Rapport de temps, interrogation temporelle
qad	—	ADVT	Certitude, incertitude
tumma	ṭumma, ṭamma	CC, ADVT, PD, Verbe	Classement (ordre)
b’d	ba’da	PREP	Temporel, Continuité
	ba’du	ADVT	Antériorité
	bu’d	Nom	Distance
	ba’ida, ba’ada, ba’uda	Verbe	Eloigner
ba’damā	—	CS, ADVT	Temporel, Continuité
ḥīnamā	—	CS, ADVT, CCOND	Temporel, condition
kullamā	—	CS, CCOND, ADVT	Temporel, Condition
sawfa	—	ADVT, PF	Promesse
ḥīnaḍāk	—	CS, ADVT	Le moment de réalisation
ḥaytumā	—	CS, ADVT, CCOND	Temporel, Condition

Al-zuḡḡāḡi : « al-ḥarf mā dalla ‘alā ma‘nan fī ḡayrihi naḥwu min wa’ilā waṭumma » (La particule est ce qui dénote un sens dans un autre mot comme min, ‘ilā, ṭumma).

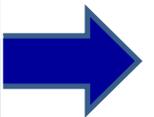
III. Notre étude des invariants lexicaux temporels.

2. Construction de corpus: Site du journal électronique « *Alwatan* » en 2004 (Ghoul, 2014).

2.1. Statistiques de notre corpus

Domaine	Nb articles	Nb phrases	Nb mots/ Nb mots différents	Nb invariants lexicaux
Culture	12	52 984	1 416 583/ 163 456	326 180
Economie	12	50 715	1 605 236/ 122 270	317 374
Religion	13	55 372	3 159 306/ 105 045	762 314
Sports	12	48 285	1 472 756/ 145 389	301 265
Total	49	207 356	7 653 881/ 466 623	1 707 133

23%



I

Contexte et objectifs.

II

Les invariants lexicaux en arabe.

III

Notre étude des invariants lexicaux temporels.

IV

L'application Kawâkib.

V

Conclusion et perspectives.

III. Notre étude des invariants lexicaux temporels.

2. Construction de corpus: Site du journal électronique « *Alwatan* » en 2004 (Ghoul, 2014).

2.1. Annotation morphosyntaxique:

- Segmentation: script Perl (Ghoul, 2013).
- Jeux d'étiquettes: 61 étiquettes (44 étiquettes pour les invariants lexicaux).
- Annotation semi-automatique en nous basant sur « TreeTagger ».

2.2. Évaluation de l'annotation:

- Calcul de l'accord inter-annotateurs AI entre deux linguistes arabes A1 et A2.
- Corpus d'évaluation : 8 135 mots / 300 phrases (~ 27 mots par phrase).
- AI =0,91% (275 phrases bien annotés validés par A1 et A2).
- P=89% (performance de l'annotation de TreeTagger).

III. Notre étude des invariants lexicaux temporels.

I
Contexte et
objectifs.

II
Les invariants
lexicaux en
arabe.

III
Notre étude
des invariants
lexicaux
temporels.

IV
L'application
Kawâkib.

V
Conclusion et
perspectives.

2. Construction de corpus: Site du journal électronique « *Alwatan* » en 2004 (Ghoul, 2014).

2.3. Correction manuelle de l'annotation: (10 personnes)

- Décomposition du corpus sur des fichiers de petites tailles (1000 fichiers).
- Correction de segmentation des phrases en unités lexicales minimales.
- Correction des frontières des phrases (de 27 mots à 15 mots par phrase) ➔ les invariants lexicaux peuvent être des marqueurs de début ou de fin de phrase.
- Correction d'annotation morphosyntaxique (attribution des étiquettes correctes aux unités lexicales).

Exemple avant correction manuelle : العالم / NMP دول / ADVT امس / ADVT صباح / NPF السلطنة / VAC شاركت / NMS / الاحتفال / NMS ب / PREP يوم / NMS التراث / NMS العالمي / NMS و / PRMS الذي / CC / النادي / PREP في / VAC اقيم / PRMS الثقافي / NMS ب / PREP القرم / NPM . /PONC

Exemple après correction manuelle : ال / NMP دول / ADVT امس / ADVT صباح / NPF سلطنة / DET ال / VAC شاركت / NMS / احتفال / DET ال / NMS ب / PREP يوم / NMS التراث / DET ال / NMS العالمي / ADJ , / PONC و / الثقافي / DET ال / NMS نادي / DET ال / NMS ب / PREP القرم / NPM . /PONC

III. Notre étude des invariants lexicaux temporels.

I
Contexte et
objectifs.

II
Les invariants
lexicaux en
arabe.

III
Notre étude
des invariants
lexicaux
temporels.

IV
L'application
Kawâkib.

V
Conclusion et
perspectives.

3. Méthodologie et modélisation par des schémas de grammaires:

3.1. Notre méthode de modélisation:

- Cadre de la théorie des « machines abstraites ».
- En nous basant sur un corpus ou échantillon des phrases.
- L'origine de cette approche est due à la représentation de la langue arabe sous forme d'un langage quotient ou squelette.
- Deux phases d'analyse:
 - Niveau linguistique: étude linguistique de fond de chaque invariant lexical et dégager les informations qui peuvent identifier l'invariant en question .
 - Niveau graphique: Codifier les différentes informations sous forme des schémas de grammaires.

III. Notre étude des invariants lexicaux temporels.

I
Contexte et
objectifs.

II
Les invariants
lexicaux en
arabe.

III
Notre étude
des invariants
lexicaux
temporels.

IV
L'application
Kawâkib.

V
Conclusion et
perspectives.

3. Méthodologie et modélisation par des schémas de grammaires:

3.1. Notre méthode de modélisation:

3.1.1. Niveau linguistique:

- Utilisation de la méthode « feedback » en traitant le contexte des invariants lexicaux à partir d'un corpus.
- Créer une fiche signalétique pour chaque invariant lexical:
 - ✓ Classe ou catégorie syntaxique.
 - ✓ Forme graphique.
 - ✓ Reconnaissance assurée d'un invariant
 - ✓ Marqueurs de genre , de nombre, aspect...
 - ✓ Suppressible ou non.
 - ✓ Position dans la phrase.
 - ✓ Gouverne un ou deux noyaux.
 - ✓ Attentes syntaxiques en n+1.
 - ✓ Analyse distributionnelle.
 - ✓ etc...

III. Notre étude des invariants lexicaux temporels.

I
Contexte et
objectifs.

II
Les invariants
lexicaux en
arabe.

III
Notre étude
des invariants
lexicaux
temporels.

IV
L'application
Kawâkib.

V
Conclusion et
perspectives.

3. Méthodologie et modélisation par des schémas de grammaires:

3.1. Notre méthode de modélisation:

3.1.1. Niveau linguistique:

- Repérage des invariants lexicaux dans le corpus.
- Traitement de l'environnement de l'invariant lexical.
- Elaborer des hypothèses sur la structure de la phrase en nous fondant sur le principe de la phrase vide.
- Dégager les différents caractéristiques de l'invariant lexical (fiche signalétique).
- Étude des attentes syntaxiques des invariants lexicaux.
- Étude sémantique des invariants lexicaux.
- Codifier les différentes informations sous forme de grammaires formelles non figées.

III. Notre étude des invariants lexicaux temporels.

I
Contexte et
objectifs.

II
Les invariants
lexicaux en
arabe.

III
Notre étude
des invariants
lexicaux
temporels.

IV
L'application
Kawâkib.

V
Conclusion et
perspectives.

3. Méthodologie et modélisation par des schémas de grammaires:

3.1. Notre méthode de modélisation:

3.1.2. Niveau graphique:

- Modéliser les grammaires sous la forme des automates à états finis (schémas des grammaires calculables).
- Ces schémas représentent les successions des attentes syntaxiques des invariants lexicaux.

3.1.3. Définition d'une grammaire:

Une grammaire $G = \{N, \Sigma, P, q_0\}$, avec N est le vocabulaire non terminaux, Σ est le vocabulaire terminaux (les différentes catégories syntaxiques d'un mot), P est l'ensemble des règles de production et $q_0 \in N$ est le non terminal initial.

3.1.4. Définition d'un schéma de grammaire:

Un schéma de grammaire $A = \{Q, \Sigma, \delta, Q_0, Q_F\}$, avec Q est l'ensemble des états, Σ est un ensemble finis de symboles (catégories syntaxiques d'un mot) δ est l'ensemble de transitions entre les différents états, Q_0 est l'état initial et Q_F est l'ensemble des état finaux.

III. Notre étude des invariants lexicaux temporels.

I
Contexte et
objectifs.

II
Les invariants
lexicaux en
arabe.

III
Notre étude
des invariants
lexicaux
temporels.

IV
L'application
Kawâkib.

V
Conclusion et
perspectives.

3. Méthodologie et modélisation par des schémas de grammaires:

3.1. Notre méthode de modélisation:

3.1.5. Limites de notre méthode:

- Notre méthode consiste en une analyse de surface de ces invariants dans le but de construire des grammaires formelles non figées qui représentent l'environnement de chaque invariant. Ces grammaires peuvent être reliées et combinées entre elles. Cependant, notre méthode présente quelques limites:
 - Construction « ad hoc » conçue pour une tâche donnée: les automates peuvent contenir des chemins inutiles.
 - Difficulté de marquer le sens de l'invariant lexical dans l'automate.
 - Nos schémas de grammaires peuvent produire dans certains cas des phrases grammaticalement correctes mais sémantiquement incorrectes (phrases non valides à cause de la présence d'une succession d'épsilons).

III. Notre étude des invariants lexicaux temporels.

3. Méthodologie et modélisation par des schémas de grammaires:

3.2. Catalogue des invariants lexicaux étudiés:

- 20 invariants lexicaux temporels.
 - Invariants lexicaux simples
 - Invariants lexicaux composés: x mā, avec x: les invariants lexicaux simples.
- Temporalité:

La temporalité désigne le temps grammatical et le temps chronologique ou sémantique. Elle s'exprime grammaticalement soit par le temps du verbe : le passé, le présent et le futur. La langue arabe exprime également la temporalité par des éléments lexicaux comme « 'indamā, ḥīna, lammā... ». En revanche, l'étude linguistique de la temporalité appelle deux catégories imbriquées : l'aspect et le temps.

➔ Les invariants lexicaux temporels de la langue arabe sont donc les éléments grammaticaux qui permettent d'assurer l'enchaînement chronologique dans un récit. Par exemple, pour marquer la durée, comme la conjonction de subordination, l'adverbe de temps, la conjonction conditionnelle, etc.

I
Contexte et
objectifs.

II
Les invariants
lexicaux en
arabe.

III
Notre étude
des invariants
lexicaux
temporels.

IV
L'application
Kawâkib.

V
Conclusion et
perspectives.

III. Notre étude des invariants lexicaux temporels.

I
Contexte et
objectifs.

II
Les invariants
lexicaux en
arabe.

III
Notre étude
des invariants
lexicaux
temporels.

IV
L'application
Kawâkib.

V
Conclusion et
perspectives.

4. Les invariants lexicaux temporels simples:

- 13 invariants lexicaux simples: ḥattā, l.m.ā, ḥīn, lam, lan....

4.1. L'invariant lexical « ḥattā »:

4.1.1. Étude linguistique:

- Extrait de 150 phrases à partir de notre corpus.

FD (n+1)	N	VAC	VIN	Particule	CH
Fréq ḥattā	39	18	38	53	2
(%)	26%	12%	25%	35%	2%
Catégorie grammaticale	Préposition ou adverbe	Subordination temporelle	Subordination de but ou temporelle	Subordination ou préposition ou adverbe	Préposition

- ➔ L'attente d'un autre invariant lexical en n+1 est plus élevée par rapport aux autres formes de discours. Ce qui montre que « ḥattā » a une relation forte avec quelques invariants lexicaux.

- ➔ ḥattā avec sa forme graphique unique, peut jouer le rôle d'une CS, PREP ou ADV.

III. Notre étude des invariants lexicaux temporels.

I
Contexte et
objectifs.

II
Les invariants
lexicaux en
arabe.

III
Notre étude
des invariants
lexicaux
temporels.

IV
L'application
Kawâkib.

V
Conclusion et
perspectives.

4. Les invariants lexicaux temporels simples:

4.1. L'invariant lexical « ḥattā »:

Noyau2: V+S

4.1.1. Étude linguistique: conjonction de subordination

Exemple:

Noyau1: V+S

سأطوف به انحاء البلاد حتى يراه الجميع « sa'aṭūfu bihi 'anḥa 'al-bilād ḥattā yarahū al-ḡamī'u » (Je vais l'emmener partout dans le monde **jusqu'à** ce que tout le monde le voie).

➤ Deux sens:

- ✓ إلى أن *ilā 'an* (**jusqu'à**) : conjonction de subordination temporelle. «Je vais faire un circuit autour du monde avec lui **jusqu'à** ce que tous le voient ».
- ✓ كي *kay* (**afin que**) : conjonction de subordination de but. « je vais faire un circuit autour du monde avec lui afin que tous le voient ».

➤ Changement de temps et d'aspect du verbe:

طاف به انحاء البلاد حتى رأه الجميع « *tāfa bihi 'anḥā' albilād ḥattā ra'āhu alḡamī'* » (Il lui a fait faire un circuit aux quatre coins du pays jusqu'à ce que tout le monde l'ait vu). → CST (*ilā 'an*).

طاف به انحاء البلاد حتى يراه الجميع « *tāfa bihi 'anḥā' albilād ḥattā ra'āhu alḡamī'* » (Il lui a fait faire un circuit aux quatre coins du pays jusqu'à ce que tout le monde l'ait vu). → CSB (*kay*).

III. Notre étude des invariants lexicaux temporels.

I
Contexte et objectifs.

II
Les invariants lexicaux en arabe.

III
Notre étude des invariants lexicaux temporels.

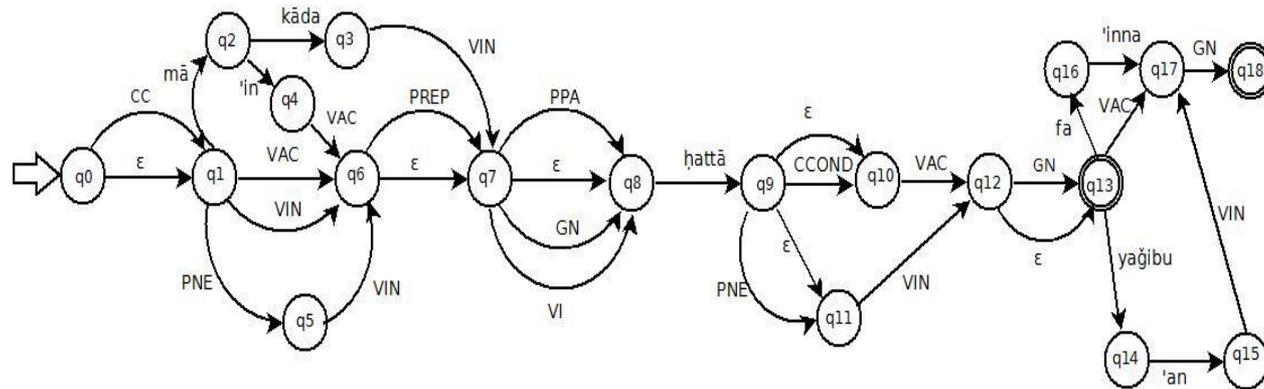
IV
L'application Kawâkib.

V
Conclusion et perspectives.

4. Les invariants lexicaux temporels simples:

4.1. L'invariant lexical « ḥattā »:

4.1.2. Schéma de grammaire: conjonction de subordination



- ➔ ḥattā dans le cas de CS attend toujours un verbe (VAC ou VIN) ou PNE, CCOND.
- ➔ ḥattā peut être discontinu sous la forme (mā kāda ...ḥattā ... ou mā 'in ... ḥattā...).
- ➔ (CC+ e) mā p kāda p VIN (PPA+ P GN+ e) p ḥattā p VAC (GN+ e).

III. Notre étude des invariants lexicaux temporels.

I
Contexte et
objectifs.

II
Les invariants
lexicaux en
arabe.

III
Notre étude
des invariants
lexicaux
temporels.

IV
L'application
Kawâkib.

V
Conclusion et
perspectives.

4. Les invariants lexicaux temporels simples:

4.1. L'invariant lexical « ḥattā »:

4.1.2. Étude linguistique: Adverbe

Exemple:

- 1) « *'ulawin ḥattā alḥā'it alladī tu'alliqu 'alyhi allawḥat.* » (Je colore **même** le mur sur lequel est attaché le tableau.)
- 2) « *wa li'adam wḡūd 'aṣḍiqā' ḥaqīqīn fī ḥayātihi fayawmahu kullahu mawā'id ḥattā al-dirāsat.* » (Vu qu'il n'a pas de vrais amis dans sa vie, toute sa journée est une suite de rendez-vous, même ses études.)
- 3) « *kāna yastaḥī' al-sayda ḥattā fī šuhūr al-sayf.* » (Il pouvait chasser même pendant les mois d'été)

→ Marqueur d'intensité

→ la suppression de « ḥattā » perturbe le sens de la phrase.

III. Notre étude des invariants lexicaux temporels.

I
Contexte et
objectifs.

II
Les invariants
lexicaux en
arabe.

III
Notre étude
des invariants
lexicaux
temporels.

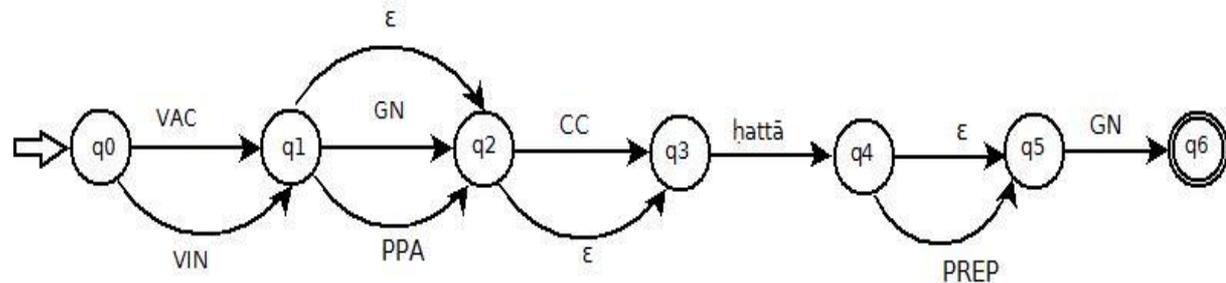
IV
L'application
Kawâkib.

V
Conclusion et
perspectives.

4. Les invariants lexicaux temporels simples:

4.1. L'invariant lexical « ḥattā »:

4.1.2. Schémas de grammaire: Adverbe



➔ Dans le cas d'un adverbe, ḥattā attend toujours un groupe nominal.

➔ (VAC+ VIN) (PPA+ p GN+ e) p (CC+ e) ḥattā p (PREP+ e) p GN.

III. Notre étude des invariants lexicaux temporels.

I
Contexte et
objectifs.

II
Les invariants
lexicaux en
arabe.

III
Notre étude
des invariants
lexicaux
temporels.

IV
L'application
Kawâkib.

V
Conclusion et
perspectives.

4. Les invariants lexicaux temporels simples:

4.1. L'invariant lexical « ḥattā »:

4.1.3. Étude linguistique: Préposition

Exemple:

يستمر حتى الثالث والعشرين من أبريل الجاري « *yastamiru ḥattā alṭalīṭa wa al 'iṣrūn min 'abrīl alġārī* » (Il se poursuit jusqu'au 23 Avril) ≡ يستمر إلى الثالث والعشرين من أبريل الجاري « *yastamiru ḥattā alṭalīṭa wa al 'iṣrūn min 'abrīl alġārī* » (Il se poursuit **jusqu'au** 23 Avril)

➔ « ḥattā » permute avec « 'ilā » qui désigne la fin d'une période.

➔ « ḥattā » se localise toujours au milieu de la phrase.

➔ La suppression de « ḥattā » entraîne la perturbation de la phrase.

➔ Il engage la globalité de la phrase.

III. Notre étude des invariants lexicaux temporels.

I
Contexte et objectifs.

II
Les invariants lexicaux en arabe.

III
Notre étude des invariants lexicaux temporels.

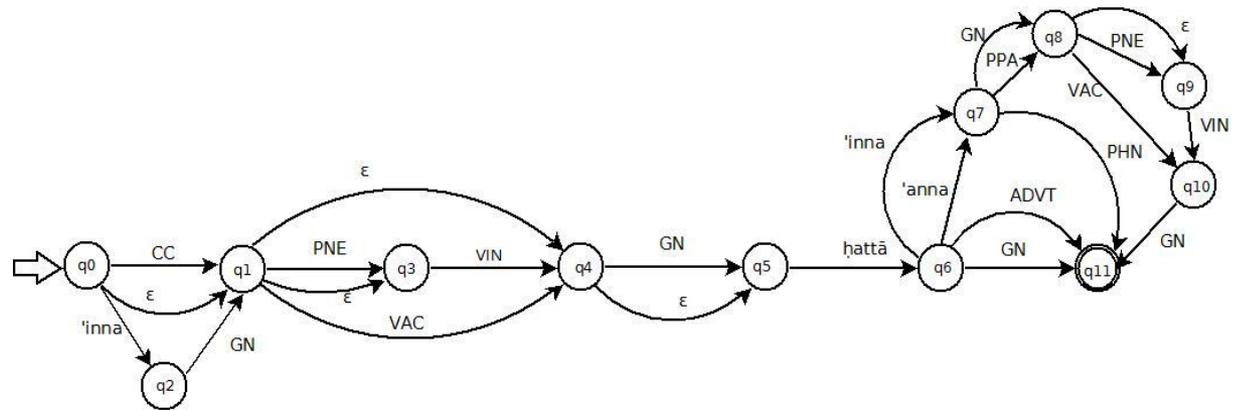
IV
L'application Kawâkib.

V
Conclusion et perspectives.

4. Les invariants lexicaux temporels simples:

4.1. L'invariant lexical « ḥattā »:

4.1.3. Schémas de grammaire: Préposition



أكلت السمكة حتى رأسها [q0, q1, q4, q5, q6, q11].

أكلت السمكة حتى رأسها [q0, q1, q2, q3, q4, q5, q6].

➔ (CC+ e) ḥattā p ('inna+ 'anna) (PPA+ p GN) p ((PNE+ e) p VIN) + VAC) p GN.

➔ ḥattā peut avoir 9 sens distincts: Ceux qui correspondent à 'ilā 'an wasalat 'ilā avec la possibilité d'effacer wasalat 'ilā, 'ilā, kay, 'ilā yawm ou 'ilā yawmanā hadā, bimā fī ḍalika, 'illā wa, 'ilā darajat, raġma 'annahom, wa.

ḥattā (حتى)	
Ambiguïté graphique	Unique
Ambiguïté grammaticale	CS, ADV, PREP
Sens	Multiple : intensité, temporel, justification...
Suppressible /insuppressible	Insuppressible
Position	Début ou milieu de la phrase
Changement de la position	Oui, mais dans certains cas
Agglutination à n-1	wa et fa
Agglutination à n+1	Jamais
Forme de la phrase	ḥattā P Q ou P ḥattā Q
Commutation avec d'autres invariants	Oui, avec : kay, 'ilā, 'ilā 'an, ...
Les attentes syntaxiques en n+1	GV, GN, PREP, ADVT, 'inna, 'anna
Gouvernance	Gouverne deux noyaux : Verbe+ sujet dans le cas d'une conjonction de subordination
Discontinus	Oui
Traduction	Jusqu'à, même, pour que...

III. Notre étude des invariants lexicaux temporels.

I
Contexte et
objectifs.

II
Les invariants
lexicaux en
arabe.

III
Notre étude
des invariants
lexicaux
temporels.

IV
L'application
Kawâkib.

V
Conclusion et
perspectives.

5. Les invariants lexicaux temporels composés: (x.mā)

- 7 invariants lexicaux complexes: baynamā, 'indamā, ba'damā....

5.1. L'invariant lexical « baynamā »:

5.1.1. Étude linguistique: bayna . mā

- 100 phrases .
- Il joue toujours le rôle d'une conjonction de subordination temporelle qui exprime divers sens.
- Au début de la phrase, il indique un rapport de temps qui annonce une surprise:

بينما أتجهز بمكة للحاق بأبي لقيتني هند بنت عتبة
« baynamā 'ataggahiz bimakkat li-alḥāq bi'abī laqītnī hind bint 'atabat »
(Pendant que je me préparais à la Mecque pour rejoindre mon père, j'ai rencontré Hend bent Ataba).

- Au milieu de la phrase, il prend une valeur d'opposition , contradiction

يخضع كل شيء لقانون الزمان والمكان بينما يتخطى القرآن نطاق هذا القانون
« yaḥḍa'u kullu šay' liqānūn alzamān wa almakān
baynamā yataḥaṭā alqur'ān niṭāq ḥaḍā alqānūn.» (Tout dépend de la loi du temps et de l'espace mais le Coran dépasse les dimensions de cette loi)

- ➔ Il engage la globalité de la phrase.
- ➔ La suppression de « baynamā » entraîne la perturbation de la structure et du sens de la phrase
- ➔ Permutation avec d'autres mots est autorisée: 'indamā kāna, fī ḥīn,

III. Notre étude des invariants lexicaux temporels.

I
Contexte et
objectifs.

II
Les invariants
lexicaux en
arabe.

III
Notre étude
des invariants
lexicaux
temporels.

IV
L'application
Kawâkib.

V
Conclusion et
perspectives.

5. Les invariants lexicaux temporels composés: (x.mā)

5.1. L'invariant lexical « baynamā »:

5.1.1. Étude linguistique: bayna . mā

➤ baynamā, attire plus loin une particule de surprise comme « 'id̲ » et « 'idā »?.

« *fabaynamā* 'anā nāzil tahta šajarat 'idā birasūl allah ṣallā allah 'alyhi wa salim 'aqbala 'ilaynā » (Tandis que j'étais assis au-dessous d'un arbre, le prophète Mohamed est venu)

« *baynamā* naḥnu yamāa gulūs 'id̲ marit binā 'agūz min 'agā'izihom taḥmalu 'alā ra'sihā qilatuu min mā' ». (Un jour pendant que nous étions assis une vieille des leurs passa ayant une cruche d'eau sur la tête

- ➔ La suppression de la particule de surprise dépend de la nature de mot que le suit.
- ➔ « baynamā » peut attirer la particule « 'idā » ou « 'id̲ » dans certains cas qui présentent une contrainte lâche de « baynamā » dans le cas où « 'idā » et « 'id̲ » est suivi d'un verbe. Cette contrainte est devenue stricte dans le cas où « 'idā » et « 'id̲ » est suivi de bi +GN.

baynamā (بينما) : PREP (bayna) . mā	
Ambiguïté graphique	Unique
Ambiguïté grammaticale	Unique : Conjonction de subordination temporelle.
Sens	Annonce d'un évènement inattendu, l'opposition : valeur contradictoire.
Suppressible /insuppressible	Insuppressible
Position	Début ou milieu de la phrase
Changement de la position	Oui
Agglutination à n-1	wa et fa, alif hamza d'interrogation
Agglutination à n+1	Jamais
Forme de la phrase	baynamā P Q ou P baynamā Q
Commutation avec d'autres invariants	Oui, avec : 'indamā kāna, fī ḥīn, fīmā
Marqueur	Accompli
Les attentes syntaxiques en n+1	VAC ou VIN ou PPI ou PD ou GN
Gouvernance	Gouverne deux noyaux : Verbe+ sujet et/ou thème+ prédicat
Discontinus	Oui : baynamā'id/'idā
Traduction	Tandis que, Pendant que, Cependant, au moment où on était Cependant ...

III. Notre étude des invariants lexicaux temporels.

I
Contexte et objectifs.

II
Les invariants lexicaux en arabe.

III
Notre étude des invariants lexicaux temporels.

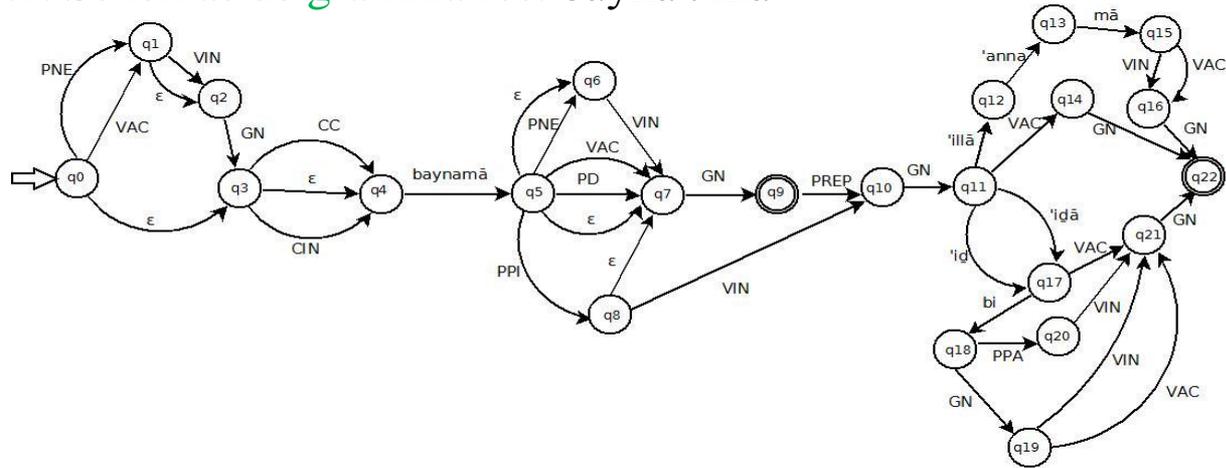
IV
L'application Kawâkib.

V
Conclusion et perspectives.

5. Les invariants lexicaux temporels composés: (x.mā)

5.1. L'invariant lexical « baynamā »:

5.1.1. Schémas de grammaires: bayna . mā



(CC+ e) baynamā p PPI p GN p PREP p GN p ('id+ 'idā) p bi (PPA+ p GN) p (VAC+ VIN) p GN.

- ➔ Les invariants lexicaux composés attendent en n+1 un groupe verbal sauf dans le cas de « baynamā » qui peut attendre en plus un groupe nominal.
- ➔ Ils gouvernent toujours deux noyaux, soit deux noyaux verbaux, soit un noyau verbal et un noyau nominal.

I

Contexte et
objectifs.

II

Les invariants
lexicaux en
arabe.

III

Notre étude
des invariants
lexicaux
temporels.

IV

L'application
Kawâkib.

V

Conclusion et
perspectives.

IV. L'application Kawâkib.

1. Présentation:

- Application Web: IFAO et MMSH (Gaubert, 2001): base de données.
- Analyser les différentes parties de discours de la langue arabe (nom, verbe, invariants lexicaux) en se fondant sur la théorie des automates.
- Traiter la langue arabe sans lexique ou avec un lexique très réduit (règles minimales) .

2. Différentes fonctions:

- Corpus : une zone qui représente une banque de données textuelles.
- Texte arabe : une zone qui sert à insérer un texte en arabe.
- Réglages et actions : une zone qui contient les différentes fonctions pour traiter le texte.
- Résultats : une zone qui contient les résultats de chaque opération d'analyse ou fonction.

Al-zuġġāġi : « al-ḥarf mā dalla 'alā ma'nān fi ġayrihi naḥwu min wa'ilā waṭumma » (La particule est ce qui dénote un sens dans un autre mot comme min, 'ilā, ṭumma).

Kawâkib Pro * كواكب

Corpus : [modifier](#) 0 textes, 0 mots

Corpus

Texte arabe : [agrandir](#) [masquer](#) [effacer](#) Kawâkib Web v 1.0b2 Denab 72 toks tmp "الكواكب"

[Analyses](#) [Expressions](#) [Corpus](#) [Critères](#)

Séquence

Next sel.

Réglages et actions : [clavier arabe](#) [aide](#) [regex](#)

Citations
Min al-
NegBal

- prep mot أن ctx
- subj mot أن mot
- mot الذي ctx
- nega أن ctx
- mot أن mot أن mot

Regex 2

color: n, v, mv filtr: mv

brut [chr](#)

Suites de tokens

Tokens

Phr ∅

- Tous
- adje: Adjectifs
- advb: Adverbes
- affi: Affirmatifs
- altr: Alternatifs

Expressions

[éditer expr.](#)

Tok temporels

Dates

Expr. temp.

Expr. argu.

Répétitions

Racine نظر (racine attestée)

Art. + waws

Art. - waws

Racines fréquentes

Débuts annex

Silences

NVT-lab

NVT-coul

Wa propers

Résultat : [agrandir](#) [droite à gauche](#) [effacer](#) [debug](#) Listok: [masquer](#) ID:76 mem: 95 Uniques

Testez-moi!

Al-zuġġāġi : « al-ḥarf mā dalla ‘alā ma ‘nan fī ġayrihi naḥwu min wa’ilā waṭumma » (La particule est ce qui dénote un sens dans un autre mot comme min, ‘ilā, ṭumma).

I

Contexte et
objectifs.

II

Les invariants
lexicaux en
arabe.

III

Notre étude
des invariants
lexicaux
temporels.

IV

L'application
Kawâkib.

V

Conclusion et
perspectives.

IV. L'application Kawâkib.

2. Rôle de « Kawâkib » dans l'analyse des invariants lexicaux:

- Détection des invariants lexicaux: Tokens.
- Détection des invariants lexicaux temporels : Tok temporels.
 - ✓ Distribution d'invariants en fonction de sa catégorie grammaticale et le taux de chaque invariant par rapport au nombre totale des invariants.

Al-zuġġāġi : « al-ḥarf mā dalla 'alā ma'nān fi ġayrihi naḥwu min wa'ilā waṭumma » (La particule est ce qui dénote un sens dans un autre mot comme min, 'ilā, ṭumma).

Kawākib Pro * كواكب

Corpus : [modifier](#) 61 textes, 237968 mots

Texte arabe : [agrandir](#) [masquer](#) [effacer](#) Kawākib Web v 1.0b2 Deneb 72 toks tmp "الراكب"

[Analyses](#) [Expressions](#) [Corpus](#) [Critères](#)

exemples temp :79

id	texte	dist
test Fatima :78	لا تفر	68
حوار مع الفنان القدير :77	و 2	64
حكايات الملك شهر	مجد	64
لماذا لا ينتجر الحكاء، :75		
عارفون على إعادة :74		
في أضرار غسيل الـ :73		

1) كلت السمكة حتى رأسها
2) سأطوف به انحاء البلاد حتى يراه الجميع
3) يعمل بشكل مكثف على مدار العشرة أيام الاخيرة حتى يتم عرض المسرحية في موعدها
4) يتعلق بعدم استطاعة الثقافة العربية حتى اليوم الوصول الى صيغة تتجاوز بها نقائضا
5) يستمر حتى الثالث والعشرين من أبريل الجاري
6) يسعى جاهدا لاتقان تلك اللهجة حتى يستطيع ان يجسد دوره بشكل صحيح
7) يتحدث بلهجة تلك المنطقة حتى لا يحدث تناقض في المشهد الدرامي
8) اللون حتى الحائط الذي تعلق عليه اللوحة

Sequence 1 6800 (25.0) 3000 Next set exemples temp - corpus thés Ghoul -

Réglages et actions : [clavier arabe](#) [aide](#) [regevl](#)

Citations

Min at NegBal

prep mot à ctx
subj mot à mot
mot fin ctx
nega à ctx
mot à mot à mot

mot à prep ctx

Reges 2 color xi v, nv filtre mv

brut dtr

Suites de tokens Tokens Phr

Expressions éditer expr

Tok temporels

Dates Expr. temp. Expr. argu

Tous adje Adjectifs adv: Adverbes aff: Affirmatifs altr: Alternatifs

Racine (racine attestée) Art + wavs Art - wavs

Racines fréquentes Débats annex

Silences NVT-Tab NVT-Couli Wa propres

Résultat : [réduire](#) [gauche à droite](#) [effacer](#) [debug](#) Listok: [masquer](#) ID:79 mem: 137 Uniques 563 lignes

txt 79 non ambigus : | ambigus : | tokens 36.82% 2579 wa,fa,bi,ka,li 11.07% 775

tok	cat	tmp	cpt
	prep		24.34 % (962)
	subj		7.67 % (303)
	rela		7.54 % (298)
	temp		7.49 % (296)
	inna		6.15 % (243)
	cond		6.07 % (240)
	intr		6.02 % (238)
	advb		5.46 % (216)
	nega		5.31 % (210)
	coor		5.01 % (198)
	verb		4.63 % (183)
	qprt		4.43 % (175)
	demo		3.47 % (137)
	detr		2.73 % (108)
	rstr		1.42 % (56)
	prop		1.39 % (55)
	altr		0.43 % (17)
	adje		0.23 % (9)
	voca		0.1 % (4)
	comp		0.1 % (4)
	affi		0.03 % (1)
	tous	Total amb.	3953
	prep		5.06% (200)
	prep		4.53% (179)
	subj	tmp: jusq	4.53% (179)
	prep		4.2% (166)

1) كلت السمكة حتى رأسها
2) سأطوف به انحاء البلاد حتى يراه الجميع
3) يعمل بشكل مكثف على مدار العشرة أيام الاخيرة حتى يتم عرض المسرحية في موعدها
4) يتعلق بعدم استطاعة الثقافة العربية حتى اليوم الوصول الى صيغة تتجاوز بها نقائضا
5) يستمر حتى الثالث والعشرين من أبريل الجاري
6) يسعى جاهدا لاتقان تلك اللهجة حتى يستطيع ان يجسد دوره بشكل صحيح
7) يتحدث بلهجة تلك المنطقة حتى لا يحدث تناقض في المشهد الدرامي
8) اللون حتى الحائط الذي تعلق عليه اللوحة
9) وتتعرف ماجدة بأنها في ذلك اليوم لم تكن تملك أي خبرة بطبيعة العمل في السينما حتى
أنها لم تكن تعرف موضع الكاميرا
10) التزمت حرفيا بكل تعليمات المخرج حتى لا تتسبب في أي خطأ
11) طلّت كذلك حتى انتهى تصوير الفيلم
12) اخفضي رأسك يا خالة حتى لا تصيبك رصاصاتهم الطائشة

I
Contexte et
objectifs.

IV. L'application Kawâkib.

2. Rôle de « Kawâkib » dans l'analyse des invariants lexicaux:

- Détection une suite des invariants lexicaux: Suites des tokens.

Suite de tokens /Entrée	Résultats
أكلت السمكة حتى رأسها	-
سأطوف به انحاء البلاد حتى يراه الجميع	-
يعمل بشكل مكثف على مدار العشرة أيام الاخيرة حتى يتم عرض المسرحية في موعدها	#حتى يتم #حتى لا
يتحدث بلهجة تلك المنطقة حتى لا يحدث تناقض في المشهد الدرامي	-
الون حتى الحائط الذي تعلق عليه اللوحة	#بأنها في ذلك #لم تكن
وتعترف ماجدة بأنها في ذلك اليوم لم تكن تملك أي خبرة بطبيعة العمل في السينما حتى أنها لم تكن تعرف موضع الكاميرا	#حتى أنها لم تكن #حتى لا
التزمت حرفيا بكل تعليمات المخرج حتى لا تتسبب في أي خطأ	

- Recherche des invariants lexicaux via des expressions régulières: Regx2

Exemple: ctx حتىctx

Al-zuġġāġi : « al-ḥarf mā dalla 'alā ma'nān fi ġayrihi naḥwu min wa'ilā waṭumma » (La particule est ce qui dénote un sens dans un autre mot comme min, 'ilā, ṭumma).

II
Les invariants
lexicaux en
arabe.

III
Notre étude
des invariants
lexicaux
temporels.

IV
L'application
Kawâkib.

V
Conclusion et
perspectives.

I
Contexte et
objectifs.

II
Les invariants
lexicaux en
arabe.

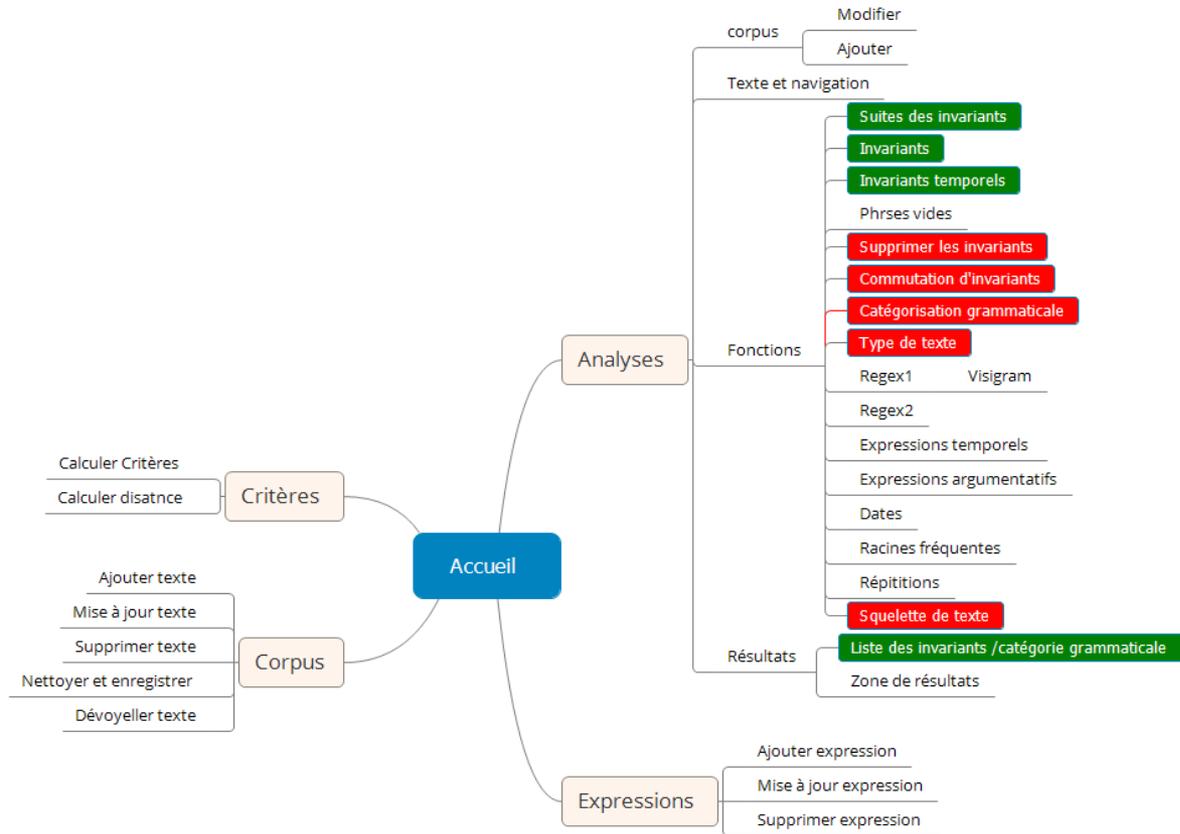
III
Notre étude
des invariants
lexicaux
temporels.

IV
L'application
Kawâkib.

V
Conclusion et
perspectives.

IV. L'application Kawâkib.

3. Quelques perspectives de l'application « Kawâkib »



Al-zuġġāġi : « al-ḥarf mā dalla ‘alā ma‘nan fi ġayrihi naḥwu min wa’ilā waṭumma » (La particule est ce qui dénote un sens dans un autre mot comme min, ‘ilā, ṭumma).

I
Contexte et
objectifs.

II
Les invariants
lexicaux en
arabe.

III
Notre étude
des invariants
lexicaux
temporels.

IV
L'application
Kawâkib.

V
Conclusion et
perspectives.

V. Conclusions et perspectives

Conclusions:

- Constitution, dans une optique de linguistique de corpus, d'un stock consistant, quoique non exhaustif, d'informations concernant l'environnement morphosyntaxique des 20 invariants lexicaux temporels.
- Ces informations recueillies, ont été ensuite systématiquement formalisées sous forme de diagrammes représentant des graphes orientés récursifs.
- Les invariants lexicaux temporels jouent un rôle très important dans la construction de la phrase: forts marqueurs de la structure générale de la phrase .
- La densité des invariants lexicaux permet de prédire la nature de texte : narratif, descriptif, ...
- La commutation d'un invariant avec un autre en gardant exactement le même sens d'origine est très rare.
- Possibilité d'analyser la langue arabe en nous fondant sur le langage squelette.
- Difficultés: séparer les critères purement syntaxiques de ceux qui sont de nature sémantique.

Al-zuġġāġi : « al-ḥarf mā dalla 'alā ma'nān fi ġayrihi naḥwu min wa'ilā waṭumma » (La particule est ce qui dénote un sens dans un autre mot comme min, 'ilā, ṭumma).

I
Contexte et
objectifs.

II
Les invariants
lexicaux en
arabe.

III
Notre étude
des invariants
lexicaux
temporels.

IV
L'application
Kawâkib.

V
Conclusion et
perspectives.

V. Conclusions et perspectives

Perspectives:

- Évaluer nos grammaires des invariants temporels sur une masse de données.
- Étudier les autres invariants lexicaux de l'arabe.
- Évoluer nos grammaires par « kawâkib ».
- Amélioration des schémas de grammaires en les rendent déterministes si possible afin de ne produire que des phrases correctes (valides).
- Transformations de ces schémas en graphe d'automates faiblement équivalents en sorte que les catégories puissent alors répondre rigoureusement au principe de commutation et de permutation qui régit la linguistique distributionnelle.
- Intégration de « sens » dans les schémas de grammaires : vers automates à états finis augmentés.
- Réalisation d'une application efficace pour le traitement automatique de l'arabe : amélioration de « kawâkib ».
- Vers une création d'un langage de manipulation de grammaires
- Vers une banque de grammaires des invariants lexicaux arabes : Modèle théorique original → Application qui traite l'arabe sans lexique. .

Al-zuġġāġi : « al-ḥarf mā dalla 'alā ma'nān fi ġayrihi naḥwu min wa'ilā waṭumma » (La particule est ce qui dénote un sens dans un autre mot comme min, 'ilā, ṭumma).

Merci pour votre attention

شكرا لكم على المتابعة والانتباه