



Al-Erfan Arabic electronocal dictionary

Nooj format

Nooj 2020 virtual conference

05 – 06 June 2020

By:

Mohamed El Hannach

Sidi Mohamed Ben Abdellah University

Fez - Morocco

And

Ahmed Bounoua

Sidi Mohamed Ben Abdellah University

Fez - Morocco



Content

- Main concepts
- Morphological database
 - N DB & V DB
 - VN DB & VA DB
 - DB Statistics
 - Morphological Netwoek
- Glides & Hamza grammar
- E-Dictionaries
 - N E-Dictionary VN E-Dictionary
 - VA E-Dictionary V E-Dictionary
- Evaluation (1): domain recognition
- Evaluation (2): Ligious proprieties recognition
- Conclusion and research perspectives
 - Idioms e-dictionary



Preamble

From Lexicon – Grammar To Nooi Platform

- Main concepts of lexicon grammar:
 - Classification, taxonomy
 - distribution an transformation
 - Large coverage data
 - Standard Arabic word components: Root and pattern
 - All our data is vocalized (diacriticized)
 - Each entry is associated to their pattern
 - Arabic basic sentence structure: V N0 W (W=: 0, N1, or N1 N2)
 - Operations: Passivation, Adjectivation, Nominalization, and Restructuration
- Morphological data base
 - V, V-n, V-a, and N Data base
- E-dictionary
 - V, V-n, V-a, and N E-Dictionary
- Local grammars:
 - V, V-n, V-a, and N graphs
- Idioms DB: More than 30.000 expressions
- Results



لـاستشارات التـرجمـة والـتطـبـيقـةـ العـصـدـيـةـ
Educational Consulting & Professional Development

Morphological part: N DB & V DB



- **N DB:**
- **21 properties, except the roots and patterns, such as:**
 - Lexicon classification: proper noun, commun noun
 - Semantic domains: 154 domains
 - Logical properties: ± Human, ± Animation; and ± Concret
 - Flexion: singular, plural, dual, diminutive, etc
 - Etc ...
- **V DB:**
 - **10 properties; Except the roots and patterns, such as:**
 - Phonological rules: Verb with ± glides
 - Syntactic structures: V N0 W
 - Syntactic description: fichers: ± transitive, reflichi
 - Etc ...



الاستشارات التربوية والتطوير المعرفي
Educational Consulting & Professional Development

Morphological part

VN DB & VA DB



- VN DB Comment:
 - 8 proprieties, except the roots and patterns, such as:
 - 5 Principal sub categories: Vn_Original, VN_mar, VN_hay, VN_mim, and VN_sin
 - Each V-n is associated to his pattern:
 - Each V-n is associated to its Verb
 - All the V-n items is vocalized
- VA DB comment:
 - 18 proprieties, more than roots & patterns, such as:
 - 12 principal sub categories
 - Each VA is associated to its pattern
 - All the V-a items is vocalized



DB statistics

Basic & compiled dictionary

VA	VN	Verbs	Nouns	Patterns	Roots
VA=: 87000 x P=: 12 = 1.044.000	VN=: 36000 x P=: 5 = 180000	V=: 26200 x P=: 10 =262000	N=:35000 x P=: 21 = 735000	1086 Patterns	12000 (3L + 4L)

Distributional process

Root	Pattern	Noun	Verb	VN	VA
+	+	+	-	+	+
+	+	-	+	-	+
+	+	+	+	+	-
+	-	-	-	+	+

E-Dictionary statistics after compilation

N	V	VN	VA	Total
21.245.000	18.250.000	15.000.725	34.250.001	88.781.000



لإسقاطات التربوية والتطوير العصري
Educational Consulting & Professional Development

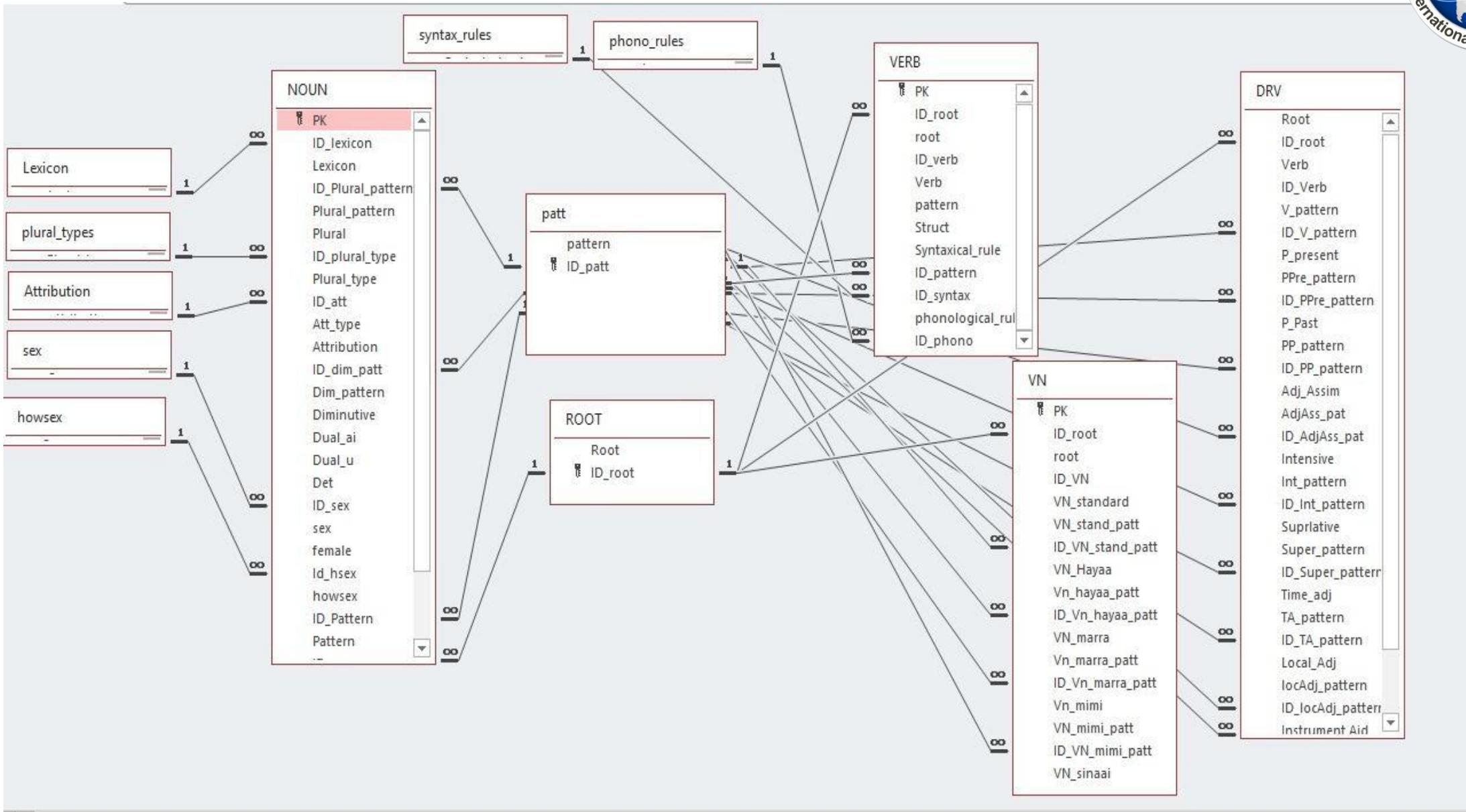
Ressources collection



- All the arabic dictionaries
 - General dictionaries
 - Specialized dictionaries: health, food, teaching, etc.
 - New arabic dictionaries
- BCA terminology database:
 - 36 dictionaries
- And others arabic documents
-

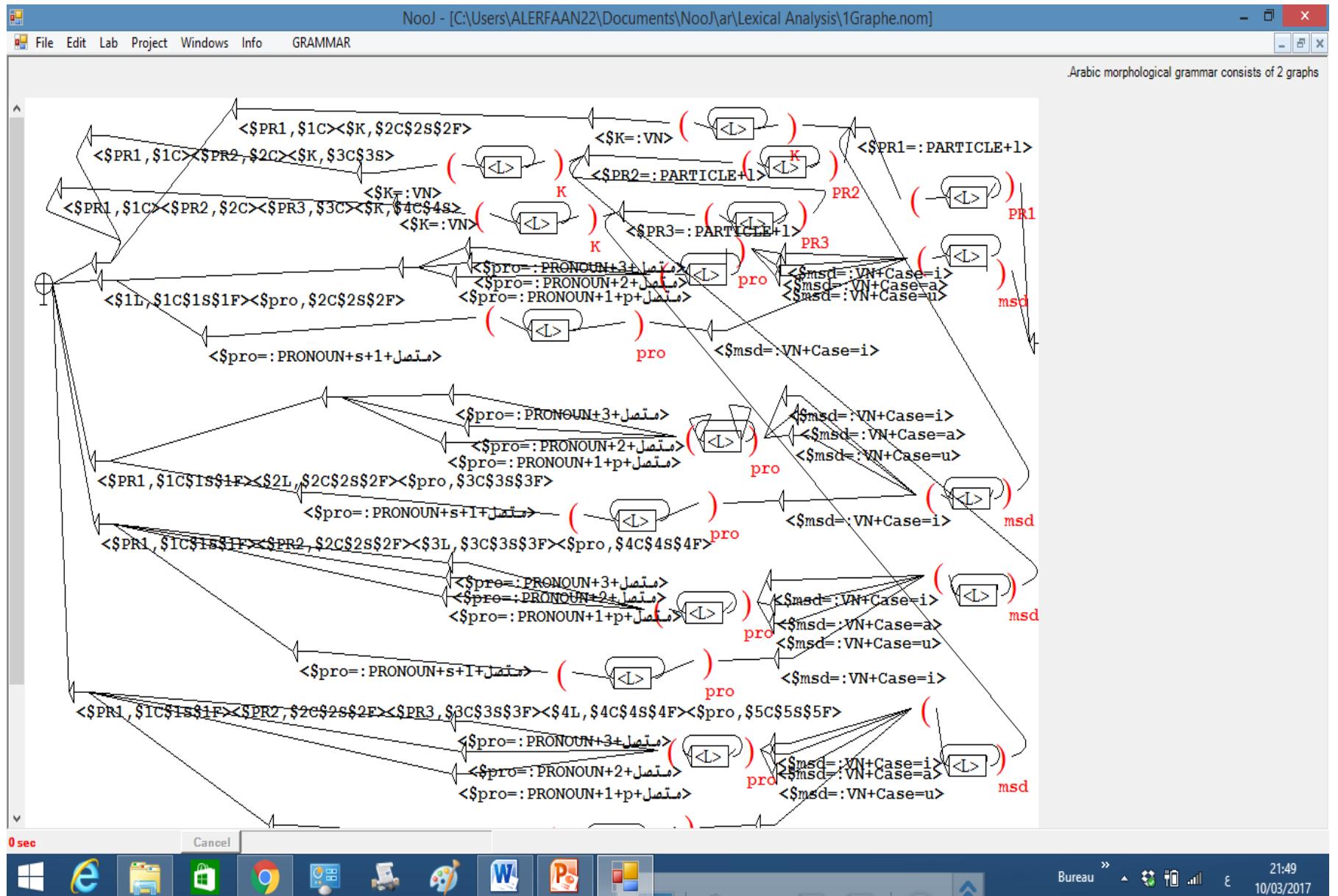


Morphological DB network





Glides and hamza graph





لإسقاطات التربوية والتطور العصبي
Educational Consulting & Professional Development

NooJ - [[Modified] Untitled]

File Edit Lab Project Windows Info TEXT

- + / 5 TU

Characters
Tokens
Digrams

Language is "Arabic(ar)".
Text Delimiter is: \n (NEWLINE)
Text contains 5 Text Units (TUs).
4 tokens including:
4 word forms
Text contains 16 annotations (16 different)

Show Text Annotation Structure

أثر
 يؤثر
 ائتج

عنى, V+Voix=Pas+Temps=PO+Pers=3+Genre=m+Nombre=s+Pattern=DRVORIGINALCAT=NW+synt_3+مُنْدَدٌ
عنى, V+Voix=Pas+Temps=PO+Pers=3+Genre=m+Nombre=s+Pattern=DRVORIGINALCAT=NW+synt_3+مُنْدَدٌ
عنى, V+Voix=Pas+Temps=PU+Pers=3+Genre=m+Nombre=s+Pattern=DRVORIGINALCAT=NW+synt_3+مُنْدَدٌ
عنى, V+Voix=Pas+Temps=PU+Pers=3+Genre=m+Nombre=s+Pattern=DRVORIGINALCAT=NW+synt_3+مُنْدَدٌ
عنى, V+Voix=Pas+Temps=PA+Pers=3+Genre=m+Nombre=s+Pattern=DRVORIGINALCAT=NW+synt_3+مُنْدَدٌ
عنى, V+Voix=Pas+Temps=PA+Pers=3+Genre=m+Nombre=s+Pattern=DRVORIGINALCAT=NW+synt_3+مُنْدَدٌ
عنى, V+Voix=Pas+Temps=PO+Pers=3+Genre=m+Nombre=s+Pattern=DRVORIGINALCAT=NW+synt_1+مُنْدَدٌ

0.1 sec Cancel

Type here to search

10:17 PM 6/4/2020





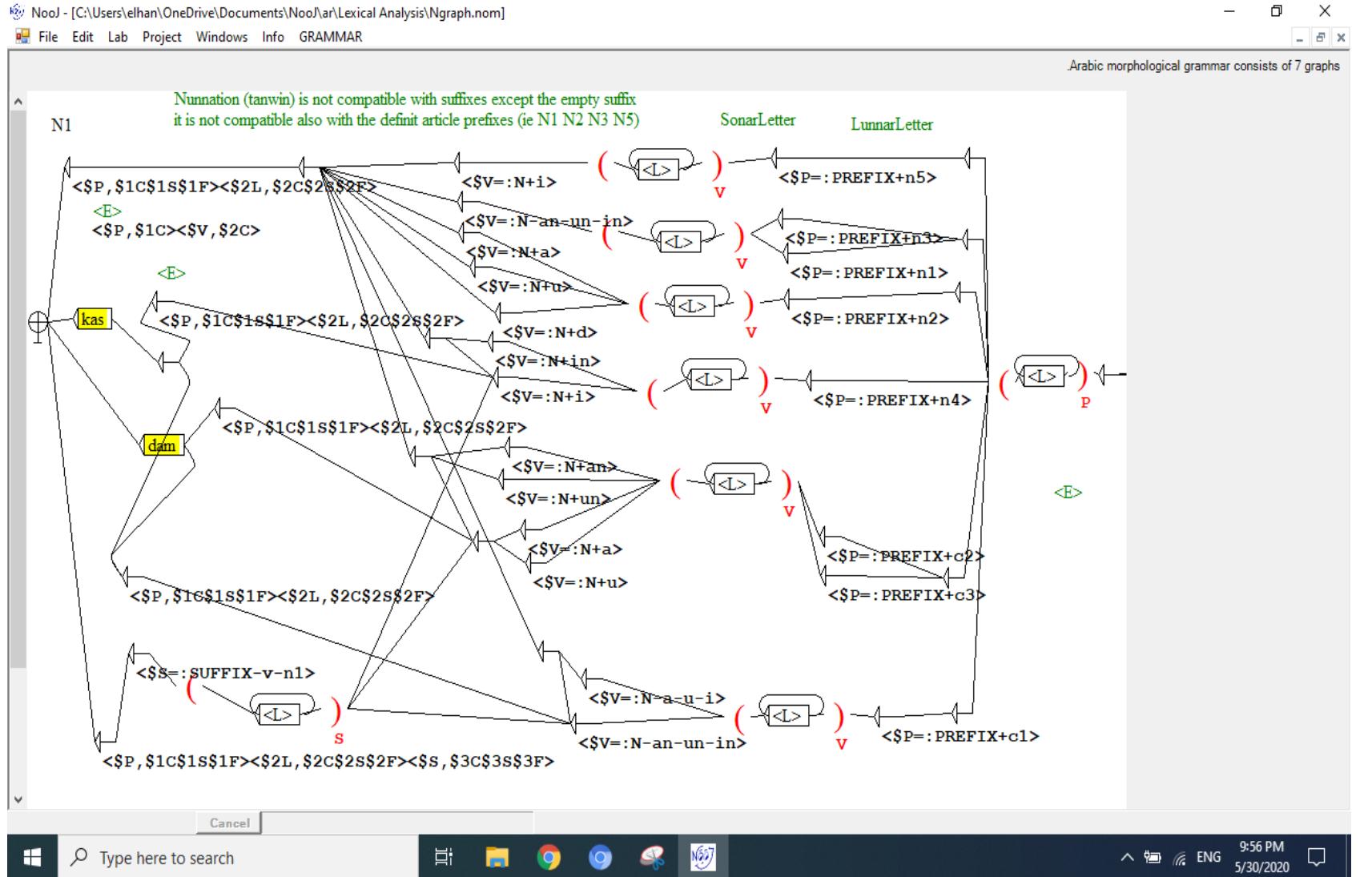
N Nooj E-Dictionary

- Proprietie: Dictionary sample

- Derivational rules: singular, dual, diminutive, plural, etc
- Text analysis by differnts ways:
 - General names: <N>
 - Morphological resarch, for example: <N+فعل>, <N+فعال>, etc ...
 - Semantical domains, for example: <N+الصحة>, <N+التعليم>; etc...
 - Flexional process, for example: <N+plural>, <N+dual>, etc...
 - Recognize the suffixes and prefixes, for example:



N Grammar



Résultats



The screenshot shows the NooJ software interface with the following details:

- Title Bar:** NooJ - [[Modified] Untitled]
- Menu Bar:** File, Edit, Lab, Project, Windows, Info, TEXT
- Toolbar:** Includes icons for Undo, Redo, Save, Print, and others.
- Text Editor Area:** Displays the word "رجل" (Rajul) in large black font at the top right.
- Annotations List:** A list of annotations for the word "رجل":
 - ن_رج_ل N+Type=Origin+ فعل+DRVORIGINALCAT=N+Lexicon=عام+Sem=Hum+Domain=جزء من_الإنسان+Sex=f+Case=un+Number=s
 - ن_رج_ل N+Type=Origin+ فعل+DRVORIGINALCAT=N+Lexicon=عام+Sem=Anim+Domain=جزء من_الإنسان+Sex=f+Case=un+Number=s
 - ن_رج_ل N+Type=Origin+ فعل+DRVORIGINALCAT=N+Lexicon=عام+Sem=Conc+Domain=جزء من_الإنسان+Sex=f+Case=un+Number=s
 - رج_ل N+Type=Origin+ فعل+DRVORIGINALCAT=N+Lexicon=الابن+Sem=Hum+Domain=الابن+Sex=m+Case=i+Number=s
 - رج_ل N+Type=Origin+ فعل+DRVORIGINALCAT=N+Lexicon=الابن+Sem=Anim+Domain=الابن+Sex=m+Case=i+Number=s
 - رج_ل N+Type=Origin+ فعل+DRVORIGINALCAT=N+Lexicon=الابن+Sem=Conc+Domain=الابن+Sex=m+Case=i+Number=s
 - رج_ل N+Type=Origin+ فعل+DRVORIGINALCAT=N+Lexicon=عام+Sem=Hum+Domain=الابن+Sex=m+Case=i+Number=s
 - رج_ل N+Type=Origin+ فعل+DRVORIGINALCAT=N+Lexicon=عام+Sem=Anim+Domain=الابن+Sex=m+Case=i+Number=s
 - رج_ل N+Type=Origin+ فعل+DRVORIGINALCAT=N+Lexicon=عام+Sem=Conc+Domain=الابن+Sex=m+Case=i+Number=s
- Information Panel:** Shows text statistics:
 - Language is "Arabic(ar)".
 - Text Delimiter is: \n (NEWLINE)
 - Text contains 92 annotations (92 different)
 - Linguistic Resources applied to the text: dic.nod(00) tokenisation.nom(00) name.nod(00) prefix.nod(00) suffix.nod(00) Ngraph.nom(00)
- Status Bar:** Shows "0.4 sec" for processing time, a "Cancel" button, and system icons for battery, signal, and date/time (1:31 PM, 6/1/2020).

Recherche du temps

Nooj

File Edit Lab Project Windows Info TEXT CONCORDANCE

Locate a pattern in Text 1

Pattern is:

- a string of characters:
- a PERL regular expression:
- a Nooj regular expression:
`<+الزمن>`
- a Nooj grammar:

Syntactic Analysis

Index Limitation

- Shortest matches
- Longest matches
- All matches

Only: 100 occ.

Reset Concordance

Concordance

Reset Display: characters before, and word forms after. Display: Matches Outputs

Text	After	Seq.	Before
أداء واجبات التلاميذ خلال فترة شهر التلائم الأخيرة، ويتمون أرباب مؤسسات	حول		وبين عدد من أشر التلاميذ
يناير وماي ويونيو، أما داعي شهر التلائم الأخيرة، ويتمون أرباب مؤسسات	شهر		أيام وأمهات التلاميذ أداء واجبات
يناير البعض منهم ينصل أبنائهم جين	حول		لأداء واجبات تدرمس أبنائهم خلال
أبريل وماي ويونيو، مما داعي شهر التلائم الأخيرة، ولكن بعض	شهر		التعليم الخصوصي بالجتمع؛ في
أبريل، ومنهم من ذهب إلى شهر التلائم الأخيرة فمع بالطرف الأول	حول		اتجاز الدافت القائم بين المدفون
الشهرى أو تأجيله أو أي حل	شهر		واجبات تدرمس أبنائهم، ابتداء من
تطلب بعض الأسر رفضها أيام	حول		لأداء واجبات تدرمس أبنائهم برس
الشهر التلائم الأخيرة بتوقف الدراسة الحضورية	شهر		مع ظله، إما بتخفيف الواجب
، لكن غالبي تصور علائى سنين	حول		يسمى بالفرضي. وجوابا عن سؤال
ورق في الإدارة المركبة والجهوية	شهر		أيام وأمهات التلاميذ خلال
	شهر		العيد من مواطن الكوى منذ
	شهر		والخدمات الرسمية المتطرورة، من خلال

Query 12/12

في المقابل، ينتي أرباب مؤسسات التعليم الخصوصي تفهمه لوضمة أولياء التلاميذ المتضررين من تداعيات جائحة كورونا، ويندون أنهما مستعدون لتسوية المشكل المتعلقة بالجانب المادي، لكنهم يكترون أنه "من غير المنطقي أن تتبع الأسر التي لم تتضرر من الجائحة واجبات تدرمس أبنائهم".

وزهاب محمد عمور، رئيس رابطة التعليم الخاص بال المغرب، إلى أنه جهات تحريض الأسر على أرباب مؤسسات التعليم الخصوصي، دون أن يحدد من هي هذه الجهات، مكتفيا بالقول: "المشكل القائم متعلق بمصطلح، وما نعيشه اليوم هو نتيجة تحريض مقصود تتف خلفه جهات مميتة لا داعي ذكرها".

0 sec Cancel

Type here to search

1:47 PM 6/1/2020 FRA



لإسقاطات التربوية والتطوير العصري
Educational Consulting & Professional Development

Verb Nooj E-Dictionary



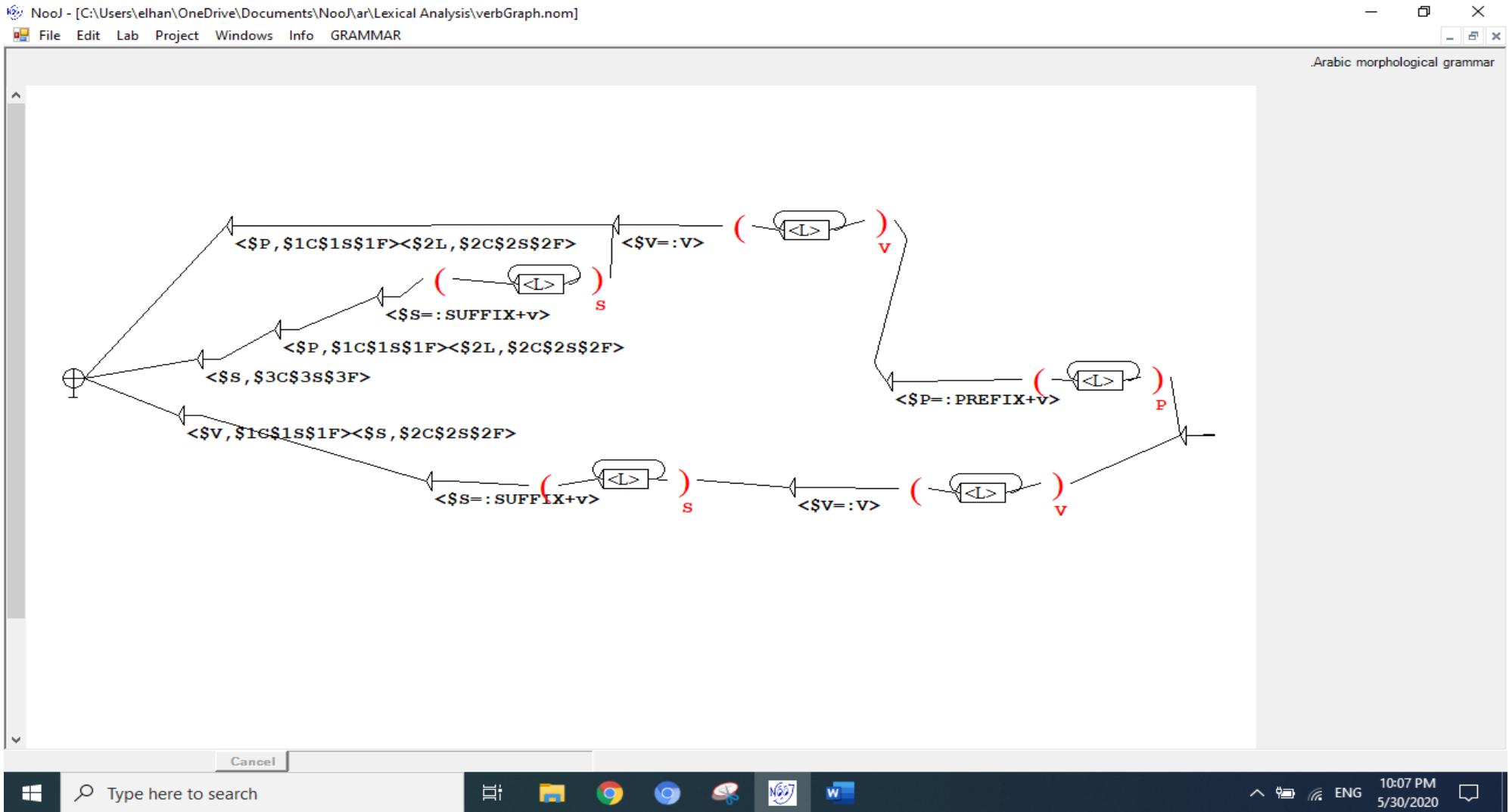
- Proprieties: [sample e-dictionary](#)
- Morphological proprieties:
 - For past and present tens: <فعُلُّ يفَعِلُ>
 - Syntactical structure: V N0 W: transitif, intransitif, refléchi; etc
 - Phonological rules using specialized tems, such as: مهُمُوز، مثَال، ناقص،
 - Text analysis by:
 - Verb entry only: <V>
 - Patterns, for example: <V+فاعل>,
 - Phonological rules, for exemple: <V+مثال>



لـاستشارات التـرجمـة والـنظـرـةـ الـعـدـدـيـ Educational Consulting & Professional Development



V grammar



Text sample

The screenshot shows a software window titled 'C:\Users\ALERFAAN22\Documents\NooJ\ar\Projects\Press1.not'. The top right corner has standard window controls (minimize, maximize, close). The top left shows a document icon and the file path. The main area has a toolbar with icons for 'Characters', 'Tokens', and 'Diagrams' (with dropdown arrows), and a status bar indicating 'Language is "Arabic(ar)". Text Delimiter is: \n (NEWLINE) Text contains 166 Text Units (TUs)'.

A 'Linguistic Analysis' context menu is open, listing:

- Locate Pattern... (Ctrl+L)
- Export colored text as an RTF document
- Export annotated text as an XML document...
- Modify Text (Ctrl+M)

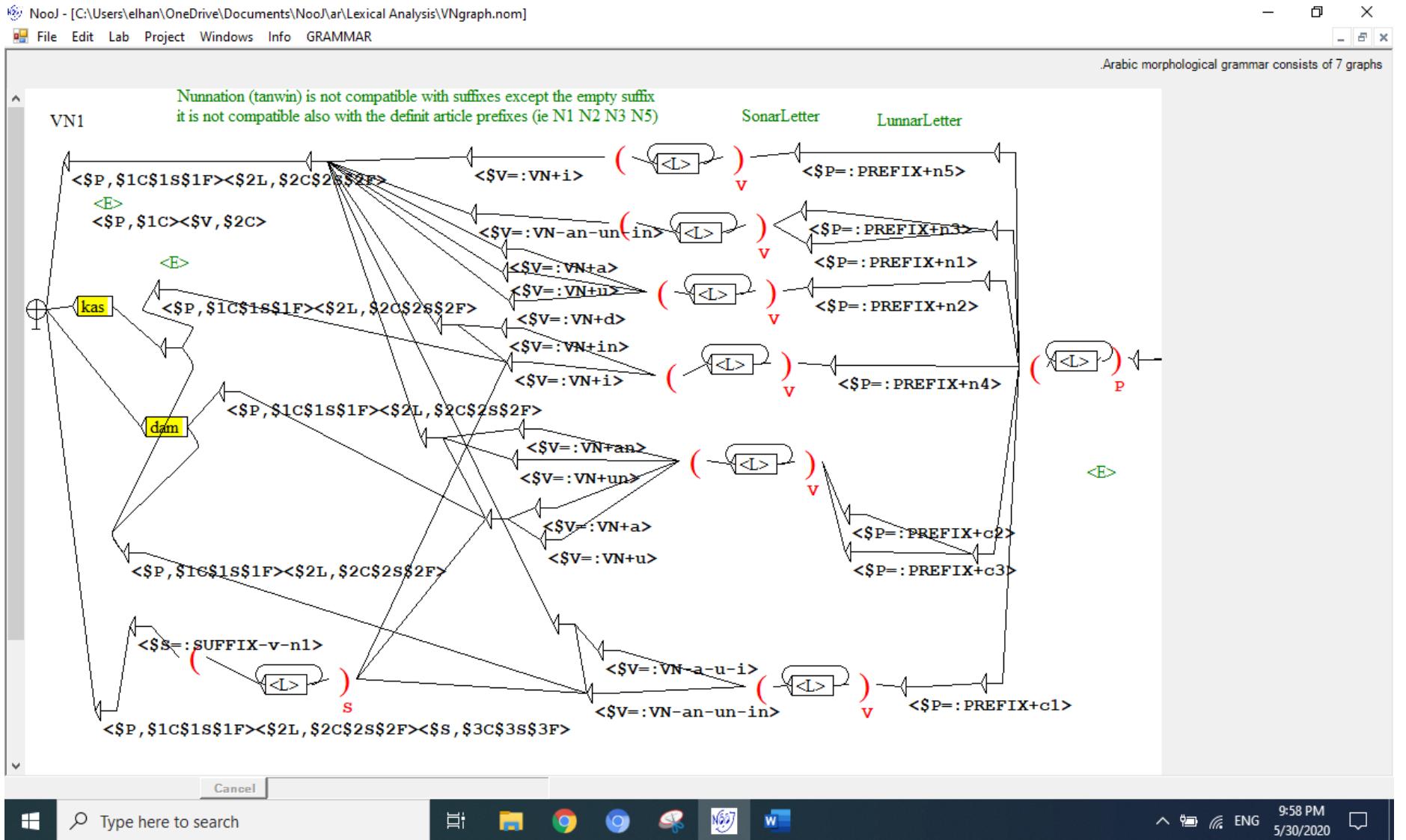
To the left, there's a sidebar with a checkbox for 'Show Text Annotation Structure' and a tree view showing hierarchical text structures. The main text area contains Arabic text, with some parts highlighted in red and blue, likely representing different linguistic features or annotations.

The text in the main area discusses political parties and their positions, mentioning the National Democratic Party, the National Bloc, and the National Salvation Force, among others, and their stances on issues like the election and the coup.

Locate V category

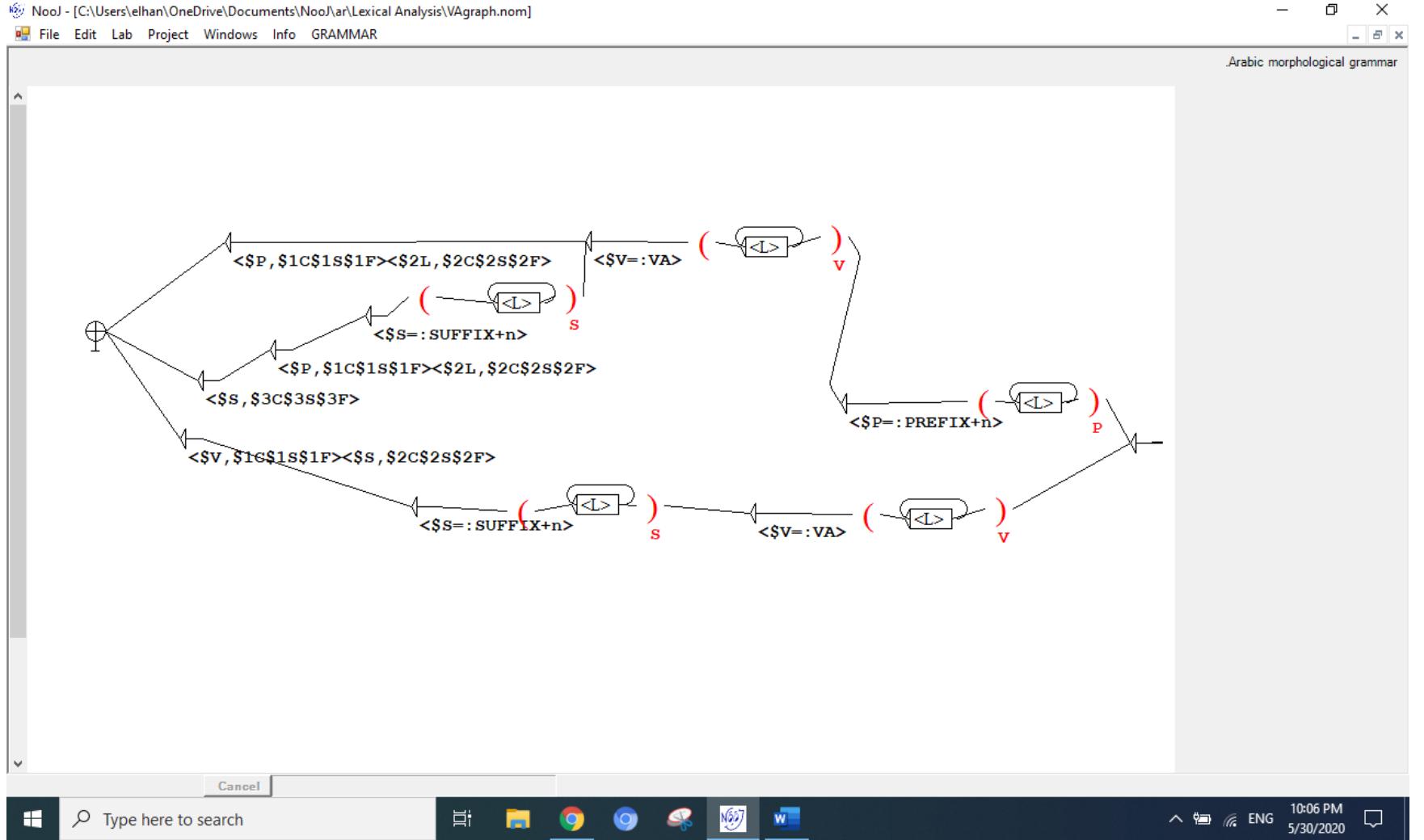


VN Grammar





VA Grammar





الاستشارات التربوية والتطوير المهني
Educational Consulting & Professional Development



Syntactical analysis

- Data collection
- Main structure: V NO W
- Verb equal basic sentence
- Tree types of sentences:
 - Basic sentence, based on main distributions
 - Idiomatic sentence
 - Sentence with support verb (Nominalization)
- Built tables: Columns & lignes based on bidirectional transformation
- More than 150 properties an sub properties
- Develop electronical dictionary
- Develop local grammar
- Apply to randomed texts



Syntactic database 2.500.000 basic sentences



Lexicon–grammar tables



Application to a LG table

التوسيع الآلي للجمل بمنصة نوج وتقانة تعلم الآلة

سبب س 0 ذعراً على	سبب س 0 فصاً ل س 1
سبب س 0 الفص ل س 1	سبب س 0 الخوف لعلى
سبب س 0 ذهول على	سبب س 0 (فص س 1)
أثار س 0 حزناً في على	أثار س 0 فصاً في س 1
أثار س 0 الفص في س 1	أثار س 0 الشجن في على
أثار س 0 أسى في على	أثار س 0 (فص س 1)
خلق س 0 المتاعب لعلى	خلق س 0 الفص ل س 1
خلق س 0 صعوبة لعلى	خلق س 0 فصاً ل س 1
أدخل س 0 البهجة على على	أدخل س 0 الفص على س 1
أدخل س 0 بهجة على على	أدخل س 0 فصاً على س 1
أحدث س 0 الألم في نفس على	أحدث س 0 الفص ل س 1
أحدث س 0 جرحاً نفس على	أحدث س 0 فصاً في س 1
حرك س 0 الغيرة في نفس على	حرك س 0 الفص في س 1
حرك س 0 أسى في على	حرك س 0 فصاً في س 1
بعث س 0 الحقد في نفس على	بعث س 0 الفص في س 1
أعطى س 0 القوة لعلى	أعطى س 0 الفص ل س 1
أضفى س 0 الجمال على على	أضفى س 0 فصاً على س 1
جلب س 0 الأدّى لعلى	جلب س 0 الفص ل س 1



Evaluation

Domain recognition

- We applied our local grammar and e-dictionary to standards arabic texts, the results were as follows:

Texts	N	Result	V	Result	VN	Result	VA	Result
Teaching	<N+ التعليم>	92.17 %	<N+ التعليم>	90.01 %	<VN+ التعليم>	86.00 %	<VN+ التعليم>	97.15 %
Health	<N+ الصحة>	87.52 %	<N+ الصحة>	89.25 %	<VN+ الصحة>	71.12 %	<VN+ الصحة>	91.00 %
Food	<N+ الغذاء>	99.00 %	<N+ الغذاء>	93.12 %	<VN+ الغذاء>	90.00 %	<VN+ الغذاء>	95.25 %
Tourism	<N+ السياحة>	94.02 %	<N+ السياحة>	90.15 %	<VN+ السياحة>	89.13 %	<VN+ السياحة>	91.25 %



Evaluation N

Logical properties recognition

N	Result	Example
<N+_+H_+A_+C>	99.01 %	رجل، ابن، أحمد
<N+_+H_-A_+C>	91.15 %	أرنب، أسد، قرد
<N_-H_+A_-C>	95.12 %	الحب، الكراهية، الفكر
<N_-H_-A_+C>	99.15 %	حائط، مدرسة، بيت
<N+_+H_-A_+C>	79.02 %	الأصبع، الأظافر، الشعر



لإسقاطات التربوية والتطوير العصري
Educational Consulting & Professional Development



Research perspectives

- E-dictionary for Arabic idioms
 - We have collected more than 30.000 expressions
 - Classification based on: Expression an mining
 - We plan to develop a e-dictionary nooj format
 - We plan to translate it to English and French language



Arabic idioms sample

أحمر أو مُهْمَرْ	عصب الأفق	عصب
أقام فيه	عصب الرجل بيته	عصب
ضم ما تفرق منها بحبل	عصب الشجرة	عصب
أطاف	عصب الغيار بالجبل	عصب
استكفت به	عصب القدم بفلان	عصب
جوعهم	عصب القوم	عصب
ضمهم واشتد عليهم	عصب القوم أمر	عصب
اجتمعوا حوله	عصب القوم بفلان	عصب
لزمه	عصب الماء	عصب
ضيق الصدر	عصب المزاح	عصب
شد فخذيها	عصب الناقة	عصب
وضع حدا	عصب سلمته	عصب
دارت حوله	عصبت الإبل بالماء	عصب



International project using Erfan E-Dictionary

- ALA

الاستشارات التربوية والتطوير المهني
Educational Consulting & Professional Development

العنوان: almojam.org

Apps Text Reverse Nouveau dossier phpMyAdmin erfan Gmail YouTube WhatsApp

محمد الحناش 204 إشعارات صفحات سجل نشاط المستخدم إحصائيات سير العمل المskوكات الجذور

المعجم الثنائي للغة العربية

رقم الملف	المؤلف	المراجعة	التاريخ	الوقت	الحروف
ث م أ	مقرر لجنة 3 - المغرب (محمد الحناش)	قيد المراجعة	05.06.2020	21:06	
ح ج ر	مقرر لجنة 3 - المغرب (أحمد صالح حسين محمد الفراشي)	قيد التنفيذ	05.06.2020	21:06	251:م غ:1
ح ج رف	مقرر عام - المغرب (عبد الفتاح الحجمري)	قيد المراجعة	05.06.2020	21:06	
ح ج ز	مقرر لجنة 3 - المغرب (محمد الحناش)	قيد المراجعة	05.06.2020	21:06	

عدد المدخلات (14) سجل تسلسل تاريخي للجذر

مقرر لجنة 2 (لجنة 3) - المغرب (علي بولعلام)

مقرر لجنة 6 (لجنة 2.3) - المغرب (أحمد صالح حسين محمد الفراشي)

مقرر عام - المغرب (عبد الفتاح الحجمري)

مقرر خبير 2 (لجنة 3) - المغرب (علي بولعلام)

بطاقة تقنية عن الو

بطاقة تقنية عن الو

Show all



Type here to search



9:22 PM

6/5/2020



الكليل للمؤسسات التعليمية
Al-Khalil for Educational Institutions



- Thank you