

النحو العربي

Arabic Syntax

النحو هو المجال اللغوي المهتم بنمذجة الطريقة التي تُرتب فيها الكلمات معاً لعمل سلاسل أكبر في اللغة. في حين يصف الصرف (المورفولوجي) تركيب الكلمات داخلياً، ويصف النحو كيف تأتي الكلمات معاً لتكوين العبارات والجمل.

الكثير من المفردات التي تناقش النحو تكون مشتركة عبر اللغات المختلفة، على سبيل المثال مفردة: الفعل، والجمله الفعلية، والفاعل، والمفعول به. غير أن هناك بعض الاستثناءات التي تتعلق بالتركيب الفريدة التي لا تظهر مشتركة لغوياً، على سبيل المثال، الإضافة والتمييز في العربية. في هذا الفصل سنناقش حسب الحاجة الشروط المحددة والعامة للنحو. وللحصول على مقدمة عامة للنحو، ننصح القارئ بالاطلاع على العديد من الأبحاث المنشورة، على سبيل المثال، [134، 135] وغيرها. وقد نُظّم هذا الفصل على النحو التالي.

أولاً: سنقدم نظرة عامة على النحو العربي في قسم ٦.١ وقسم ٦.٢، ثم نستعرض ثلاثة مشاريع لبنوك شجرية (treebanking) عربية ونقارن بين المناهج المختلفة التي تستخدمها. وأخيراً، يلخص قسم ٦.٣ الجهود البحثية في التحليل النحوي للغة العربية.

٦,١ نظرة عامة على التراكييب النحوية العربية

في هذا القسم ، سنقدم مسحاً عاماً للظواهر النحوية في اللغة العربية. وللحصول على معرفة جيدة في هذا المجال ، ننصح بالاطلاع على بعض من المراجع العديدة في قواعد اللغة العربية [78، 7، 77، 76، 81، 82].

٦,١,١ مذكرة حول النحو والصرف

قد تكون العلاقة بين النحو والصرف معقدة خاصة بالنسبة للغات الثرية نحويًا، حيث يتم التعبير عن الكثير من الظواهر النحوية ليس فقط من ناحية ترتيب الكلمات ولكن أيضاً من ناحية الصرف. على سبيل المثال، في العربية، الحالة الإعرابية للفاعل هي الرفع، ومعدلات الصفة للأسماء (adjectival modifiers of nouns) تتفق مع حالة الاسم الذي تعدله. ثراء الصرف العربي مكنه من أن يكون لديه قدر من الحرية في ترتيب الكلمات وذلك لأن الصرف يمكن أن يُمثل بعض العلاقات النحوية. لكن، وكما في العديد من اللغات الأخرى، فإن الاستخدام الفعلي للغة العربية أقل حرية من حيث ترتيب الكلمات مما يمكن أن يكون من حيث المبدأ.

يتداخل الصرف اللصقي مع العديد من التراكييب النحوية. وعلى هذا النحو، في البنوك الشجرية العربية عادة ما تقطع الكلمات لفصل جميع اللواحق (مع استثناء أُل التعريف، الذي لا يؤثر في البنية النحوية). في هذا الفصل، جميع الأمثلة ستُقطع بأسلوب بنك بنسلفانيا الشجري للتحليل النحوية [9، 136، 82].

٢, ١, ٦ تركيب الجملة

في اللغة العربية هناك نوعان من الجمل : جملة فعلية ويرمز لها (V-Sent) وجملة اسمية ويرمز لها (N-Sent). أيضاً تسمى الجملة الاسمية بالجملة الربطية/الجملة ذات الطرفين (copular/equational sentences).

الجملة الفعلية

التركيب النموذجي للجملة الفعلية هو فعل - فاعل - مفعول به (Verb-Subject-Objects)، ويمكن تمثيلها بعدة أشكال. يتكون أبسط نموذج للجملة الفعلية من فعل فقط مع فاعل الضمير مستتر فيه^(١). يعبر الفعل عن المخاطب والجنس والعدد للفاعل.

(٦,١) جملة فعلية: فعل + فاعل (Verb+Subj)

katab+a كَتَبَ

wrote+3MS

'He wrote'

katab+nA كَتَبْنَا

wrote+1P

'We wrote'

ويظهر الفاعل للاسم الظاهر (Non-pronominal subjects) بعد الفعل. ويتفق الفعل مع الفاعل في الجنس (مذكر أو مؤنث) والمخاطب (غائب) ولكن ليس في العدد

(١) تعتبر قواعد النحو العربي التقليدي ضمائر الرفع المتصلة هي الفاعل ما عدا لواصق المفرد المذكر/المؤنث الغائب التي تستخدم لمطابقة العدد في الجملة الفعلية. ويتم عرض مفهوم ضمير مستتر مفرد مذكر/مؤنث غائب لشرح استتار الضمير المفرد المذكر/المؤنث الغائب.

الذي يتحول افتراضياً إلى مفرد (S_G). وفي الجملة الفعلية، الفعل مع فاعل غير ضميري من غير الممكن أن يكون جمعاً (P_L). وتكون الحالة الإعرابية للفاعل مرفوعاً.

(٦،٢) جملة فعلية: فعل فاعل مرفوع (Verb Subject^{Nom})

katab+a *Al+walad+u/Al+ÂwlAd+u* كَتَبَ الْوَلَدُ/الْوَلَدَاتُ
wrote+3MS the+boy+Nom/the+boys+Nom
'The boy/boys wrote'

katab+nA *Al+bint+u/Al+banAt+u* كَتَبَتِ الْبِنْتُ/الْبَنَاتُ
wrote+3FS the+girl+Nom/the+girls+Nom
'The girl/girls wrote'

كما شاهدنا سابقاً في قسم ٤.٢.١، تظهر ضمائر المفعول به كجزء من اللواحق الفعلية بصرف النظر عن الفاعل إن كان ضميراً أو لا. وفيما يلي المثالان السابقان مع إضافة ضمائر مفعول به:

(٦،٣) جملة فعلية: فعل + فاعل — مفعول به (Verb+Subj -Obj)

katab+a *-hA* كَتَبَ -
wrote+3MS -it هَا
'He wrote it'

katab+nA *-hA* كَتَبْنَا -هَا
wrote+1P -it
'We wrote it'

(٦،٤) جملة فعلية: فعل - مفعول به فاعل مرفوع (Verb -Obj Subject^{Nom})

katab+a *-hA* *Al+walad+u* كَتَبَ - هَا الْوَلَدُ
wrote+3MS -it the+boy+Nom
'The boy wrote it'

عادة ما يتبع المفعول به للاسم الظاهر (Non-pronominal verb objects) الفاعل. ومن ثم تكون الحالة الإعرابية للمفعول به النصب.

(٦.٥) جملة فعلية: فعل + فاعل مفعول به منصوب (Verb+Subj Object^{Acc})

katab+a Al+qiS~ah+a كَتَبَ الْقِصَّةَ
wrote+3MS the+story+Acc
'He wrote the story'

(٦.٦) جملة فعلية: فعل فاعل مرفوع مفعول به منصوب (Verb Subject^{Nom})

katab+a Al+walad+u Al+qiS~ah+a كَتَبَ الْوَلَدُ الْقِصَّةَ
wrote+3MS the+boy+Nom the+story+Acc
'The boy wrote the story'

ونظراً إلى أن نهايات الحالة الإعرابية لا تكتب بالعادة، فإن هناك غموضاً شائعاً مرتبطاً بالتسلسل [Verb+3S NounPhrase] عندما يتفق الفعل مع العبارة الاسمية بالجنس: (أ) العبارة الاسمية هي الفاعل أو (ب) الفاعل ضمير إما مؤنث غائب (3FS) أو مذكر غائب (3MS) والعبارة الاسمية هي المفعول به.

وكما هي في اللغات الأخرى، للغة العربية أفعال لازمة (intransitive) لا تأخذ أي مفعول به، وأفعال متعدية (transitive) تأخذ مفعولاً به واحداً، وأفعال تأخذ مفعولين (ditransitive) وهنا تكون الحالة الإعرابية للمفعول به المباشر وغير المباشر النصب.

(٦.٧) جملة فعلية: أ. فعل فاعل مرفوع مفعول به غير مباشر منصوب مفعول به

مباشر منصوب (Verb Subject^{Nom} IOBJECT^{Acc} DOBJECT^{Acc})

Aaṣ Taý Al+walad+u Al+bint+a kitAb+Aã اعطى الولد البنت كتاباً
give+3MS the+boy+Nom the+girl+Acc book+Acc
'The boy gave the girl a book'

وكما هي في اللغة الإنجليزية، فإن بنية الأفعال المتعدية لمفعولين لها شكل آخر حيث يظهر المفعول به غير المباشر كمفعول به لحرف الجر ل- (1-) وبجالة جر إعرابياً: (٦,٨) جملة فعلية: ب. فعل فاعل مرفوع مفعول به مباشر منصوب ل- مفعول به

غير مباشر مجرور (Verb Subject^{Nom} DObject^{Acc} 1- IObject^{Gen})

اعطى الولدُ كتاباً ل- *Al-walad+u kitAb+Aã I- Al+bint+i*
 give+3MS the+boy+Nom book+Acc to- the+girl+Gen
 البنت
 'The boy gave a book to the girl'

يظهر ضمير المفعول به غير المباشر كلاحقة فعلية:

(٦,٩) جملة فعلية: فعل + فاعل - مفعول به غير مباشر مفعول به

مباشر منصوب (Verb+Subj -IObj DObject^{Acc})

اعطينا - ها كتاباً *Aã Tay+nA -hA kitAb+Aã*
 give+1P -her book+Acc
 'We gave her a book'

(٦,١٠) جملة فعلية: فعل - مفعول به غير مباشر فاعل مرفوع مفعول به

مباشر منصوب (Verb -IObj Subject^{Nom} DObject^{Acc})

اعطى - ها الولدُ كتاباً *Aã Tay -hA Al-walad+u kitAb+Aã*
 give+3MS -her the+boy+Nom book+Acc
 'The boy gave the girl a book'

إلا أنه إذا كان المفعول به المباشر وغير المباشر كليهما ضمائر، فإن المفعول به المباشر يظهر كضمير منفصل مباشر بعد الفاعل.

(٦.١١) جملة فعلية: فعل + فاعل - مفعول به غير مباشر مفعول به مباشر

(Verb+Subj -IObj DOBJect)

Aaṣṣ Tay+nA -hA AyA+hu اعطيتنا - ها إيأه
 give+1P -her it
 'We gave her it'

(٦.١٢) جملة فعلية: فعل - مفعول به غير مباشر فاعل مرفوع مفعول به مباشر

(Verb -IObj Subject^{Nom} DOBJect)

Aaṣṣ Tay' -hA Al+walad+u AyA+hu اعطى -ها الولد إيأه
 give+3MS -her the+boy+Nom it
 'The boy gave her it'

الجملة الاسمية

التركيب النموذجي للجملة الاسمية (N-Sent) هو مبتدأ وخبر (Subject-Predicate أو Topic-Complement)، ويطلق عليها أحياناً بنية تعتمد على فعل رابط/مساعدة أو جملة معادلة.

أنواع الجملة الاسمية: الجملة الاسمية بشكلها البسيط: يكون المبتدأ في العادة اسماً معرفاً أو اسم علم أو ضميراً في حالة الرفع والخبر يكون اسماً غير معرف أو اسم علم أو صفة تتفق مع الفاعل في الجنس والعدد.

(٦.١٣) جملة اسمية: مبتدأ مرفوع خبر مرفوع (Subject^{Nom} Predicate^{Nom})

Al+kitAb+u jaaiyd+ũ الكتابُ جديّد
 the+book+MS+Nom+Def new+MS+Nom+InDef
 'The book is new'

hiya salma'y
she Salma
'She is Salma'

هي سلمى

Al+i: jul+Ani *kAtib+Ani*
the+man+DU+Nom+Def author+DU+Nom+InDef
'The two men are authors'

الرَّجُلَانِ كَاتِبَانِ

في بقية هذا القسم ، سنعمل على الحد من التوضيح في الخصائص الصرفية (المرفولوجية) للحد الأدنى المطلوب. بالإضافة إلى الشكل الأساسي للخبر، يمكن للخبر أن يكون شبه جملة من الجار والمجرور (PP) .

(٦.١٤) جملة اسمية: المبتدأ مرفوع خبر - جار ومجرور (PP- Subject^{Nom})

(Predicate

Al+i: jul+u *iyy Albayt+i*
the+man+Nom in the+house+Gen
'The man is in the house'

الرَّجُلُ فِي الْبَيْتِ

كما يمكن للخبر أن يكون جملة اسمية أخرى. في هذه البنية ، يعد المبتدأ في الجملة الاسمية العليا بمثابة موضوع. أما الخبر في الجملة الاسمية العليا عادة ما يشير إلى الموضوع باستخدام الإحالة الضميرية.

(٦.١٥) جملة اسمية: مبتدأ^١ مرفوع [جملة اسمية مبتدأ^٢ مرفوع خبر مرفوع^١]

(Subject₁^{Nom} [N-Sent Subject₂^{Nom} Predicate₂^{Nom}]₁)

Al+bayt+u *bAb+u -hu* *jadiyd+ũ*
the+house+Nom door+Nom -its new+Nom
'The house, its door is new'

الْبَيْتُ بَابٌ -هُ جَدِيدٌ

ومع ذلك ، ربما البنية الخبرية الأكثر إثارة هي تلك التي تتضمن جملة فعلية. ومن الشائع ، أن ينتج هذا البناء ترتيباً شبيهاً بـ (فاعل - فعل - مفعول به) في العربية عندما يكون الفاعل للجملة الفعلية الإسنادية المضمنة يشير إلى الفاعل في الجملة الاسمية الرئيسية. عندها يتفق الفعل والفاعل تماماً في المخاطب والجنس والعدد بدلاً من الاتفاق في الجنس والعدد كما في الجملة الاسمية الاعتيادية. ويطلق على هذا البناء أحياناً بالجملة المعقدة. قارن المثال التالي مع بديله بالجملة الفعلية الأساسية.

(٦.١٦) جملة اسمية : مبتدأ_١ مرفوع [جملة فعلية فعل + فاعل_١ مرفوع مفعول به منصوب

أخبر (Subject₁^{Nom} [V-Sent Verb+Subj₁^{Nom} Object^{Acc}] Predicate)

Al+AwlAd+u katab+uwA Al+qiSaS+a الأَوْلَادُ كَتَبُوا الْقِصَصَ
the+boys+Nom wrote+3MP the+stories+Acc
'The boys wrote the stories' (lit. 'The boys, they wrote the stories')

katab+a Al+AwlAd+u Al+qiSaS+a كَتَبَ الأَوْلَادُ الْقِصَصَ
wrote+ 3MP the+boys+Nom the+stories+Acc
'The boys wrote the stories'

ونتيجة لذلك ، هناك ثلاثة أنواع من البنى الفعلية عندما يتعلق الأمر بكيفية التعبير عن الفاعل في اللغة العربية وهي : فعل - فاعل (Verb-Subject) ، و فاعل - فعل (Subject-Verb) ، و فعل + فاعل (Verb+Subj) .

يمكن الإشارة للفاعل في الجملة الاسمية الرئيسية بواسطة مواضيع وملاحق أخرى بداخل الجملة الفعلية الإسنادية ، مثل المفعول به في الجملة الفعلية أو المفعول به لأحد حروفه للجر.

(٦.١٧) جملة اسمية : مبتدأ_١ مرفوع [جملة فعلية فعل فاعل مرفوع مفعول به_١ خبر

(Subject₁^{Nom} [V-Sent Verb Subject^{Nom} Object₁] Predicate)

Al+kAtib+u *katab+a* *-hu* *Al+kitAb+u* الكِتَابُ كَتَبَ -هُ الكَاتِبُ
 the+book+Nom wrote+3MP -it
 the+author+Nom
 'The author wrote the book' (lit. 'The book, the author wrote it')

(٦, ١٨) جملة اسمية : مبتدأ^{مرفوع} [جملة فعلية فعل فاعل^{مرفوع} جار ومجرور^١] خبر

(Subject₁^{Nom} [V-Sent Verb Subject^{Nom} Prep Object₁]^{Predicate})

hāḍa *Al+bayt+u* *katab+a* *Al+kitAb+u* *ṣan* *-hu* هذا البيتُ كَتَبَ
 this the+house write the+author about -it
 'The author wrote about this house' (lit. 'This house, the author wrote about it')

لاحظ في المثال السابق ، أن المبتدأ في الجملة الاسمية العليا ، أي الموضوع ، حالته الإعرابية مرفوعة بصرف النظر عن إحالته الضميرية بداخل الجملة الفعلية. تسمح اللغة العربية بأشكال مختلفة لبنية الجملة في المثال السابق حيث جعل المفعول به هو الموضوع ، أي حُرِّك من دون التأثير في الحالة الإعرابية. في هذه البنية لا توجد حاجة إلى إحالة ضميرية في الجملة الفعلية. وهذه البنية غير شائعة.

(٦, ١٩) جملة فعلية : مفعول به^{منصوب} فعل فاعل^{مرفوع}

(Object^{Acc} Verb Subject^{Nom})

Al+kitAb+a *katab+a* *Al+kAtibu* الكِتَابُ كَتَبَ الكَاتِبُ
 the+book+Acc wrote+3MP the+author+Nom
 'The author wrote the book'

ملاحظة أخيرة ، إذا كان الفاعل غير معرف (نكرة) فإن ترتيب المبتدأ والخبر يُعكس. ويحصل هذا غالباً مع العبارات الجرّية الإسنادية.

(٦.٢٠) جملة اسمية : جار ومجرور - خبر مبتدأ مرفوع

(PP-Predicate Subject^{Nom})

īy Albayti

ī jul+ū

في البيت رجلٌ

in the+house+Gen man+InDef+Nom

'There is a man in the house' (lit. 'a man is in the house')

ç ind-y kitAb+ū

at -me book+InDef+Nom

عند-ي كتابٌ

'I have a book' (lit. 'a book is at me')

٦, ١, ٣ تركيبية العبارات الاسمية

أبسط العبارات الاسمية (N-Phrase) عبارة عن اسم أو صفة بأداة التعريف أو

بدونه.

(٦.٢١) عبارة اسمية : اسم (Noun)

kitAb+ū

كتابٌ

book+InDef

'a book'

(٦.٢٢) عبارة اسمية : أداة التعريف + اسم (DET+Noun)

AL+kitAb+u

الكتابُ

the+book+Def

'the book'

سنميز فيما يلي عدة أنواع من المعدلات الاسمية.

التعديلات الوصفية

تتبع الصفات في العربية الأسماء التي تقوم بتعديلها. ودائماً ما يتفق الاسم مع الصفة في التعريف والحالة الإعرابية. الصفات للأسماء العاقلة تتفق أيضاً في الجنس والعدد. أما صفات جمع التكسير فهي مفردة شكلياً وبصيغة جنس مخصص (ad hoc form-based gender)، لكنها وظيفياً جمع (انظر قسم ٤.٢.٢). على سبيل المثال كلمة المَهْرَةُ (Al+maharaḥ+u) مفرد مؤنث بالشكل لكنها وظيفياً مذكر وجمع.

(٦,٢٣) عبارة اسمية: اسم-عاقلة صفة (RATIONAL-NOUN ADJECTIVE)

kAtibū mAhirī أ. كَاتِبٌ مَاهِرٌ
author+MS clever+MS
'a clever author'

kAtibaḥū mAhiḥāḥū ب. كَاتِبَةٌ مَاهِرَةٌ
author+FS clever+FS
'a clever author'

kAtibAni mAhiAni ج. كَاتِبَانِ مَاهِرَانِ
author+MS clever+MS
'two clever authors'

AiḳAtibAtu AlmAhiAtu د. الْكَاتِبَاتُ الْمَاهِرَاتُ
the+authors+FP the+clever+FP
'the clever authors'

Aiḳut~Abu Almaharaḥu هـ. الْكُتَّابُ الْمَهْرَةُ
the+authors+MP the+clever+MP
'the clever authors'

Aiḳut~Abu AlmAhiḥuwna و. الْكُتَّابُ الْمَاهِرُونَ
the+authors+MP the+clever+MP
'the clever authors'

بينما الصفات للأسماء غير العاقلة تتفق مع الأسماء في الجنس والعدد عندما تكون الأسماء مفردة أو مثنى ، والغريب أن الصفات لجمع الأسماء غير العاقلة تكون مفرداً مؤنثاً.

(IRRATIONAL-NOUN ADJECTIVE) عبارة اسمية: اسم-غير العاقل صفة (٦,٢٤)

maktabū jādīdū أ. مَكْتَبٌ جَدِيدٌ
office+MS new+MS
'a new office'

maktabahū jādīdahū ب. مَكْتَبَةٌ جَدِيدَةٌ
library+FS new+FS
'a new library'

maktbAni ađiyaAni ج. مَكْتَبَانِ جَدِيدَانِ
offices+MD new+MD
'two new offices'

makAtibu jādīdahū د. مَكَاتِبُ جَدِيدَةٌ
offices+MP new+FS
'new offices'

maktbAtū jādīdahū هـ. مَكْتَبَاتٌ جَدِيدَةٌ
libraries+FP new+FS
'new libraries'

بُنية الإضافة

تعتبر بنية الإضافة بنية ملكية/مجرورة تربط بين اسمين: الاسم الأول المضاف الذي يت رأس الجملة نحويًا ويملك دلاليًا الاسم الثاني الذي يسمى المضاف إليه. يكون المضاف في حالة المبني والمضاف إليه في حالة الجر. ولهذا البناء العديد من المماثلات في

اللغة الإنجليزية مثل : اسم_١ اسم_٢ (Noun₁ Noun₂) يمكن ترجمته إلى اسم_١ للاسم_٢ (Noun₁ of Noun₂) واسم_٢ للاسم_١ (Noun₂ 's Noun₁) أو كمركب اسم_٢ اسم_١ (Noun₂ (Noun₁)).

(٦.٢٥) عبارة اسمية : اسم_١ مبنى اسم_٢ مجرور (NOUN₁ construct NOUN₂ Gen)

maĀtiyHu Al+say~Araħi مفاتيحُ السَّيَّارَةِ

keys the+car

'the keys of the car' or

'the car's keys' or 'the car keys'

يشكل الاسمان معاً عبارة اسمية يمكن أن تكون الجزء الثاني من بنية إضافة مختلفة. ويمكن لذلك أن يمتد بشكل متكرر لخلق ما يسمى بسلسلة الإضافة. كل الكلمات في سلسلة الإضافة تكون مجرورة إلا الكلمة الأولى. وجميع الكلمات ما عدا الكلمة الأخيرة لا بد أن تكون في حالة المبني.

(٦.٢٦) عبارة اسمية : سلسلة إضافة (Idafa Chain)

إِبْنُ عَمِّ جَارٍ رَئِيسِ مَجْلِسِ إِدَارَةِ الشَّرِكَةِ

Aibonu Eam-i jAri raġiysi maġlisi ĀidAraħi Alš-arikahī
son uncle neighbor chief committee management the-company
'the cousin of the CEO's neighbor'

والصفات التي تعدل رأس بنية الإضافة تتفق معها في الحالة الإعرابية ، لكنها تتفق معها بناء على تعريفها.

(٦.٢٧) عبارة اسمية : اسم_١ مبنى اسم_٢ مجرور صفة_١ (ADJ1 NOUN₁ construct)

(NOUN₂ Gen)

bAb+u Al+say~Araħi Al,jađiyd+u بَابُ السَّيَّارَةِ الْجَدِيدُ

door the+car+GEN the+new+NOM

'the car's new door'

bAb+u say~Arāḥi jadīyd+ū
 door car+GEN new+NOM
 'a car's new door'

بَابُ سَيَّارَةٍ جَدِيدٍ

بالإضافة إلى تراكيب الإضافة، نستخدم بنية الإضافة في العديد من البنى اللغوية في العربية مثل:

- البنى الكمية مثال ذلك كل الكتب (*kuṭ-u Alkutubī*) وخمسة كُتُب (*xamsaḥu kutubī*).
- البنى الظرفية الشبيهة بحروف الجر مثال ذلك قرب البيت (*qurba Albayti*).
- الإضافة الوصفية أيضاً تسمى إضافة غير حقيقية مثل طويل القامة (*AlqAmāhiTawiyli*).

بنية التمييز

ترتبط بنية التمييز بين اسمين. الاسم الأول يطلق عليه المميّز (the specified) يترأس ويحكم الاسم الثاني الذي يطلق عليه المميّز (the specifier) والذي يحدد الاسم الأول. المميّز دائماً ما يكون مفرداً في العدد وحالته الإعرابية منصوب. ويستخدم التمييز في مجموعة متنوعة من البنى اللغوية في اللغة العربية:

- المقارنة والتفضيل مثل: أكثرُ بياضاً (*Akṭaru bayADAā*).
- محدد للمقاييس مثل: كيلو لحماً (*kylwlaHmAā*)، أو الاستفهام الشائع: كم كتاباً؟ (*kam kitAbAā*).
- بنية بعض الأعداد مثل: خمسون كتاباً (*xamsuwna kitAbAā*).
- محدد للنوع مثل: خاتمٌ فضةً (*xAtinū tiD-ahā*).

بُنية البدل

ترتبط بنية البدل (apposition construction) بين عبارتين اسميتين تشير لنفس الكيان. رأس العبارتين الاسميتين تتفقان في الحالة الإعرابية، مثال ذلك الرئيس الأمريكي، باراك أوباما (Alḥiysu AlĀmriykiy~u, bAlAk AwbAn). ومن أبنية البدل الشائعة في العربية اسم الإشارة (demonstrative pronoun)، الذي يسبق عادة الاسم الذي يعدله على الرغم من أنه يمكن أن يلحقه أيضاً، مثال: هذا الكتاب (hāḏA Alkitāḏ).

عبارات الوصل

تعديل عبارات الوصل (Relative Clauses) الاسم الذي يترأسها. فإذا كان الاسم في رأس الجملة معرفة، فإن جملة الوصل تقدم وتسبق باسم موصول. عندما تكون موجودة، فإن الاسم الموصول يتفق مع الاسم الذي يعدله في الجنس والعدد وذلك باتباع قواعد توافق الصفة (الأسماء غير العاقلة تحصل على توافق استثنائي).

(٦.٢٨) عبارة اسمية : اسم معرفة اسم موصول جملة (NOUN^{def inite} RELPRO)

(SENTENCE)

الكتابُ الذي أُحِبُّ - هُ
 AkitAbu Al-ḏiy [ĀuHib~u -hu]
 the-book which [I+love -it]
 'the book [which] I love'

إذا كان الاسم في المقدمة نكرة، فإن عبارة الوصل (تسمى في هذه الحالة جملة صفة) لا تقدم مع اسم موصول.

(٦.٢٩) عبارة اسمية : اسم نكرة جملة (NOUN^{indef inite} SENTENCE)

كتابُ أُحِبُّ - هُ
 kitAbū [ĀuHib~u -hu]
 book [I+love -it]
 'a book [which] I love'

عبارة الوصل المعرفة المترأسة باسم موصول يمكن أن تكون بمفردها في العربية كعبارة اسمية.

(RELPRO SENTENCE) عبارة اسمية : اسم موصول جملة (٦.٣٠)

Al-ḍiy [*ĀuHib-u -hu*] الذي أُحِبُّ - هُ
 who/which [I+love -it]
 'the one/thing [whom/which] I love'

الموضوعات الاسمية

أسماء الأفعال في العربية مثل المصدر (deverbal nouns) واسم الفاعل (active participles) تتصرف مثل الأفعال بحيث يمكنها أن تأخذ كمحدد مفعولاً به منصوب ومعدلات فعلية أخرى. يسمح شكلها الاسمي بالمشاركة أيضاً في بعض البنى الاسمية التي ناقشناها سابقاً مثل الإضافة.

(٦.٣١) عبارة اسمية : مصدر مبني اسم مجرور عبارة اسمية - مفعول به منصوب

(MASDAR^{construct} NOUN₁^{Gen} NP-OBJ^{Acc})

maṣūfāhu *Al-juū* *AlHaqiyqaḥa* مَعْرِفَةُ الرَّجُلِ الْحَقِيقَةَ
 knowing+Nom the+man+Gen the+truth+Acc
 'the man's knowledge of the truth'

٤, ١, ٦ العبارات الجرّية

تتكون العبارات الجرّية في العربية من حرف جرّ يتبعه جملة اسمية. ورأس الجملة الاسمية في حالة جر.

(٦.٣٢) جملة جر: حرف جر اسم مجرور (PREP NOUN^{Gen})

liy *Al+bayt+i* في البيت
 in the-house
 'in the house'

٦,٢ البنوك الشجرية العربية

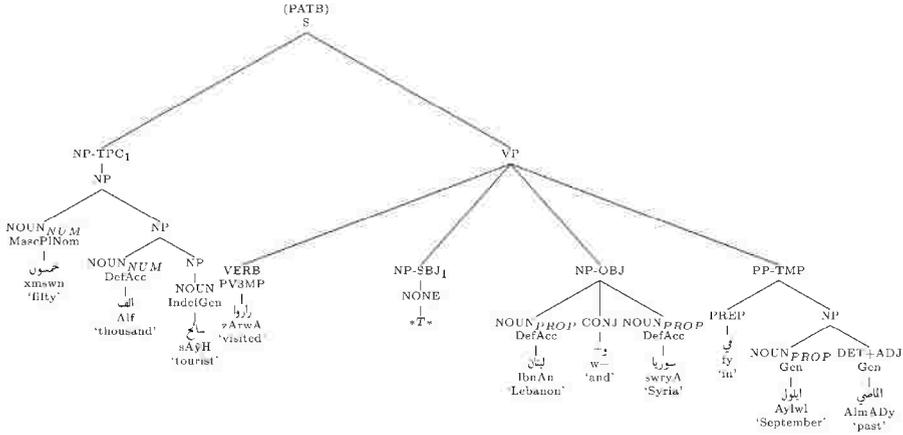
تمثل مجموعة من جمل التحليلات النحوية المراجعة يدوياً، أو ما تسمى البنوك الشجرية، مصدراً مهماً لبناء المحللات الإحصائية وتقييمها عامةً. فقد استخدمت البنوك الشجرية الثرية بالحواشي في العديد من التطبيقات مثل تقطيع النص، والتشكيل، وتعيين أقسام الكلام، وفك الغموض الصرفي، ومقطع العبارات الأساسية، ووسم الأدوار الدلالية. وتحت قيود الوقت، فإن إنشاء بنك شجري يواجه مفاضلة بين الثراء اللغوي وحجم البنك الشجري. وهذا هو الحال خاصة للغات المعقدة صرفاً - نحوياً مثل العربية والتشيفية. يقدم تمثيل الثراء اللغوي العديد من (جميع) الخصائص اللغوية التي قد تكون مفيدة لمجموعة متنوعة من التطبيقات. يأتي هذا على حساب بطء عملية التحشية وذلك نتيجة لطول الإرشادات والتدريب المكثف للحشّاء (annotator)^(١). وكنتيجة لذلك، كلما كانت التحشية ثرية، فإن عملية التحشية تكون أبطأ وأيضاً حجم البنك الشجري يكون أصغر. ومن ثم، هناك بيانات أقل لتدريب الأدوات.

في حالة اللغة العربية، يوجد هناك مجهودان مهمان لتحشية ثرية لبنوك شجرية هما: بنك بنسلفانيا الشجري للتحليل النحوية PATB [9, 136, 82] وبنك براغ للتحليل الشجرية النحوية التبعية للغة العربية PADT [137,138]. ويوظف كلا المجهودين تمثيلات لغوية معقدة وغنية جداً تتطلب الكثير من التدريب البشري. أما عن كمية التفاصيل المحددة في التمثيلات فهي مثيرة للإعجاب. إن بنك بنسلفانيا الشجري للتحليل النحوية لا يقدم فقط تقطيعاً للنص، وسمات معقدة لأقسام الكلام، والتركيب النحوي، بل يقدم أيضاً تصنيفات فارغة وتشكيل واختيار المادة المعجمية

(١) يقصد بها الشخص الذي يقوم بعمليات التحشية (توضيح من المترجمة).

وبعض السمات الدلالية. وتسمح هذه المعلومات بالبحث في التطبيقات العامة لمعالجة اللغة الطبيعية، غير أن الكثير من هذه الحواشي غير مستخدم حالياً في أبحاث التحليل العربي [119] حيث يعتبر عادة مشتقاً من نتائج التحليل نفسه. على سبيل المثال، الحالة الاسمية، التي يمكن تحديدها كمعيار ذهبي للتحليل النحوي وبدقة عالية [9]، لا يمكن التنبؤ بها جيداً في خطوة ما قبل تحليل تعيين أقسام الكلام [116، 135]. لمعالجة هذه المشكلة، قدم مؤخراً بنك تحاليل نحوية ثالث (بنك كولومبيا الشجري للتحاليل النحوية الشجرية للغة العربية (Columbia Arabic Treebank (CATiB) بهدف تسريع عملية التحشية عن طريق تبسيط التمثيل [112، 121].

في هذا القسم، سنناقش بإيجاز كلاً من هذه الموارد الثلاثة ونقارنها ببعضها.



الشكل رقم (٦، ١). التمثيل النحوي لجملة (خمسون ألف سائح زاروا لبنان وسوريا في أيلول الماضي)

(xmswn Alf sA 'yH zArwA IbnAn wswryA fy Aylwl AlmADy) في بنك بنسلفانيا

الشجري للتحاليل النحوية.

٦,٢,١ بنك بنسلفانيا الشجري للتحليل النحوية

بدأ مشروع بنك بنسلفانيا الشجري للتحليل النحوية (PATB)^(١) عام ٢٠٠١م في مركز اتحاد البيانات اللغوية (Linguistic Data Consortium) وجامعة بنسلفانيا (مهد بنك التحليل النحوية للغة الإنجليزية والصينية والكورية) [139، 9، 71، 47، 136، 82]. وحتى وقت كتابة هذا الكتاب، نُشرت ثلاثة أجزاء للعام من بنك بنسلفانيا الشجري للتحليل النحوية عن طريق مركز اتحاد البيانات اللغوية (قراية ٦٥٠ ألف كلمة) وأربعة أجزاء أخرى، بما فيها بنك التحليل النحوية للعربية الشامية [140]، التي طورت للمشروع الممول من (DARPA)^(٢). كل جزء من بنك بنسلفانيا الشجري للتحليل النحوية صُدر في إصدارات مختلفة وبدرجات مختلفة من التحسينات.

ويحشى بنك بنسلفانيا الشجري للتحليل النحوية بالمعلومات النحوية، والمقابل بالإنجليزية، والتراكيب النحوية في نمط (جملة - تركيب) في بنك بنسلفانيا الشجري للتحليل النحوية للغة الإنجليزية (PTB) [118]. وقد عُدلت إرشادات (PTB) للتعامل مع اللغة العربية. على سبيل المثال، الفاعل بالعربية يُحلل داخلياً كعبارة فعلية (VP) التي تتبع الفعل.

يعرض الشكل رقم ٦.١ مثلاً على شجرة في بنك بنسلفانيا الشجري للتحليل النحوية. للتمثيل الفعلي للأشجار في بنك بنسلفانيا الشجري للتحليل النحوية، انظر الشكل رقم ٦.٤. وتُوسع وسوم تركيب الجملة في بنك بنسلفانيا الشجري للتحليل النحوية بسمات الشرطة (dash tags) التي تحدد الأدوار النحوية والدلالية مثل: الموضوع (TPC) والفاعل (SBJ) والزمن (TMP). وبعض العبارات يتم مشاركة

(١) لكونه أول بنك تحاليل نحوية فإن بنك بنسلفانيا الشجري للتحليل النحوية في الغالب يشار إليه

بالبنك الشجري العربي (Arabic Treebank (ATB).

(٢) يمكن إتاحة بعض من هذه الموارد لعمامة الناس في المستقبل.

فهرستها باستخدام امتداد رقم الفهرس لوسم العبارة. وتمثل الشجرة كاملة جملة (وسمها S) تحتوي على عبارتين: عبارة اسمية (NP) تعمل كموضوع (NP-TPC) وعبارة فعلية (VP). يتكون الموضوع من رقم يتبعه جملة اسمية (NP)، وهو بدوره يتكون من رقم يتبعه جملة اسمية (NP). وأعمق جملة اسمية تحتوي على اسم مفرد. ويستخدم (PATB) الإعدادات التالية (NP NOUN NP) لعمل بنية الإضافة والتمييز. ولا يمكن تمييز البنيتين إلا باستخدام الحالة الصرفية للاسم في الجملة الاسمية المضمنة. تحتوي العبارة الفعلية على فاعل (NP-SBJ) ومفعول به (NP-OBJ) وعبارة جريّة زمنية (PP-TMP). الفاعل في هذه الحالة يعتبر تصنيفاً فارغاً (ضميراً مستتراً) مشيراً إلى أن الفاعل عبارة عن ضمير ومضاف لشكل الفعل. أيضاً فإن الفاعل يشارك في الفهرسة مع الموضوع، مشيراً بذلك إلى أن الموضوع في البنية هو نفسه الفاعل للفعل. بهذه الطريقة تمثل (PATB) الجملة الاسمية بوجود جمل فعلية خبرية. يتكون المفعول به من اسمي علم تربط بالرابط "w+ +". وعبارة جريّة (PP) تحتوي على حرف جر (PREP) يتبعها مفعول به مكونة من جملة اسمية بحرف جر. تتكون الجملة الاسمية من اسم علم معدلة من قبل صفة. وتتم الإشارة إلى الصرف الشكلي الكامل لجميع الكلمات في الشجرة. والوصف الكامل لبنك بنسلفانيا الشجري للتحاليل النحوية (PATB) موجود كجزء من دليل التحشية في (PATB) [82]. معظم الجمل في (PATB) تُرجمت للإنجليزية أو لها ترجمة مسبقة مرتبطة بها. كما أن بعضها حُللت نحويّاً للإنجليزية، مكونة بذلك مورداً فريداً للتحاليل النحوية الموازية.

يعتبر إنشاء (PATB) إنجازاً كبيراً لمجال معالجة اللغة العربية، فقد أصبح مورداً مهماً للعديد من الأبحاث في التحليل الصرفي، وفك الغموض، وتعيين أقسام الكلام، والتقطيع، ولا ننسى بالطبع التحليل. وكل البنوك الشجرية التي أنشئت بعد (PATB)

قامت باستخدامه أو طورت بعض الأدوات له. على سبيل المثال ، فإن كلا من بنك براغ للتحليل الشجرية النحوية التبعية للغة العربية وبنك كولومبيا الشجري للتحليل النحوية الشجرية للغة العربية قاما بتحويل (PATB) لتمثيلاتها الخاصة بالإضافة إلى تحشية معلومات إضافية. ومثال آخر على أهمية (PATB) هو أن طريقته في التقطيع أصبحت بحكم الواقع المعيار المطبق لمعظم الجهود في بناء البنوك الشجرية.

٢,٢,٦ بنك براغ للتحليل الشجرية النحوية التبعية للغة العربية

يدار بنك براغ للتحليل الشجرية النحوية التبعية للغة العربية (PADT) بواسطة معهد اللغويات المقننة والتطبيقية (Institute of Formal and Applied Linguistics) بجامعة تشارلز في براغ. تحتوي (PADT) على وصف متعدد المستويات يضم الصرف الوظيفي ، والنحو التبعي التحليلي ، وتمثيل التكتيكات النحوية (tectogrammatical) للمعاني اللغوية. وتعتمد هذه الحواشي اللغوية على نظرية الوصف التوليدي الوظيفي (Functional Generative Description) [141] ومشروع شجرة براغ التبعية للغة العربية [142].

تختلف التحشية النحوية والصرفية في (PADT) كلية عن بنك بنسلفانيا الشجري للتحليل النحوية. فحواشي أقسام الكلام ممثلة بمجموعة سمات صرفية وظيفية مطورة كجزء من ElixirFM [143] (انظر قسم ٥.٢.٥). فالتحشية النحوية تمثل بتركيب تبعي معين وبمستويين من المعلومات : التحليلي والتكتيك النحوي. ولن نناقش التمثيل التكتيكي النحوي هنا.

الوصفية ، فهي تسمى جميعاً علاقات وصفية (Atr). الابن الثاني للفعل يترأس اسمي علم بعلاقة مركبة (Obj_Co) مما يدل مباشرة على أن اسمي العلم تم التنسيق بينهما (Co) بواسطة الأب وكلاهما مفعول به (Obj) للفعل. الابن الأخير للفعل وهو حرف الجر "fy" الذي يترأس اسم العلم "ايلول/Ay/wl" مع العلاقة الظرفية (Adv) والذي بدوره يترأس الصفة في العلاقة الوصفية (Atr). تبين العلاقة الظرفية كيف أن اسم الشهر يعدل على الفعل الرئيسي على الرغم من ظهور حرف الجر بين الاثنين. وهذا يسلط الضوء على جانب مهم من جوانب تمثيل التحليل النحوي في (PADT) ، وهو أنها أعمق وأكثر دراية دلالياً (تحديداً خبيراً/حملكياً) من محملات نحوية أخرى. وللمزيد من المعلومات عن (PADT) ، انظر [144 ، 145 ، 143 ، 67 ، 138].

احتوت النسخة الأولية من (PADT) [145] على قرابة مائة ألف كلمة. وقد استخدمت (PADT) في المهام المشتركة لكل من (CoNLL 2006) و (CoNLL 2007) في التحليل التبعي [146] وبياناته الصرفية استخدمت لتدريب المعنونات الأتوماتيكية [114]. والنسخة الحالية من (PADT 2.0) تحتوي على قرابة المليون وحدة لفظية من شجرات (PATB) المحولة وشجرات نحوية محشاة بـ (PADT) مباشرة.

٦،٢،٣ بنك كولومبيا الشجري للتحاليل النحوية الشجرية للغة العربية

بدأ مشروع جامعة كولومبيا للبنك الشجري للتحاليل النحوية الشجرية للغة العربية (كاتب) (CATiB) عام ٢٠٠٨م. ويتناقض مع سابقه من طرق مشاريع البنوك الشجرية في محاولته لجعل التركيز على الإنتاجية الأسرع مع وجود بعض القيود على الشراء اللغوي [112 ، 121]. هناك فكرتان أهما منهجية كاتب. الأولى ، يتجنب كاتب تحشية المعلومات اللغوية المكررة. على سبيل المثال ، الحالة الاسمية والظرفية في العربية

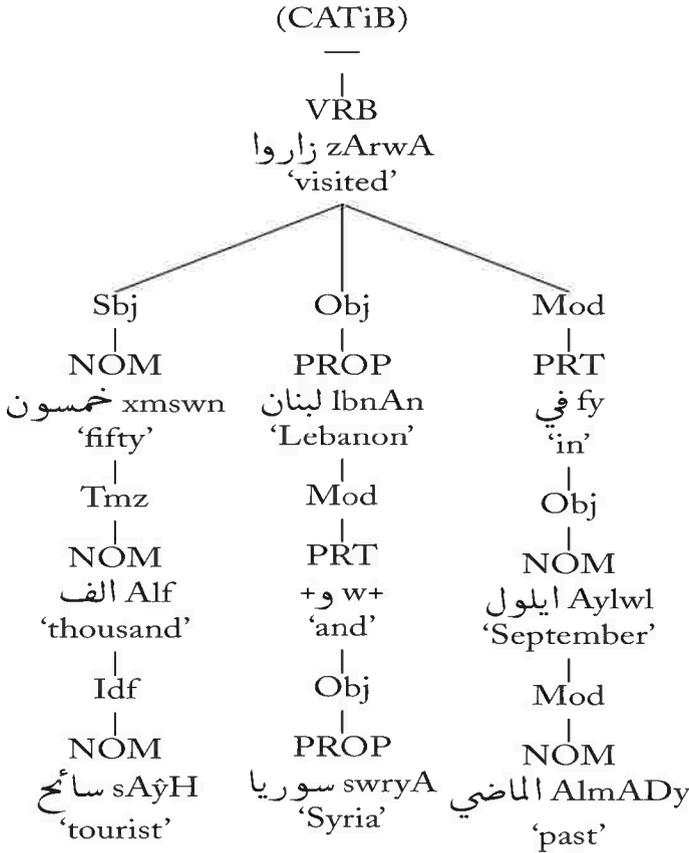
تحدد تلقائياً من التحليل النحوي والصرفي للكلمات ولا تحتاج أن تكون محشاة بشرياً. وبالطبع ، ليس من السهل استرجاع بعض المعلومات في كاتب مثل التبويب العباري (phrasal co-indexation) وفك الغموض الكامل للمادة المعجمية. الثانية ، يستخدم كاتب تمثيلاً بديهيّاً لتركيب التبعية والوسوم العلائقية (relational labels) مستوحاة من قواعد اللغة العربية التقليدية مثل التمييز والإضافة ، بالإضافة إلى الوسوم المعروفة للفاعل والمفعول به والمميز. مما يجعله من السهل تدريب الشخص الذي يقوم بوضع الحواشي والذي لا يحتاج لشهادة في اللغويات.

هناك ثماني علاقات نحوية مستخدمة لوسم المكملات التابعة (dependency attachments) في كاتب ، وهي : فاعل (SBJ) ومفعول به (OBJ) وخبر (PRD) وموضوع (TPC) وإضافة (IDF) وتمييز (TMZ) ومميز (MOD) والساكن (—). يمثل (SBJ) الفاعل النحوي الصريح للأفعال (المعلوم أو المجهول) ، بغض النظر عما إذا كانت تظهر قبل أو بعد الفعل والفاعل للجمل الاسمية. يقتصر (TPC) على المبتدأ لجملة اسمية معقدة يكون مكملها فعل مع فاعل مختلف. عادة هناك ضمير مفعول به يشير إلى الموضوع. إن استخدام (SBJ) و (TPC) في كاتب يختلف عن (PATB). تعتبر (MOD) العلاقة الأكثر شيوعاً التي تستخدم لتحديد جميع التعديلات الوصفية مثل تعديل الأسماء وتعديل الظرف وتعديل العبارات الجريّة للأسماء والأفعال. وتمثل العلاقة الساكنة تركيبية متعددة للكلمة التي لا يمكن تفسيرها باستخدام أي من العلاقات السابقة. وتعتبر الأجزاء المختلفة من اسم العلم الحالة الأكثر شيوعاً ، على سبيل المثال ، الاسم الأخير في علاقة ساكنة مع الاسم الأول.

يتضمن كاتب حوالي مليون وحدة لفظية (Token) مقسمة إلى : ٢٧٠ ألف وحدة لفظية لنصوص وكالات الأنباء محشاة ، بالإضافة إلى شجرات (PATB) محولة

(الأجزاء ١ و٢ و٣). وبما أن (PATB) يحوي معلومات أكثر، فإن التحويل إلى كاتب يعتبر مجد بدرجة جيدة من الصحة [112]. وتكون جميع جمل كاتب المحشة مأخوذة من مدونة عربي-إنجليزي متوازية وبذلك فإن الجمل لها ترجمات مرتبطة بها.

يستعرض الشكل رقم ٦,٣ مثلاً لجملة في كاتب. للتنسيق الفعلي لشجرات كاتب، انظر الشكل رقم ٦,٤. إلى حد ما، فإن تمثيل التبعية ماثلة لتلك المستخدمة في (PADT) ولكن مع بعض الاختلافات الهامة جداً (التي سناقشها في القسم التالي). رأس الجملة هو الفعل زاروا (zarwa) ولديه ثلاثة أبناء هم: الفاعل (SBJ) والمفعول به (OBJ) والمعدل الخبري (MOD). يحتوي الفاعل على تعبير عددي معقد يحتوي على علاقات إضافة وتمييز. ويترأس المفعول به أداة حرف عطف يرأس بدوره المشتقات الوصفية. الابن الثالث للفعل هو حرف الجر ويحكم المفعول به (OBJ) الذي بنفسه معدل بواسطة جملة اسمية وصفية. هذه البساطة والطبيعة غير المشذبة للعلاقات المستخدمة هي علامة تفرق بين تحشية كاتب مقارنة بطرق لبنوك شجرية أخرى.



الشكل رقم (٦,٣). التمثيل التبعي في كاتب للجملة (خمسون الف سائح زاروا لبنان وسوريا في ايلول الماضي) (xmswn Alf sAÿH zArwA lbnAn wswryA fy Aylwl AlmADy) .

٦,٢,٤ مقارنة: PATB و PADT و CATIB

عند مقارنة كاتب و (PADT) و (PATB) يمكننا التمييز بين جانبيين عاميين رئيسيين هما: التمثيل النحوي والمحتوى اللغوي. من حيث التمثيل النحوي، فإن (PATB) يستخدم تركيب العبارات (PS) بينما كل من كاتب و (PADT) يستخدمان التركيب التبعي (DS). انظر للأشكال رقم ٦.١ و ٦.٢ و ٦.٣. تركيب العبارات (PS) هو تمثيل

شجري يظهر الكلمات في الجملة كأوراق والعقد الداخلية تصنف نحوياً على أنها عبارات اسمية (NP) أو فعلية (VP). التركيب التبعي يعتبر أيضاً شجرة إلا أن الكلمات في الجملة هي العقد في الشجرة [147]. ومن حيث المحتوى اللغوي، يمكننا أن نميز المزيد من الفئات التالية من المحتوى.

التركيب النحوي: يقوم كاتب و (PADT) بتحشية الرأس صريحاً وامتدادات العبارات ضمناً، بينما يقوم (PATB) بتحشية الامتدادات بشكل صريح والرأس بشكل ضمني. ويستخدم (PATB) الإسقاطات الوسيطة، مثل العبارات الفعلية، لتمثيل حقائق نحوية معينة. تستخدم البنوك الشجرية ذات التركيب التبعي وكاتب و (PADT) وسائل أخرى مثل الوسوم المرفقة لتمثيل نفس الحقائق. يقترب كاتب و (PADT) من بعض التراكيب بطريقة مختلفة. على سبيل المثال، في (PADT) تترأس حروف العطف على العناصر المختلفة التي تربطها على عكس الطريقة التي نفذت في كاتب. انظر كيف مثلت "لبنان و+سوريا" (*IbnAn w+ swryA*) في (PADT) وكاتب في الشكلين رقمي ٦.٢ و ٦.٣.

الوظائف النحوية والدلالية: يستخدم (PADT) حوالي عشرين (dashtags) لوضع علامات على الوظائف النحوية والدلالية. تحتوي وسم الشرطة (dashtags) النحوية على (TPC-) و (OBJ-) أما السمات الدلالية فتحتوي على الزمن (TMP-) والمكان (LOC-). بعض وسم الشرطة تخدم غرضاً مزدوجاً، دلالي/نحوي، مثل: (SBJ-) التي يمكن أن تعلم الفاعل النحوي للفعل والفاعل الدلالي للأسماء المشتقة من الأفعال.

مثال من بنك بنسلفانيا الشجري (Penn Arabic Treebank)				
(S (NP-TPC-1 (NOUN_NUM+NSUFF_MASC_PL_NOM xmswn) (NP (NOUN_NUM+CASE_DEF_ACC Alf) (NP (NOUN+CASE_INDEF_GEN sA}H))))))				
(VP (PV+PVSUFF_SUBJ:3MP zArwA) (NP-SBJ-1 (-NONE- *T*)) (NP-OBJ (NOUN_PROP+CASE_DEF_ACC lbnAn) (CONJ w+) (NOUN_PROP+CASE_DEF_ACC swryA)) (PP-TMP (PREP fy) (NP NOUN_PROP+CASE_DEF_GEN Aylwl (DET+ADJ+CASE_DEF_GEN AlmADy))))))				
مثال من بنك كولومبيا الشجري (Columbia Arabic Treebank)				
1	xmswn	NOM	4	SBJ
2	Alf	NOM	1	IDF
3	sA}H	NOM	2	IDF
4	zArwA	VRB	0	---
5	lbnAn	PROP	4	OBJ
6	w+	PRT	5	MOD
7	swryA	PROP	6	OBJ
8	fy	PRT	4	MOD
9	Aylwl	PROP	8	OBJ
10	AlmADy	NOM	9	MOD

الشكل رقم (٤، ٦). التمثيل الداخلي لشجرة نحوية في تركيب العبارة (تحديداً تمثيل إصدارة (PATB) والتركيب التبعي (تحديداً تمثيل إصدارة كاتب). وترتبط هذه الأمثلة مع الأمثلة في الأشكال رقم ٦، ١ و ٦، ٣. وتطبع عادة أشجار (PATB) على سطر واحد. أما أشجار كاتب فتتمثل في خمسة أعمدة تشير إلى مؤشر الكلمة (word index)، والكلمة (word)، وأقسام الكلام (POS)، والمؤشر الرئيسي للكلمة (parent word index)، والعلاقة (relation). وتمثل الكلمات العربية بمخطط بكوالتر للنقل الكتابي [88]. وتعتبر جميع المقابلات الإنجليزية إضافية.

لا يعمل (PATB) على تحشية الـ (dashtags) صريحة ففي بعض الحالات مثل جار المفعول أو بنية التمييز/الإضافة. وتُميز ضمناً من خلال تراكيب نحوية. التمييز والإضافة متطابقان في (PATB) باستثناء معلومات الحالة الصرفية التي يمكن استخدامها للتمييز بينها. وتعلم وسوم العلاقات في كاتب الوظيفة النحوية فقط. واستخدام الوسوم النحوية (SBJ) و (TPC) مختلفة بين كاتب و (PATB). ففي (PATB) يستخدم (TPC) لتعليم الفاعل أو المفعول به عندما يظهران قبل الفعل. كذلك يتم استخدام الفهرسة المشاركة (co-indexation) لتحديد دور (TPC) داخل العبارة الفعلية. انظر إلى كيف يتم معالجة الفاعل في الشكل رقم ٦.١ و ٦.٣. الفاعل الجملة اسمية (غير معقدة) يتم تعليمها بـ (SBJ) في كل من كاتب و (PATB). يستخدم (PADT) حوالي عشرين وسماً، على الرغم من وظائفها المختلفة عن (PATB) وكاتب. بشكل عام الوسوم التحليلية في (PADT) تكون أعمق من كاتب حيث إن الهدف منها أن تكون نقطة انطلاق نحو مستوى التكتيكات النحوية لـ (PADT). مثلاً، تُوضع علامة على الكلمات المعتمدة على حروف الجر مع علاقتها بالعقدة التي تحكم حرف الجر (العقدة الجد). على سبيل المثال، في الشكل رقم ٦.٢ كلمة ايلول (Ay/w) تعلم كوصف (Adv) للفعل الرئيسي زاروا (zArwA). بالمثل، العناصر المرتبطة لبنان و+ سوريا (IbnAn w+ swryA) علمتا على أنهما مرتبطتان (Co) بعلاقة مفعول به (Obj) مع الفعل الحاكم. ولا يفرق (PADT) بين الأشكال المختلفة للمعدلات الاسمية، أي أن الصفات والإضافة والتمييز (بالعدد) تعلم جميعاً كخصائص (Att).

الضمائر الفارغة (Empty Pronouns): في (PATB) الضمائر الفارغة محشاة، ولكن ليست في (PADT) أو كاتب. يمكن اعتبار الأفعال من دون فاعل صريح في كاتب و (PADT) ضميراً مستتراً (تحشية ضمنية).

الإحالة (Coreference): في (PATB) مؤشرات الإحالة محشاة، لتتبع الضمائر وإظهارها. يقوم (PADT) بتحشية الإحالة بين الضمائر الصريحة فقط وما يحال إليها. أما كاتب فلا يقوم بعمل تحشية لأي مؤشرات للإحالة.

صرف الكلمة: يستخدم كاتب المخطط الأساسي لتقطيع الكلمات المستخدمة في (PATB) و (PADT). أما بالنسبة لأقسام الكلام، فيستخدم (PATB) أكثر من ٤٠٠ سمة محددة لكل جانب من جوانب الصرف للكلمة العربية مثل: التعريف، والجنس، والعدد، والشخص، والإعراب، وصيغة بناء المجهول، والحالة الإعرابية. وصرف (PADT) أعقد بكثير من صرف (PATB). على سبيل المثال، فإنه يفرق بشكل أكثر تعقيداً بين الجمل الاسمية والوصفية المعرفة/المبينة، والعدد والجنس. على النقيض من ذلك، فإن كاتب يستخدم ست سمات فقط لأقسام الكلام. من المهم أن نشير إلى أنه في معظم أعمال تحليل اللغة العربية، تُستخدم مجموعة صغيرة من سمات أقسام الكلام، وذلك بتقليص ما يقارب من ٤٠٠ سمة في (PATB) إلى مجموعة بين ٢٠ و ٤٠ سمة [119]. يذكر في [110] أنه استخدم امتداداً يعتمد على التعبيرات المنتظمة (regular-expression) لمجموعة سمات كاتب ونتجت عنها نتائج تنافسية. ويمكن استرجاع بعض المعلومات الصرفية الغنية غير المدرجة في مجموعات سمات أقسام الكلام المخفضة، مثل الحالة الاسمية، من تركيبية الشجرة لأنها معرفة نحوياً [79].

على الرغم من وجود العديد من الاختلافات، يمكن القيام بالتحويل بين هذه التمثيلات المختلفة وبدرجة جيدة من النجاح باعتبار أن المعلومات متاحة في الشجرة على الرغم من تمثيلها بشكل مختلف. وبما أن كاتب يحتوي على معلومات أقل من (PATB) و (PADT)، قد يكون من الأسهل التحويل من هذين التمثيلين إلى كاتب وليس العكس.

٦,٢,٥ غابة من البنوك الشجرية

عموماً، هناك امتدادات عديدة للأعمال على البنوك الشجرية. نقدم هنا مؤشرات على الجهود المبذولة في البنوك الشجرية العربية والممتدة في ثلاثة أبعاد هي: النوع، والتمثيل، والعمق.

- قام الفريق وراء (PATB) بتمديد محتواها لتشمل أنواعاً إضافية مثل اللغة العربية المستخدمة في بث الأخبار والمحادثات، والمحادثات الهاتفية، والمدونات. وأحد الأمثلة على ذلك هو البنك الشجري الشامسي العربي (Levantine Arabic Treebank [140, 148].

- يحتوي مشروع مدونة القرآن الكريم في جامعة ليدز على جهود عمل بنك شجري موجه للقرآن الكريم. التمثيل المستخدم في هذا المشروع هو مزيج من تراكيب العبارة والتبعية وهي الأقرب إلى وصف قواعد اللغة العربية التقليدية [10، 149].

- هناك إصدار من (PATB) في جامعة دبلن (Dublin University) تستخدم تمثيل نحوي معجمي وظيفي (LFG) lexical functional grammar. وقد حُوّل البنك الشجري آلياً ولا يحوي على تحشيات إضافية [150].

- يستخدم بنك الأبنية الحملية (Propositional Bank) [151] ومشروع (OntoNotes) لتحشية المعلومات العربية دلاليًا. سنناقش موضوع بنك الأبنية الحملية في الفصل القادم.

٦,٣ التطبيقات النحوية

أكثر التطبيقات النحوية وضوحاً هو التحليل النحوي الذي يهدف إلى تعيين التركيب النحوي لسلسلة من الكلمات. وهذه التقنية يمكن استخدامها واستخدمت فعلاً في تطبيقات عليا متنوعة ومهمة لمعالجة اللغة الطبيعية مثل الترجمة الآلية والتلخيص الآلي. وبشكل عام، تم القيام بالعديد من الأعمال في مجال التحليل.

هناك العديد من المحللات العصرية التي استخدمت لتحليل العربية منها: محلل بايكل (Bikel parser) لتحليل تركيب العبارات [153، 119، 154]، محلل مالت (Malt parser) لتحليل التبعية [146، 112، 110]، ومحلل ستانفورد (Stanford parser [155، 156]، وغيرها. جميع المحللات التي ذكرت سابقاً بحاجة لبنك شجري. ولا تعتبر هذه مشكلة في اللغة العربية المعاصرة، إلا أن اللهجات العربية فقيرة من ناحية البنوك الشجرية. وفي ورشة صيفية قريبة بجامعة جون هوبكنز (Johns Hopkins) استعرض كيف يمكن للهجات العربية أن تحلل باستخدام موارد من اللغة العربية المعاصرة [148]. من بين هذه الجهود تحليل اللغة العربية المعاصرة بالاعتماد على القواعد، انظر [157]، [158].

وهناك عملية ذات علاقة بالتحليل النحوي ولكنها سطحية تدعى التقطيع للعبارات الأساسية (BPC). وهي العملية التي من خلالها يتم تجميع سلسلة من الكلمات المتجاورة معاً لتشكيل عبارات نحوية مثل عبارات اسمية (NPs) وعبارات فعلية (VPs [117، 120]). وقد درس باحثون مختلفون جدوى (BPC) لتطبيقات مختلفة مثل الترجمة الآلية [130] والتعرف على أسماء الأعلام (NER [129]).

٤, ٦ المزيد من القراءات

تواجه اللغة العربية تحديات مماثلة لغيرها من اللغات الغنية صرفياً عندما يتعلق الأمر بالتحليل. ومن ورش العمل القريبة جداً التي ركزت على مثل هذه التحديات وهي ورشة عمل حول التحليل الإحصائي للغات الغنية صرفياً (Workshop on Statistical Parsing of Morphologically Rich Languages (SPMRL – NAACL 2010) . في الملحق د، ذكرنا مصادر وأدوات أخرى ذات صلة بالنحو.