

قضايا بحثية 1 - التكاليف 5

Prepared by: **Abobakr Mohammed Baras**
ID: **253002656**

Nov 2025



الورقة البحثية المنتقدة

Enrichment of Arabic WordNet Using Machine Translation and Transformers

Link: <https://aclanthology.org/2023.icnlp-1.36.pdf>

Chapter
01

ملخص سريع للورقة

المخرجات:

- توليد 90,127 Synsets مرشحة
- دقة تحقق يدوي لعينة 1,000 Synsets
- تطوير منصة ويب للتحقق البشري

هدف الورقة:

- إغناء Arabic WordNet (AWN) اعتمادًا على:
 - Princeton WordNet (PWN)
 - الترجمة الآلية
 - نماذج Transformers و BERT/AraBERT

نوع الورقة:

عمل تطبيقي لبناء مورد لغوي (Lexical Resource) أكثر من كونه بحثًا نظريًا بحثًا

Chapter 01

أهداف البحث

الهدف العام:

- تحسين غطاء وجودة الشبكة الدلالية للغة العربية عبر استغلال PWN وتقنيات التعلم العميق.

أهداف فرعية:

- اضافة مرادفات (Synsets) عربية جديدة مطابقة لبنية PWN.
- دمج تعريفات وأمثلة عربية مُستخرجة/مترجمة لكل Synset لرفع قابلية الاستخدام.
- استعمال عدة محركات ترجمة ثم تنقية النتائج بواسطة Transformers (مثل MT5 / AraBERT) لتحسين الدقة السياقية.
- توفير منصة ويب للتحقق البشري والتصديق على المرشحات قبل إدراجها.

تتناول مشكلة حقيقية: فقر الموارد الدلالية العربية وخصوصًا AWN

توظيف تقنيات حديثة:

- Transformers, BERT, AraBERT
- استغلال أنظمة ترجمة متعددة

تغطية مراحل متكاملة:

- استخراج الكلمات
- ترجمة المفاهيم والتعاريف والأمثلة
- تصفية/اختيار باستخدام نماذج لغوية
- بوابة للتحقق البشري

عرض نتائج رقمية (إحصاءات، أعداد Synsets، نسب دقة)

- اعتماد كبير على الترجمة الآلية
→ قد يؤدي إلى أخطاء دلالية و صياغية
- عدم توضيح الفجوة بدقة → أرقام AWN غير منسقة بين الإصدارات
- تجارب محدودة
→ التقييم على 1000 فقط من 90,127
→ لا توجد مقارنة مع أعمال سابقة أو أنظمة فعلية
- لا توجد تطبيقات عملية تثبت الأثر
→ لم يتم اختبار النسخة المُعنّاة في نظام NLP فعلي

- استخدام التشابه العددي (cosine similarity) فقط لا يكفي للجودة المعجمية
- الأمثلة مُستخرجة من Reverso (معظمها ترجمة آلية غير أصيلة)
- اختيار كلمة واحدة/مثال واحد لكل Synset قد لا يعكس جميع المعاني
- منصة التحقق مفيدة لكن غير مغطاة بتفاصيل كافية (عدد المراجعين، السرعة، الجودة)

الدقة:

- الكلمات: 79%
- التعاريف: 95%
- الأمثلة: 81%

التقدير بأن Synsets 68,000 ستكون صحيحة افتراض متفائل

لا يوجد تحليل لأسباب الأخطاء أو توزيعها على أنواع الكلمات

الورقة تقدّم خطوة تطبيقية مهمّة لتوسيع Arabic WordNet

لكنها تعاني من:

→ نقص التقييم التجريبي الحقيقي

→ اعتماد كبير على الترجمة الآلية

→ غياب تحليل الأخطاء والمقارنات



مجموعة جامعة
الجميع الذكية
للتعليم القابضة

MIDOCEAN
UNIVERSITY



Thank You

