

كيف نعلم الذكاء  
الاصطناعي فهم  
اللهجة الشامية؟



يهدف هذا المشروع إلى توثيق المصطلحات الشامية ضمن إطار لغوي قابل للستخدام في تطبيقات الذكاء الاصطناعي ومعالجة اللغة العربية.

01



# لماذا هذا المشروع مهم؟

- بعد السنوات الأخيرة، ازداد اهتمام العالم باللهجة والثقافة السورية.
- الحاجة إلى بيانات لغوية دقيقة للهجات العربية لاستخدامها في البحث، التعليم، والذكاء الاصطناعي.
- معظم أدوات NLP لا تتعامل مع اللهجات جيداً بسبب نقص البيانات المنظمة.

02



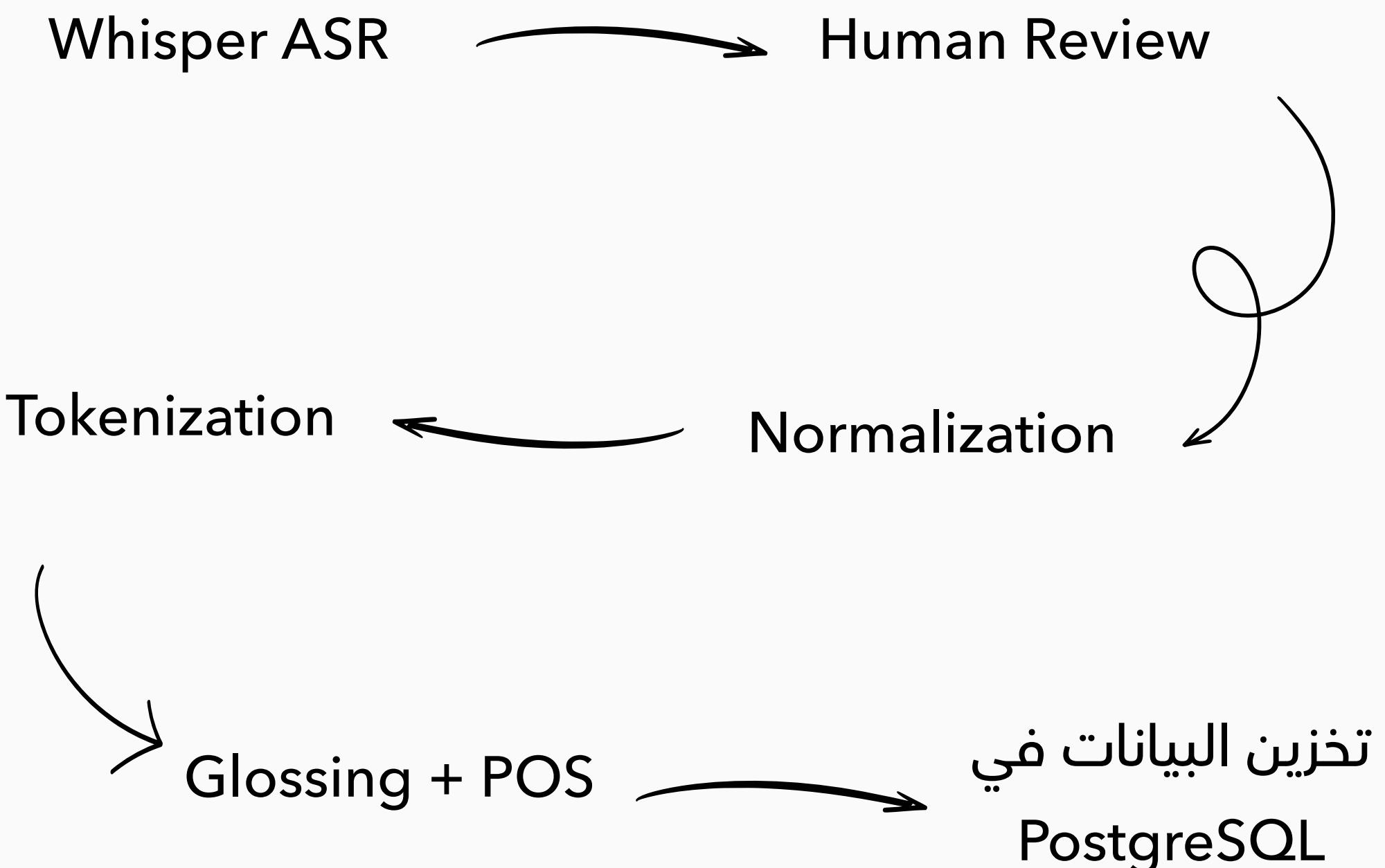
## الهدف النهائي

- بناء قاعدة بيانات لغوية منظمة تشمل:
- المصطلح الشامي
- كتابته الموددة (Normalized)
- معناه العربي الفصيح
- دوره النحوی (POS)
- السياق الذي يظهر فيه
- تخزينه في قاعدة PostgreSQL للستخدام المستقبلي

03



# مخالك العمل (Pipeline)



04



# من الصوت إلى نص

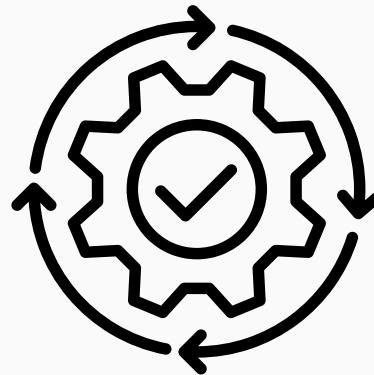


1. استخدام Whisper لتحويل التسجيلات إلى نص
2. مراجعة بشرية لتصحيح الأخطاء
3. توحيد الكتابة (Normalization)

05



# معالجة النص



- CAMeL Tools باستخدام Tokenization
- إزالة التكرار الحرفية والضبابي
- إضافة المعاني
- تحليل نحوبي باستخدام Stanza

06



# الناتج النهائي



قاعدة بيانات لغوية منظمة للهجة الشامية

تدعم:

1. تدريب نماذج NLP

2. بناء Chatbots

3. تطوير أدوات تعليمية

4. دراسات لغوية للهجات السورية

06



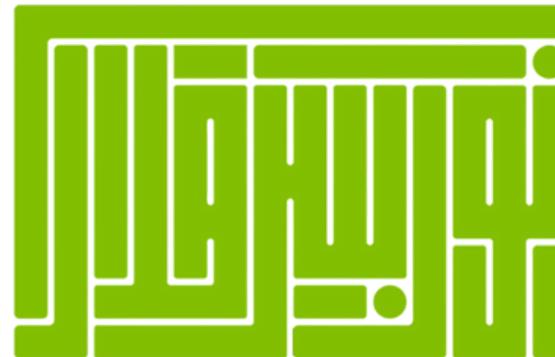
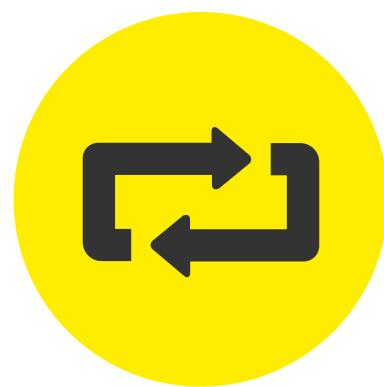


ما هي المصطلحات الشامية التي تعتقد أنها  
منتشرة بكثرة، ولكن لم يتم تضمينها بعد في  
أدوات الذكاء الاصطناعي؟

07



تابعوني لنتساعد على نشر  
المعلومات التقنية باللغة  
العربية



Eng. Nour Berkdar