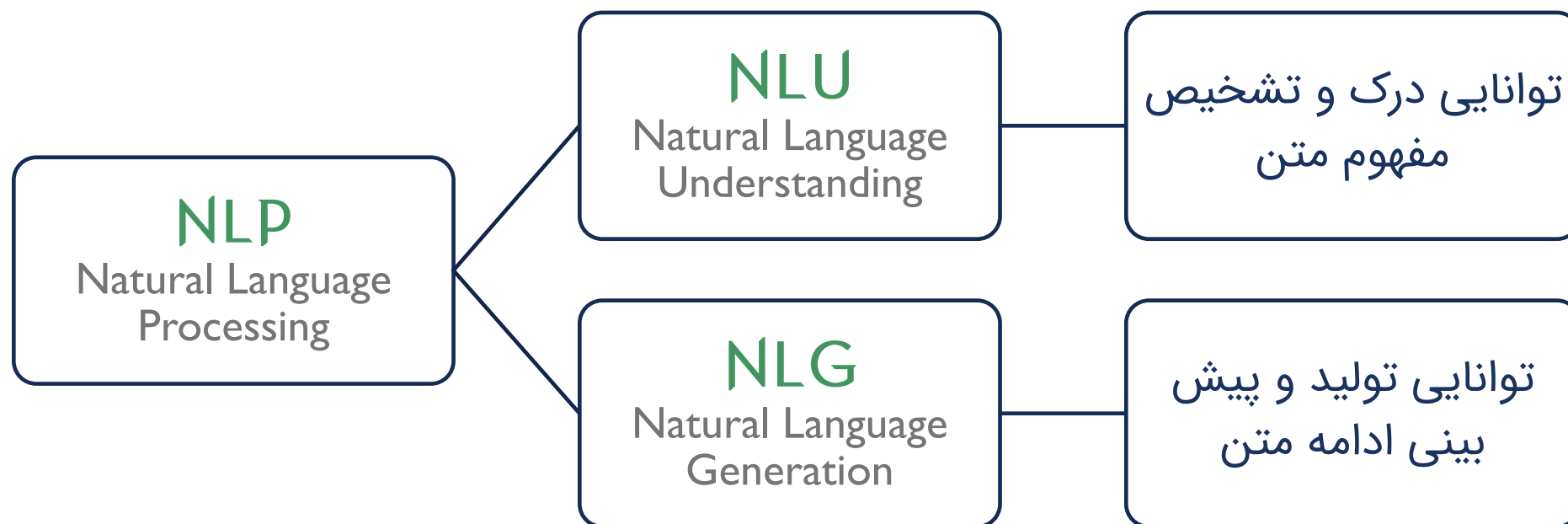




۵- پردازش زبان طبیعی (NLP)







قوانین و قواعد ثابت و مشخص نداریم

- قوانین سطح پایین: نقطه، علامت، فعل، سوال و ...
- قوانین سطح بالا: طنز، کنایه، اشاره و ...

لغات دارای معانی متعدد و متفاوت

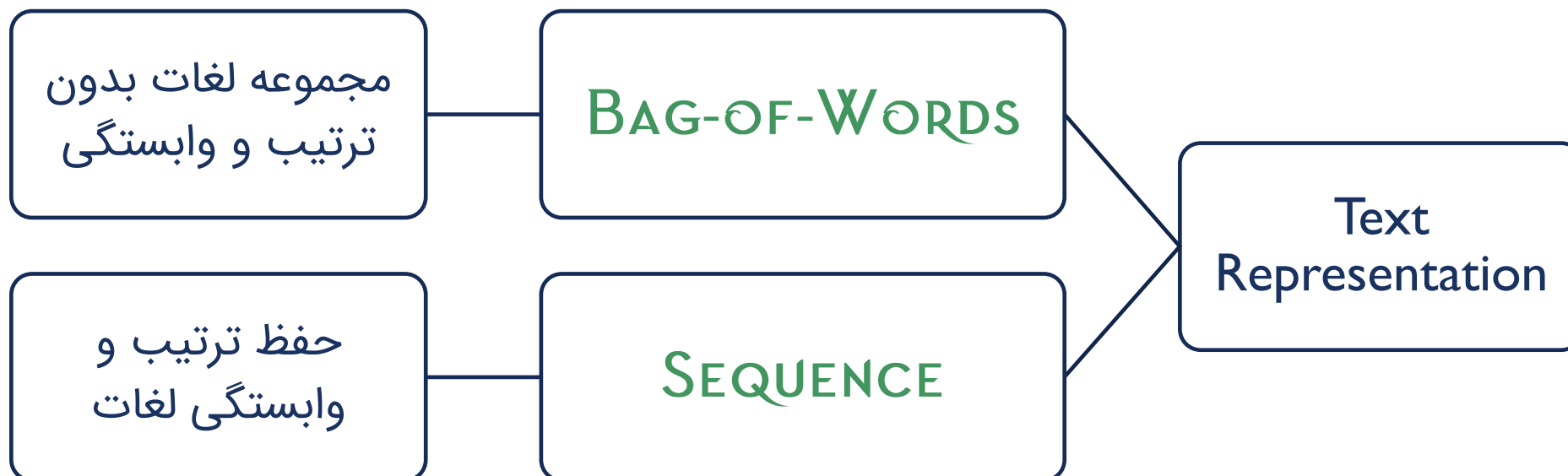
- مثلاً لغت match را در دیکشنری ببینید !

شکل های مختلف متن

- ایمیل، مقاله، نظر، محاوره ای، کتابی و ...

داده های غیر عددی و بدون ساختار

- مدل یادگیری عمیق فقط با داده عددی آموزش می بیند



پیش پردازش متن

- استانداردسازی / جداسازی / حذف کلمات اضافی / ریشه یابی

تبدیل نوع داده

- تبدیل متن به بردارهای عددی

طراحی لایه ها و مدل های متناسب

- لایه ها و مدل های یادگیری عمیق مناسب برای آنالیز متن

مراحل پیش پردازش متن در روش‌های BAG-OF-WORDS

