چالش های تحلیل متن فارسی:

>> پیش پردازش صحیح. نرمالایز کردن

>> اینکه در چه سطحی توکن کنیم خیلی مهمه. توکن ها نه زیاد باید طویل باشن نه خیلی به سوی اتومیک پیش رفتن. مثلا به این جمله توجه کنید:

دکتر حسن روحانی به نیویورک سفر می کند.

در این جمله، اگر توکنایز بتواند "دکتر حسن روحانی " را دتکت کند برای مدل سازی بسیار خوب است. ولی اگر هر کدام ازکلمات را یک توکن در نظر بگیرد کمک چندانی نمیکند.

>> چالش دیگر چالش تشخیص افعال ترکیبی است. مثلا : سفر می کند. غذا می خورد.

راه حل: روکیرد فازی