



# **DEVELOPING LANGUAGE RESOURCES FOR ENGLISH/বাংলা MACHINE TRANSLATION**

**Rabia Sultana Ummi, Fahmina Huda**

## SOME DEFINITIONS..

- **Machine Translation-** Used to translate text or speech from one natural language to another

<http://translate.google.com/>

<http://www.apertium.org/>

- **Parallel Corpora-**Multilingual corpora (same set of texts in different languages) specially formatted for side by side comparison

- **Statistical Machine Translation-** Generates translation using statistical methods based on bilingual corpora.



# APPROACHES

- Rule based
- Example based
- Statistical



# RESOURCES

- Grammar/Rules
- Bilingual corpus



# WHAT ARE THE RESOURCES DO WE HAVE FOR MT?

- Parallel corpus (Brown corpus)
- Tagset: 56 tags
- Completed 350 strings.



## AN EXTRACT FROM THE CORPUS

- Pierre Vinken , 61 years old , will join the board as a nonexecutive director Nov. 29 .
- ৬১ বছর বয়স্ক পিয়েরে ভিনকেন অনির্বাহী পরিচালক হিসেবে নভেম্বরের ২৯ তারিখে পরিষদে যোগদান করবেন।
- Mr. Vinken is chairman of Elsevier N.V. , the Dutch publishing group.
- ভিনকেন সাহেব ওলন্দাজ প্রকাশনা সংস্থা এলসেভিয়ার এন.ভি এর চেয়ারম্যান।



# POS- TAGGING THE CORPUS

## 56 TAGS IN THE BANGLA TAGSET

৬১/QFNUM বছর/NN বয়স্ক/JJ পিয়েরে/NNPC ভিনকেন/NNP অনির্বাহী/NNC পরিচালক/NN হিসেবে/ON নভেম্বরের/NNP+SF\$ ২৯/JJ তারিখে/NN+SFON পরিষদে/NN+SFON যোগদান/NNV করবেন/VB|/.

Level 1	Level 2	Tag	Examples
Noun	Proper	NNP	ভিনকেন
Noun	Compound Common Noun	NNC	অনির্বাহী পরিচালক
Noun	Compound Proper Noun	NNPC	পিয়েরে ভিনকেন
Adjective	Simple	JJ	২৯ নভেম্বর, বয়স্ক
Verb	Main Finite Verb	VB	করবেন
Suffixes	Adpositional	SF\$	এর
Punctuation Marks	Sentence Final Punctuation	.	

Thanks

