

Unifying Speech Resources: A Case of Ibibio

MOSES EKPENYONG

SUMMARY

- × Introduction
- × Ibibio
- × Nsibidi
- × Ibibio speech sounds
- × Ibibio language technology development
- × Some useful byproducts
- × The unified model
- × Conclusion

INTRODUCTION

- ✘ Implementing linguistic knowledge on computers has posed numerous challenges to scientists and linguists, due to the dynamic nature of natural languages
 - + Languages evolve and diversify over time, borrow words from other languages
 - + All language has some element of uniqueness
 - + Some are resource-rich while others are resource-limited
- ✘ NLP approaches have emerged to cope with these challenges . The problems associated with NLP relate to
 - + generation and understanding of linguistic features of the processed language, i.e., the ability to model for instance morphology (the structure of words).

INTRODUCTION

- ✘ Speech processing is becoming increasingly common in the world today. Speech resources benefit four categories of experts. These categories are:
 - + Core tools and data: for speech scientists, phoneticians and phonologists.
 - + Training and test data: for speech recognition and synthesis researchers
 - + Corpora and corpora processing tools: for researchers working on dialog systems
 - + Corpora from spoken domain: computational linguists working with text.

INTRODUCTION

- ✘ The by-products from these categories (for instance e-dictionaries, translation systems, etc.) are useful to everyone else, most especially the language communities. Otherwise, the fairness of such resources is questionable.
 - + Fairness enhances efficiency and longevity of such language products

IBIBIO (ISO: 639-3, ETHNOLOGUE IBB)



ekpo



- ✘ Tone language
- ✘ Spoken by approx. 4,000,000 speakers
- ✘ In Akwa Ibom State
- ✘ Southeast coastal region of Nigeria
- ✘ West Africa
- ✘ From the (new) Benue Congo language family

abang



ekpe



NSIBIDI WRITING SYSTEM

- ✘ Nsibidi (also known as nsibiri, nchibiddi or nchibiddy) is a system of symbols indigenous to what is now southeastern Nigeria
- ✘ that is apparently ideographic, though there have been suggestions that it includes logographic elements
- ✘ The symbols are at least several centuries old: Early forms appeared on excavated pottery as well as what are most likely ceramic stools and headrests
- ✘ from the Calabar region, with a range of dates from between 400 and 1400 CE

"Nsibidi"	"Calabash"	"Matchet"
"Welcome"	"Big drum"	"Woman"
"Two men talking"	"Etak Ntaña Nsibidi — Nsibidi's bunch of plantains. When the head of the house wants plantains he sends this sign to the head boy on the farm." ^[17]	"Man"
"Door"	"Umbrella"	"Moon"
		"Tortoise" ^{[1][15][17]}
"Gun"		
"Crossbow"		



IBIBIO SPEECH SOUNDS

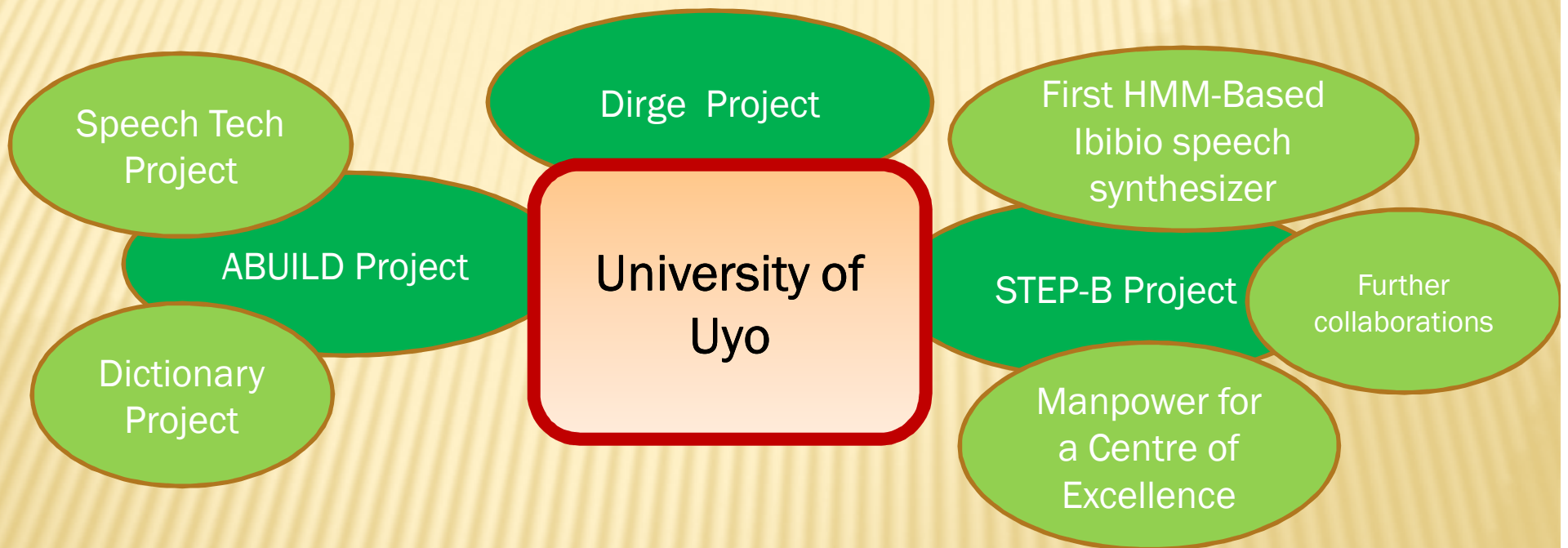
Table 1. Ibibio vowel system

		Vowel		
		front	central	back
Vowel height	High	i, ii	ɨ	ɯ
	Mid	e, ee	ə, ʌ	o, oo, ɔ, ɔɔ
	Low			a, aa

Table 2. Ibibio consonant system

		Places of articulation					
		labial	alveolar	palatal	velar	glottal	uvular
Manners of articulation	Stops	voiceless	kp, p, t, tt		k, kk		
		voiced	b	d			
Manners of articulation	Non stops	nasal	m, mm	n, nn		ŋ, ŋm, ŋn, ŋw, ɲ	
		fricative	f	s		h	
		tap					ɓ
		trill		r			
		approximant	w		j		

IBIBIO LANGUAGE TECHNOLOGY DEVELOPMENT



SOME USEFUL BY-PRODUCTS

- ✘ (1) Nwed usem Ibibio (Ibibio Dictionary) – approved for primary and secondary schools in Akwa Ibom State

a,A

aa [áá] *interj* indicates surprise; áá, áké dámmá idáhá ákě

aa [áâ] *interj* indicates surprise, áâ, ntórò ké ádó

aafó [àáfò] *n* [Eng half] half penny; ñkporé émi iyóhóké áâfò.

aafu [àáfù] *n* [Eng half] half penny; nkédèp bésìn ódò ké àáfù.

aak [á!ák] *interj* please (emphatic, persuasive, very polite); á!ák, kùyat ésit.

aakporo [ààkporó] *n* named child (someone named after another, each is *akporo* to the other); ààkporó m̄fò?

aba edem [ábâ èdèm] *n* the back (anat) (*edem* - back, side) ñkporé áfít ányé ke ábâ èdém.

aba¹ [àbà] *n* cavity, hollow, hole;ékpú áfèghé ádák àbà ùsàñ.

aba² [àbà] *num* forty;ékáēté ódò ówó ísùà àbà.

aba³ [àbà] *adv* again, anymore, still (used most commonly with a negative verb); kúnám úyóm àbà.

abaak [àbáák] *aj* young, not full grown (*abaak* can be used with the name of any domestic animal); édíñí ódó ódò àbáák

abaakañ [ábààkàñ] *n* a kind of tree which often grows very large in the bush, its succulent leaves are used for goat fodder, african nutmeg (adams), wild nutmeg (dalziel), pycnanthus kombo (dalziel); ábààkàñ m̄mò ádá ké ádáñáíkót ódò.

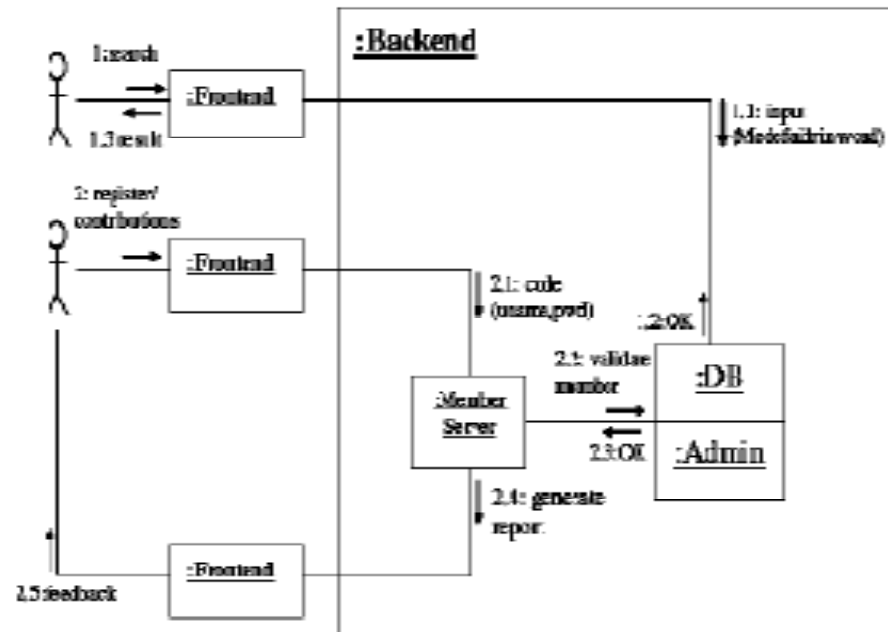
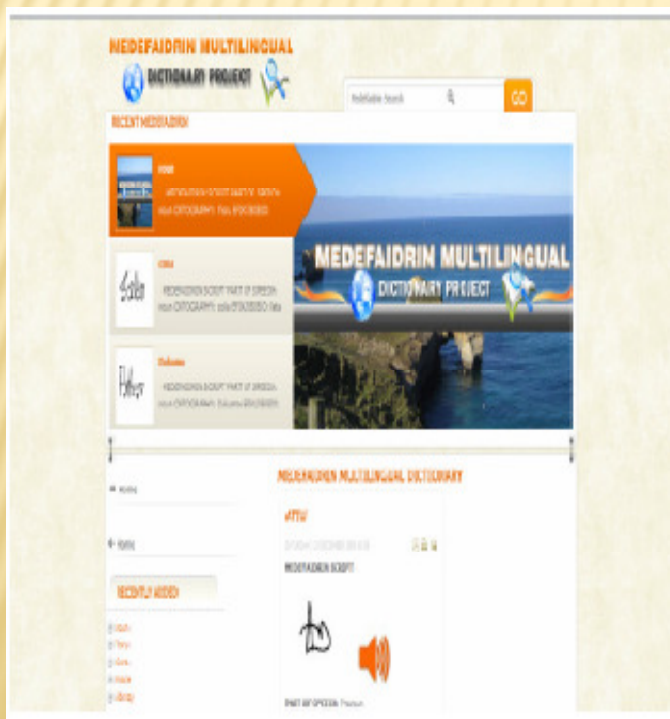
abada [ábádá] *n* [yoruba] agbada, men's robes; èbè ñdò àkèsìne ábádá.

abai [ábái] *n* pillar; ówòùdèèn ódò ábái úfok.

abakañ [ábákàñ] *n* a kind of tree which often grows very large in the bush, its succulent leaves are used for goat fodder, african nutmeg (adams), wild nutmeg (dalziel), pycnanthus kombo (dalziel); ábákàñ m̄mò ádá ké ádáñáíkót ódò.

SOME USEFUL BY-PRODUCTS

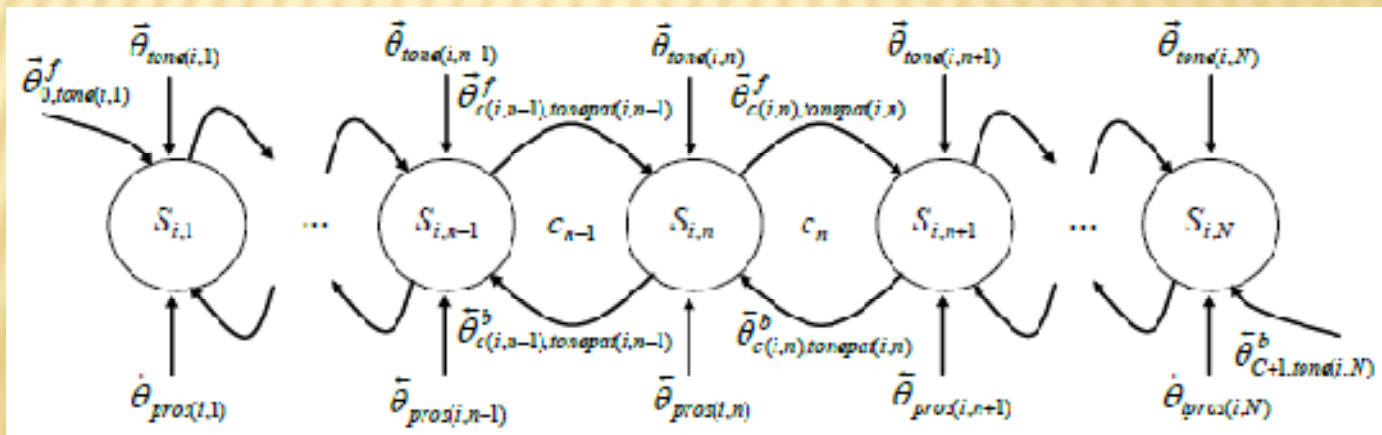
- ✘ (2) The Medefaidrin talking dictionary



S/N	Script	Translit.	POS	IBIBIO	ENGLISH	Sound
1	Ati	Atiu	Pron.	Ami	I	
2	Ọ́máwọ́zír	Nimazid	Verb	Wed	Write	
3	Colia	Colia	Noun	Leta	Letter	
4	Capium	Capium	Conj.	Nyun	And	
5	Trek	Trek	Verb	Dia	Eat	
6	Dukuma	Dukuma	Noun	Ikpon; mkpon	Cocoyam	
7	Abinzy	Abinzy	Noun	Bia	Yam	
8	Fosra	Fosra	Noun	Garri	Garri	

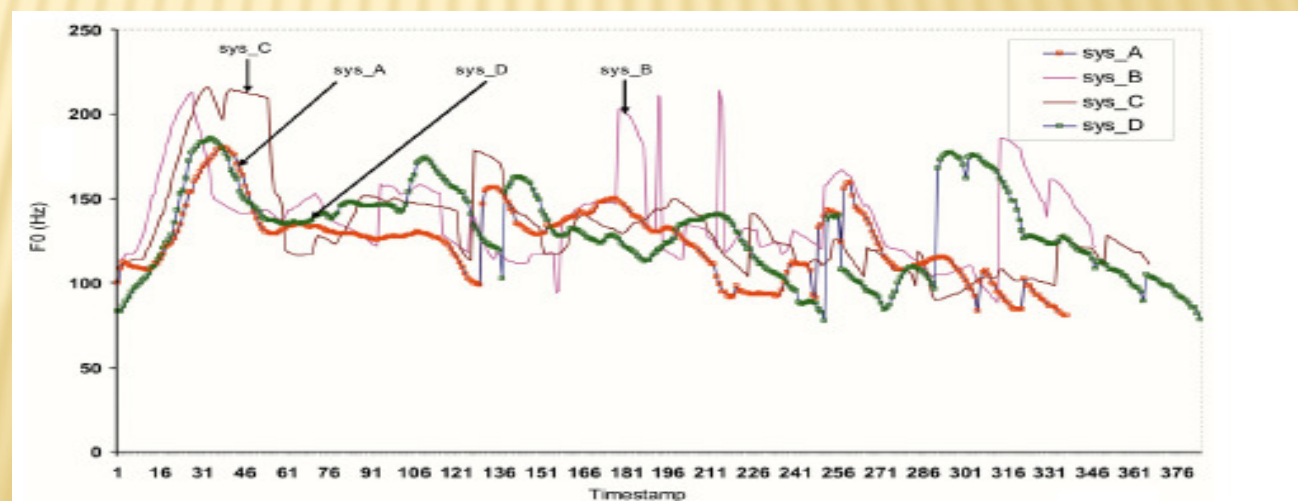
SOME USEFUL BY-PRODUCTS

- ✘ (3) First (HMM-Based) - statistical parametric speech synthesis system for Ibibio
 - + Front-end modules/scripts
 - + Generic and replicable for other Nigerian languages
 - + Useful for context dependent modelling of tone and prosody



SOME USEFUL BY-PRODUCTS

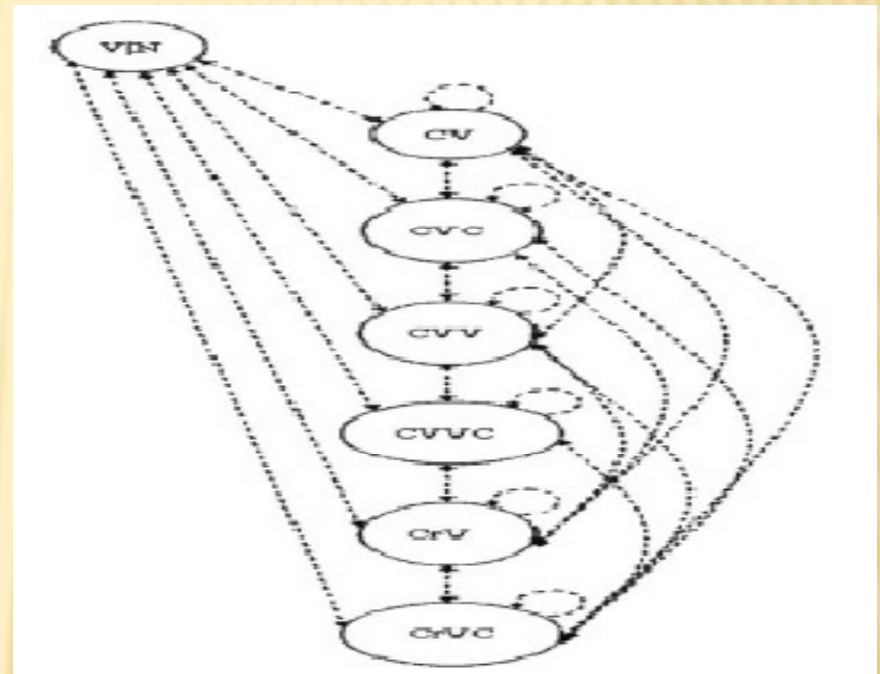
- ✘ Important front-end components
 - + Phoneme inventory table definition
 - + Syllable table definition
 - + HRG generation/utterance syllabification
 - + Generation of tone-tagged equivalent of syllabified corpus
 - + Running the associated scripts for label files, HTS labels for prosody/tone features addition



SOME USEFUL BY-PRODUCTS

✘ The syllable FST

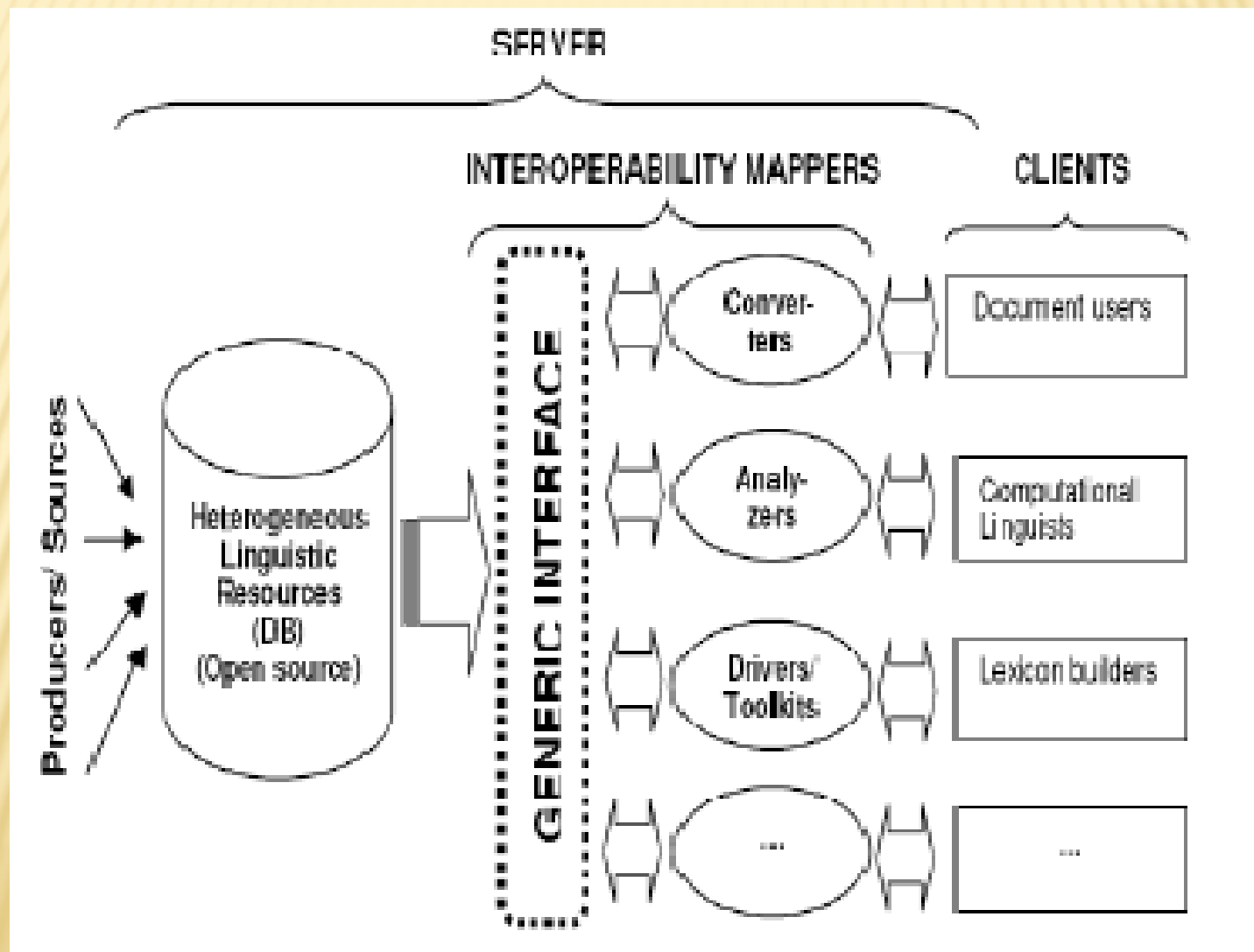
Sample input sentences fed to the FST
1. owo ibarake ikaN ikeene daNa udVk ikOt OJOON
2. eebOppo ke idIt keetekeet anaasIne Nkpo siben ufOn isOk ibaan
3. ekanba ase Nkpo mbaNa kooperatif defedopmen bank
4. mbaak Omno ekeme ediwuO nda ke usVN OfOnno
5. ikiduOUNO ke iJan
6. nO eJe je ikOt utom OmO eewOt nJIn Nkem Omno OjOhOOjOhO
Sample output generated from the FST
1. #o-wo#i-ba-ra-ke#i-kaN#i-kee-nee#da-Na#u-dVk#i-kOt#O- JOON
2. #ee-bOp-pO#ke#i-dIt#kee-te-keet#a-naa-sI-nee#N- kpO#si-ben#u-fOn#i-sOk#i-baan
3. #e-kam-ba#a-se#N-kpO#m-ba-Na#koo-be-ra-tif#de-fe- dop-men#baN-k
4. #m-baak#Om-mO#e-ke-me#e-di-wuO#n-da#ke#u-sVN#O-fOn- no
5. #i ki duO NO#ke#i Jan
6. #nO#e-Je#je#i-kOt#u-tom#O-mO#e-wOt#n-JIn#N-kem#Om- mO#O-jO-hOO-jO-hO



THE UNIFIED MODEL

- ✘ A client-server model that will permit interoperability of linguistics resources
- ✘ The server side has a database of heterogeneous resources and interoperability mapping tools.
- ✘ The clients on the other hand interface the server through applications mapping.
- ✘ The proposed model is generic and will encourage collaboration, content exchange and management of resources.
- ✘ It will also enforce adherence to standards and agreeable formats.
- ✘ The clients are plug-and-plays to the server and details behind the generic interface are made transparent to the clients.

THE UNIFIED MODEL



✘ Thank you