

EPSRC

Engineering and Physical Sciences
Research Council



Edinburgh – Cambridge – Sheffield

Lecture processing

Initial work on Domain Independence

Ghada AlHarbi and Thomas Hain

<http://www.natural-speech-technology.org>



- Main resource - Liberated Learning Consortium - <http://liberatedlearning.com>

Corpus	#Words	# Uniq. Words
In-Domain sources		
LLC	1.4M	24.5k
BASE	1.5M	30k
MICASE	600k	17k
TED	72k	4k
Out-Of-Domain sources		
Wed(CHIL)	68M	300k
Web(RT07-lect)	40.5M	250k
HUB4-LM96	131M	200k
SWB(CTS)	3.5M	27k
Web(Fisher-conv)	500M	700k
Web(ami-rt05)	78M	300k

Data sets

- Acoustic training data - 133 hours, recordings from 10 universities in US/UK/Australia. s
- Development set (dev) = 13.5 hours of speech
- Evaluation set (eval)

Discipline	Dur.	# W	# Uniq. W
Arts & humanities	0.34	2854	722
Biological Science	0.34	4004	666
Business	0.34	3950	832
Education	0.26	2677	737
Engineering	0.34	3942	855
English	0.64	6414	1130
Medicine	0.60	6130	1281
Physical Science	0.34	3746	547
Psychology	0.34	3999	994
ALL	3.59	37649	4138

Vocabulary

Wordlist Source	# Word	<i>dev</i>	<i>eval</i>
Lect	36116	0.7	0.8
LLC	24487	0.7	0.8
BASE	30321	1.1	1.4
MICASE	17296	1.8	2.2
TED	4243	4.1	6.2
CTS	27912	6.2	7.1
HUB4	216418	0.01	1.1

- Rather surprisingly low OOV rates even though wide range of disciplines

Perplexities on domain specific LMs

dev set	dev	eval	dev _{<i>d</i>}	eval _{<i>d</i>}
Arts & humanities	132.7	112.4	121.9	102.1
Biological Science	104.3	90.1	98.2	82.7
Business	111.6	109.9	110.1	105.3
Education	115.7	109.2	115.4	108.1
Engineering	111.7	97.7	107.0	101.3
English	135.2	111.2	125.0	112.3
Medicine	112.0	105.3	-	-
Physical Science	79.1	100.6	68.1	91.0
Psychology	98.4	95.7	97.8	96.1

- Same wordlist (derived from Lect data)
- Subscript *d* denotes domain dependent

Domain dependent interpolation weights

dev set	lect	chil	rt07	hub	ami	fish.
Arts&Hum.	0.75	0.05	0.01	0.03	0.02	0.12
Bio._Sci.	0.75	0.07	0.04	0.02	0.05	0.07
Business	0.51	0.08	0.05	0.10	0.06	0.20
Education	0.57	0.06	0.04	0.15	0.04	0.14
Engineering	0.53	0.07	0.05	0.03	0.15	0.16
English	0.59	0.04	0.01	0.02	0.03	0.30
Medicine	0.51	0.06	0.02	0.05	0.03	0.33
Physical_Sci.	0.85	0.09	0.02	0.00	0.01	0.03
Psychology	0.65	0.05	0.02	0.03	0.03	0.22

- Surprising fluctuation

Results using standard models

- ML acoustic models
- 3gram interpolated LMs

LM	PPL <i>dev</i>	PPL <i>eval</i>	%WER
Lect	128.9	129.9	37.2
Lect+Wed(chil)	114.6	114.1	35.1
Lect+Web(rt07)	118.0	117.5	35.5
Lect+HUB4	118.3	117.8	35.5
Lect+SWB(cts)	127.9	128.3	36.9
Lect+Web(fisher)	112.8	114.9	35.2
Lect+Web(ami)	116.0	110.9	34.6
ALL	111.7	110.4	30.2

Results by Discipline

Category	%WER	SR	%H	%FS
Arts & humanities	26.4	2.40	1.8	0.4
Biological Science	14.4	3.47	0.7	0.5
Business	26.8	3.37	0.6	0.5
Education	43.7	3.08	0.4	0.4
Engineering	24.7	3.29	1.9	0.3
English	37.0	2.86	2.0	1.1
Medicine	39.6	-	-	-
Physical Science	20.4	3.08	0.2	0.2
Psychology	32.0	3.41	1.1	0.4

- SR = speaking rate
- %H = percentage of hesitation words
- %FS = percentage of false starts

Temporal structure

- Split all lectures into beginning, middle and ending
 - Beginning = first 10 min
 - Ending = last 5 minutes

Lecture Discipline	Start	Middle	End
Arts & humanities	148.5	144.6	118.3
Biological Science	144.2	126.1	113.3
Business	131.0	124.4	125.0
Education	107.4	122.5	140.4
Engineering	151.8	117.1	129.2
English	156.3	124.8	156.7
Medicine	110.4	111.5	115.2
Physical Science	83.9	82.7	75.4
Psychology	103.3	95.8	103.8

Structure

- Perplexities with varying size of beginning section
- Speech settles after 5 minutes

