CS674, Spring 2002
Lecture Handout, 3/11/02

These examples are all taken from David Yarowsky, "Decision lists for lexical ambiguity resolution: Application to accent restoration in Spanish and French", *Proceedings of the ACL*, 1994, pp. 88–95. Excerpts were produced by slightly editing the source text available from the E-print Arxiv, `http://xxx.lanl.gov/format/cmp-lg/9406034`.

1. Sample de-accented ambiguities, and baseline data.

De-accented Form	Accent Pattern	%	Number
cesse	cesse	53%	669
	cessé	47%	593
cout	coût	100%	330
couta	coûta	100%	41
coute	coûté	53%	107
	coûte	47%	96
cote	côté	69%	2645
	côte	28%	1040
	cote	3%	99
	coté	<1%	15
cotiere	côtière	100%	296

2. Example training instances.

Pattern	Context		
côté	du laisser de	*cote*	faute de temps
côté	appeler l' autre	*cote*	de l' atlantique
côté	passe de notre	*cote*	de la frontiere
côte	vivre sur notre	*cote*	ouest toujours verte
côte	creer sur la	*cote*	du labrador des
côte	travaillaient cote a	*cote*	, ils avaient

3. An abbreviated decision list.

|LogL|	Evidence	Classification
8.28	PREPOSITION QUE *terminara*	⇒ terminara
7.24	de que *terminara*	⇒ terminara
7.14	para que *terminara*	⇒ terminara
6.87	y *terminara*	⇒ terminará
6.64	WEEKDAY (within $\pm k$ words)	⇒ terminará
5.82	NOUN QUE *terminara*	⇒ terminará
5.45	domingo (within $\pm k$ words)	⇒ terminará