

Macrosyntactic Segmenters of a French spoken corpus

Ilaine Wang¹

Sylvain Kahane¹

Isabelle Tellier²

MoDyCo (CNRS)
Université Paris Ovest Nanterre

LaTTiCe (CNRS)
Université Paris 3

Introduction

- > Hypothesis : automatic segmentation of speech in **sentence-like units** is possible using prosodic and syntactic information together
- > Data : Rhapsodie 33,000-word treebank of **spoken French**, including syntactic and prosodic tags as well as a macrosyntactic-level annotation based on speech acts simplified for this study

donc quand vous arrivez sur la place de Verdun / la Préfecture est sur votre droite // vous / vous continuez tout droit //
en fait / euh / ben / là / euh / vous passez / euh / ben / en fait / c'est la ligne du tram / toujours / là //
' so when you arrive on place de Verdun / the prefecture is on your right // you / you continue straight away //
in fact / er / well / there / you pass / er / well / in fact / it's the tram line / still / there //'

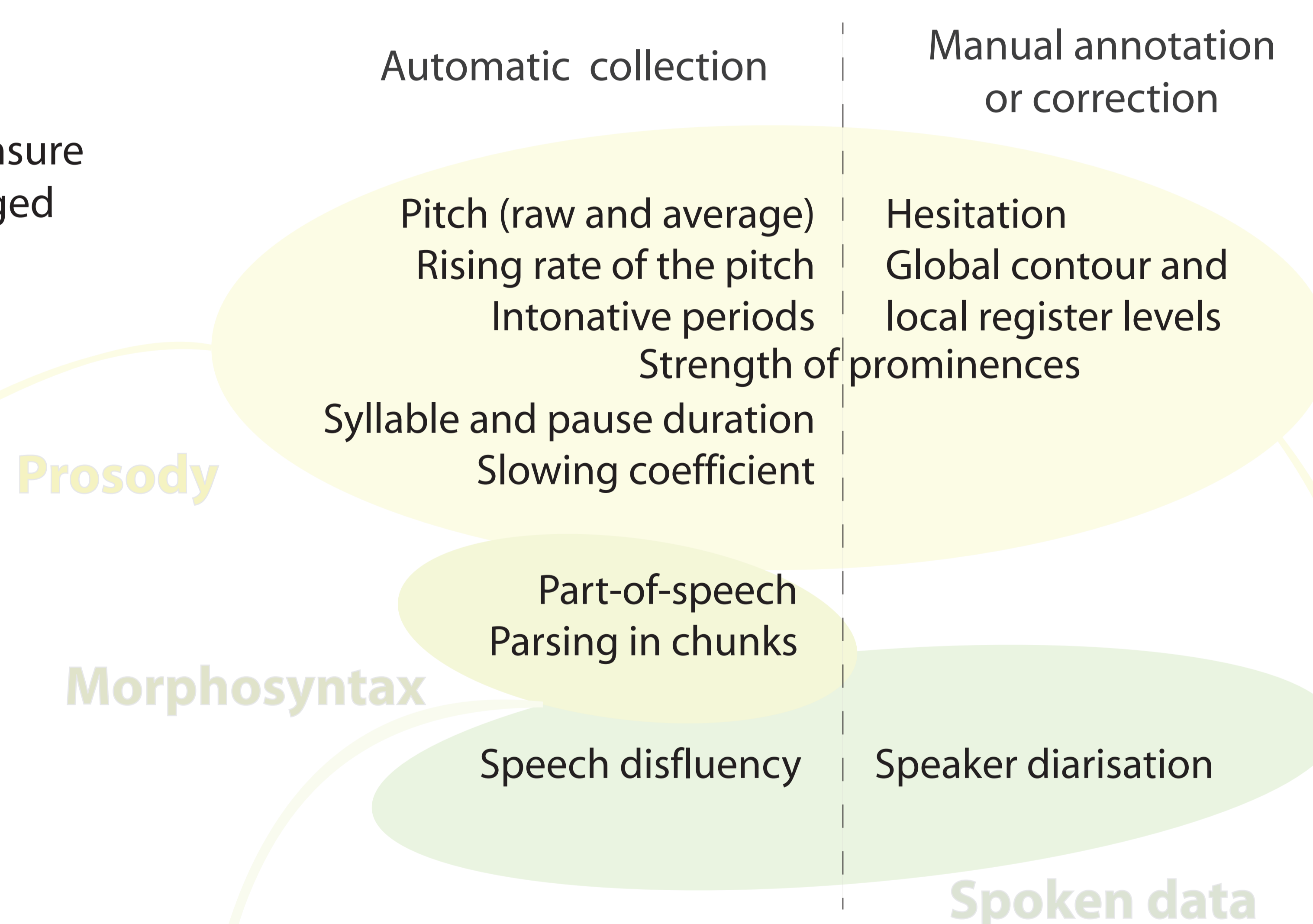
Objectives

- > determine the **cost of manual effort** necessary to segment spoken data into macrosyntactic units
- > measure the **discriminating power** of prosodic and non-prosodic features for this task
- > create a **fully automated segmentation process** to ensure application to large non-tagged corpora

Methodology

- > identifying **macrosyntactic boundaries** (major // and minor /)
- > using **prosodic and syntactic tags** either from the Rhapsodie treebank or from state-of-the-art tools as features
- > with **supervised classification** methods (Machine Learning), in this case Decision Trees J48 algorithm

Which cues ?



Training of 6 different segmenters using different sets of features

Results

Model	Annot.	P	R	F1	Acc.
All	//	0.62	0.55	0.58	84.3%
	/	0.6	0.35	0.44	
	none	0.89	0.96	0.92	
Prosody only	//	0.57	0.39	0.46	79.2%
	/	0.36	0.08	0.16	
	none	0.82	0.96	0.89	
No prosody	//	0.64	0.49	0.55	84.5%
	/	0.68	0.31	0.43	
	none	0.87	0.98	0.92	

(a) Both automatic and manual features

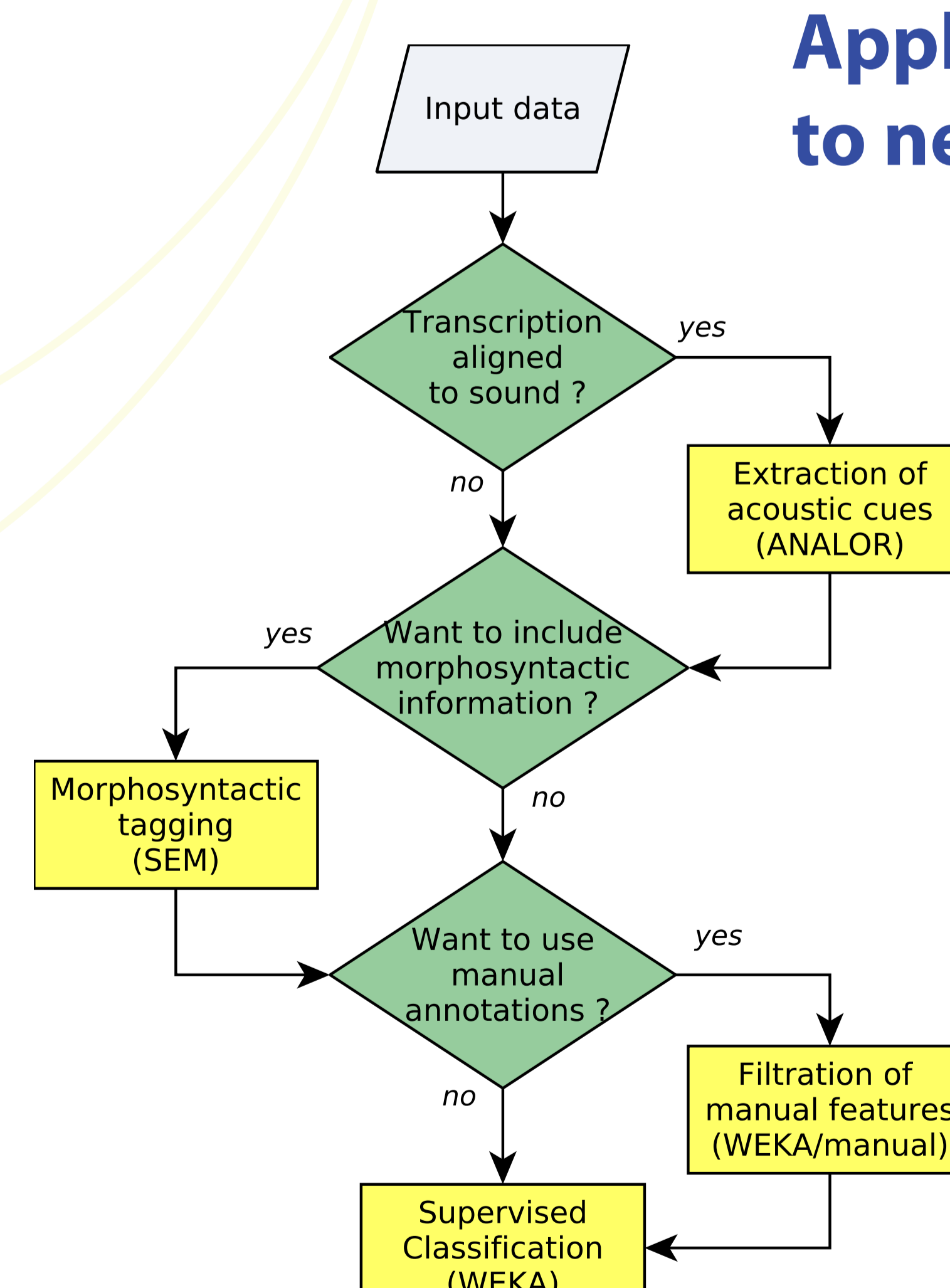
Model	Annot.	P	R	F1	Acc.
All	//	0.61	0.55	0.58	84.4%
	/	0.61	0.36	0.45	
	none	0.89	0.96	0.93	
Prosody only	//	0.51	0.37	0.43	77.8%
	/	0.27	0.09	0.14	
	none	0.83	0.94	0.88	
No prosody	//	0.63	0.47	0.54	84.3%
	/	0.68	0.31	0.42	
	none	0.87	0.98	0.92	

(b) Automatic features only

Conclusions & further exploration

- > all models **above baseline**, and latest model scores 85.2% accuracy with more disfluency cues
- > further **examination of model trees** to get details on the use of the different cues and their efficiency in discriminating boundaries
- > future experiments including the type of speech, as the **heterogeneity of samples** in the Rhapsodie corpus (monologs, dialogs, interactive or semi-interactive speeches) might have misled classifiers

Applicability to new corpora



Freely available online :

<http://www.lattice.cnrs.fr/sites/itellier/TEOK.html>