Transcriber driving strategies for transcription aid system

Grégory Senay, Georges Linarès,
Benjamin Lecouteux, Stanislas Oger

Laboratoire Informatique d'Avignon

LREC'2010 - May 2010

Overview

- Introduction
- What is interactive decoding ?
- Driving strategies
- Experiences and results
- Conclusion

Introduction

Current situation

- Automatic Speech Recognition system performance:
  - $\Rightarrow$ accurate on defined domains (ex: Broadcast news)
  - $\Rightarrow$ decreases, if the conditions are changed
- Manual transcriptions are needed to provide a perfect transcription
- Recent projects use transcriptions provided by a speech recognition system
  - $\Rightarrow$ they only use the one-best hypothesis [Bazillon LREC08]

Objective

- Reduce the cost of the global transcription
- Correction efficiency
- Computer and Human can work together

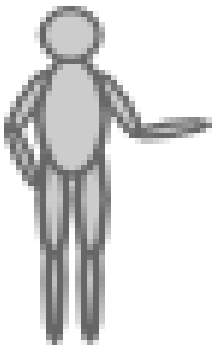
Interactive decoding

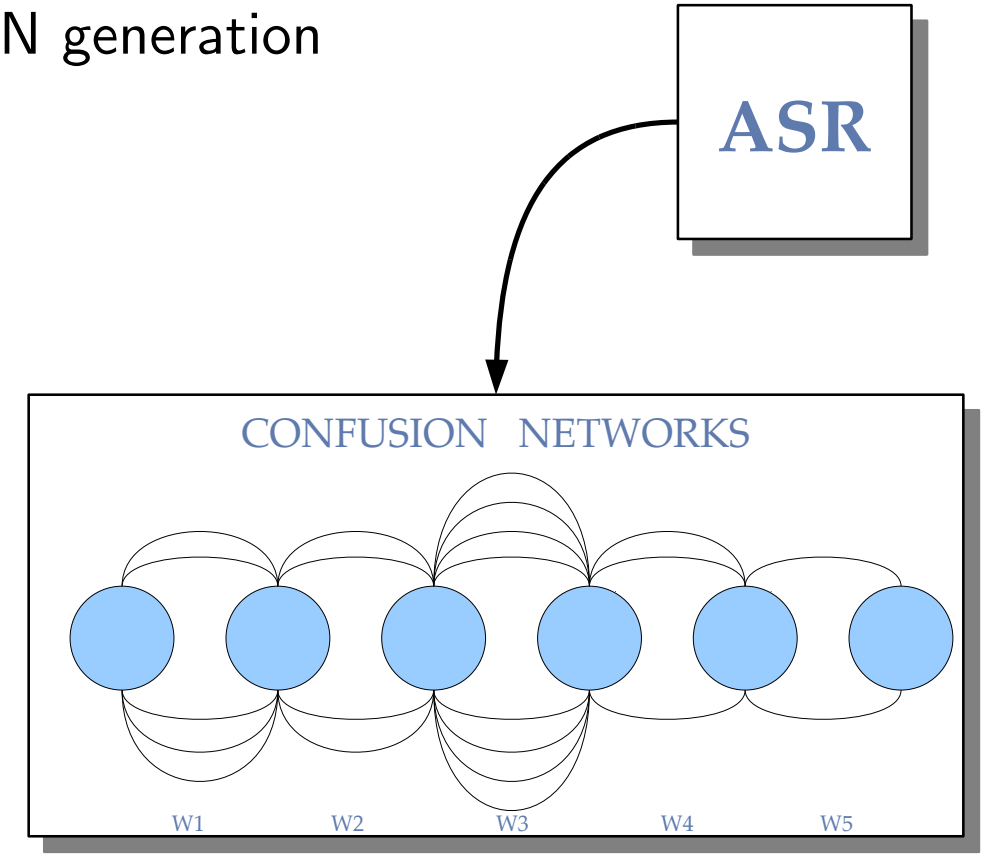
Description

- It is a semi automatic transcription task, in 2 steps:
  - human correction
  - a fast decoding pass
- ASR system evaluates a lot of alternatives paths
- Different alternatives could be proposed to the transcriber
- We use Confusion Network: more readable than lattice

Interactive decoding

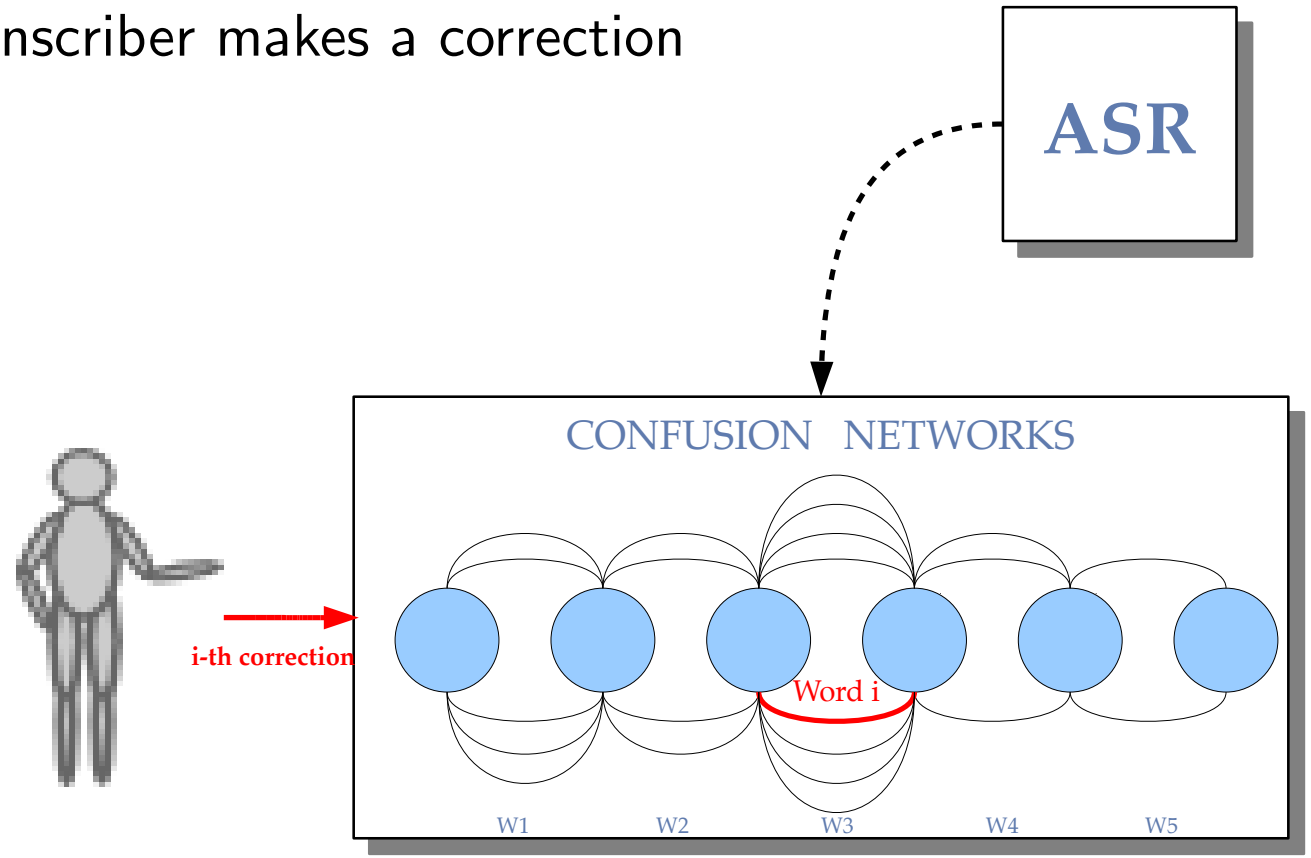
⇒First pass and CN generation





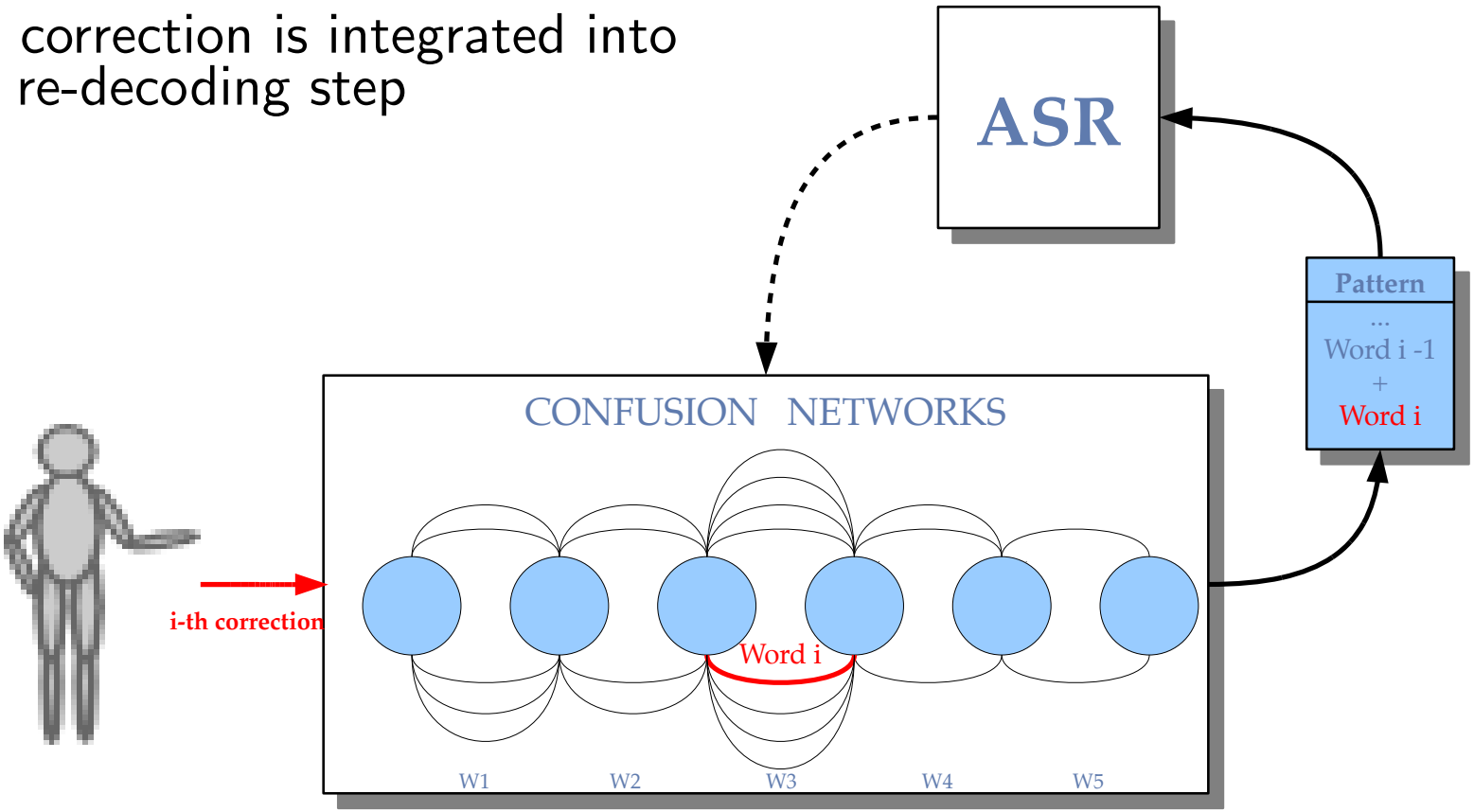
Interactive decoding

⇒Transcriber makes a correction



Interactive decoding

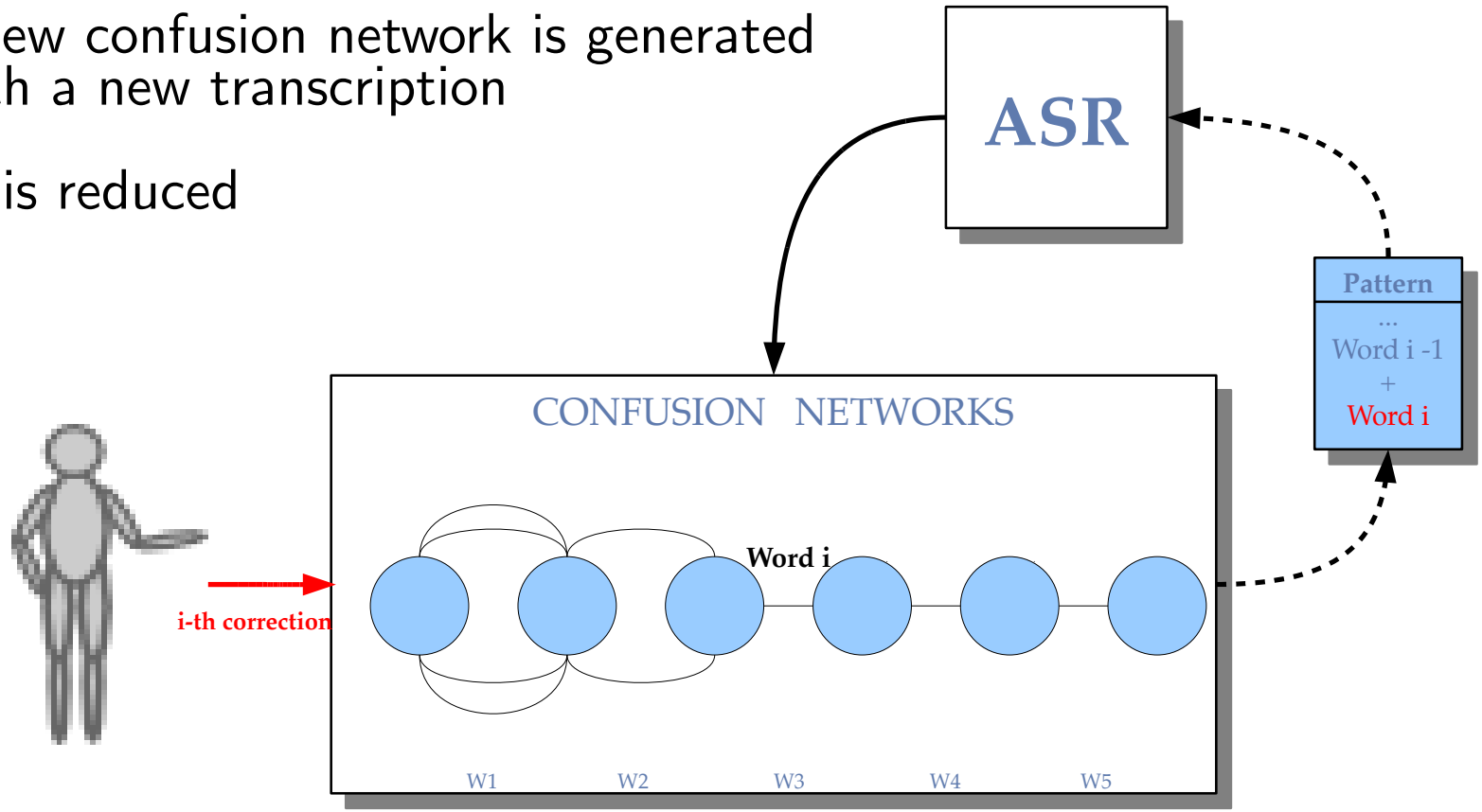
⇒The correction is integrated into the re-decoding step



Interactive decoding

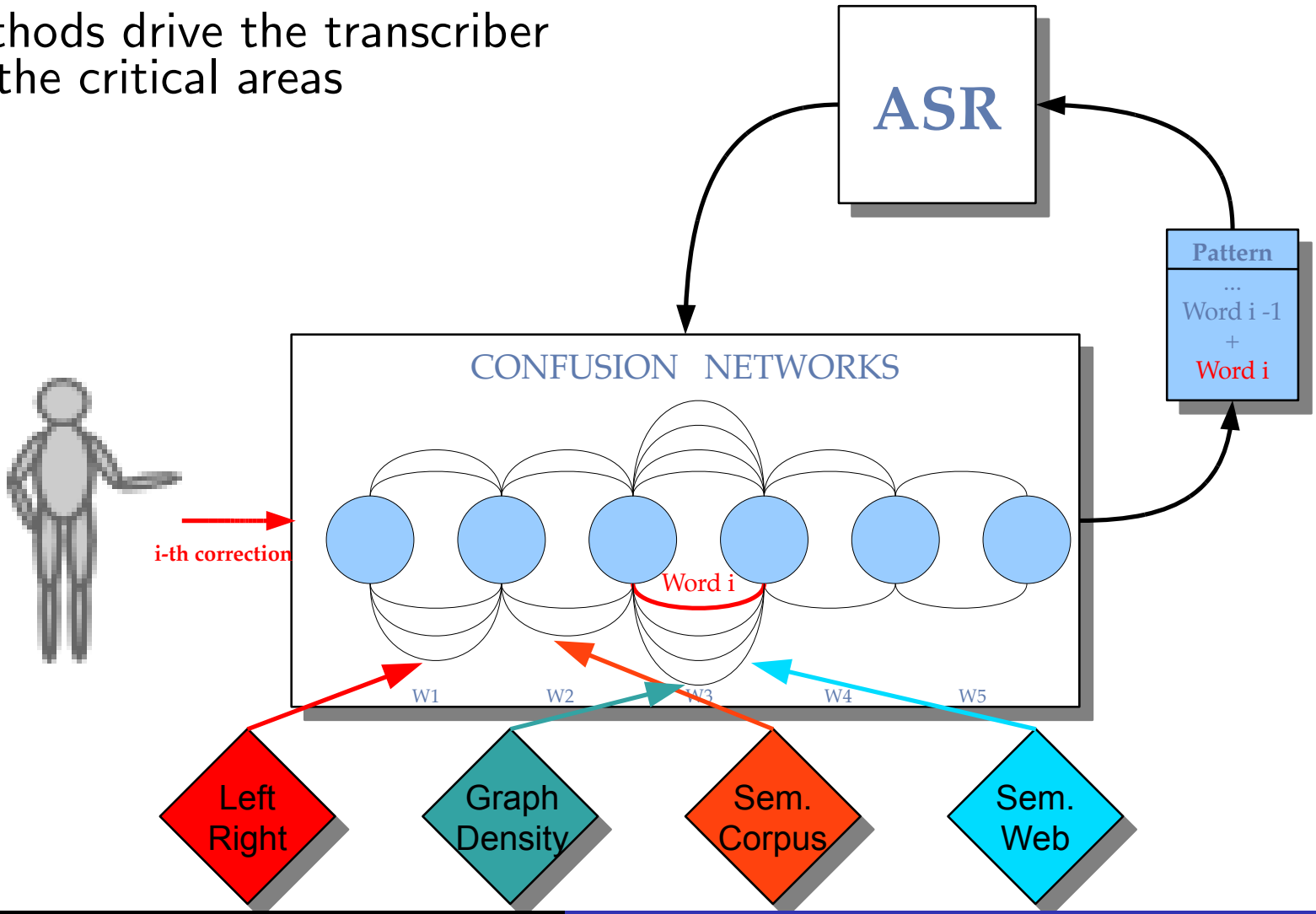
⇒A new confusion network is generated with a new transcription

⇒CN is reduced



Interactive decoding with driving strategies

⇒Methods drive the transcriber to the critical areas



Driving Strategies

Left-Right

- In the reading direction
- A normal strategy for the transcriber
- Drives on the left to the right

Graph density

- Numerous methods use graph density as a confidence measure
- The deepest part of a graph is a critical area where system has trouble to choose between a large number of hypotheses
- Graph density drives toward the widest section of the Confusion Network

Driving Strategies - Semantic consistency

2 methods are used: based on Corpus and Web

$\Rightarrow$ Each segment is split in small windows (10 relevant words)

$\Rightarrow$ The transcriber is driven to the lowest score window

Corpus criterion

- Principle: find in the corpus the closest newswire
- Based on a large corpus of newswires: Gigaword
  - 2 millions of newswires - 250 millions of sentences
- Corpus score is performed by the *Cosine* metric

Web criterion

- Web has a large language coverage
- Each Web documents is regarded as a bag-of-words
- Web score: words co-ocurrence probability on the Web

Experiments - Protocol

Broadcast news system

- LIA broadcast news system: SPEERAL
- Development framework of the ESTER campaign
  - 8 hours from 4 different radio stations
- System on first pass: 32.6% Word Error Rate
  - 2 x Real Time
  - without speaker adaptation
  - first pass produces Confusion Networks
- Transcription is automatically split according to:
  - speaker turns
  - silence areas
  - length (30 seconds maximum)

Experiments - Protocol

Interactivity

- Corrections are simulated by *Sclite*
- $WER = \frac{confusion\ +\ insertion\ +\ deletion}{\#word\ \ number}$
- Re-decoding on Real Time system

Results

- Corrections start from the ASR transcriptions
- The baseline: **Human only** (without interactive decoding)
- Global WER evaluated for each correction
- 2 classes: below and above 40% WER

Profit WER according to the manual correction

WER of corrections for initial transcriptions of WER below 40%.

# c/segment	1	3	10	20
Human only	25.22	22.98	17.23	9.44
LR-ID	24.28	**20.82**	**11.88**	**5.26**
GD-ID	26.58	25.38	16.62	11.76
Corp-ID	**23.90**	21.15	13.93	8.51
Web-ID	24.33	21.10	12.21	7.40

ID: Interactive Decoding

Profit WER according to the manual correction

WER of corrections for initial transcriptions of WER below 40%.

# c/segment	1	3	10	20
Human only	25.22	22.98	17.23	9.44
LR-ID	24.28	**20.82**	**11.88**	**5.26**
GD-ID	26.58	25.38	16.62	11.76
Corp-ID	23.90	21.15	13.93	8.51
Web-ID	24.33	21.10	12.21	7.40

ID: Interactive Decoding

Profit WER according to the manual correction

WER of corrections for initial transcriptions of WER below 40%.

# c/segment	1	3	10	20
Human only	25.22	22.98	17.23	9.44
LR-ID	24.28	20.82	11.88	5.26
GD-ID	26.58	25.38	16.62	11.76
Corp-ID	**23.90**	21.15	13.93	8.51
Web-ID	24.33	21.10	12.21	7.40

ID: Interactive Decoding

Profit WER according to the manual correction

WER of corrections for initial transcriptions of WER above 40%.

# c/segment	1	3	10	20
Human only	55.91	54.05	47.81	40.14
LR-ID	54.95	49.77	37.71	**25.36**
GD-ID	57.51	53.52	44.05	36.99
Corp-ID	54.19	49.37	39.06	29.54
Web-ID	**51.88**	**48.32**	**37.49**	29.49

ID: Interactive Decoding

Profit WER according to the manual correction

WER of corrections for initial transcriptions of WER above 40%.

# c/segment	1	3	10	20
Human only	55.91	54.05	47.81	40.14
LR-ID	54.95	49.77	37.71	**25.36**
GD-ID	57.51	53.52	44.05	36.99
Corp-ID	54.19	49.37	39.06	29.54
Web-ID	51.88	48.32	37.49	29.49

ID: Interactive Decoding

Profit WER according to the manual correction

WER of corrections for initial transcriptions of WER above 40%.

# c/segment	1	3	10	20
Human only	55.91	54.05	47.81	40.14
LR-ID	54.95	49.77	37.71	25.36
GD-ID	57.51	53.52	44.05	36.99
Corp-ID	54.19	49.37	39.06	29.54
Web-ID	**51.88**	**48.32**	**37.49**	29.49

ID: Interactive Decoding

Conclusion

Interactive decoding conclusion

- Effectiveness of interactive strategies
- Global cost reducing
- Driving methods:
  - Graph density is rather inefficient
  - Left-Right is the best way to produce a perfect transcription
  - Semantic methods are effective for massively erroneous transcriptions
- Improvement of the semantic quality using semantic strategies
- Efficient way of correcting transcriptions dedicated to:
  - speech indexing
  - speech understanding

Conclusion

Thanks you for your attention !