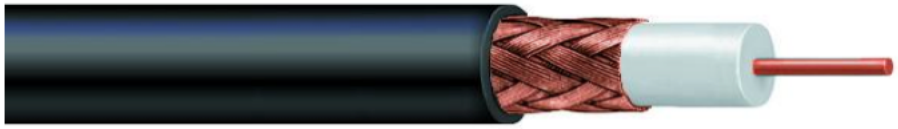
Engaging Content
Engaging People

Chat and chunk phases in conversation, and what they tell us about how to talk

Emer Gilmartin,
Speech Communication Lab / ADAPT Centre
Trinity College Dublin

Motivation



To better understand the bundle of signals in conversation

The Problem: Building social dialogue systems entails understanding of casual social dialogue but…

- Much linguistic theory is based on language similar to writing but highly unlike talk
  - regards spoken interaction as debased, chaotic
- SDS technology based on
  - Practical Dialogue Hypothesis (Allen, 2000)
  - Constraint introduced to make dialogue modelling tractable
- Much corpus study of spoken interaction based on Task-based Dialogue
  - Information gap activities – MapTask (HCRC), DiaPix (Lucid)
  - Meetings – AMI, ICSI
  - These are not corpora of casual or social talk

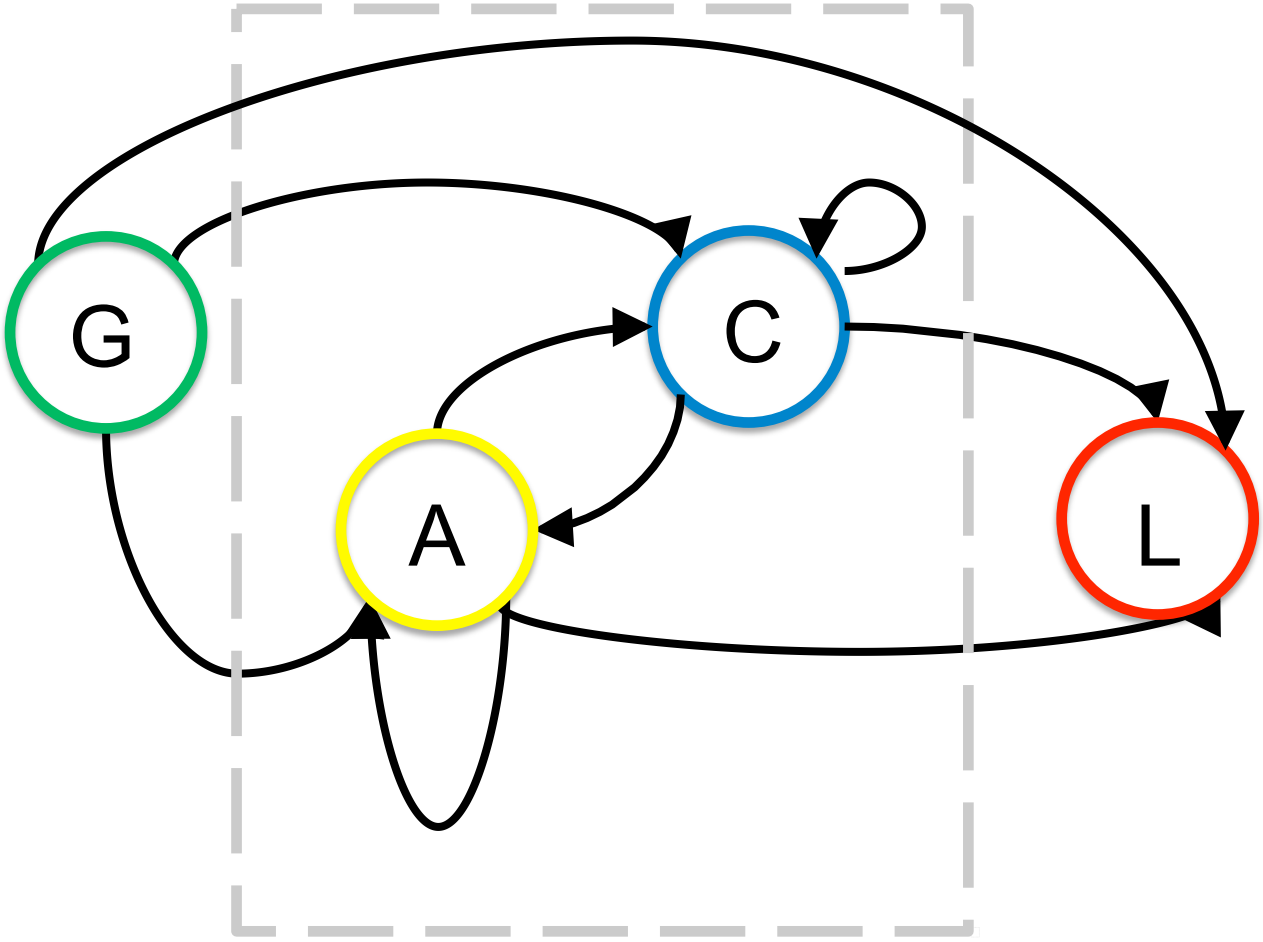
Transactional v Interactional Conversation

- Ordering a pizza (transactional)
  - performing a well-defined task
  - content (‘What?’) vital for success
- Chat with neighbour (interactional)
  - building/maintaining social bonds
  - social (‘How?’) very important
- Longer form (c 1 hr) casual conversation
  - ‘continuing state of incipient talk’
- Interested in the structure of such conversations

Social Talk

- Spoken interaction as social activity
  - Malinowski, Dunbar, Jakobsen, Brown and Yule
- Structure and Content
  - Smalltalk at the margins (Laver)
  - Chat and chunks (Slade & Eggins)
    - chat – highly interactive, many speakers contributing
    - chunks – gossip, narrative, dominated by one speaker
  - Phases – greetings, approach, centre, leavetaking (Ventola)
  - Multiparty (Slade)
- Problems:
  - much of this is theory, analysis by example
  - based on orthographical transcriptions
  - corpus based studies on transactional dyadic interaction, phonecalls…

Anatomy of casual conversation (Ventola model)



Engaging Content
Engaging People

Annotation of Greeting and Leave-taking in Social Text Dialogues Using ISO 24617-2

Emer Gilmartin, Brendan Spillane, Maria O'Reilly, Christian Saam, Ketong Su, Killian Levacher, Loredana Cerrato, Benjamin R. Cowan, Leigh M. H. Clark, Arturo Calvo, Nick Campbell, Vincent Wade

Outline

- Dialogue Act Annotation of Interactional Talk
- ADELE Corpus – Collection
- ADELE Corpus – Annotation
- GIL Acts added
- Discussion

ADELE Corpus

- Purpose
  - Training data for SDS
- Scenario
  - Dyadic text interaction
- Data Collection
  - 37 participants (26M/11F, age range 18-43)
  - native English speakers or IELTS 6.5
  - working/studying and living in Ireland
  - 193 completed dialogues were collected.
- Data
  - 40,297 words over 9231 turns or ‘utterances’ (~200, 50)
  - 7811 or 84.7% tagged with a single label
  - 1209 (13%) - two tags, 181 (2%) - three tags
  - 26 (0.3%) and 3 utterances had four and five tags respectively.

Annotation of social acts

- Many schemes include social acts
- In a survey of 14 schemes, Petukova found
  - 10 included greeting functions, 4 included introductions, 6 had goodbyes, 5 included apology type functions, and 5 contained thanking
- The Social Obligations Management dimension of the ISO standard contains nine communicative functions
  - initialGreeting, initialSelfIntroduction, returnSelfIntroduction, apology, acceptApology, thanking, acceptThanking, initialGoodbye, and returnGoodbye.

Annotation

- Used ISO Standard (with additions)
- Lexical tags for topic – PropQuestion[hobby]
- Informs that were not first mentions tagged as comments
- Noticed problems with SOM – greetings, introductions, leavetaking
- Greeting sections were marked as beginning with the first utterance of the conversation, and ending with the last production of a formulaic greeting/introduction or greeting/introduction response.
- leave-taking sequences from the first attempt to close the conversation to the final utterance of the conversation.

Greetings and Introductions

A: **Hi**

B: **Hello**, I'm Ann. I'm from Mexico City. Yourself?

A: **Hi** Ann, nice to meet you. I'm John.

B: **Hey** John, nice to meet you too. How are you today?

A: Good, good. You? I'm from Paris, living in London now.

B: I'm in good form!.

Greetings and Introductions

A: Hi

B: Hello, I'm Ann. I'm from Mexico City. Yourself?

A: Hi Ann, **nice to meet you**. I'm John.

B: Hey John, **nice to meet you too**. **How are you today?**

A: **Good, good. You?** I'm from Paris, living in London now.

B: **I'm in good form!.**

Greetings and Introductions

A: **Hi**

B: **Hello**, I'm Ann. I'm from Mexico City. Yourself?

A: **Hi** Ann, nice to meet you. I'm John.

B: **Hey** John, nice to meet you too. How are you today?

A: Good, good. You? I'm from Paris, living in London now.

B: I'm in good form!.

Additional GIL Acts

Table 1: Acts introduced for the ADELE annotation and common surface forms

Act	Common Examples	Functional Area
ntmy	Nice to meet you	Greeting
	Good to talk to you	Greeting
repNtmy	Nice to meet you too	Greeting
	Good to talk to you too	Greeting
hay	How are you?	Greeting
	How's it going?	Greeting
repHay	Fine	Greeting
greet	Hello	Greeting
	Hi	Greeting
wntmy	It was lovely to meet you	leave-taking
	Nice talking to you	leave-taking
repWntmy	It was nice to meet you too	leave-taking
	Likewise	leave-taking

Distribution of GIL acts

Table 2: Greeting, Introduction, and Leavetaking (GIL) Acts in ADELE corpus

Description	Count	% Corpus
All acts included in GIL sequences (GILseq)	2336	21.5
GILA: Only GIL Acts: GILseq Acts - Interloper Acts	1820	16.7
GILB: Only GIL acts without LeaveTaking Introductions: GILA - Leavetaking Introductions	1626	15
Social Obligation Management Acts (SOM) other than GIL	198	2

Interloping acts

A: Hi

B: Hello, I'm Ann. **I'm from Mexico City.** Yourself?

A: Hi Ann, nice to meet you. I'm John.

B: Hey John, nice to meet you too. How are you today?

A: Good, good. You? **I'm from Paris, living in London now**.

B: I'm in good form!.

LeavetakingIntro: Functional Segment Annotation

Great people here	Great people here		c
I will kind of miss it there	I will, kind of miss it there	71	acceptInvite
I will kind of miss it there	I will, kind of miss it there	71	c
Anyway Rob it's been great speaking with you	**Anyway Rob**		**leavetakingIntro**
Anyway Rob it's been great speaking with you	it's been great speaking with you		wntty
Sure, thanks, to you too. Talk to you again, soon.	sure		autopos
Sure, thanks, to you too. Talk to you again, soon.	thanks to you too	76	repwntty
Sure, thanks, to you too. Talk to you again, soon.	talk to you again soon		ttyl
Byes	Byes		goodbye

Current Work

- Validate GIL acts
- Work on centre phases
- Creation of spoken corpus – scenario

Engaging Content
Engaging People

Exploring Multiparty Casual Talk for Social Human-Machine Dialogue

Emer Gilmartin, Ben Cowan, Carl Vogel, Nick Campbell

Speech Communication Lab
Trinity College Dublin

Genre differences in spoken interaction?

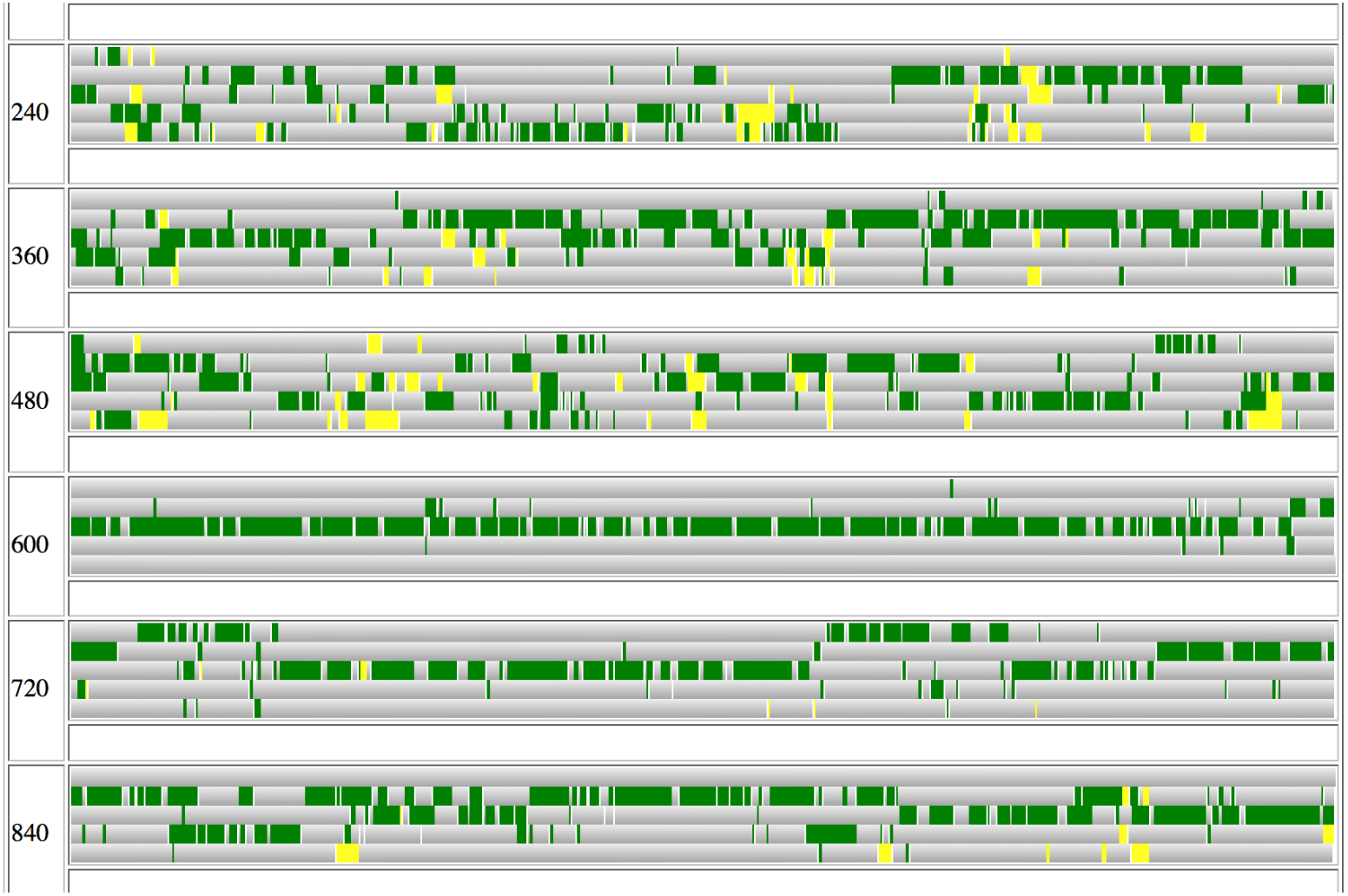
- Spoken interaction is situated
  - ‘speech-exchange systems’ (SSJ),
  - communicative activities (Allwood)

- Some low level mechanisms may follow universal patterns
- It is also possible that even basic interaction mechanisms such as turn-taking vary with the type and parameters of different interactions

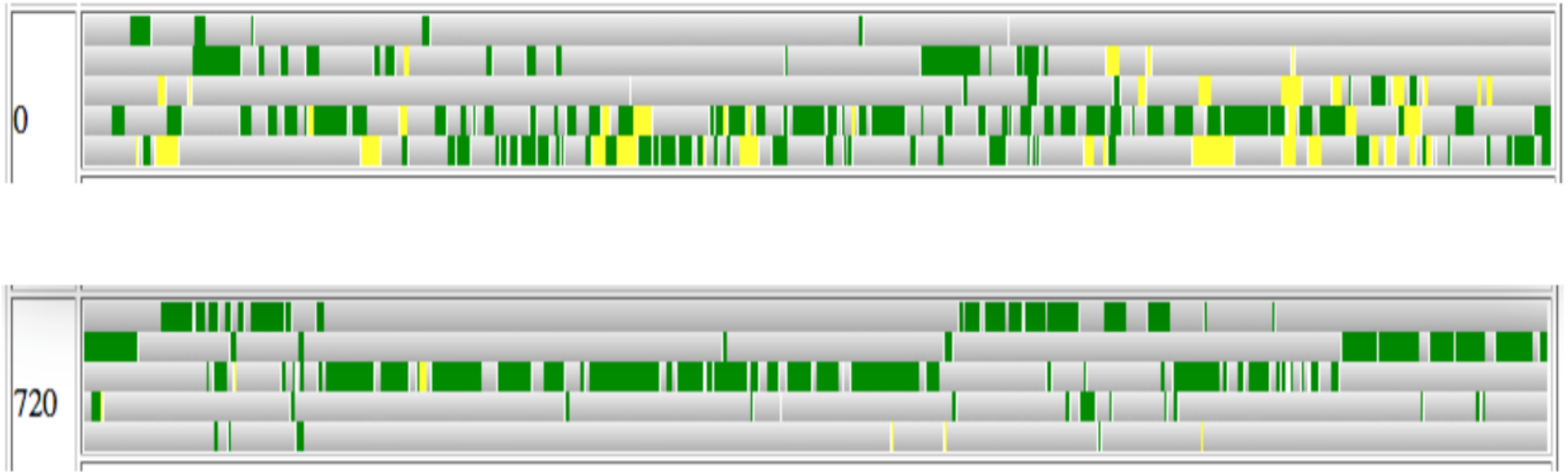
- What might vary?
  - Utterance/turn characteristics
  - Distribution of pauses/gaps/overlaps
  - ‘Disfluencies’, VSU’s, laughter…

- Explore different genres and use knowledge to inform design of interfaces

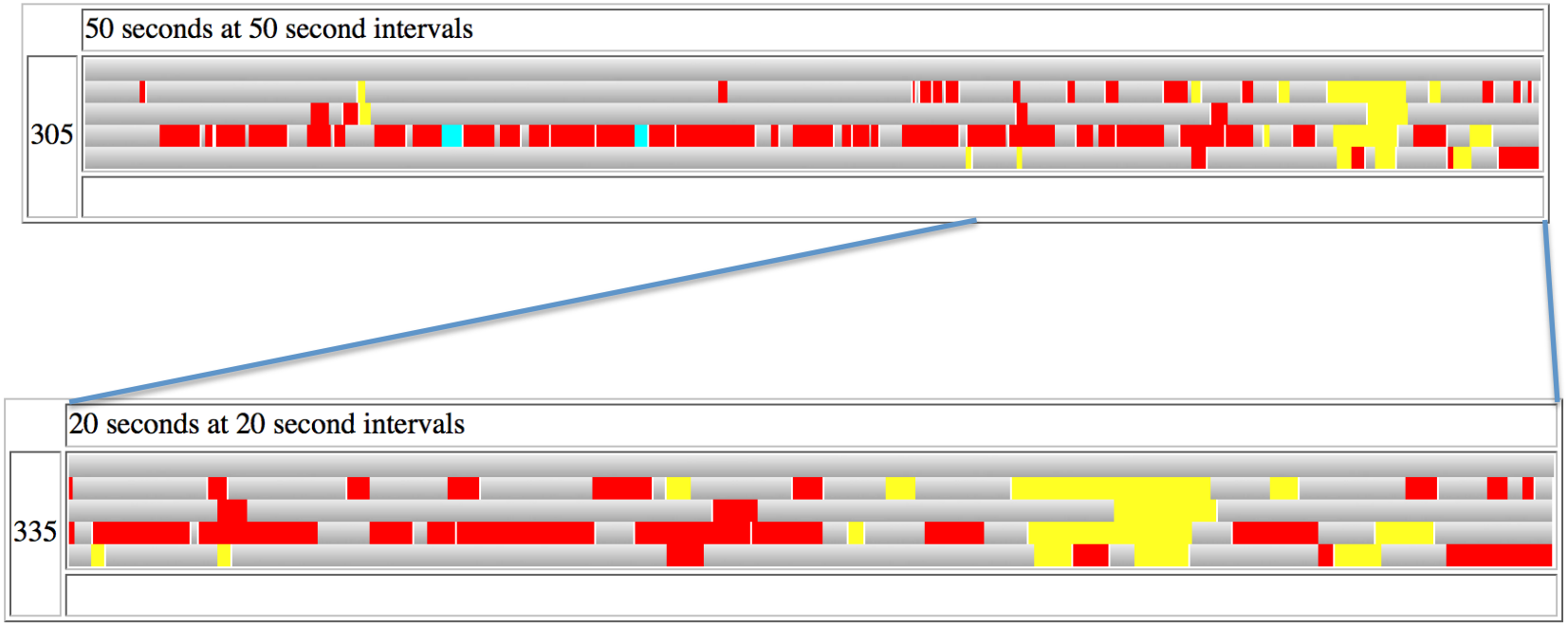
12 minutes from a 5-party casual conversation showing chat (240s-480s and chunk 480 – end) phases
Green-speech, yellow-laughter, grey-silence



Chat and Chunk



Chunk to Chunk Transition – more interaction and laughter at end of chunks



Question

Can chat and chunk phases be classified using acoustic/ discourse features?

Data and annotation

Corpus	Participants	Gender	Duration (s)
D64	5	2F/3M	4164
DANS	3	1F/2M	4672
DANS	4	1F/3M	4378
DANS	3	2F/1M	3004
TableTalk	4	2F/2M	2072
TableTalk	5	3F/2M	4740

Table 1. Source corpora and details for the conversations used in dataset

Chat/Chunk Results

Significant differences in:

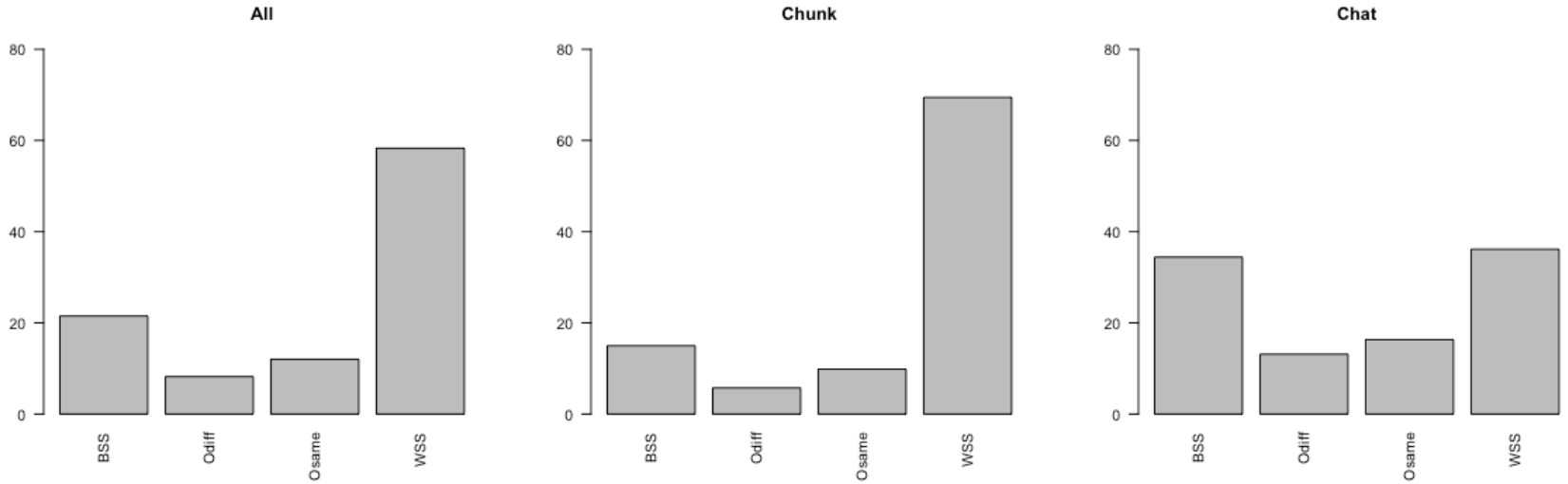
- Length – (chat more variable) gmean ~ 28s, chunk ~ 30s
- Distribution, more chat at beginning – c.8 minutes
- Laughter – over twice as much in chat – 9.7 vs 4%
- Gap lengths and distribution – WSS most common overall, more BSS in chat
- Overlap – more in chat, particularly more multiparty overlap
- Disfluency distribution, especially fp in chunks by role

Overlap and gap results

Speaker change: Between speaker silence (BSS) and between speaker overlap (Odiff)

Turn retention: Within speaker silence (WSS) and within speaker overlap (Osame)

Distributions differ between chunk and chat



Answer

Looks likely

Discussion

Important because;

- Need different timing modules for different phases
  - Many within speaker pauses in chunks are longer than between speaker pauses in chat so need different turntaking policies
- Suit different tasks – companion applications
  - System can recognise when to listen to a story (chunk)
- Aid comprehension – design educational dialogue in chunks

Current and Future Work

- Stochastic model
  - Preliminary results promising
- Goals
  - online classifier
  - incorporate in social dialogue system. CALL applications

Engaging Content
Engaging People

CARAMILLA

Emer Gilmartin
Jaebok Kim
Yong Zhao
Alpha Ousmane Diallo
Neasa ní Chiaráin
Ketong Su
Yuyun Huang
Benjamin R. Cowan
Nick Campbell

SLATE 2017

Introduction

Learning a language involves:

- acquisition and integration of a range of skills
- exposure and noticing

Language learning is *personal*

- neither classes nor ‘self-study’ are optimal– ‘spiky’ profiles

A human tutor aids learners by scaffolding

- providing a framework of engaging tasks suitable to the learner’s needs
- giving relevant and timely feedback
- monitoring learner progress, adapting tasks and delivery style to suit
- providing a speech model and a source of speaking practice

Learner Groups

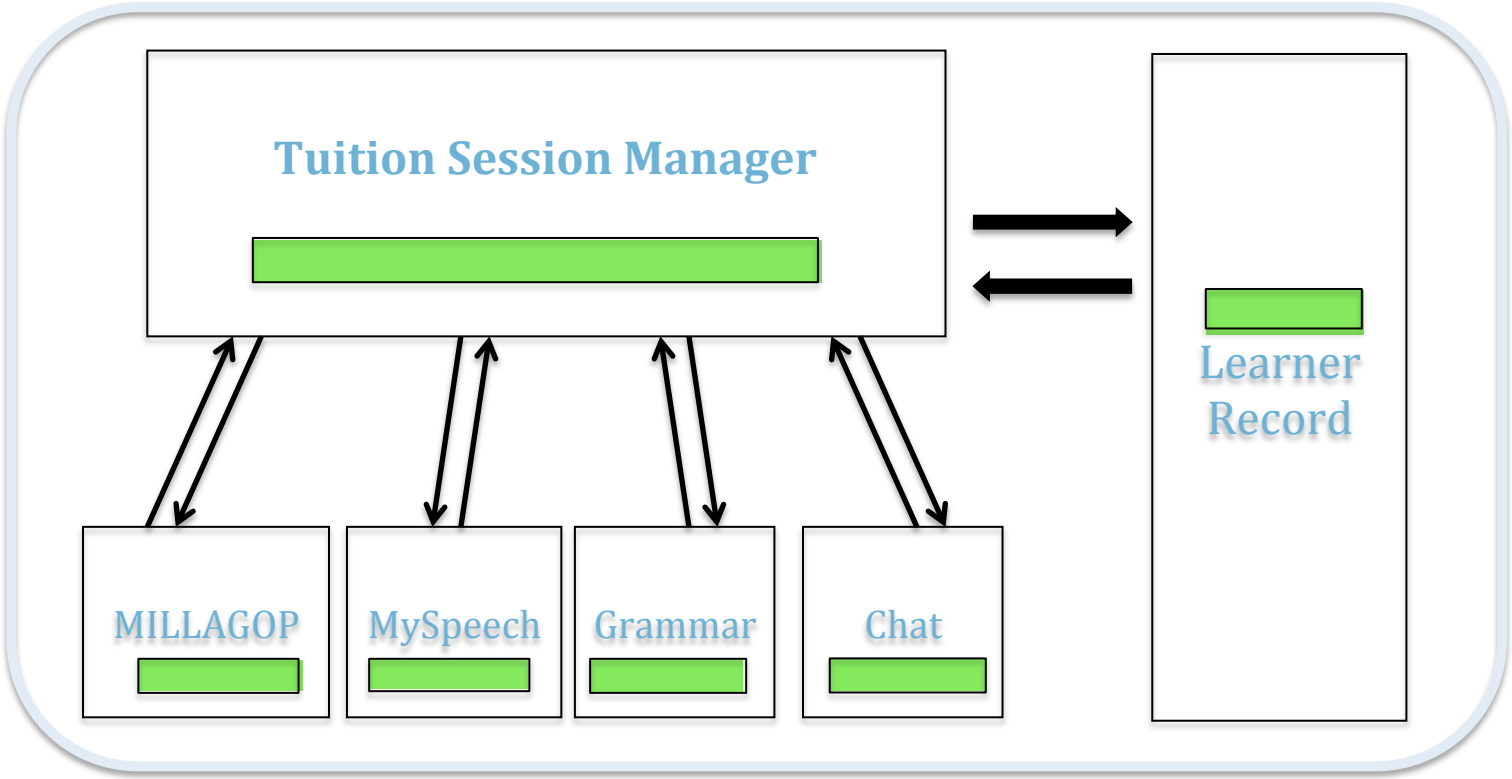
Adult Migrants/Refugees

- Very diverse group
- Spiky profiles
- Needs

Learners of Irish

- School age, teachers

MILLA – Enterface 2014



CARAMILLA

- Enterface 2016 – Twente
- 4 postgrad, 1 postdoc
- Two new modules
  - Pronunciation
    - Improved GOP
    - Common mistakes in English by language background
  - Expert knowledge from the website: http://www.tedpower.co.uk/
  - Made up 50 sentences: v|vet|w|wet|The vet likes wet weather
- Spoken Dictogloss
  - Highly adaptable tried and tested teacher led activity
  - Integrates reception and production skills

CARAMILLA Pronunciation

GOP corpus:

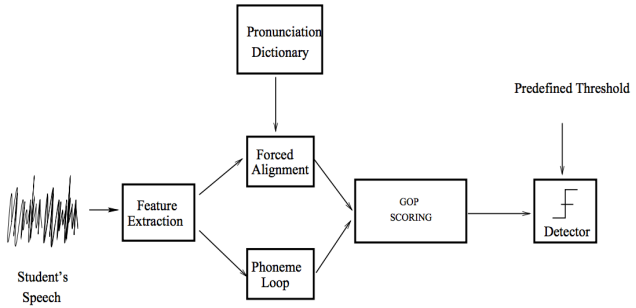
- Common mistakes in English by language background
- Expert knowledge from the website: http://www.tedpower.co.uk/
- Made up 50 sentences: v|vet|w|wet| The vet likes wet weather

Synthesis

- Text To Speech
  - English synthesizer: Cerevoice
  - Abair Irish synthesizer: Phonetics and Speech TCD

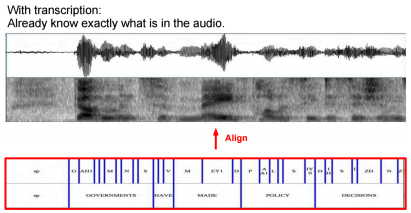
Summary of GOP

Calculate a distance between forced alignment and phone recognizer

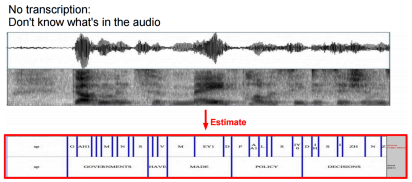


$$GOP(n) = \left|LLF(n) - LLP(n)\right|$$

Forced alignment



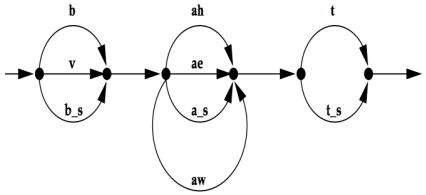
Automatic speech recognition



Using scores of both general pronunciation and explicit error model

Each mother language has its own error models (network of phones)

Chinese, Korean, Spanish, German, Arabic…



GOP(qi) + w * GOPE(qi)
[Witt 1999, Strik 2009]

Improved GOP (Jaebok Kim)

Previous works used duration of phones or sum of likelihoods of frames (general GoP)

Duration model does not give any ideas how good pronunciation is

General GoP score does not explain mistakes

We use likelihood of frames using explicit error and stress models

A recognized phone is further categorized into correct, insert, and deleted

It gives not only overall goodness of pronunciation

but also mistakes depending on mother tongue

Spoken Dictogloss

- What is it?
- Why is it useful?
- Implementation
  - Java-based
  - Voices
    - English - Cereproc Caithlin
    - Irish – ABAIR voices
  - Texts
    - Irish - Pre-loaded texts
    - English - Webscraping – Simple English Wikipedia

Dictogloss Example

	A	E	I	O	U
1	I	went	to	see	
2	Vincent				
3					
4		seniors	at	the	cinema
5				crunching	popcorn

Affective status recognition in CARAMILLA

Offline experiments using Theano

5 emotional speech corpora
LDC, VAM, SEMAINE, ENTERFACE, EMODB

LOSOCV for each corpus

365 acoustic feature set (IS09)

High vs. low arousal, negative vs. positive valence

DNN (365-1024-1024-2)

Impressive performances but huge variance over speakers
(Accuracy 100% - 30%)

Is it due to over-fitting or variation of speaker?

Spoken Dictogloss

Testing of Pronunciation and Dictogloss

Academic year 2017/18

Add tutor / peer learner avatars

Extend activities

Incorporate pre-existing curricula drawn from relevant ELPs (European Language Portfolios)

Free web resource

Engaging Content
Engaging People

Thank You

Questions?