

**"I want to tell you a joke...  
Are you ready?"**

## An introduction to the STANDUP Project

Graeme Ritchie

University of Aberdeen

Annalu Waller

Rolf Black

Dave O'Mara

University of Dundee

Helen Pain

Ruli Manurung

University of Edinburgh



# Outline

---

1. Background
2. Aim of Project
3. Functional requirements
4. User requirements
5. The Development of the Lexicon
6. Designing the Interface
7. STANDUP demonstration



## 1.1 Background

---

- Conversational narratives (story)
  - Different types of story, e.g. jokes
  - Punning riddles have question-answer format
- Role of jokes in language development
  - pragmatics ⇒ turn taking, initiation etc.
    - ⇒ early development
  - vocabulary acquisition
    - ⇒ word play (ambiguity)
    - ⇒ phonetic and semantic awareness



## 1.2 Background

---

- Humour research
  - laughability vrs understanding
  - comprehension studies begin at 5 years
- Computational humour
  - JAPE
  - STANDUP



## 1.3 Background

---

- Humour and Augmentative and Alternative Communication (AAC)
  - prestored jokes
  - pragmatics
  - little opportunity for independent vocabulary acquisition and word play
  - research mainly into enjoyment and fun



## 2 Aim of the project

---

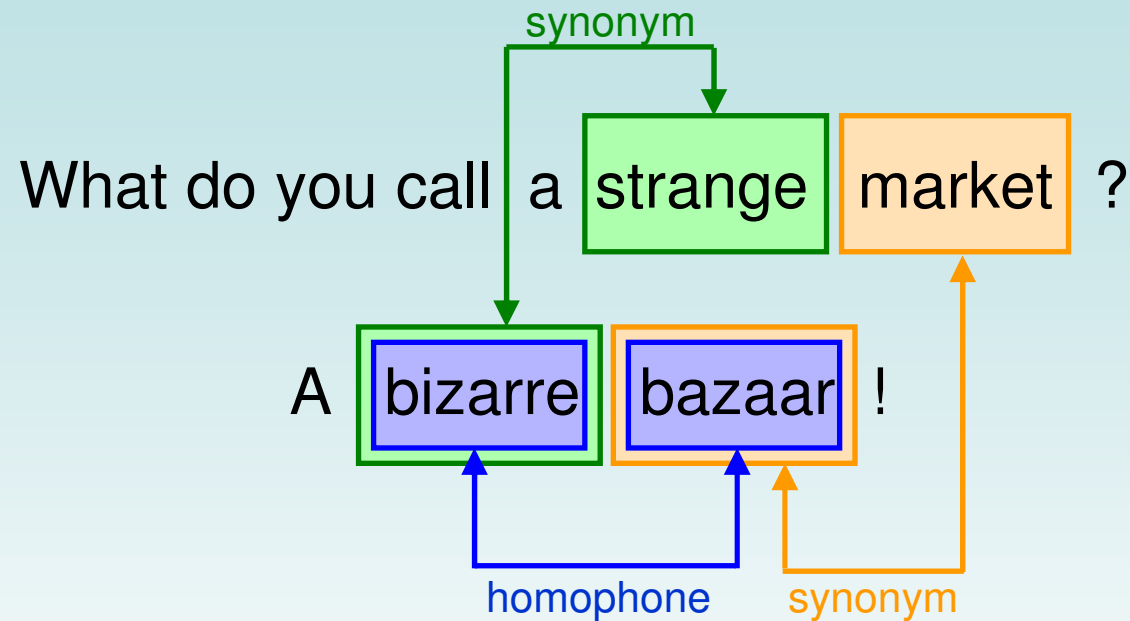
### System To Augment Non-speaker's Dialogue Using Puns

- to provide a language playground through the generation of novel puns!
- could we develop an interface to a joke generator for children with complex communication needs (CCN)?



### 3 Functional requirements

#### Joke Generation System



- Based on JAPE developed by artificial intelligence researchers
- System creates new jokes (not pre-stored)
- Jokes can be saved by user

## 4 User requirements - General

---

- Accessible to wide range of users
  - Scanning & direct access
  - Reduced selections
  - Recovery – e.g. “go back”, “go home”
- Different levels of access to manage language skills and possible progressions:
  - Task difficulty (keyboard input harder than simple selection of words)
  - Joke type (partial word matching harder than homophone substitution)
  - Vocabulary (measured by word frequency)



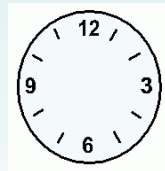


## 4 User requirements - Vocabulary

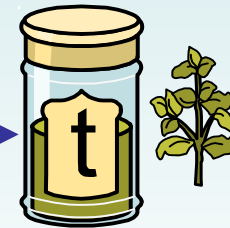
---

- Appropriate for Young Children
  - No Unsuitable Words
- Appropriate for Children with Emerging Literacy
  - Preference for Familiar Words
  - Speech output
  - Symbol support using Rebus and PCS symbol libraries e.g.:

*“time”*



*“thyme”*



- Access to jokes using subjects – lexicon grouped into subject-areas (topics) and clustered into a hierarchy

# 5 The Lexicon: Requirements

---

## Lexicon Information

- Part-of-speech (POS) tags
- Phonetic spelling, for computing:
  - homophones                      time                      thyme
  - rhyme                              pub                      tub
  - spoonerism                      bare/spank                      spare/bank
- Compound nouns and their components  
e.g. long time, traffic jam
- Distinct senses of a word/phrase,  
e.g. match=sporting event, match=ignition stick
- Semantic relations:
  - synonyms                      strange                      bizarre
  - hypernyms                      thyme                      herb
  - meronyms                      traffic                      car



# 5.1 The Lexicon: Resources

---

## Word collections / Lexica

- **WordNet**: 200k word senses, synonyms (**synsets**), hypernym hierarchy, meronyms.
- **Unisyn**: pronunciation dictionary, assigning phonetic strings to >115k word forms. Edinburgh accent used.
- **SemCor**: subset of Brown corpus with >230k WordNet sense-tagged words. >35k WordNet entries have SemCor frequency>0.

## Problems:

- **Unsuitable**
- **Unfamiliar**
- **Americanisms (e.g. baseball information)**



## 5.2 The Lexicon: Additional Resources

---

- **MRC Psycholinguistic Database:** various ratings relevant to familiarity.
- **BNC Spoken Corpus:** frequency ratings for compound nouns.
- **Widgit conceptcodes:** >11k concepts linked to >6k Widgit Rebus symbols, >4k Mayer-Johnson PCS symbols.
- **Schonell spelling lists:** spelling list of >3k words for children aged 7-12. Used as preferred source of “familiar” words.



## 5.3 The Lexicon: Data Preparation

- **Database**

Lexical resources

- **WordNet+Unisyn**

Disambiguation

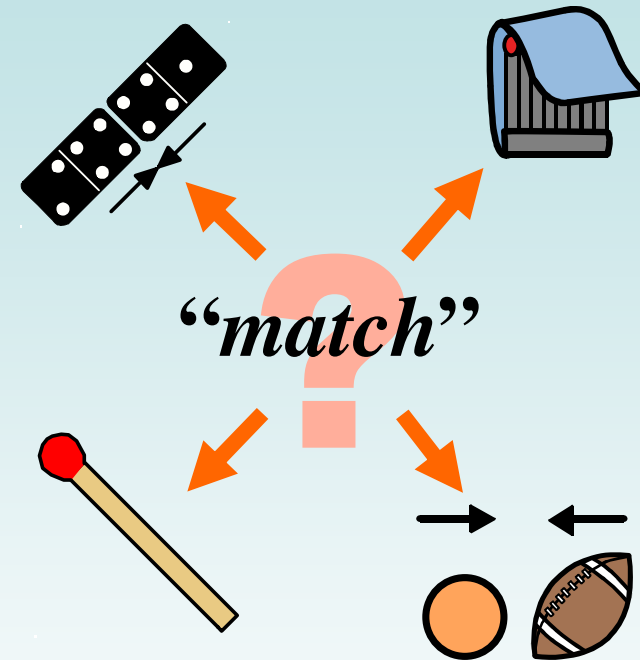
- **Phonetic relations**

Similarity, rhymes,  
spoonerisms

- **Familiarity scoring**

Word-sense: “F-score”

Prioritising and combining sources  
(MRC>Schonell>Widgit>Semcor)



## 5.4 Data Preparation (2)

Auto disambiguation:

- **Widgit & Schonell**
- **MRC** database

Manual disambiguation:

- **Widgit** conceptcodes
- **Schonell** spelling list

Custom authoring tools were used.

The screenshot shows the SymbolToolNoSQL application interface. It is divided into several panels:

- Widget details:** Contains fields for 'No.' (2212 out of 2311), 'Widget concept code' (3090040047000000), 'Widget part of speech' (noun), 'Widget topic (> subtopic)' (non category > misc items), 'Widget words' (match, matches), and 'Widget symbols' (showing a matchstick image).
- STANDUP details:** Contains 'STANDUP wordforms' (match(wf060945), matches(wf060954)) and a table of 'STANDUP concepts'.
- Concept details:** Contains 'Concept ID' (cn103589363), 'Gloss' (lighter consisting of a thin piece of wood or cardboard tipped with combustible chemical; ignites with friction; "he always carries matches to light his pipe"), and 'Lexemes' (friction match(n,k035585), lucifer(n,k035584), match(n,k035583)).

At the bottom, there is a 'Currently loaded file' field showing 'C:\Documents and Settings\ruli\Desktop\checked\_cc\_5.xml' and a 'Controls' section with buttons for 'Load...', 'Prev <', 'Jump To...', 'Next >', 'Save...', and 'Exit'.

Match	Flag	Concept ID
<input checked="" type="checkbox"/>		cn103589363
<input type="checkbox"/>		cn103589691
<input type="checkbox"/>		cn103589862
<input type="checkbox"/>		cn103586675
<input type="checkbox"/>		cn107014432
<input type="checkbox"/>		cn107506211
<input type="checkbox"/>		cn109019701
<input type="checkbox"/>		cn109270000
<input type="checkbox"/>		cn112830718
<input type="checkbox"/>		cn200404886

## 5.5 The Lexical Knowledge Base (1)

---

### Core lexicon:

- 130k lexemes
  - ✓Sense (synset)
  - ✓Part of speech
  - ✓F-score (in [0,1])
- 79k wordforms
  - ✓Orthography
  - ✓Phonetic spelling
- 32k compound nouns
  - ✓Head
  - ✓Modifier
- 85k concepts
  - ✓WordNet gloss
- 65k hypernym pairs, 7.5k meronym pairs
- 10k Wigit-to-WordNet matches  
(>8k POS matched)
- >500k phonetic similarity ratings (in [0.75,1])

### Lexical relations stored in additional cache tables:

- Syntactic:  
noun, verb, adj, mod, compound
- Semantic:  
synonym, hypernym, meronymy, alternate meaning
- Phonetic:  
homophone, rhyme, spoonerism, prefix, suffix

## 6.1 Designing the Interface


---

- User Centred Design
- Clinicians and adults who use AAC
  - Focus groups with clinicians
  - One to one sessions with adults
- Requirements gathering
  - Paper prototypes
- Design
  - Paper prototypes
  - Low fidelity mock ups





# Highly literate prototype



Return

Type in your joke keyword...

Your system selected these words please choose one:

Bees

sting  
wasp  
yellow  
black  
honey

Your system has suggested three puns-choose your favourite and try it out on friends

**Pun 1**  
Q. Why do bees have sticky hair?  
A. Because they use honey combs.

**Pun 2**  
Q. Why do bees eat sticky cookies?  
A. Because they use honey jars.

**Pun 3**  
Q. Why are bees sweet talkers?  
A. Because they are full of honey.

Speak Save **Speak** Save Speak Save

# J evaluating paper Prototype



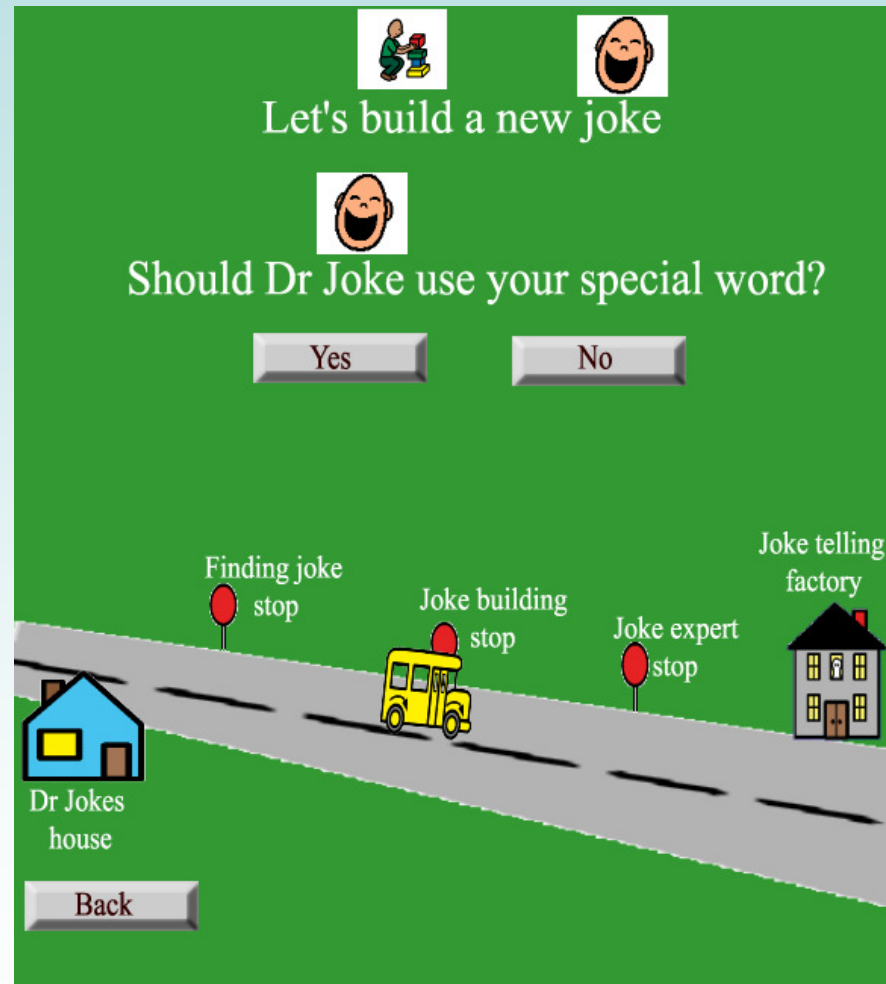
---

## Results Paper prototypes

- suggested too much reliance on text
- needed picture language interface
- suggested various ways of use



# Pictorial Journey Metaphor



# Pictorial Journey Metaphor

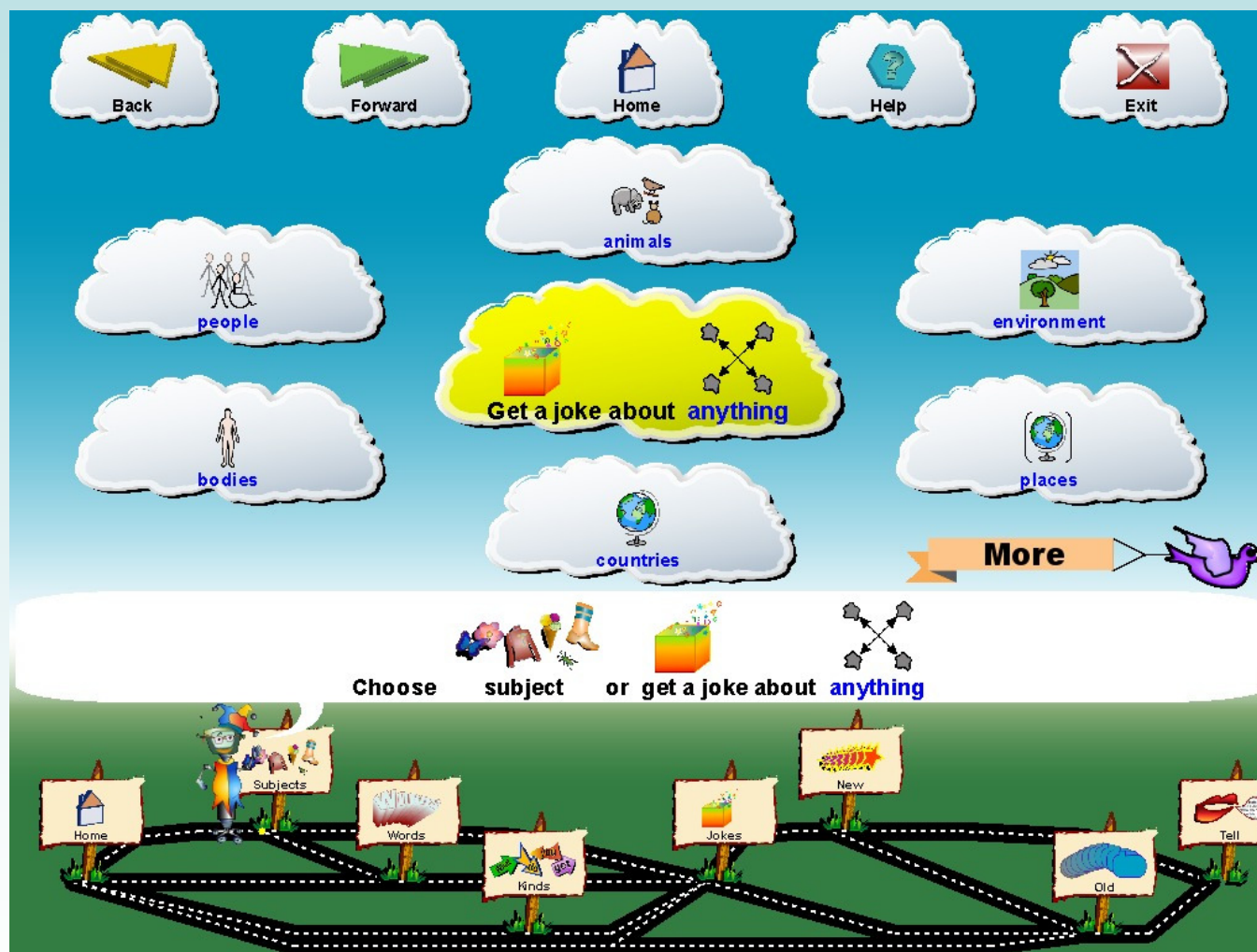




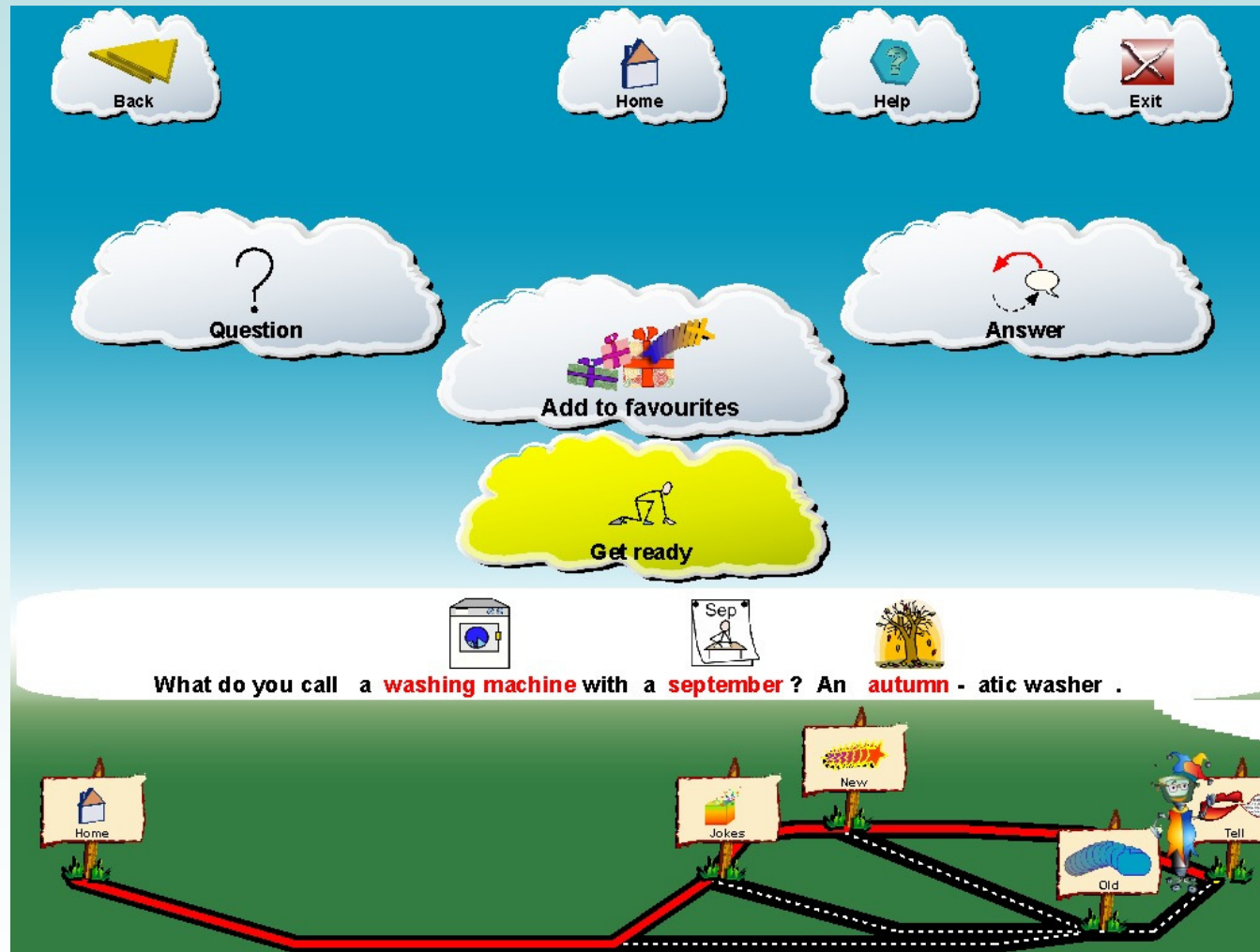
# K evaluating metaphor design



## 6.2 Using STANDUP – Screen layout



## 6.3 Using STANDUP - "Are you ready?"





## 6.3 Using STANDUP - “Are you ready?”



## 6.3 Using STANDUP - "Are you ready?"



An **autumn** - atic washer .



Say again

OK

## 6.4 Using STANDUP - Scanning

