Marco Aldo PICCOLINO BONIFORTI
KF University – Graz

# TEXT-TO-SPEECH: the Linguistic Perspective

Graz University of Technology, 29th October 2003

# Outline

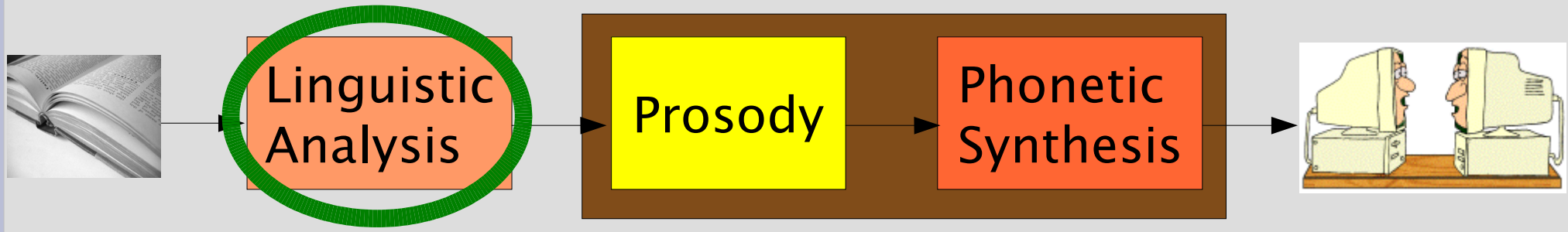
- Speech Synthesis: 2 ways
- TTS: Basic Components
- TTS: Complexity of Analysis
- Linguistic Analysis
  - Lexical Analysis
  - Morphological Analysis
  - Word Context Analysis
  - Phonological & Accent Analysis
- Summing Up
- Resources

# Speech Synthesis: 2 ways

- **Text-to-Speech (TTS)**
- Input: ortographical text
- Method: conversion to speech of any kind of text
  - General text
  - Document structure
  - Markup
- Features:
  - Flexible
  - Exposed to error
- Usage: general speech synthesis purposes

- **Concept-to-Speech (CTS)**
- Input: linguistic representation
- Method: conversion to speech of concepts
  - Semantics
  - Pragmatics
  - Discourse knowledge
- Features:
  - Specific
  - Reliable (esp. prosody)
- Usage: dialogue systems, machine translation, etc.

# TTS: Basic Components




- Each component may consist of different modules

# TTS: Complexity of Analysis

- "Bei der Wahl am 12.3.1998 gewann Tony Blair ca. 52% der Wählerstimmen."

- st should be realized as [ʃt] and not as [st] (see "Erstimpfung")
- Tony Blair should be recongnised as foreign name entity
- 52%, ca., 12.3.1998 should be treated as regular words
- Punctuation (.) has here 3 different meanings:
  - 12.3.1998 part of date
  - ca. abbreviation
  - Wählerstimmen. sentence boundary

- **Text-to-Speech conversion is NOT a trivial task: Linguistic Knowledge is necessary**

# Linguistic Analysis: Relevant Components

- Lexical Analysis
- Morphological Analysis (Derivation, Composita)
- Word Context Analysis
  - Syntactic Agreement
  - Syntactic Phrases/Sentences
  - Prosodic Phrases/Sentences
  - Sentence mode
- Phonological & Accent Analysis (Out-of-Lexicon words)

# Lexical Analysis

- **Lexicon**: a dictionary. It contains, for each entry, relevant informations such as:
  - **Part of speech** (POS): name, verb, adjective etc.
  - **Phonetic transcription**
  - **Relevant grammatical categories**: number, gender, etc.

- Different kinds of lexica:
  - **Whole form**: for each lexeme, all possible word forms are listed (e.g. “gehe, gehst, geht, ...”)
  - **Word stem:** for each lexeme, just the basic form and a general paradigm to be followed are listed (e.g. “geh-” + regular verbal flexion).

# Lexical Analysis: Special Items

- Word-stem lexica: special word lists are created for:
  - Non-flectional words
  - Geographical nouns, proper nouns, foreign words and other special categories
  - Abbreviations, acronyms etc.
  - Numbers. They are associated with more or less complex linguistic models

# Lexical Analysis: Lexical Entries

- Lexical entries: two examples

**Festival 1.4.0**

```
( "walkers" n ((( w oo ) 1) (( k @ z ) 0)) )
( "present" v ((( p r e ) 0) (( z @ n t ) 1)) )
( "monument" n ((( m o ) 1) (( n y u ) 0) (( m @ n t ) 0)) )

( "lives" n ((( l ai v z ) 1)) )
( "lives" v ((( l i v z ) 1)) )
```

**CELEX**

SHOW

| Headword | PhonStrsCPA | MorphStructure | MorphC | Cla | Freq |
|---|---|---|---|---|---|
| celebrant | ´sE.lI.br@nt | ((celebrate),(ant)) | Vx | N | 6 |
| celebrated | ´sE.lI.bre/.tId | ((celebrated)) | V | A | 158 |
| celebration | "sE.lI.´bre/.Sn, | ((celebrate),(ion)) | Vx | N | 221 |
| celibacy | ´sE.lI.b@.sI | ((celibate),(cy)) | Ax | N | 13 |
| celibate | ´sE.lI.b@t | ((celibate)) | A | N | 2 |
| cell | ´sEl | (cell) | N | N | 1216 |
| cellar | ´sE.l@r* | (cellar) | N | N | 225 |
| cellarage | ´sE.l@.rIJ/ | ((cellar),(age)) | Nx | N | 0 |
| cellist | ´T/E.lIst | ((cello),(ist)) | Nx | N | 6 |
| cello | ´T/E.lO/ | (cello) | N | N | 36 |

V

START GOTO ZOOM HIDE COUNT PRINT SAVE QUERY

Page: 1 (2) Columns: 6 (6) Tempo: 10 Count:

# Morphological Analysis

- Relevant for inflectional languages (e.g. German) as well as polysynthetic and agglutinative languages
- **2** different processes:
  - **Derivation**: word stem + adfixes

    Example: schlag / vor-ge-schlag-en
  - **Composition**: 2 ore more word (stem)s are joined

    Example: Kopf + Hörer / Kopfhörer

# Morphological Analysis: Composition

- Languages like German are very productive in word composition - no lexicon could include every possible realisation: morphological analysis is needed.

- Morphological analysis is problematic: more than one (more or less plausible) analysis is possible.
  - Example: Wählerstimmen
    - wähl [Vb-stem] + erst [Adj-stem] + imme [Nom-stem] + n [pl]
    - wähler [Vb-stem] + st [2per-sg] + imme [Nom-stem] + n [pl]
    - wähler [Nom-stem] + stimme [Nom-stem] + n [pl]
  - Solution:
    - Statistical methods: use of language corpora
    - Consideration of the syntactic context

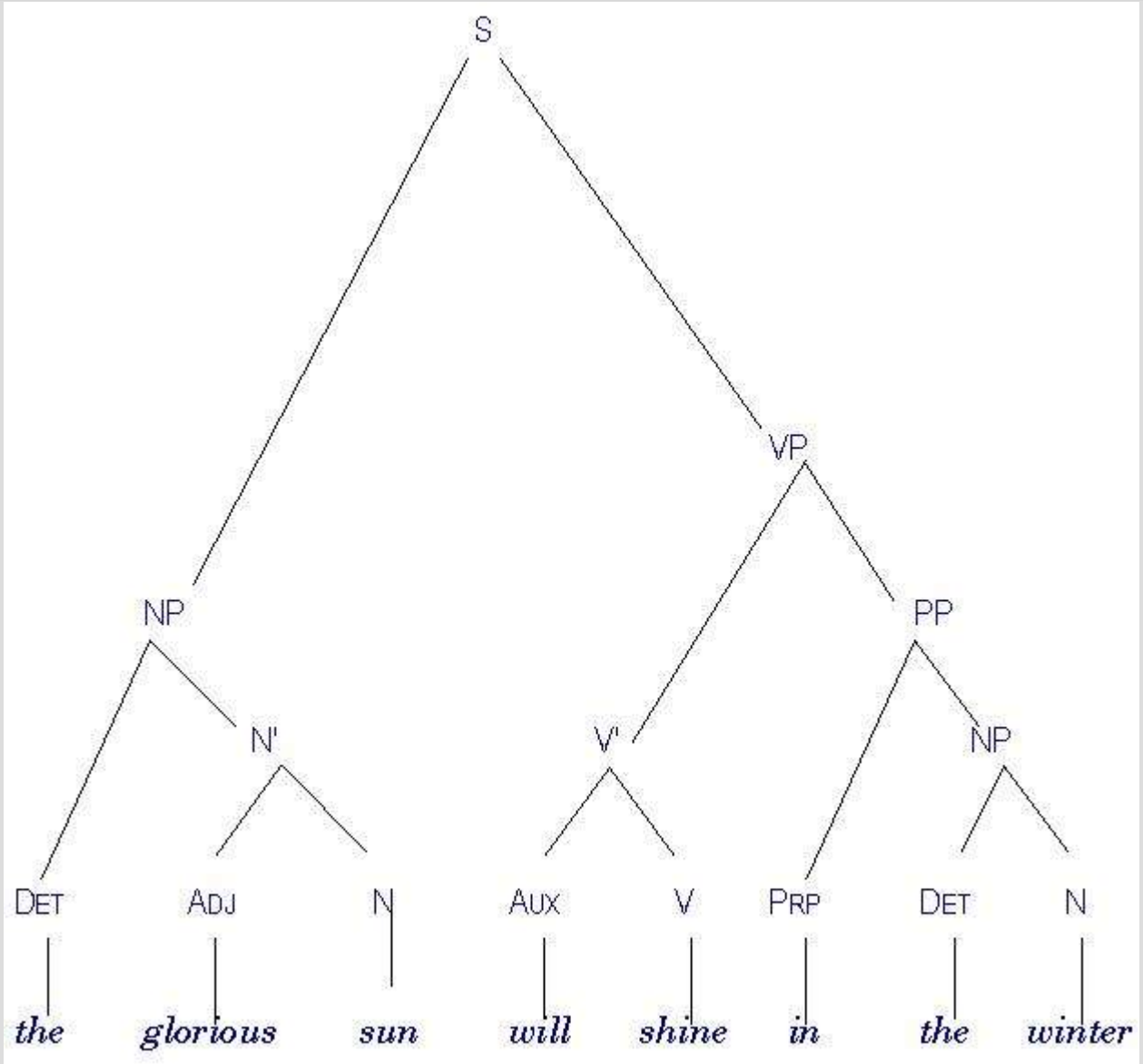
# Word Context Analysis: Syntax

- **Syntactic agreement:**
  checks for grammatical congruence between linked words (e.g. Det + Adj + Nom)

  Example: der [Art.Sing.Masc.N.] + Hund [Nom.Sing.Masc.N.]

- **Syntactic phrases/sentences:**
  phrase and sentence boundaries are individuated through punctuation and syntactic structure. The last one is usually realized in form of hierarchical representations

# Word Context Analysis: Syntactic Analysis

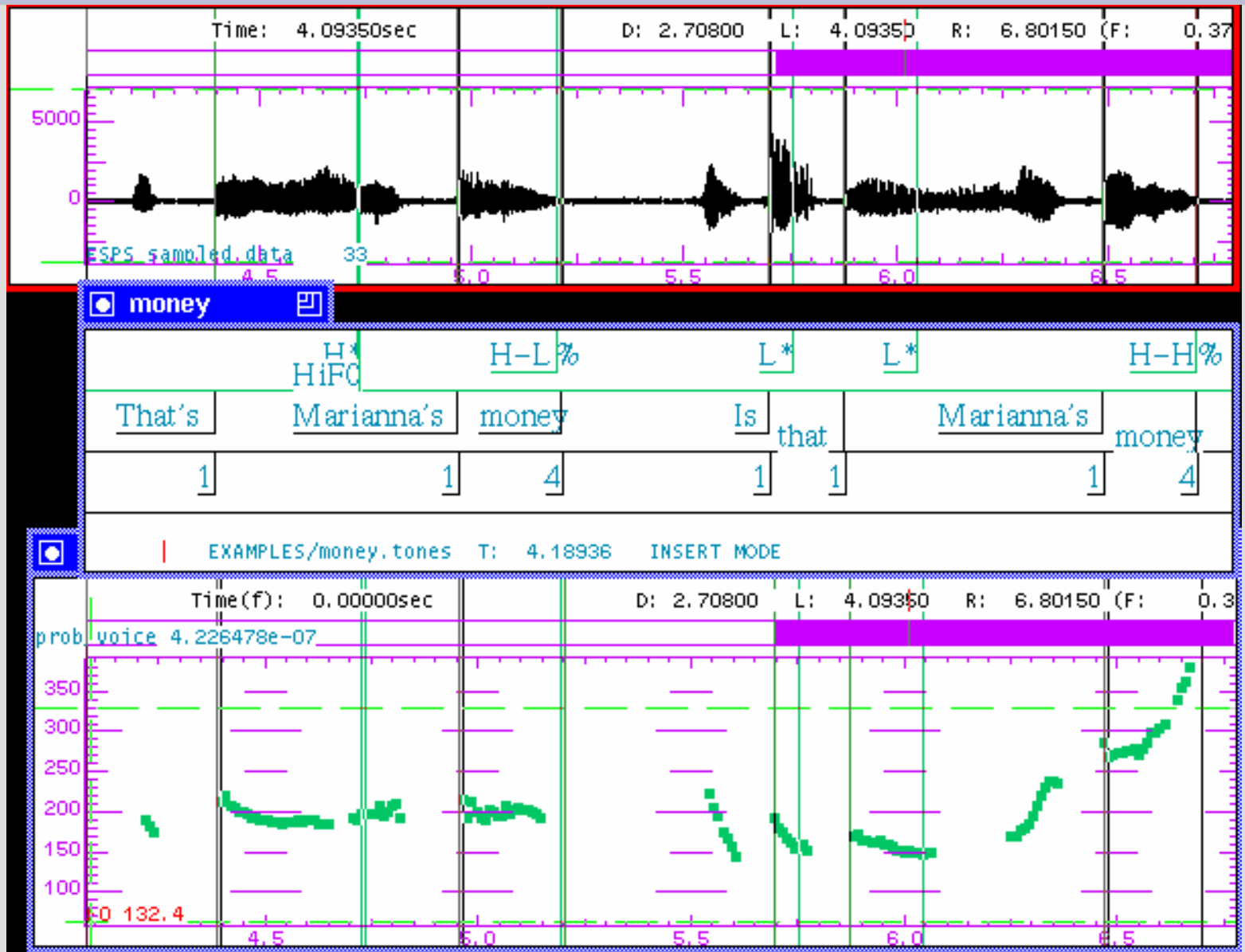


# Word Context Analysis: Prosody

- **Prosodic phrases/sentences:**
  syntactical and lexical information is used to determine prosodic boundaries in order to build intonation and prominence models

- **Sentence modes:**
  retrieved prosodic information will determine the sentence mode, basing on intonation and prominence

# Word Context Analysis: Sentence Mode



# Phonological & Accent Analysis

- In texts we oft encounter out-of-lexicon words, for which there's no pronunciation information available
- For such words we need to build phonological rules, which associate certain phonemes to certain graphemes
- In most languages word accent has also to be determined
- In many languages morphological analysis of unknown words may help to find the right pronunciation

# Phonological & Accent Analysis: Phonological Rules

- For languages with a more or less 1:1 grapheme-phoneme relationship simple conversion rules may be sufficient

**Example: spanish (Festival 1.4.3)**

```
( [ a ] = a )
( [ e ] = e )
( [ i ] = i )
( [ o ] = o )
( [ u ] = u )
( [ "'" a ] = a1 )
( [ "'" e ] = e1 )
( [ "'" i ] = i1 )
( [ "'" o ] = o1 )
( [ "'" u ] = u1 )
( [ b ] = b )
( [ v ] = b )
( [ c ] "'" EI = th )
( [ c ] EI = th )
( [ c h ] = ch )
( [ c ] = k )
( [ d ] = d )
( [ f ] = f )
( [ g ] "'" EI = x )
( [ g ] EI = x )
( [ g u ] "'" EI = g )
( [ g u ] EI = g )
( [ g ] = g )
( [ h u e ] = u e )
( [ h i e ] = i e )
( [ h ] =  )
( [ j ] = x )
( [ k ] = k )
( [ l l ] # = l )
( [ l l ] = ll )
( [ l ] = l )
( [ m ] = m )
( [ ~ n ] = ny )
( [ n ] = n )
( [ p ] = p )
( [ q u ] = k )
( [ r r ] = rr )
( # [ r ] = rr )
( LNS [ r ] = rr )
( [ r ] = r )
( [ s ] BDGLMN = th )
( [ s ] = s )
( # [ s ] C = e s )
( [ t ] = t )
( [ w ] = u )
( [ x ] = k s )
( AEO [ y ] = i )
( # [ y ] # = i )
( [ y ] = ll )
( [ z ] = th )
```

```
(# 0.0 0.250)
(a 0.0 0.090)
(e 0.0 0.090)
(i 0.0 0.080)
(o 0.0 0.090)
(u 0.0 0.080)
(b 0.0 0.065)
(ch 0.0 0.135)
(d 0.0 0.060)
(f 0.0 0.100)
(g 0.0 0.080)
(j 0.0 0.100)
(k 0.0 0.100)
(l 0.0 0.080)
(ll 0.0 0.105)
(m 0.0 0.070)
(n 0.0 0.080)
(ny 0.0 0.110)
(p 0.0 0.100)
(r 0.0 0.030)
(rr 0.0 0.080)
(s 0.0 0.110)
(t 0.0 0.085)
(th 0.0 0.100)
(x 0.0 0.130)
```

# Summing Up

- TTS for most natural languages needs quite complex **linguistic analysis** to perform a good job: **linguistic models** help improving system's performance
- **Linguistic components** of a standard TTS system include a lexicon, morphological and context rules (syntactic and prosodic) as well as phonological rules
- **Each** linguistic **component** is **strictly correlated** with the others: they all concurr to build a **complete linguistic representation** for prosodic and phonetic synthesis

# Resources


- **Möbius**, Bernd: *Sprachsynthesesysteme.* In: Computerlinguistik und Sprachtechnologie. Eine Einführung. Heidelberg-Berlin 2001

- **Black**, Alan W. et al.: *Festival Speech Synthesis System. System Documentation. Edition 1.4.* 1999 http://www.cstr.ed.ac.uk/projects/festival/manual/festival_toc.html

- **Cole**, Ronald A. et al.: *Survey of the State of the Art in Human Language Technology: Spoken Output Technologies.* 1996 http://cslu.cse.ogi.edu/HLTsurvey/ch5node2.html

- *IPA Alphabet*. http://www.arts.gla.ac.uk/IPA/ipachart.html
- *SAMPA Alphabet*. http://www.phon.ucl.ac.uk/home/sampa/home.htm