Language Science & Technology:

Linguistic Foundations

WS 2007-2008 (14.11.2007 & 16.11.2007)

PD Dr. Tania Avgustinova

avgustinova @ coli . uni – saarland . de

Central questions of language research

- **LINGUISTIC KNOWLEDGE**
  - What are the contents and structures of this knowledge?

- **LANGUAGE PROCESSING**
  - How do we produce and comprehend linguistic utterances?

- **LANGUAGE ACQUISITION**
  - How does the child learn his mother tongue?

- **LANGUAGE CHANGE**
  - How do languages (dialects, sociolects) emerge, change, evolve?

Language science and its components

- Variants of language science
  - Traditional Grammar
  - Theoretical Linguistics
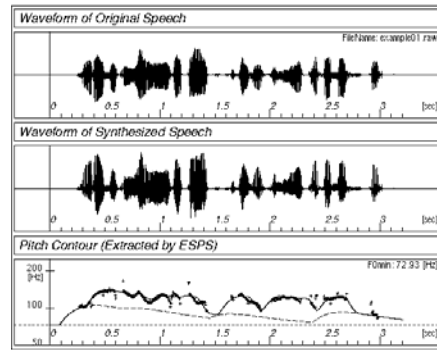 - Computational Linguistics

- The components of grammar
  - Phonology: Science of language sounds
  - Morphology: Science of word form structure
  - Lexicon: Listing analyzed words
  - Syntax: Science of composing word forms
  - Semantics: Science of literal meanings
  - Pragmatics: Science of using language expressions

Simplified Big Picture

Phonology

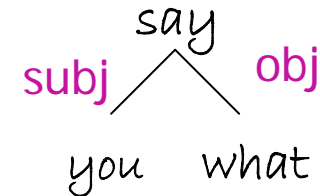$\Leftrightarrow$ /waddyasai/

Morphology

/waddyasai/ $\Leftrightarrow$ what did you say

Syntax

what did you say $\Leftrightarrow$

```
        say
subj  /    \  obj
   you    what
```

Semantics

```
        say
subj  /    \  obj
   you    what
```
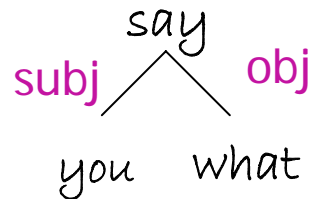
$\Leftrightarrow$ P[λx. say(you, x)]

Units of Language – Subfields of Linguistics

	Grammar	**Semantics**	**Pragmatics**
Sound	Phonetics/ Phonology	---	---
Word	Morphology	Lexical Semantics	---
Sentence	Syntax	Compositional Semantics	Pragmatics
Text & Discourse	Text & Discourse Grammar	Discourse Semantics	Pragmatics
	Structure	**Meaning**	**Use**

UNIVERSITÄT DES SAARLANDES
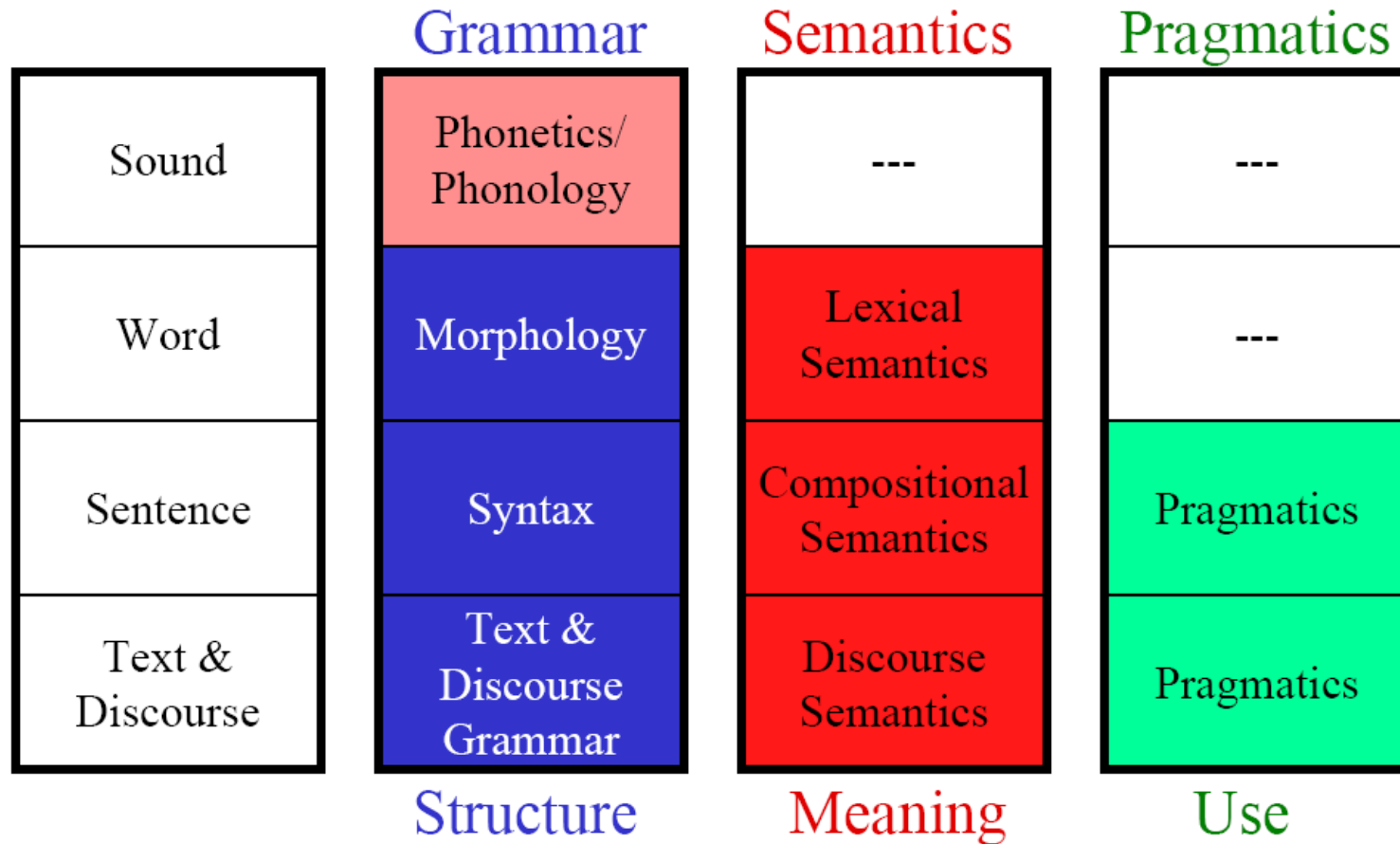
Combination principles of morphology

- **Inflection** is the systematic variation of a word with which it can perform different syntactic and semantic functions, and adapt to different syntactic environments.
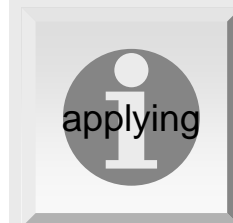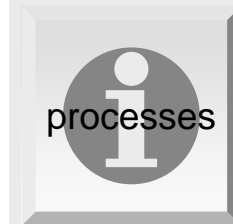
  Examples: *learn, learn/s, learn/ed, learn/ing*

- **Derivation** is the combination of a word with an affix.

  Examples: *clear/ness, clear/ly, un/clear*

- **Composition** is the combination of two or more words into a new word form.

  Examples: *gas/light, hard/wood, over/indulge, over-the-counter*

basics

processes

applying

Introduction to Morphology

Morphology is the study of form and structure.

In linguistics, it generally refers to the study of form and
structure of words.

There are two main usages of the term *word*:

1. Surface form (spoken or written represenation)
2. Abstract form (lemma or dictionary entry,
   e.g. bare infinitives in English, nominative single form of
   nouns in Latin)

   The class of forms representing a word in different contexts
   is called a **lexeme**
   e.g. sing = {*sing, sings, sang, sung, singing*}

A definition of words?

Words can be described as units of language (either sequences of sounds, or signs) that function as meaning bearers. But this is a fuzzy notion, e.g.:

- *sang* expresses both "singing" and past tense.
- Is *more or less* one word, or are there three words?

A structuralist solution: **morphemes**

A language:

> 11-112 phonemes

$\downarrow$

> 4,000-10,000 morphemes

$\downarrow$

> An infinite number of sentences

Morphemes and Morphological analysis

- **Morphemes**
  - Morphemes are minimal meaning-bearing units:
    e.g. *talked* contains two morphemes: *talk* and *-ed* (past).
  - Form-function pairs (sound/sign-meaning)
  - Basic units of morphology
  - The realisations of morphemes are called *morphs*:
    e.g. English plural morpheme:
    [NUMBER pl]: -s, -es, -en, -$\emptyset$
    boy-s, box-es, ox-en, sheep
  - These different realisations of the same morpheme are
    called **allomorphs**.

- **Morphological analysis**
  - Segmentation of expressions into basic units (mostly
    starting from word-level).
  - Classification of these basic units according to function.

- **Free Morphemes**
  Free morphemes can occur independently. Free morphemes are common in both English and German.

  e.g. *boy*, *sing*

- **Bound Morphemes**
  Bound morphemes must be attached to another morpheme, and cannot be used independently.

  e.g. [NUMBER pl] *-s* → *boys*

  Typical bound morphemes are:
  - **affixes** (*boy+s*, *talk+ed*)
  - **clitics** (French: *je ne sais pas*, *je* and *ne* cannot occur without a verb)
  - **roots** (Spanish *habl-* needs an ending indicating person, number, mode, etc.)

Morphemes are form-meaning pairs, but not all segmentable forms have an identifiable meaning:

- **Formatives** are forms without identifiable meaning

  e.g. Linking elements in German compounds:
  *Geburt+**s**+tag* (Birthday), *Schwan+**en**+hals* (swan neck).

- **Pseudo-morphemes** or **cranberry morphemes** are special cases of formatives.
  They are segmentable part of a complex word, but do not have an independent meaning:

  e.g.
    - **cran**+berry, **rasp**+berry
    - re+**ceive**, con+**ceive**

Morphology can refer to three different things

a Description of the behaviour of morphemes and how they
  are combined.

b Derivational, inflectional and compositional processes of
  word formation occurring in a specific language.
  e.g. "German has a richer morphology than English"

c Description of such word formation processes.

- **Root**: an unanalysable form, expressing the basic lexical content of a word. Also defined as 'what is left of a complex form when all affixes are stripped'.
- **Stem**: consists of at least a root.
  It can contain (an) derivational affix(es).
  In inflectional morphology, *stem* is generally defined as the root + a thematic vowel.
- **Base**: a form to which an affix may be added. A base may be simplex (root) or complex (root + affixes).

We distinguish:

- **Word forming**:
    - Derivational morphology
    - Compounding
- **Inflection**

- allows to build complex words by combining bound and free morphemes.
- Derivational operations are per definition optional, i.e. not required by syntactic criteria.
- They change
  - a semantics,
    e.g. [*clear*] → [*un*+[*clear*]] = unclear
  - b syntactic category,
    e.g. [*derive*]$_V$ → [[[*derive*]$_V$+*ation*]$_N$ +*al*]$_{Adj}$ = derivational
  - c valency of a verb,
    e.g. [*qaw*] 'it breaks' → [*t*+[*qaw*]] 'he breaks it' (Havasupai)
  - d several from the above, e.g. [understand]$_V$ →
    [[understand]$_V$+able] = understandable

Compounding

- allows to build complex words by juxtaposition of free morphemes.
  [[*sale*]+s+[*man*]], [[*dish*]+[*washer*]].
- Productive compounding results in an infinite lexicon.
  $\left\{\begin{array}{l} \textit{English} \\ \textit{German} \\ \textit{Havasupai} \end{array}\right\} \left\{\begin{array}{l} \textit{phonetics} \\ \textit{phonology} \\ \textit{morphology} \end{array}\right\} \left\{\begin{array}{l} \textit{teacher} \\ \textit{researcher} \\ \textit{student} \end{array}\right\}$
- Compounds are "referential islands".

Inflectional Morphology

- Inflection is required by syntactic criteria, e.g. an English verb must have tense.
- It marks grammatical (=morphosyntactic) distinctions:
    - Conjugation (verbal categories):
        1. person, number, gender
        2. tense, aspect, mood, agreement
    - Declination (nominal categories)
        - case, number, gender, degree, definiteness
- Meaning or, at least, the general concept is (generally) not changed, though *when, who* or *what* and sometimes *where, how* and *whether* may be specified by inflectional morphemes.
- There are bound and free inflectional morphemes:
  go [TENSE past]: *went*
  go [TENSE future]: *will go*

Inflectional morphology is typically organised in **paradigms**.

Paradigm

"A set of forms having the same root/stem, one of which must be selected in a certain syntactic environment" (definition based on Crystal (1997:277) and Payne (1997: 26)

For instance, German conjugation:

present	NUMBER		*past*	NUMBER	
	singular	*plural*		*singular*	*plural*
1.	dehn-e	dehn-en	1.	dehn-te	dehn-te-n
2.	dehn-st	dehn-t	2.	dehn-te-st	dehn-te-t
3.	dehn-t	dehn-en	3.	dehn-te	dehn-te-n

Latin declination of a noun of the first declination:

case	NUMBER	
	singular	*plural*
NOM	puella	puellae
GEN	puellae	puellarum
DAT	puellae	puellis
ACC	puellam	puellas
ABL	puella	puellis

We observe both:

- **syncretism**: the same form is used to express different feature combinations.
  Here: -*ae*: GEN or DAT singular, or NOM plural, -*a* NOM or ABL singular, -*is*: DAT or ABL plural.
- **exponence**: the relation between form and function is **m:n**:
  - **multi-exponence** (cumulation): one form expresses several functions.
    Here: -*am* expresses both accusative and singular
  - **Extended exponence**: in *ge-dehn-t*, *ge-* and -*t* express one function together.

Morphological Properties — Synthesis

Synthesis: the number of morphemes that tend to occur within a word.

- In **isolating** languages words tend to consist of only one morpheme. (e.g. Chinese languages)
- **Polysynthetic** languages are known for the large number of morphemes that may occur in a single word. For instance, the Quechua and Inuit languages. The following example is from Yup'ik:

  (1)   tuntussuqatarniksaitengqiggtuq
            tuntu-ssur-qatar-ni-ksaite-ngqiggte-uq
            reindeer-hunt-FUT-say-NEG-again-3gg-IND
            'He had not yet said again that he was going to hunt reindeer'

(Payne, 1997:28)

Fusion: the number of meaning units that are found in one morphological shape:

- **Agglutinative** languages have little fusion: each meaning component is represented by its own morpheme (e.g. Turkish).
- **Fusional** languages have morphemes that express many meaning units: e.g. *-ó* in Spanish *habló* expresses indicative mode, 3rd person, singular, past tense and perfect aspect.

In English, both examples of agglutinative morphemes, and fusional ones can be found:

- **agglutinative**: anti+dis+establish+ment+arian+ism
- **fusion**: vowel change in plural forming (*goose/geese*) and strong verbs (*sing/sang*).
  Individual morphemes (root and number/tense) cannot be segmented in chunks, therefore these forms are fusional.

Morphology related applications in computational linguistics are:

1. Analysing complex words, defining their component parts:

   anti+dis+establish+ment+arian+ism

2. Analysis of grammatical information, encoded in words:

   sings
   sing[PERSON 3, NUMBER singular, TENSE present]

Morphological processes

❏ **Segmental processes**
  ❍ Affixation
  ❍ Modification
    − Substitution of segments (umlaut, ablaut, suppletion)
    − Subtractive morphology (deletion of segments)

❏ **Suprasegmental**
  ❍ Stress
  ❍ Tone

Affixation

- ❏ **Recursive process**
- ❏ **Affixes are bound morphemes**
- ❏ **Affixes are positionally fixed with respect to the base**
  - ❍ prefix
    - – un+happy
  - ❍ suffix
    - – happy+ly
- ❏ **Root**
  - ❍ Part of a morphologically complex form after all affixes are stripped
- ❏ **Stem**
  - ❍ Root + thematic vowel in inflectional morphology
- ❏ **Base**
  - ❍ Part of a morphologically complex form to which an affix can be added
  - ❍ A base may be simplex (i.e. a root) or complex (root + affixes)

Affixation

❏ **Order of application is meaningful**

   [un [[do] able]]   vs.   [[un [do]] able]

❏ **Words can have internal structure**
❏ **Morphotactics describes constraints on morpheme order**
❏ **Morphotactics can be determined by**
   ○ word syntax
   ○ non-syntactic factors, e.g. lexical strata

   e.g.: *non-impartial* vs. **in-non-partial*

Types of affixation processes

affixation

    constant string   copied string

continuous base   discontinuous base   reduplication

Prefix   Suffix   Circumfix   continuous affix   discontinuous affix

infix   transfix
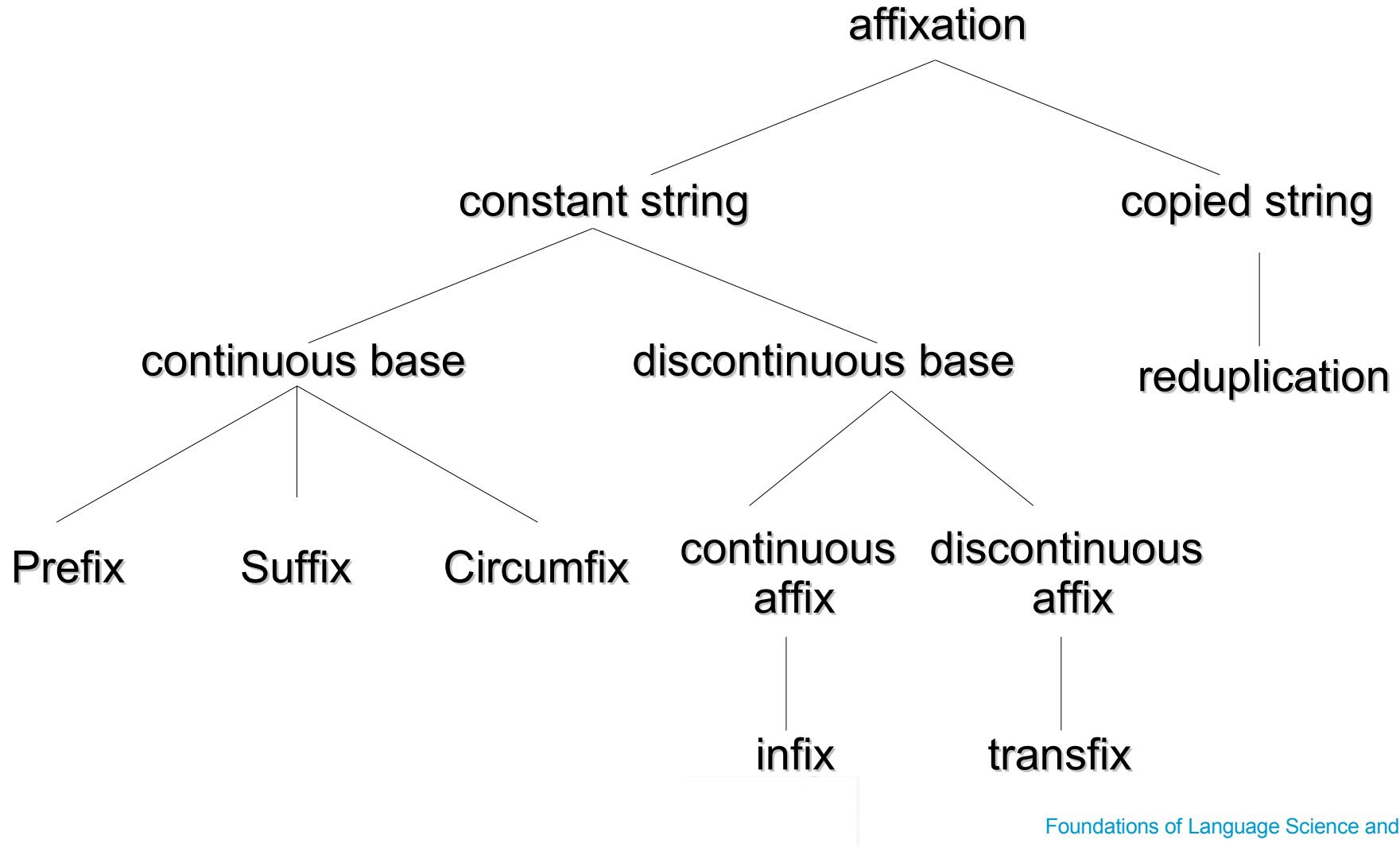
Prefixation, Suffixation, Circumfixation

❏ **Prefixation and suffixation are crosslinguistically predominant affixation processes**

❏ **In English and German, most inflectional and derivational affixes are suffixes**

❏ **In Bantu languages, such as Swahili, prefixation is dominant**

❏ **Circumfixation can be described as simultaneous addition of pre- and suffixes**

❏ **Ex: German regular past participles**

ge+arbeit+et `worked'

Infixation

❏ **Infixes are affixes which are inserted into the base, thereby leading to discontinuous bases**

❏ **The infix itself is continuous**

❏ **Infixation is rare in European languages**

❏ **Infixation can be motivated by prosodic factors**

  ○ e.g. Tagalog *um + sulat = s-um-ulat,* (vs. *um + aral = um-aral*)

  ○ Avoidance of closed syllables (consonant-final syllables)

  ○ Prosodic conditioning of infixation extensively studied in Optimality Theory (McCarthy and Prince)

❏ **Infixation can also be purely morphologically conditioned**

  ○ e.g. Udi infixation (Harris 1997)

| Root	Transitive		Intransitive	
box	*bo-**ne**-x-sa*	boils	*box-**ne**-sa*	boils
uk	*u-**ne**-k-sa*	eats	*uk-**ne**-sa*	is edible

Transfixation

❑ **Transfixation is an affixation where the segmental material of root and affix gets interleaved**
  ○ i.e. both the root and the affix are discontinuous

❑ **Transfixation is widely attested in Semitic languages, e.g. Arabic and Hebrew**

❑ **Ex.: forms of the Arabic root *ktb***

Binyan	ACT (a)	PASS (u i)	Template	Gloss
I	*katab*	*kutib*	CVCVC	write
II	*kattab*	*kuttib*	CVCCVC	cause to write
III	*kaatab*	*kuutib*	CVVCVC	correspond

❑ **Theoretically modeled by means of multidimensional representations (Autosegmental Phonology), associating consonantal and vocalic tiers to a CV skeleton**

Transfixation

❏ **Theoretically modeled by means of multidimensional representations (Autosegmental Phonology), associating consonantal and vocalic tiers to a CV skeleton**

Consonantal root tier

k   t   b

C — V — C — C — V — C

u   i

Vocalism tier
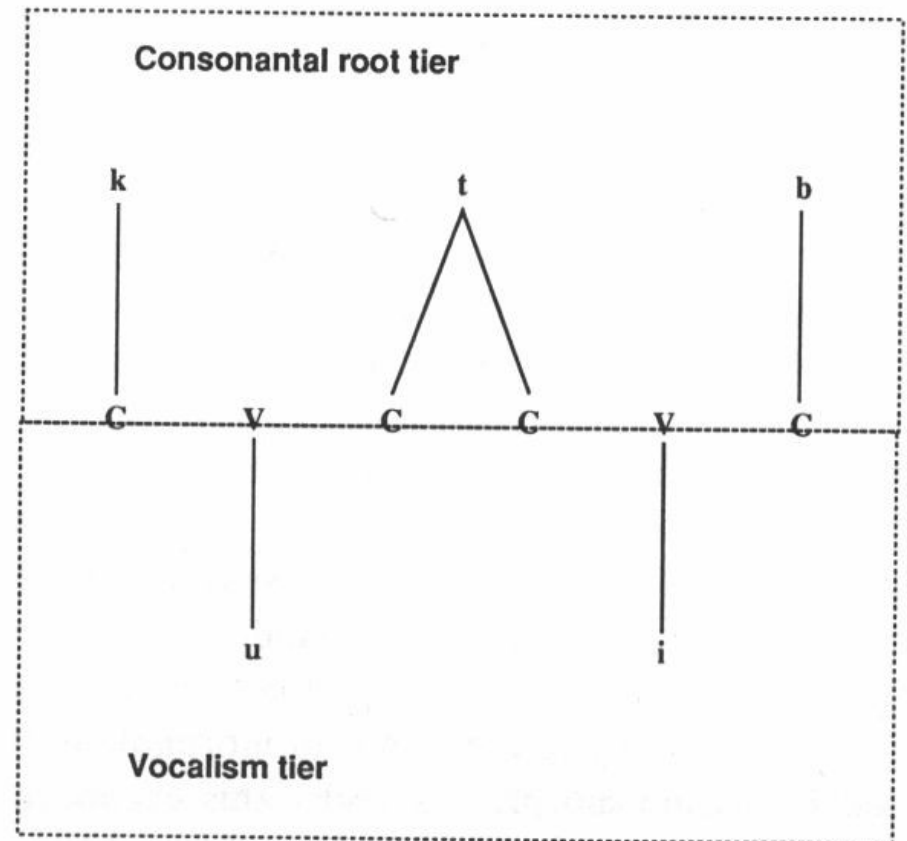
Modification

❏ **Morphological process affects stem-internal segments**
❏ **Typical examples include "ablaut" and "umlaut" in German and English**
❏ **Umlaut:**
  ○ Phonologically predictable segmental alternation (e.g. fronting in German): a → ä, o → ö, u → ü
  ○ *Mutter* (sg)→ *Mütter , Wald* (sg)→ *Wälder* (pl)*, Tod* (N)→ *tödlich* (A)
  ○ Umlaut in German is morphologically conditioned: e.g. Futter (sg)
❏ **Ablaut:**
  ○ Phonologically unpredictable segmental alternation
  ○ *gehen – ging – gegangen* vs. *sehen – sah – gesehen*

Subtractive morphology

❑ **Process which marks morphological category by removing segments from the base**

❑ **Shape of the base cannot be predicted from the shape of the derived form**

❑ **Subtractive morphology presents severe foundational problem for morpheme-based theories of inflection and derivation**

❑ **Ex: Koasati**

singular	plural	gloss
pit**af**+fi+in	pit+li+n	to slice up the middle
las**ap**+li+n	las+li+n	to lick something
acokcan**a:**+ka+n	acokan+ka+n	to quarrel with someone

Suprasegmental marking

❏ **Stress shift**

   ❍ English verb-noun derivation:

     prod**u**ce (V) – pr**o**duce (N)
     perm**i**t (V)   – p**e**rmit (N)
     imp**o**rt (V)   – **i**mport (N)
     ins**u**lt (V)   – **i**nsult (N)
     disc**ou**nt (V) – d**i**scount (N)

❏ **Tone**

   ❍ Kanuri (North-eastern Nigeria)
     lezè (subjunctive) – lezé (optative) 'gehen'
     tussè (subjunctive)– tussé (optative) 'ruhen'

Reduplication

- ❏ **Morphological process where (part of) the base is copied**
- ❏ **Often used to express categories such as plurality, iterativity, habituality etc.**
- ❏ **Total reduplication**
  - ○ entire base is copied, e.g. Indonesian
    orang `man' – *orang orang* `men'
  - ○ redup[lication can interact with segmental changes, e.g. Javanese
    bali `return' – *bola+bali* `return repeatedly/habitually'
- ❏ **Partial reduplication**
  - ○ segmental material is partially copied, typically, a prosodic constituent, like a syllable or a foot, e.g. Yidin[y]
    mulari   *mula+mulari*   `initiated man'
    gindalba   *gindal+gindalba*   `lizard'
- ❏ **Autosegmental Phonology assumes affixation of CV templates and spreading (copying) of segments to skeleton slots**

Morphophonology

- **Morphological process can trigger phonological or graphemic alternations**
- **Phonological alternations at the juncture between morphemes are highly frequent (internal Sandhi**
- **Sandhi can also occur at word boundaries (external sandhi)**
- **Morphophonological alternations**
  - Assimilation
    - Homorganic nasal assimilation
      iN+possible = impossible [imp...]
      iN+complete = incomplete [iŋk...]
    - Voicing assimilation
      cat+s = [...ts]
      dog+s = [...gz]
  - Epenthesis: *wish+s = wishes [wišiz]*
  - Deletion
- **Graphemic alternations**
  - *y + s ~ ies*

Harmony processes

❏ **Phonological processes can also apply long-distance**

❏ **Harmony processes require identity of segments (typically vowels) with respect to some feature**

E.g. Finnish front/back vowel harmony

[back +] vowels: a, u, o
[back -] vowels: ä, y, ö
neutral vowels: i, e

taivas (NOM)   –   taivas+ta (PART)   –   *taivas+tä
lyhyt (NOM)   –   lyhyt+tä (PART)   –   *lyhyt+ta

❏ **Number of interacting harmony processes highly restricted**
  ○ typically 1, at most 2 (Warlpiri)
  ○ Low number may be correlated with set of distinct features (Koskenniemi)

Morphological processing systems

❏ **Inflection:**
  - ○ lemmatisation/stemming
  - ○ extraction of grammatical (morphosyntactic) features (preprocessing for parsing)
  - ○ reduction in lexicon size (1:2 for English, 1:5 for German, >1>200 for Finnish/Turkish)
  - ○ Finite state technology is state of the art

❏ **Derivational morphology**
  - ○ Semi-productivity and semantic opaqueness still pose problems
  - ○ Rule-based approaches may suffer from overgeneration
  - ○ Lexicalisation of complex forms useful

❏ **Compound analysis**
  - ○ indispensible for languages with productive compounding (e.g. German)
  - ○ Issues: bracketing

Combination principles of syntax

- Correlation of morphology and syntax in different types of language: Some natural languages compose meaning mainly in the syntax and others mainly in morphology.

- Differences between natural languages

- Natural languages are all based on the same time-linear derivational order.

- They differ only in their language specific handling of valency structure (lexicalization), agreement, word order

typology

UNIVERSITÄT
DES
SAARLANDES

- How languages differ (linguistic diversity)

- How languages are alike (linguistic homogeneity)
    - Every language distinguishes nouns from verbs
    - Every language combines words into phrases and sentences

Identifying Word Classes

Three types of criteria:

1. Distributional: Where does it occur?

2. Morphological: What forms can it have?

3. Functional: What work does it perform?

Grammatical Categories

- Form:
    - Inflection
        - Affix indicates grammatical category
    - Closed class words

- Types
    - Inherent categories
        - Properties a word has or doesn't have
    - Agreement categories
        - Show syntactic links between words
    - Relational categories
        - Mark the relationship a word or phrase has to the whole sentence

- **Nouns**
    - Inherent: number, gender or noun class, definiteness
    - Relational : case

- **Verbs**
    - Inherent: tense, aspect, mood, transitivity
    - Relational: voice
    - Agreement: agreement with arguments

- **Adjectives**
    - Inherent: degree of comparison (equative, comparative, superlative)
    - Agreement: agreement of attributive adjectives with head noun; agreement of predicative adjectives with subject.

Heads and their dependents

- Properties of heads
  - Head bears most important semantic information of the phrase.
  - Word class of head determines word class of entire phrase.
    - [$_{NP}$ very bright [$_N$ sunflowers]]
    - [$_{VP}$ [$_V$ overflowed] quite quickly]
    - [$_{AP}$ very [$_A$ bright]]
    - [$_{AdvP}$ quite [$_{Adv}$ quickly]]
    - [$_{PP}$ [$_P$ inside] the house]

  - Head typically has same distribution as the entire phrase.
    - Go inside the house.
    - Go inside.
    - Kim likes very bright sunflowers.
    - Kim likes sunflowers.

  - Heads normally can't be omitted.
    - *Go the house.
    - *Kim likes very bright.

  - Heads select dependent phrases of a particular word class.
    - The soldiers released the hostages.
    - *The soldiers released.
    - He went into the house.
    - *He went into.
    - bright sunflowers
    - *brightly sunflowers

- Heads often require dependents to agree with grammatical features of head.

- Heads may require dependent NPs to occur in a particular grammatical case.

→

Head-Marking and Dependent- Marking Languages
- Syntactic relationships between heads and dependents

Head	Dependent
postposition/preposition	object NP
verb	arguments (subject, object)
(possessed) noun	possessor NP
noun	adjective

English
- **in** [$_{NP}$ the shower]   (**P** + NP)
- Kim **loves** Lee   (Su + **V** + Obj)
- Kim's **house**   (possessor NP + **N**)
- red **book**   (modifying A + **N**)

- Head preposition/postposition and its NP object

Dependent-marking
German: prepositions 'govern' the case of their object
- Für meinen Freund   mit   meinem   Freund
  for my:ACC friend   with   my:DATIVE friend
  'for my friend'   'with my friend'

Head-marking
Tzutujil
- ru-ma   ri-achin
  3SG-because.of   the-man
  'by the man'
Welsh
- arna   i   arno   fo   arni   hi
  on:1SG   me   on:3M:SG him   on:3F:SG her
  'on me'   'on him'   'on her'

The clause: a head verb and the arguments of the verb

Dependent-marking
Japanese
- Taroo-**ga**   tegami-**o**   kaita
  Taroo-NOM   letter-ACC   wrote
  'Taroo wrote a letter.'

German
- Der   Hund sah   den   Vogel
  the:NOM   dog   saw   the:ACC   bird
  'The dog saw the bird.'
- Den   Vogel   sah   der Hund.
  The:ACC   bird   saw   the:NOM dog
  'The dog saw **the bird**.'

Head-marking
Kambera
- Hi **ku**-palu-**ya**
  so   1SG:SU-hit-3SG:OBJ
  'So I hit him.'
- I   Ama, **na**-kei-**ya**   na ri   muru
  the father 3SG:SU-buy-3SG:OBJ the vegetable   green
  'Father buys the green vegetables.'
  Lit., 'Father he-buys-it the green vegetable'

Cakchiquel
- Per ma   x-**e**-**r**-komsaj-ta
  but NEG   CMPL-3PL:OBJ-3SG:SU-kill-IRREALIS
  'but he didn't kill them'

Head noun and dependent possessor NP

- **Dependent marking**
- English
  - Kim**'s** house
- Finnish
  - tytö-**n**　　kissa
    girl-GEN　　cat
    'girl's cat'

- **Head-marking**
- Saliba
  - Sine　　natu-**na**
    woman　child-3SG
    'the woman's child'

Head noun and dependent AP

- **Dependent-marking**
- Spanish: adjective agrees with noun in gender
  - el　　　　niño　pequeño
    the:MASC　boy　　small:MASC
    'the small boy'
  - la　　　　niña　pequeña
    the:FEM　girl　　small:FEM
    'the small girl'

- **Head-marking**
- Persian: noun is marked as having a dependent
  - kûh-**e**　　boländ
    mountain　high
    'high mountain'

- **Head-marking languages**
  - Abkhaz, Mayan (Jacaltec, Tzotzil, Cakchiquel), Athabaskan, (Navajo), Iroquoian (Mohawk, Cherokee), Algonquian (Cree, Blackfoot), Siouan (Crow, Lakhota), Salish (Squamish)
- **Dependent-marking languages**
  - Indo-European (German, Greek, Armenian, Slavic [Russian, Polish, Czech, Bulgarian]), Pama-Nyunngan (Dyirbal, Yidiny), Northeast Caucasian (Chechen), Dravidian (Malayalam).
- **Neither head-marking nor dependent-marking**
- Chinese
  - **Wo** changchang jian **ta**
    I   often   see he
    'I often saw him'
  - **Ta**   changchang jian **wo**
    he   often   see   I
     'He often saw me'
- English: a little dependent-marking
  - Kim**'s** house   Possessor marker 's
  - **He** met **him**   Case-marking in pronouns
  - **these** books   Determiner-noun number agreement
- But also a little head-marking
  - Bill smoke**s**   Subject-verb agreement
  - I **am**, she **is**, we **are**   Subject-verb agreement
- Mixtures are not unusual: German is dependent-marking with subject-verb agreement
  - Ich   **sehe**   den   Vogel
    I:NOM   see:PRES:1SG   the:ACC   bird
    'I see the bird.'
  - Wir   **sehen**   den   Vogel
    we:NOM see:PRES:1PL   the:ACC   bird
    'We see the bird.'

→

Relationships within the clause

▪ All languages have intransitive sentences, with one participant:
  o John sneezed. -> *John* is subject

▪ All languages have transitive sentences, with two participants
  o John saw Mary. -> *John* is subject, *Mary* is object

▪ To distinguish subjects from objects (**core arguments**), languages use one or more of three strategies:

  o Word Order
  o Case Marking
  o Agreement Marking

How do we identify constituents?

Discovering the structure of sentences

Evidence of structure in sentences
- Structural ambiguity
  - Black cab drivers went on strike yesterday
  - **Black [cab drivers]** went on strike.
  - **[Black cab] drivers** went on strike.
  - The boy and the girl's uncle stayed to dinner.
  - **[The boy and the girl]'s** uncle stayed.
  - The boy and **[the girl]'s** uncle stayed.
- Sometimes intonation distinguishes the two readings.
- Constituent
  - A group of words that forms a phrase in a sentence
- Constituent Structure
  - A particular grouping of words
- A sequence of words which form a constituent in one environment, need not in another
  - The students wondered how simple textbooks could be obtained.
  - The students wondered how simple textbooks could be.
- We need to manipulate the sentence to discover constituency, using formal constituency tests.
  - The students wondered how **they** could be obtained.
  - The students wondered how simple **they** could be.